Molecular analysis of leukaemia genomes during malignant transformation and leukaemia progression

Philippa Charlotte May

Imperial College London

Department of Haematology

Submitted for the degree of Doctor of Philosophy

December 2017

Declaration of originality

This thesis is my own work. All contributions of others have been referenced and acknowledged.

Copyright declaration

The below figures are from the sources referenced in the legends of each figure and permission for reproduction has been obtained via the licenses below:

Figure 1-2: #4173070141752 from Nature Reviews Genetics

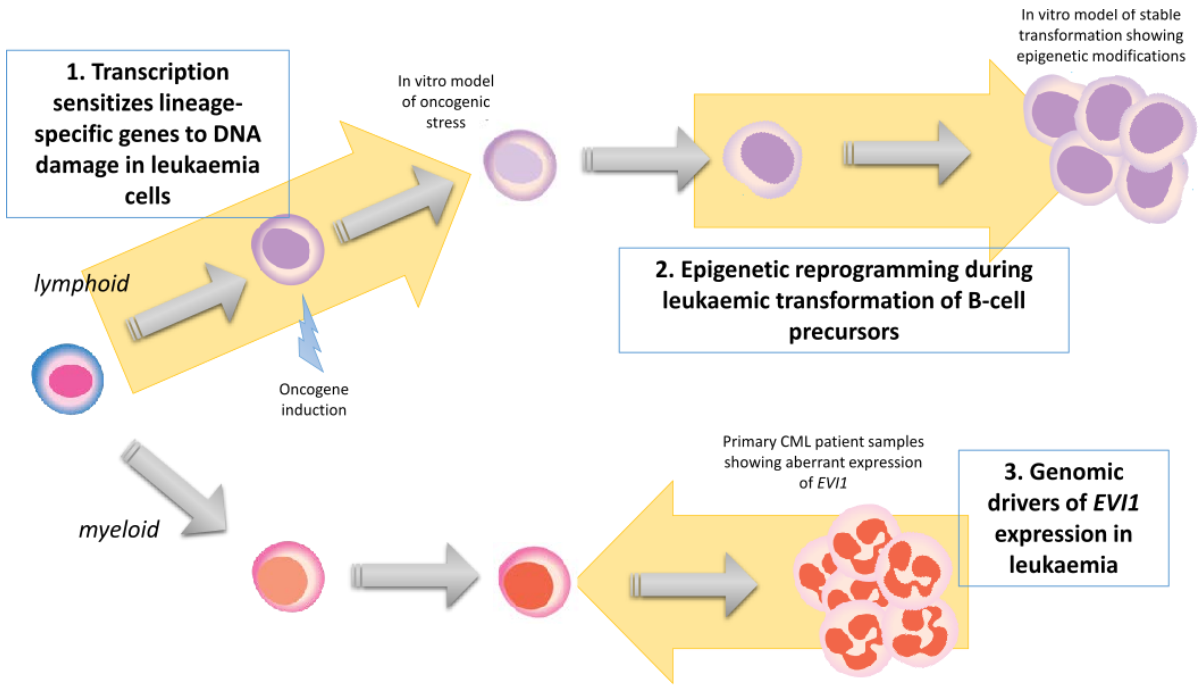
Chapter 2: Cell Press states that permission is not required to use images or information that an individual has themselves previously published in a Cell Press journal; they ask only that the original publication is cited (Canela et al., 2017)

Figure 4-17: #4243050537417 from Nature Protocols

Abstract

This thesis details a tripartite approach to understanding genetic and epigenetic modifications in human leukaemia. First, an *in vitro* mouse model of precursor B-cell leukaemia was established by retroviral transduction of primary B-cell precursors with the oncogenes *BCR-ABL1* or *MYC*. The model was used to study DNA damage incurred during the initial phase of oncogene activation i.e. oncogenic stress. Analysis of γH2AX (a mark of DNA damage) and the H3K27 acetylation histone modification (a mark of active enhancers and promoters) by chromatin immunoprecipitation sequencing (ChIP-seq) and mRNA expression (RNA-seq) were combined to establish that lineage specific transcription predisposes lineage specific genes to DNA damage. This was functionally validated with a lineage-switch model and technically validated using FISH.

Second, the same cells were maintained during their recovery from the period of acute oncogenic stress until stably transformed. Re-analysis of the transformed populations using the same H3K27Ac ChIP-seq and RNA-seq techniques showed that the stable transformation of these cells is paralleled by activation of a subset of intergenic enhancers which induce an aberrant transcriptome enriched for cell activation and metabolic pathways.



Graphical abstract showing tripartite approach to understanding genetic and epigenetic modifications in human leukaemia.

Finally, the techniques used in these *in vitro* experiments and the knowledge of oncogenic enhancers gained from them were then applied to the clinical setting. *EVI1*, a key haematopoietic transcription factor that regulates differentiation, is aberrantly expressed in

a subset of myeloid leukaemia where it is universally associated with a poor prognosis. A patient sample biobank was established and suitable samples were filtered using a combination of RT-qPCR and FISH. Thus, H3K27AC ChIP-seq and newly optimised chromosome conformation capture sequencing (3C-seq or 4C) was used to understand the epigenetic interactions which result in aberrant expression. This has revealed several interactions between the *EVI1* promoter with up- and down-stream intergenic regions which must be further investigated.

Acknowledgements

I will be forever grateful to both my supervisors, Dr Alistair Reid and Dr Niklas Feldhahn. To Alistair, for giving me the opportunity to study for this PhD and for opening the door to so many opportunities outside of Clinical Science. To Niklas, for seeing beyond my FISH abilities and enabling me to join his group and finish my studies in the most unconventional of circumstances.

The Feldhahn Group (past and present) have changed me for the better and I am so fortunate to have worked with all of them. They have all supported me in so many ways, it is difficult to specify. Dr Bryant Boulianne taught me everything I needed to know about ChIP-seq which has been pivotal in the work presented here. Dr Andrew Innes and Dr Matthias Pfeifer have both been unquestioningly generous with their time, advice and good humour. Dr Mark Robinson is uncomfortable with direct praise which is unfortunate because he deserves a huge amount. It will all remain unwritten but I am confident he knows the scale of my gratitude. Together, they have seen me through some very stressful times and all the while made me laugh more than I have in years. I am proud to be a part of Feldhahn Group.

I would like to thank those members of Imperial Molecular Pathology (past and present) who have supported me or helped out with samples; I'm sorry that circumstances meant I couldn't stay. I am grateful for the support of Professor Jane Apperley, Dr Dragana Milojkovic, Professor Irene Roberts, Dr Andy Porter and Dr Jamshid Sorouri Khorashad.

Although not directly involved in this research, I would also like to acknowledge the support given to me by Antony Senner and Ved Ramnani at the Royal Free London NHS Trust. They took a chance on me in a new role and thus enabled me to finish this work.

I could not have completed this project without the patience and love of my husband, Phil, and thus I dedicate this thesis to him. He has supported me throughout: accepting my ridiculous working patterns and my inability to take time off, tolerating my lack of focus on anything else and, of course, keeping me well-nourished with his amazing cooking.

I am grateful to Bloodwise for supporting the Feldhahn group and to Marie Curie Cancer Care for supporting my PhD. Lastly, I would like to acknowledge the patients at Imperial College Healthcare NHS Trust who have given permission for their samples to be used for research, and the clinicians who help coordinate that. Their generosity for others is greatly appreciated.

Contents

List of Figures

¶

List of Tables

Abbreviations

- (n)RPM: (Normalised) Reads counts Per Million
- AID: Activation-Induced Deaminase
- ALL: acute lymphoid leukaemia
- AML: acute myeloid leukaemia
- AP: Accelerated Phase (of CML)
- APML: acute promyelocytic leukaemia
- ATLL: Adult T-cell Leukaemia/Lymphoma
- ATLL: adult T-cell leukaemia/lymphoma
- BC: Blast Crisis (of CML)
- ChIA-PET: Chomatin Interaction Analysis with Paired End Tags
- CHIP: Clonal Haematopoiesis of Indeterminate Potential
- ChIP-seq: Chromatin Immuno Precipitation sequencing
- CML: chronic myeloid leukaemia
- CMML: Chronic Myelomonocytic Leukaemia
- CP: Chronic Phase (of CML)
- CRISPR(i): Clustered Regularly Interspaced Short Palindromic Repeats (interference)
- CSR: Class Switch Recombination
- Ct: Cycle Threshold
- DAPI: 4',6-diamidino-2-phenylindole
- DDR: DNA-damage response
- DMSO: DiMethy Sulfoxide
- DNA: deoxyribose nucleic acid
- DRIP-seq: DNA:RNA ImmunoPrecipitation sequencing
- DSB: Double Strand Break
- ERFS: Early Replicating Fragile Site
- eRNA: enhancer RNA
- EV: Empty Vector
- FBS: Fetal Bovine Serum
- FC: Fold Change
- FCS: Fixed Cell Suspension
- FISH: Fluorescence *In Situ* Hybridisation
- FPKM: Fragments Per Kilobase of transcript per million reads
- GFP: Green Fluorescent Protein
- GO: Gene Ontology
- GRO-seq: Global Run-On sequencing
- H3K27ac: acetylation of histone H3 lysine 27
- H3K27me3: tri-methylation of histone H3 lysine 27
- H3K4me1: mono-methylation of histone H3 lysine 4
- H3K4me3: tri-methylation of histone H3 lysine 4
- HAT: Histone Acetyl Transferase
- HDAC(i): Histone Deacetylase (inhibitor)
- HIV: Human immunodeficiency virus
- HR: Homologous Recombination
- HTLV: Human t-cell leukaemia virus
- HU: HydroxyUrea
- ICHTB: Imperial College Healthcare Tissue Bank
- IHC: ImmunoHistoChemistry
- KEGG: Kyoto Encyclopedia of Genes and Genomes
- MACS: Magnetic Activated Cell Sorting
- MDS: Myelodysplastic Syndrome
- MPD: Myeloproliferative Disorder
- MPN: Myeloproliferative Neoplasm
- mRNA: Messenger RNA
- NDR: Nucleosome-Depleted Region

- NFR: Nucleosome-Free Region
- NHEJ: Non-Homologous End Joining
- OIS: Oncogene Induced Senescence
- OS: Oncogenic Stress
- PCR: polymerase chain reaction
- PET: Paired End Tags
- PR: Positive Regulatory
- qPCR: quantitative polymerase chain reaction
- RAEB: Refractory Anaemia with Excess Blasts
- RAG: Recombination Activation Gene
- RLFS: R-Loop Forming Sequence
- RNA: ribose nucleic acid
- RNA-seq: RNA sequencing
- ROS: Reactive Oxygen Species
- RSS: Recombination Signal Sequence
- RT: room temperature
- RT-qPCR: reverse transcriptase quantitative polymerase chain reaction
- sAML: secondary AML
- SCT: Stem Cell Transplant
- SDS-PAGE: Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis
- SEM: Standard Error of the Mean
- SHM: Somatic HyperMutation
- SICER: Statistical model for Identification of ChIP-Enriched Regions
- SNP: Single Nucleotide Polymorphism
- STAR: Spliced Transcripts Alignment to a Reference
- T-ALL: T-cell acute lymphoblastic leukaemia
- TAR: Thrombocytopenia with Absent Radii
- TKI: Tyrosine Kinase Inhibitor
- TRAIL: Tumour necrosis factor Related Apoptosis Inducing Ligand
- UCSC: University of California, Santa Cruz
- WBC: White Blood Cell

1 Introduction

Haematological cancers such as leukaemia are relatively well studied compared to solid tumours. In the decade between 1995 and 2004, publications relating to haematological cancers comprised 12% of basic research studies, compared to just 4.1% in the field of solid tumours (Chizuka et al., 2006). The reasons for our advanced understanding of leukaemia compared to solid tumours are varied; blood and bone marrow are both relatively easy to access for repeated sampling, and the relative simplicity of the leukaemia genome compared to that of solid tumours are two contributing factors.

Many fundamental discoveries which were made in the field of leukaemia research are universally applicable to cancer in general. For example, the first report of genomic instability in cancer came from Theodor Boveri's observations of leucocytes in his 1914 monograph *On the Problem of the Origin of Malignant Tumours* (Boveri, 1914). The first clinical description of leukaemia is generally attributed to John Hughes Bennet (1812-1875), a pathologist at the Royal Infirmary Edinburgh (Bennett, 1845). In 1960, Peter Nowell and David Hungerford identified a shortened chromosome 22 in patients with chronic myeloid leukaemia (CML); the toponymous Philadelphia chromosome, or Ph. (Nowell and Hungerford, 1960a; Nowell and Hungerford, 1960b). This was the first consistent chromosomal abnormality to be associated with a cancer. Janet Rowley later established that the Ph was actually a derivative chromosome formed via translocation of the long arms of chromosome 9 and 22; this was the first chromosomal translocation associated with a leukaemia (Rowley, 1973).

Blood cancer also yielded the first fusion gene: *IGH-MYC* (despite the identification of the Philadelphia translocation, the genes involved were not identified until later (Deklein et al., 1986; Groffen et al., 1984)). Thus, the discovery of the *IGH-MYC* fusion gene rearranged via the t(8;14) found in Burkitt lymphoma kick started the discovery of oncogenic fusion genes which goes on to this day (Taub et al., 1982). Burkitt lymphoma was also the first cancer to be linked to an infectious virus, Epstein Barr (Epstein et al., 1964). Similarly, leukaemia was the first to be studied by genome wide expression analysis (Golub et al., 1999), the first to have copy number analysis (Mullighan et al., 2007) and the first to be subjected to whole genome sequencing (Ley et al., 2008).

Most dramatically, perhaps, is the paradigm of leukaemia therapy. Treatment of Ph+ leukaemia was revolutionised by the development of imatinib in the late 1990s. Originally known as CGP 57148B, trialled as STI571 and eventually marketed by Novartis as Glivec (in the UK) or Gleevec (elsewhere), imatinib mesylate was the first rationally designed drug, a tyrosine kinase inhibitor (TKI) designed to specifically to inhibit the kinase activity of the BCR-ABL1 fusion protein (Druker et al., 2001; Druker et al., 1996) and was hailed as a 'magic bullet' for CML (Henderson, 2003).

Other therapeutic firsts for leukaemia include the Nobel Prize-winning stem cell transplantation developed for leukaemia and now undertaken in a wide variety of cancers (Thomas, 1999), the introduction of monoclonal antibodies for targeted elimination of a B-cell lymphoma (Miller et al., 1982) and the use of rationally designed chimeric antigen receptors to trigger an immunotherapeutic response and thus eliminate leukaemia (Maude et al., 2014).

Observations of leukaemia have led researchers to make hypotheses which are applicable to cancer in general. This includes models of clonal evolution made by cytogeneticists in the 1970s (Nowell, 1976) which are still relevant to next generation sequenced solid tumours in the present day (Gundem et al., 2015). The concept of cancer as a clonal disease (Fialkow, 1974) arising from a stem cell (Fialkow, 1978) was borne in leukaemia, and is still not fully understood in solid tumours. Thus leukaemia remains a worthy subject for basic cancer research.

1.1 Epidemiology and aetiology of leukaemia

Leukaemia is a cancer of the blood or bone marrow, characterised by the accumulation of immature or abnormal leucocytes. This leads to a progressive suppression of normal blood production, and with it, reduced capacity of the blood and bone marrow to undertake its normal functions. This can result in anaemia, immune deficiency and clotting problems.

There are various factors which affect the likelihood of developing leukaemia. One is gender: men are more likely than women to develop virtually any leukaemia and at any age. The single exception to this a slight increase in myeloid leukaemia in females in their early adulthood (Cartwright et al., 2002). As shown in Figure 1-1, the incidence of leukaemias also increases with age. One exception to this is acute lymphoblastic leukaemia which is predominantly a disease of childhood (Pui and Evans, 1998).

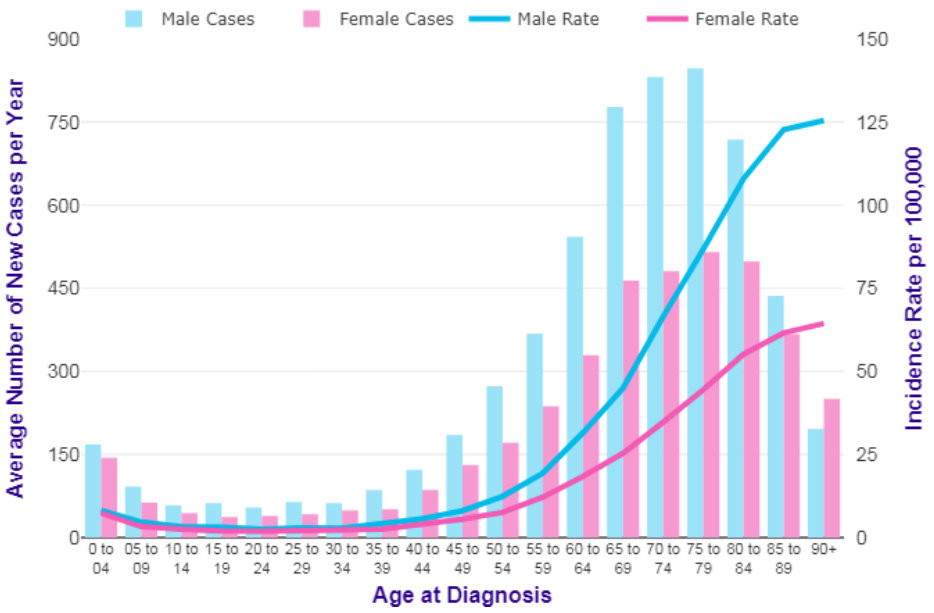


Figure 1-1: Bar chart showing age and gender at diagnosis of leukaemia. ¶Average number of new cases per year and age-specific incidence rates per 100,000 population, UK, 2013-2015. Figure reproduced from Cancer Research UK with permission.

There are some known environmental risk factors for developing leukaemia. A link between exposure to ionising radiation was first noted in 1911 – the excess relative risk of leukaemia in monitored radiation workers was found to be 2.96 per Gray (Deininger et al., 1998). Leukaemia is the second most prevalent cancer (after thyroid cancer) in survivors of the atomic bomb, which has a short latency, and can occur just two years after exposure (Moysich et al., 2002).

Although the cause is not understood, there is a well-established association between cigarette smoking and adult leukaemia – a meta-analysis conducted in 1993 found that up to 14% of adult leukaemia in the United States was attributable to smoking (Brownson et al., 1993). Exposure to benzene and other petrochemicals leads to accumulation of toxic metabolites and increases oxidative stress (McHale et al., 2011); the United States Environmental Protection Agency evaluated the risk of occupational benzene exposure and concluded that employees should not be exposed to more than 1 part per million parts of air (ppm) over an 8-hour shift (Snyder, 2012).

Whilst rare, there are known associations of leukaemia with transmissible viruses. The human T-cell leukaemia virus type 1 (HTLV-1), which is endemic in some areas including Japan, South America and Africa, causes aggressive adult leukaemia/lymphoma (ATLL) (Mahieux and Gessain, 2003). Patients with human immunodeficiency virus (HIV) have an approximately 2.5x increase in leukaemia (Patel et al., 2008).

Genetic predisposition can also underlie the development of leukaemia. These include the bone marrow failure syndromes such as Diamond Blackfan Anaemia, Shwachmann-Diamond syndrome and dyskeratosis congenita – reviewed in (Godley and Shimamura, 2017). Individuals with constitutional trisomy 21 (Down syndrome) have a 150x increased risk of some types of leukaemia, thought to be a consequence of acquired mutations in *GATA1,* which is located on chromosome 21 (Wechsler et al., 2002). Familial leukaemia is rare but recognised in families with mutations of genes such as *CEBPA* (Sellick et al., 2005), *GATA2* (Hahn et al., 2011) and *RUNX1* (Walker et al., 2002); some of these mutations are recurrent but others may be unique to a pedigree. In addition, inherited mutations of DNA damage repair genes can result in a predisposition to leukaemia. This is discussed in more detail in section 1.3.3.

1.2 Classification and genetics of leukaemia

In 1976, the first classification system for acute leukaemia was proposed, based on the morphological and cytochemical techniques in common use (Bennett et al., 1976). It was termed the FAB system, after the group of seven French, American and British authors who authored the paper. The classification system incorporated three main categories: lymphoblastic leukaemias (with three sub-categories termed L1, L2 and L3), myeloid leukaemias (with six subcategories M1 to M6, which were acknowledged to be broadly similar to a prior classification of myeloid disease designed by (Galton and Dacie, 1975)) and lastly 'dysmyelopoietic' syndromes, comprising Refractory Anaemia with Excess Blasts (RAEB) and Chronic Myelomonocytic Leukaemia (CMML). The same group published various updates and additions to the classification, such as the inclusion of myelodysplastic syndromes (Bennett et al., 1982), the acute leukaemias of the megakaryocytic lineage (M7) in 1985 (Bennett et al., 1985) and minimally differentiated leukaemia (M0) as recently as 1991 (Bennett et al., 1991).

The FAB classification system was surpassed by the World Health Organisation's Classification of Tumours of Haematopoietic and Lymphoid Tissues (Norris and Stone, 2008) that made great strides to improve characterisation and standardisation for more accurate diagnosis. It incorporated morphologic and heamatologic evaluation, as with FAB, but also added cytogenetic and molecular genetic studies (see Table 1-1).

Table 1-1: The WHO myeloid neoplasm and acute leukemia classification. Major categories are shown in bold, selected sub-categories relevant to this thesis are shown indented.

Myeloproliferative neoplasms (MPN)
Chronic myeloid leukemia (CML), *BCR-ABL1*+
Myeloid/lymphoid neoplasms with eosinophilia and rearrangement of *PDGFRA*, *PDGFRB*, or *FGFR1,* or with *PCM1-JAK2*
Myelodysplastic/myeloproliferative neoplasms (MDS/MPN)
Myelodysplastic syndromes (MDS)
Acute myeloid leukemia (AML) and related neoplasms
AML with recurrent genetic abnormalities
AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2*, *MECOM*
AML with myelodysplasia-related changes
Therapy-related myeloid neoplasms
AML, NOS
Blastic plasmacytoid dendritic cell neoplasm
Acute leukemias of ambiguous lineage
B-lymphoblastic leukemia/lymphoma
B-lymphoblastic leukemia/lymphoma with recurrent genetic abnormalities
B-lymphoblastic leukemia/lymphoma with t(9;22)(q34.1;q11.2);*BCR-ABL1*
B-lymphoblastic leukemia/lymphoma with t(v;11q23.3);*KMT2A* rearranged
B-lymphoblastic leukemia/lymphoma with t(12;21)(p13.2;q22.1); *ETV6-RUNX1*
B-lymphoblastic leukemia/lymphoma with hyperdiploidy
B-lymphoblastic leukemia/lymphoma with hypodiploidy
Provisional entity: B-lymphoblastic leukemia/lymphoma, BCR-ABL1–like
Provisional entity: B-lymphoblastic leukemia/lymphoma with iAMP21
T-lymphoblastic leukemia/lymphoma

1.2.1 Precursor lymphoid neoplasms

The precursor lymphoid neoplasms, more commonly known as acute lymphoblastic leukaemias (ALL), are a heterogeneous group of haematological cancers, characterised by the accumulation of poorly differentiated blasts in the blood, bone marrow and extramedullary sites. This in turn causes anaemia, thrombocytopenia and leukopenia which manifest as fever, weight loss, night sweats, easy bruising and infection (Terwilliger and Abdul-Hay, 2017). Around 20% of patients also have lymphadenopathy, splenomegaly or hepatomegaly due to the involvement of extramedullary sites (Alvarnas et al., 2015).

ALL affects less than 1 in 100,000 per year (McNally et al., 1999) in the UK, with a bimodal distribution – the first peak occurs in childhood, where ALL represents approximately 75% of childhood leukaemia, and the second occurs in adulthood, where ALL comprises about 20% of adult leukaemia (Jabbour et al., 2005).

The presence of 20% or more lymphoblasts in the bone marrow or peripheral blood is sufficient for a diagnosis, but further refinement of prognosis is undertaken by morphological evaluation, flow cytometry, immunophenotyping and genetic testing. Immunophenotyping classifies ALL into precursor B-cell ALL, mature B-cell ALL and T-cell ALL with various subdivisions conferred by expression of different cell surface markers (Bassan and Hoelzer, 2011). The genetic subtypes of B-ALL, which are now the basis of the World Health Organisation classification, include hyper- and hypo-diploidy, *ETV6-RUNX1* translocations, *MLL* translocations, an intrachromosomal amplification of chromosome 21 (iAMP21), *BCR-ABL1* translocations and more recently '*BCR-ABL1*-like' ALL (Swerdlow et al., 2008). This subset shares a similar gene expression profile, and the high risk prognosis, with *BCR-ABL1*-positive ALL (Tasian et al., 2017). Notably, each of the categories defined by a recurrent genetic abnormality individually comprise less than 10% of the total, which makes cohort size a confounding factor in clinical trials (Moorman et al., 2007).

1.2.1.1 Precursor B-cell leukaemia/lymphoma with t(9;22)(q34;q11.2); *BCR-ABL1*

With the evolution of chromosome banding techniques in the 1970s, Janet Rowley established that the Ph was a derivative chromosome formed via translocation of the long arms of chromosome 9 and 22 (Rowley, 1973). The translocation was shown to involve the proto-oncogene v-Abelson murine leukaemia viral oncogene homolog 1 (*ABL1*) at 9q34 (Deklein et al., 1986) and, later, the breakpoint cluster region (*BCR*) at 22q11.2 (Groffen et al., 1984). The *BCR-ABL1* fusion gene results in the majority of the *ABL1* moiety falling under the control of the *BCR* promoter, and consequently, constitutive activation of *ABL1*. Whilst the presence of the *BCR-ABL1* fusion gene is the definitive feature of CML (discussed in more detail in section 1.2.3.1) it is also found in patients with precursor B-cell

leukaemia/lymphoma and commonly termed Ph+ ALL.

Interestingly, the *BCR* and *ABL1* breakpoints are variable, resulting in fusion proteins comprising different exons and of different molecular weights. In CML and approximately 25% of Ph+ ALL, translocation of *BCR* exons 1 to 12 or 13 to *ABL1* exon 2 results in a fusion protein of 210kDa. In the remaining 75% of Ph+ ALL and very rarely in CML, however, translocation of *BCR* exon 1 only results in a fusion protein of just 190kDa.

Approximately 3% of children and 25-30% of adults with ALL have Ph+ ALL (Moorman et al., 2007). The Ph+ subtype is correlated with age, so by the age of 55 years, the proportion has increased to over 50% (Appelbaum et al., 2006). Unlike CML, where a patient is likely to present with the *BCR-ABL1* translocation alone, Ph+ ALL will have additional mutations. Whilst ALL is now classified by the defining cytogenetic subgroups, translocations like t(9;22) (*BCR*-ABL1) and t(12;21) (*RUNX1-RUNX1T1*) alone are insufficient to cause a full acute leukaemia, thus cooperating lesions are required (Greaves and Wiemels, 2003). As described above in section 1.3.2, *IKZF1* is deleted in 83.7% of paediatric ALL (Mullighan et al., 2008) and around 67% have additional cytogenetic abnormalities at diagnosis (Moorman et al., 2007).

Historically, Ph+ ALL has had a dismal prognosis; a retrospective review of *de novo* paediatric ALL treated between 1986 and 1996 showed that disease free survival at five years was 20% in the worst prognostic cohort and only 49% in the most favourable (Arico et al., 2000).

As discussed in section 1.1, whilst imatinib was hailed as a 'magic bullet' for CML (Henderson, 2003), the equivalent early studies for Ph+ ALL were disappointing. In distinct contrast to CML, where a minority of patients with intolerance and/or resistance began to emerge (Druker et al., 2001), the use of TKI as a monotherapy for Ph+ ALL was generally agreed to be a failure – patients had complete haematological response rates of just 20% which were only sustained for a month in 6% (Ottmann et al., 2002).

However, the development of TKI combination therapy and the evolution of second generation TKIs have improved the outlook for Ph+ ALL significantly. Nilotinib (AMN107; Tasigna, Novartis Pharmaceuticals) and Dasatinib (BMS-354825; Sprycel, Bristol-Myers Squibb) are second generation TKIs with more potent *BCR-ABL1* inhibition and less vulnerability to resistance-conferring mutations of the kinase domain (Lombardo et al., 2004) and were initially approved as a second-line treatment for those patients who were resistant or intolerant to imatinib (Hochhaus et al., 2008; Shah et al., 2008; Shah et al., 2010). More recently, the third generation of TKIs, such as Ponatinib (AP24534; Ariad Pharmaceuticals) has efficacy against mutated *BCR-ABL1* – even the so-called gatekeeper mutation,

T315I (Cortes et al., 2012). Higher remission rates meant that more patients could be treated with haematopoietic stem cell transplant, which was at the time, the only definitive treatment option (Ribera, 2013). Notably, mutations conferring resistance to TKI may pre-date treatment and may be linked to the increased number of mutations present at diagnosis which is not the case in CML (Ottmann et al., 2002; Ribera et al., 2010).

It is now recognised that upfront and continuous exposure to TKI with monitoring of residual disease is the standard of care for Ph+ ALL (Bassan et al., 2010; Thomas et al., 2004; Wassmann et al., 2006), with some early trials achieving a 94% complete remission rate and a predicted two year overall survival of 64% (Ravandi et al., 2010). Most recently, the combination of chemotherapy with ponatinib as front line therapy, followed by ponatinib-only maintenance resulted in a two-year event free survival rate of 81% in a cohort of 34 Ph+ adult patients (Jabbour et al., 2015). Summarising these collective improvements, a recent paper which compared Ph+ and Ph- ALL in the pre- and post-TKI eras actually found that the negative prognostic impact of the *BCR-ABL1* fusion gene no longer exists (Igwe et al., 2017).

An understanding of the pathways which lead to leukaemic transformation is essential to the rational design of targeted therapies. As described above, loss of *IKZF1* expression is a common phenomenon in B-ALL, resulting in a stem cell-like phenotype with increased adhesion. A systematic search for drugs that could reverse this phenotype identified a synthetic retinoid which activated retinoid X receptors; this reduced adhesion, restored *IKZF1* expression and increased sensitivity to dasatinib in a mouse model (Churchman et al., 2015).

Despite these advances, resistance to treatment and disease relapse remain the two most important causes of treatment failure in Ph+ ALL (Soverini et al., 2014). The prognosis for relapsed Ph+ ALL remains poor, with an overall survival at five years of just 7% (Fielding et al., 2007). Understanding the genetic (in particular, the secondary mutations and how they arise) and the epigenetic co-operating aberrations (especially histone modifications other than methylation, such as acetylation) remains an important area for research and could hold the key to new therapeutic targets.

1.2.2 Acute myeloid leukaemia

AML is the most common leukaemia in adults, with an incidence of 3-5 per 100,000 adults, and increases with age (Siegel et al., 2015). It is characterised by the clonal expansion of immature myeloid progenitors which accumulate in the bone marrow and eventually spill out in to the blood or other tissues. A diagnosis is reliant on 20% or more blasts in the peripheral blood or bone marrow (Arber et al., 2016) The hallmark of the disease, and that which differentiates it from chronic myeloid disease, is the severity of the block in myeloid differentiation. In the majority of cases it arises *de novo*, however, it

also occurs in patients with previous exposure to alkylating or topoisomerase II therapy, radiation or a prior haematological disease, and these tend to have a poorer prognosis (Sill et al., 2011).

Acute myeloid leukaemia is an extremely heterogeneous disease, with some well characterised subtypes and others still largely undefined. Subclassification began with the French-American-British system in the 1970s, with cellular and morphological examination defining eight subtypes designated M0-M7, and more recently with the World Health Organisation's classification in which attempts have been made to employ the underlying genetic cause of disease (White et al., 2010).

The World Health Organisation classification now includes various cytogenetically and genetically defined subgroups. The core binding factor leukaemias – those with inv(16), t(16;16) or t(8;21), encoding *CBFB-MYH11* and *RUNX1-RUNX1T1*, respectively – have a good prognosis (Burnett et al., 2010; Dohner et al., 2017). Others, such as those with *FLT3* mutations or those with a complex karyotype (i.e. those with multiple different cytogenetic abnormalities) have a poor prognosis (Mrózek et al., 2012). Acute promyelocytic leukaemia (APML) is defined by *RARA* translocations, most commonly *PML-RARA* formed via t(15;17) – and this subtype benefits from therapy with arsenic trisoxide and retinoic acid and is the only subtype to have seen a significant improvement in outcome in AML (Lo-Coco et al., 2013).

Despite the growing number of defined subtypes, most therapy still relies on various combinations of chemotherapy and haematopoetic stem cell transplant where it can be tolerated. There are some novel therapies in development and trial; for example, sorafenib is a TKI with action against *RAF, cKIT* and the *FLT3* internal tandem duplication (Zhang et al., 2008). Chimeric antigen receptor T-cells (CART) can be created from host-derived cells to target an antigen of choice; preliminary results suggest they can confer anti-tumour activity without the burden of long term myelosupression (Kenderian et al., 2015).

1.2.3 Myeloproliferative neoplasms

Myeloproliferative neoplasms (MPN), sometimes known as myeloproliferative disease (MPD) are a diverse group of clonal stem cell disorders which are characterised by the proliferation of one or more of the myeloid lineages. Notably, the proliferation is of mature cells. MPNs may present with a subtle phenotype but symptoms will progress and up to 20% may continue to bone marrow failure and acute transformation (Jamieson et al., 2008).

The latest WHO categorisation includes chronic neutrophilic leukaemia, polycythaemia vera, primary myelofibrosis, essential thrombocytopenia, chronic eosinophilic leukaemia not otherwise specified and mastocytosis (Jaffe et al., 2008). These diagnoses are reliant on the exclusion of *BCR-ABL1;* if the *BCR-ABL1* fusion gene is present the MPN is, by definition, a chronic myeloid leukaemia.

1.2.3.1 Chronic myeloid leukaemia

CML was the first leukaemia to be described, almost simultaneously by Craigie, Bennet and Virchow in the 1840s (reviewed in (Goldman, 2010)) and the first human malignancy to be associated with a consistent chromosomal abnormality: the Philadelphia chromosome or Ph, as described in section **Error! Reference source not found.**. It is a distinct subtype of the more diverse group of MPN, all of which are characterised by clonal expansion of mature myeloid cells, but lack the Ph.

CML affects 1-2 per 100,000 adults, and comprises approximately 10-15% of all adult leukaemia (Siegel et al., 2012). The disease is characterised by progression through two, or sometimes three, stages with a median survival – if left untreated - of around 4 to 5 years. An initial chronic phase (CP) is relatively indolent and the patient may even be asymptomatic. Indeed, a proportion of patients are only diagnosed having undertaken routine blood tests. Symptomatic patients may present with increased white blood cell (WBC) counts of over 100 x 10^9/litre. Splenomegaly is the most common finding. When present, symptoms are mainly the consequence of splenomegaly and anaemias, such as fatigue, weight loss and malaise. Left untreated, patients inevitably progress to blast crisis (BC), akin to an acute leukaemia, with worsening symptoms plus bleeding, fevers and infections. Some patients progress to BC with no warning signals, while others progress through a transitional accelerated phase (AP) with progressive headaches, bone pain, arthralgia or fever (Swerdlow et al., 2008).

ABL1, located on chromosome 9 at band q34 is a proto-oncogene with homology to the Abelson murine leukaemia viral oncogene. It is 225Kb and expressed as a 6 or 7Kb mRNA, depending on which of two alternative first exons are spliced to the common exons 2 to 11 (Westin et al., 1982). *BCR*, found at 22q11.2, is a protein kinase with both auto- and transphosphorylation activity. Interestingly, both functions are encoded by exon 1, the only exon of *BCR* which features in all known variants of the *BCR-ABL1* translocation (Maru and Witte, 1991). A novel fusion protein, *BCR-ABL1*, with constitutive kinase activity is encoded by regions of both of these genes following chromosomal translocation.

The majority of CML patients have a somatic translocation between the long arms of chromosomes 9 and 22 - designated t(9;22)(q34;q11.2) (International Standing Committee on Human Cytogenetic et al., 2009) - while a smaller proportion have variant and/or cryptic rearrangements still leading to the same *BCR-ABL1* fusion gene (Swerdlow et al., 2008).

CML was almost always fatal until the introduction of stem cell transplantation (SCT) in the 1970s. SCT offered the only possible 'cure' for CML patients, albeit with considerable risks; one group reported an average 5 year survival rate of 57%, rising to 74% in the optimal patients under 50 years old and transplanted within a year of diagnosis (Hansen et al., 1998).

As introduced in section **Error! Reference source not found.**, the treatment of CML was revolutionised by the development of imatinib in the late 1990s. TKI therapy transformed the outlook of CML from a median survival of about 5 years to one in which more than 90% of patients are well more than 9 years later (Cortes et al., 2013; Giles et al., 2013; Jabbour et al., 2014). With the second generation of TKIs, the astonishing success of TKI therapy mean most patients diagnosed with CML in 2013 or later can expect to have a normal life expectancy (Bower et al., 2016).

Unlike other adult leukaemias, at diagnosis, the majority of CML patients have a single abnormality – the *BCR-ABL1* fusion gene. In a large study of 1151 German patients, only 6.9% were found to have additional cytogenetic abnormalities and half of these were cases with isolated loss of the Y chromosome (Fabarius et al., 2011).

1.2.4 Myelodysplastic syndromes

The myelodysplastic syndromes (MDS) have been summarised by the WHO as simply a 'clonal disease characterized by morphological dysplasia, ineffective haematopoiesis leading to cytopenias and risk of transformation to acute myeloid leukaemia' (Arber et al., 2016). In practice, MDS is a broad term which covers a diverse array of myeloid disorders, from Clonal Haematopoiesis of Indeterminate Potential (CHIP; (Jaiswal et al., 2014), to those which overlap with MPNs, to the early stages of secondary AML (sAML) which has a significantly pooper prognosis than *de novo* AML. MDS is diagnosed at a median age of 71-76 years (Sperling et al., 2016), and has an incidence of 3.4 per 100,000 in the United States (Sekeres, 2011). Notably, prior exposure to chemotherapy or radiation results in an increased risk of therapy-related MDS (tMDS) and thus cases of tMDS are increasing and will continue to do so, in parallel with improved outcomes for cancer treatment in general (Sekeres, 2011).

To date, more than 50 genes have been identified as recurrently mutated in MDS and these have diverse roles in splicing, epigenetic regulation and gene transcription (Sperling et al., 2016). This includes some rare but recurrent inherited bone marrow failure syndromes which have an inherent risk of transformation to MDS, such as *GATA2, RUNX1* and *ETV6* (Liew and Owen, 2011).

The majority of patients present with anaemia-related fatigue, thrombocytopenia-related bleeding and/or neutropenia-related fevers or recurrent infections. MDS is remarkably heterogeneous, however, and some asymptomatic patients may be detected by incidental finding of cytopenias. Similarly, some patients will live for many years with cytopenia while others will progress rapidly to sAML (Sperling et al., 2016).

Various attempts to classify MDS have been made: the International Prognostics Scoring System (IPSS), the MD Anderson Comprehensive Scorning System (Kantarjian et al., 2008) and the WHO-based

Prognostic Scorning System (Malcovati et al., 2011) have all been surpassed by the most recent revision of the IPSS: the IPSS-R (Greenberg et al., 2012). The IPSS-R system assigns a prognostic score between 0 and 4 for five categories assessed at diagnosis, including cytogenetics subtype, blast percentage, haemaoglobin, platelets and neutrophil count.

1.3 DNA damage and leukaemia

1.3.1 Genomic instability

Genomic instability – resulting in an increased frequency of abnormalities in a genome and including mutations, chromosomal rearrangements and aneuploidy - is a hallmark of cancer (Lengauer et al., 1998), and both a cause and consequence of disease progression (Vogelstein et al., 2013).

Normal cells suffer DNA damage on a daily basis. This is estimated to be in the region of 10^5 lesions per day per genome (Collins, 1999). Every cell therefore has a strict schedule of surveillance, maintenance and checkpoints in place to recognise damage, correct damage and/or to prevent a damaged cell from multiplying. DNA double strand break repair pathways include homologous recombination (HR), which employs the undamaged DNA strand as a template for accurate repair, and non-homologous end joining (NHEJ), which can re-join broken ends independently of homologous sequence. If cellular checkpoints are triggered by persistent DNA damage (i.e. unrepaired or unrepairable), however, one of two protective mechanisms is induced: a complete growth arrest called cellular senescence (Hayflick and Moorhead, 1961) or programmed cell death via apoptosis (Kerr et al., 1972). Senescent cells are characterised by an active DNA damage response, expression of senescence-associated acidic β-galactosidase and high levels of p16 (Dimri et al., 1995; Gorgoulis et al., 2005). Apoptosis is a controlled dismantling of the cellular contents with minimal damage to neighbouring cells (Kerr et al., 1972).

When a cell is subject to any kind of hyper proliferative stimuli that puts concomitant strain on DSB repair mechanisms or the checkpoints, the risk of genomic instability is increased. A cell that receives hyper proliferative stimuli and escapes a checkpoint may therefore suffer genomic instability. Many of the pathways leading to genomic instability are related or overlap.

The mechanism of genomic instability resulting in hereditary cancers and sporadic cancers is quite distinct. Inherited defects in the DNA damage repair pathways is discussed in more detail in section 1.3.3 but briefly, these result in a dramatically increased risk of developing cancer in a lifetime; cancers are often detected earlier than the same cancer in individuals without an inherited defect, and are often recurrent (Knudson, 1996). In contrast, sporadic cancers are caused by the propagation of improper repairs following DNA damage.

There are multiple sources of genomic instability; some are long-recognised and well understood, others newly recognised and still controversial. Telomere attrition was one of the first recognised causes of genomic instability. Primary cells cultured *in vitro* can only divide a defined number of times and when cells reach this number of divisions, eponymously termed the Hayflick limit (Hayflick and Moorhead, 1961), they senesce. Telomere attrition leaves the ends of chromosomes uncapped, bearing molecular markers of DNA double strand breaks, which induces a DNA damage response. This block can only be lifted by subsequent inactivation of the cell cycle checkpoint kinases (d'Adda di Fagagna et al., 2003).

A degree of oxidative stress arises from normal cellular metabolism which releases reactive oxygen species (ROS). These molecules have an unpaired electron and are thus highly reactive to DNA (Phillips, 1956). A sudden increase in the amount of endogenous ROS can be caused by the 'respiratory burst' from antimicrobial leukocytes following inflammation, ultra violet irradiation or increased apoptosis and can damage DNA in the neighbouring cells (Hussain et al., 2003).

Replicative stress results from impediments to DNA replication and includes the stalling or slowing of replication forks leading to their collapse (Bartkova et al., 2005; Gorgoulis et al., 2005). Replicative stress can be caused by deregulation of licensing and/or firing of origins of replication, which are usually tightly controlled with the cell cycle (Nishitani and Lygerou, 2002). Replicative stress has been frequently associated with breakage of so-called fragile sites, and genomic lesions associated with fragile sites are found in cancer including haematopoietic malignancies. A recently identified class of fragile sites, named early replicating fragile sites (ERFS, identified by Barlow et al. in 2013), has been shown to frequently associate with genomic alteration in B-cell lymphoma. ERFS occur at CpG dinucleotides within highly transcribed gene clusters, and accumulate spontaneous DNA damage during conditions of replicative stress (Barlow et al., 2013).

Akin to replicative stress, transcriptional stress arises from the requisite physical interaction of the transcriptional machinery and the DNA. The nascent RNA strand can hybridise to the complementary DNA strand, displacing the other DNA strand, and forming a stable three-stranded RNA:DNA hybrid called an R-loop (Aguilera and Gomez-Gonzalez, 2008). R-loop Forming Sequences (RLFS) can be predicted computationally because they contain three features: a short G-rich cluster for initiation, a non-specific linker sequence and a long, downstream G-rich elongation zone (Jenjaroenpun et al., 2015). Though tolerated transiently, the accumulation of R-loops hinders resolution and DSBs can arise as a consequence (Sollier and Cimprich, 2015).

Some endogenous sources of DNA damage are in common to both replicative and transcriptional stress. Both require unwinding of the helix, which leads to positive and negative supercoiling and concomitant

torsional stress. This can be overcome by topoisomerase via the introduction of single or double strand breaks. Torsional stress can be exacerbated by simultaneous transcription of overlapping genes or genes in close proximity (Barlow et al., 2013). Exposure of single stranded DNA during replication and transcription permits the formation of secondary DNA structures such as hairpins (Chou et al., 2003), triplexes (Agazie et al., 1996) and G-quadruplexes (Huppert and Balasubramanian, 2005), particularly when the single stranded DNA has repetitive sequence. These non-B DNA structures correlate with double strand breaks and thus chromosomal rearrangements (Zhao et al., 2010).

1.3.2 Genomic instability in human leukaemia

It is now clear that solid tumour genomes have a much higher frequency of abnormalities than the genomes of leukaemia; Mullighan and colleagues (Mullighan et al., 2007) found an average of 6.46 somatic copy number alterations per genome in a cohort of 242 paediatric acute lymphoblastic leukaemias (ALLs), in contrast to the data compiled by Vogelstein et al (Vogelstein et al., 2013) which showed ~50 abnormalities per genome in common adult solid tumours such as ovarian cancer, ~150 in cancers resulting from exposure to mutagens such as lung cancer, and over 500 in tumours with DNA repair defects.

Leukaemias have historically been the most well studied group of cancers, since it is relatively easy to access and culture tumour tissue, and it has thus long been observed that genes that are frequently rearranged by chromosomal abnormalities tend to be affected in a lineage-specific fashion. Examples include the dic(9;12) in ALL which was first reported in 1987 (Carroll et al., 1987), in which *PAX5* was implicated in 1992 (Behrendt et al., 1995) and in which involvement of *PAX5* was confirmed to define the dic(9;12) entity in 2003 (Strehl et al., 2003). Similarly, the eponymously titled *Acute Myeloid Leukaemia 1* gene (*AML1*, now known as *RUNX1*) was cloned from a series of 16 patients with t(8;21) AML (Miyoshi et al., 1991). A rare exception to this rule is the gene, *Mixed Lineage Leukaemia* (*MLL,* now known as *KMT2A*), located at 11q23, which earned its name as a consequence of being found to be rearranged in a subset of both myeloid and lymphoid leukaemia patients, including those who were diagnosed with malignancy in one lineage and relapsed with disease affecting the other (Mirro et al., 1986a; Mirro et al., 1986b; Mirro et al., 1985).

Mullighan et al used single nucleotide polymorphism (SNP) arrays to investigate copy number changes in the cohort of paediatric ALLs described above to identify the cooperating mutations which induced leukaemia in patients with translocations or recognised cytogenetic subtypes such as hyper- and hypodiploid karyotypes (Mullighan et al., 2007). The group found infrequent gains and frequent deletions, but most notably, they found that 40% of the abnormalities affected the genes involved in the regulation of B-cell differentiation. The authors undertook extensive validation to demonstrate that the abnormalities were specific to the relevant gene (some deletions were confined to the gene body),

pathogenic (for example, by using luciferase-based reported assays of mutant proteins to assess their activity) and somatically acquired (by assessing paired copy number and loss of heterozygosity in matched remission samples where available). Overall, the mutations identified resulted in demonstrably lower overall levels of B-cell transcription factors, or the production of altered forms, many of which have since been demonstrated to result in dominant negative activity (Kurahashi et al., 2011). Subsequently, the group established that deletions of *IKZF* were present in 84% of *BCR-ABL1* –positive ALL but not in chronic myeloid leukaemia (CML; a chronic leukaemia with the same *BCR-ABL1* fusion gene) (Mullighan et al., 2008) and that they conferred a poor prognosis (Mullighan et al., 2009b) further emphasising a lineage specific pattern of lineage aberrations.

Following its identification in the dic(9;12) the B-cell transcription factor *PAX5* (paired box 5) was found to be essential for B-lineage commitment and maintenance (Nutt et al., 1999). Underlining its key role in B-cell development, *PAX5* was found to be subject to copy number changes (monoallelic loss, biallelic loss or internal amplification) and/or mutations in 31.7% of B-progenitor ALLs (Mullighan et al., 2007). All of these aberrations were predicted to result in loss of *PAX5* function or altered DNA binding capacity.

Similarly, the myeloid transcription factor *RUNX1* (runt-related transcription factor *1*) is mutated in 9% of acute myeloid leukaemia (AML) and is translocated to contribute half of the chimeric fusion protein *RUNX1-RUNX1T1* (arising from the translocation t(8;21)) in a further 12-15% (Speck and Gilliland, 2002). The fusion protein retains DNA-binding capabilities and thus acts as a dominant negative inhibitor of the wild type protein.

An association between the specific developmental importance of particular genes and the presence of oncogenic mutations within those genes in haematological neoplasms of a related lineage has subsequently been reported by other groups. For example, Chan et al used chromatin immunoprecipitation and RNA sequencing to identify a novel pathway regulating glucose, and thus the energy supply, available to developing lymphoid cells; *PAX5* and *IKZF* normally repress the pathway to ensure cellular levels remain insufficient for malignant transformation but dominant negative mutants of the same transcription factors and their transcriptional targets (such as *TXNIP*) bypassed this crucial block and resulted in cellular levels of ATP which permit transformation (Chan et al., 2017).

Somewhat counterintuitively, some DNA damage is *essential* for normal cell maturation: programmed double strand breaks, induced by carefully regulated enzymes, are essential for normal B cell development following activation. Activation-induced deaminase (AID) is required for both class switch recombination (CSR) and somatic hypermutation (SHM) during maturation of B-cells (Muramatsu et al., 2000). AID deaminates cytosine to uracil in single stranded DNA (Bransteitter et al., 2003) which is undergoing active transcription (Chaudhuri et al., 2003); the resulting U:G mismatch can be processed

by any one of several different pathways all of which frequently cause a mutation. Alternatively, the mismatch may remain unrecognised; in which case replication proceeds obliviously resulting in a simple C→T transition in one daughter strand. If the mismatch is recognised by uracil-DNA glycosylase it is excised; the resulting abasic site can be filled by a random nucleotide or cleaved to form a double strand break (Di Noia and Neuberger, 2007). Lastly, if the mismatch is recognised by DNA mismatch repair machinery it can be repaired, albeit by error-prone polymerases. Generally, AID mismatches in Ig switch regions are processed to form double strand breaks and result in recombination and deletion which facilitate CSR (Chaudhuri et al., 2003) whilst mismatches in the variable exons undergo error-prone repair resulting in an accumulation of point mutations which increase antibody affinity by SHM (Di Noia and Neuberger, 2007). The activity of AID is generally restricted to these precise regions of the immunoglobulin locus in germinal centre B-cells. However, AID activity is recognised in non-Ig genes, such as BCL6 and MYC, for example (Ramiro et al., 2004). Here, the U:G mismatches can be repaired by base excision or mismatch repair pathways which have a much higher fidelity (Ramiro et al., 2006). AID is essential for the formation of some B-cell lymphoma translocations and the activity of this enzyme is thus a considerable risk to genomic stability (Ramiro et al., 2004). AID has been further associated with leukaemia as oncogenic and inflammatory stimuli can induce aberrant AID expression in malignant haematopoietic cells (Feldhahn et al., 2007; Klemm et al., 2009; Swaminathan et al., 2015).

Likewise, the lymphocyte-specific endonucleases encoded by recombination activation genes (RAG1 and RAG2) have been implicated as a potential cause of genomic alterations found in human leukaemia. These endonucleases, whose function is to introduce diversity by somatic recombination during the tightly controlled development of lymphocytes, target the V(D)J sites via recombination signal sequence (RSS) motifs. The RSS motifs comprise one highly conserved heptamer (CACAGTG) separated by 12 or 23bp from a less conserved nonamer (ACAAAAACC) (Fugmann et al., 2000). The RAG endonucleases generate two blunt ends and two hairpin ends by binding to the RSS and cutting between it and the flanking sequence. Crucially for the development of a healthy immune system, the resolution of the two hairpin ends by terminal deoxynucleotidyl transferase (TdT) is imperfect, and introduces non-template base pairs which further increase diversity in the region (Komori et al., 1993). However, RAG can also be aberrantly recruited by non-B DNA structures (Raghavan et al., 2005), by the presence of cryptic (but functional) heptamers and nonamers outside a conserved RSS motif (Zhang and Swanson, 2008) and by deaminated methyl CpGs (Tsai et al., 2008).

So whilst lineage specific abnormalities have been observed at the cytogenetic level for some time, and have been refined to lineage-specific molecular lesions in the post-genomic era, significantly, the mechanisms underlying the aetiology and pathogenesis of these aberrations was lacking until recently. In addition to the work shown here (Boulianne, Robinson and May et al., 2017 *Cell Reports*), during the

course of this PhD thesis a Finnish group used global run on sequencing (GRO-seq) to determine that convergent transcription and PolII stalling can be found at breakpoints of genomic alterations in B-ALL (Heinaniemi et al., 2016). Notably, however, their models associated DNA damage with potential targeting by RAG and AID. The laboratory of Eva Petermann (UK) further established that replicative stress upon *CYCLIN E* or *HRAS* oncogene expression in fibroblasts (and associated DNA damage) requires transcription of the affected loci (Jones et al., 2013; Kotsantis et al., 2016).

1.3.3 Inherited predisposition to DNA damage in leukaemia

Important insights into types of DNA damage and how they cause leukaemia were first made in inherited chromosome breakage syndromes. Fanconi anaemia is an autosomal recessive bone marrow failure syndrome caused by an inability to repair DNA crosslinks. Individuals with Fanconi anaemia are predisposed to various types of cancer but predominantly suffer from AML (Alter, 1996). Fanconi anaemia can be caused by any of at least eight different complementation groups, indicating the complexity of the pathways involved in homologous recombination and the vulnerability to mutation that a deficiency in any one of them will convey (D'Andrea and Grompe, 2003).

Similarly, Bloom syndrome is an autosomal recessive disorder caused by mutations in the *BLM* gene, a RecQ helicase which facilitates the unwinding of DNA for various forms of DNA repair (German, 1993). Patients lacking *BLM* undertake an excessively high level of sister chromatid exchange to compensate for the lack of proper DNA repair mechanisms and as a consequence have a high prevalence of various cancers, predominantly leukaemia and lymphoma, at a young age (Cunniff et al., 2017).

Acute lymphoblastic leukaemias lacking subtype defining recurrent translocations have long been grouped together based on their ploidy. Hypodiploid ALL (those with less than 46 chromosomes), for example, has a poor prognosis in childhood ALL (Pui et al., 1987) Recently, however, genomic profiling of 124 hypodiploid ALL cases revealed that one specific subset – namely the low hypodiploid ALL (those with 31-39 chromosomes) were characterised by germline, and not somatically acquired, *TP53* mutations. These patients are thus presenting with manifestions of Li-Fraumeni syndrome and not *de novo* leukaemia as in all the other sub-types (Holmfeldt et al., 2013).

1.3.4 DNA damage signatures

Recent large scale analyses of various cancers (in particular, the work of the Catalogue of Somatic Mutations in Cancer (COSMIC) groups) including leukaemia have revealed distinct ‘mutational signatures’ which hint at the particular cause of damage, enzymatic modification and/or type of deficient DNA repair (Alexandrov and Stratton, 2014).

The simplest signature, which is found in all cancers including leukaemia, demonstrates clock-like behaviour in that the number of mutations correlates with the age at which cancer is diagnosed, suggesting a stepwise accumulation of mutations (Alexandrov et al., 2015). This has been proposed to be due to the spontaneous deamination of 5-methycytosine which triggers transition mutations (Pfeifer, 2006).

Other signatures are more tightly defined. For example, 25% of paediatric B-ALLs have the *ETV6-RUNX1* fusion gene as a consequence of the t(12;21) chromosomal translocation. It is now recognised that the translocation occurs *in utero* (Greaves and Wiemels, 2003), and can be detected in archived neonatal samples of individuals who do not progress to an overt leukaemia, thus a further cooperating event is required (Mori et al., 2002). In those cases which progress to overt leukaemia, a characteristic mutational profile is consistent with aberrant RAG activity – clustering of breakpoints near the recombination signal sequences and enriched in promoter and enhancer regions, together with the motifs suggestive of TdT activity (Papaemmanuil et al., 2014). This ongoing RAG activity is due to aberrant expression as a consequence of the developmental arrest induced by the *ETV6-RUNX1* fusion gene (Hübner et al., 2004).

A stepwise accumulation of mutations leading to progressively more aggressive cancer is the generally accepted mechanism of cancer development. Contrary to this is a phenomenon termed chromothripsis, first detected in chronic lymphocytic leukaemia, and characterised by an isolated 'chromosome shattering' event in which tens or hundreds of DSBs affect a small number (often just one) of chromosome regions or chromosomes (Stephens et al., 2011). Unusually, this damage occurs in an isolated event, and to survive such catastrophic damage, the cell must repair the fragmented chromosome region quickly. This results in multiple rearrangements (deletions, tandem duplications and inversions) that oscillate between two copy numbers reflecting the ploidy of the cell (usually 2n and thus rearrangements are heterozygous or show loss of heterozygosity).

1.4 Epigenetics and leukaemia

Epigenetic modifications shape the genome; chromatin can be heterochromatic or euchromatic. Heterochromatin is transcriptionally inert and densely compacted whilst euchromatin is much more loosely compacted, accessible and thus transcriptionally active (Jenuwein and Allis, 2001; Luger et al., 1997b).

1.4.1 Histones and nucleosomes

In both humans and mice, compaction of DNA into chromatin is facilitated by the wrapping of 146bp of DNA around a nucleosome (Pruss et al., 1995) formed from eight histone proteins (Luger et al., 1997a). There are four core histone families - H2A, H2B, H3 and H4 – contributing two proteins each. The

histone families are encoded by several different genes and these genes are usually expressed in synchronisation with replication (Heintz, 1991). A second level of compaction is mediated by a linker histone H1 between the nucleosomes (Widom, 1989). The nucleosome is present approximately every 200bp throughout the genome and is the main factor that dictates accessibility of DNA to transcription machinery (McGhee and Felsenfeld, 1980). The H2B, H3 and H4 histones differ only slightly from each other (Baxevanis and Landsman, 1996) whereas histone H2A has evolved into three distinct subfamilies called H2A1-2, H2AZ and H2AX (Thatcher and Gorovsky, 1994; West and Bonner, 1980). Histone proteins can be modified during the cell cycle (Bradbury et al., 1974; Gurley et al., 1978).

1.4.2 Posttranslational modifications

A histone consists of a globular domain and a flexible tail which protrudes from the nucleosome and this tail is subject to various post translational modifications. These modifications include acetylation, methylation, and phosphorylation (Wang et al., 2007). The combination of modifications dictate the conformation of chromatin in the region; an epigenetic code, comparable to the genetic code. This code is read, written and erased by enzymatic proteins which triggers transcriptional activity or DNA damage repair, for example (Jenuwein and Allis, 2001). Histone modification can be permanent (termed imprinting), as is the case for X chromosome inactivation, or transient.

1.4.3 The role of γH2AX in DNA damage response

Following lethal and sub-lethal doses of ionising radiation to cultured mammalian cells, novel compounds (collectively referred to as gamma) were noted in the H2AX region of two-dimensional gels. This was identified as specific phosphorylation at serine 139 of the H2AX sub-family of H2A histone family [subsequently termed γH2AX] (Rogakou et al., 1998). This serine is present in the C-terminal motif of H2AX, and it is the sequence of the C-terminal motif which distinguishes H2AX from the other members of the H2A protein family. It is phosphorylated quickly and accurately in response to the DNA damage caused by ionising radiation: maximal amounts of γH2AX were noted 10 minutes post-irradiation and half of this after just 1 minute. A dose of 1Gy induced the phosphorylation of about 1% of H2AX, or 2Mb per DSB, indicating there is considerable ‘spread’ of this histone modification in response to DNA damage (Rogakou et al., 1998).

It was later established that γH2AX is also induced by controlled V(D)J recombination in maturing thymocytes (Chen et al., 2000), by meiotic recombination (Mahadevaiah et al., 2001), by ultraviolet irradiation and by inhibition of replication by hydroxyurea (Ward and Chen, 2001) indicating that γH2AX plays an important role in a more general DNA damage response. The mechanisms governing the removal of γH2AX from the chromatin following repair completion are yet to be understood; it is unclear if the trigger is the repair itself or the return to the standard DNA compaction (Mah et al.,

2010).

Phosphorylation of H2AX has therefore been used experimentally to successfully identify regions of DNA damage including early replicating fragile sites (Barlow et al., 2013), DNA damage hotspots (Seo et al., 2012) and alternative DNA structures formed by DNA-drug interactions (Rodriguez et al., 2012). Petermann et al. (Petermann et al., 2010) reported that γH2AX co-localises with a single stranded DNA binding protein called replication protein A (RPA) and this has been confirmed experimentally by others; Barlow et al. (Barlow et al., 2013) confirmed that 93% of their γH2AX ChIP-seq regions overlapped with RPA ChIP-seq regions.

Epigenetic abnormalities have been targeted by novel therapies. The chromatin reader BRD4 binds to acetylated histones and recruits further co-activators. This binding can be inhibited by bromodomain inhibitors, such as JQ1 and iBET, which have had some success in various haematological malignancies, such as multiple myeloma (Delmore et al., 2011)

1.4.4 Promoters, enhancers and acetylation of H3K27

The genomic sequence of the core promoter - in the area around the transcription start site - is sufficient to recruit and assemble the transcriptional machinery necessary for transcription to begin. However, for transcription to occur at the necessary level, additional epigenetic input is required from regulatory regions as depicted in (Figure 1-2) (Shlyueva et al., 2014). Chromatin can be either closed or open to DNA binding proteins, as dictated by the density of nucleosomes to which it is bound (Figure 1-2a). Core promoter sequences (Figure 1-2b) have open chromatin and are flanked by nucleosomes with H3K27ac and H3K4Me3 modifications (Heintzman et al., 2007). The modifications of promoters are largely invariant across different cell types (Heintzman et al., 2009).

Identification of enhancers is much more complex by comparison - nucleosomes are dynamic and can partially disassemble and move (Narlikar et al., 2013; Zhou et al., 2016). Enhancers are frequently present in the intergenic regions of the genome which are much less studied; a meta-analysis conducted in 2009 of 1,170 single nucleotide polymorphisms that had been reported as highly significant by various genome wide association studies showed that in 472 (40.3%) cases, the SNP was not overlapped by an exon or an associated haplotype block (Visel et al., 2009). Importantly, however, an enhancer cannot be reliably predicted from the sequence or chromatin architecture alone, so a combination of approaches is required.

The first reported enhancer – a DNA sequence which could increase the rate of transcription from a target gene due to transcription factor binding sites – was a 72bp sequence derived from the SV40 virus which was found to increase the transcription of a beta globin reporter gene in HeLa cells by several

hundred fold (Banerji et al., 1981). The same group later found the first animal enhancer: a lymphoid specific enhancer of the immunoglobulin heavy genes located downstream of the gene (Banerji et al., 1983).

Active enhancers are typically found in open stretches of DNA which lack nucleosomes (sometimes termed a nucleosome-free (NFR) or nucleosome-depleted (NDR) region), and unlike the histone modifications affecting promoters, are affected by cell-type-specific modifications which correlate with cell-type-specific gene expression (Heintzman et al., 2009). The neighbouring nucleosomes (see Figure 1-2b) are likely to have histone modifications such as mono-methylation of histone H3 lysine 4 (H3K4me1) and acetylation of histone H3 lysine 29 (H3K27ac) at their N-terminus (Creyghton et al., 2010). Active enhancers can also function at great distance from their target promoter by looping out of the DNA – a process mediated by cohesin. They function independent of context (hence the experimental reporter gene constructs are functional) and the effect of an enhancer is additive.

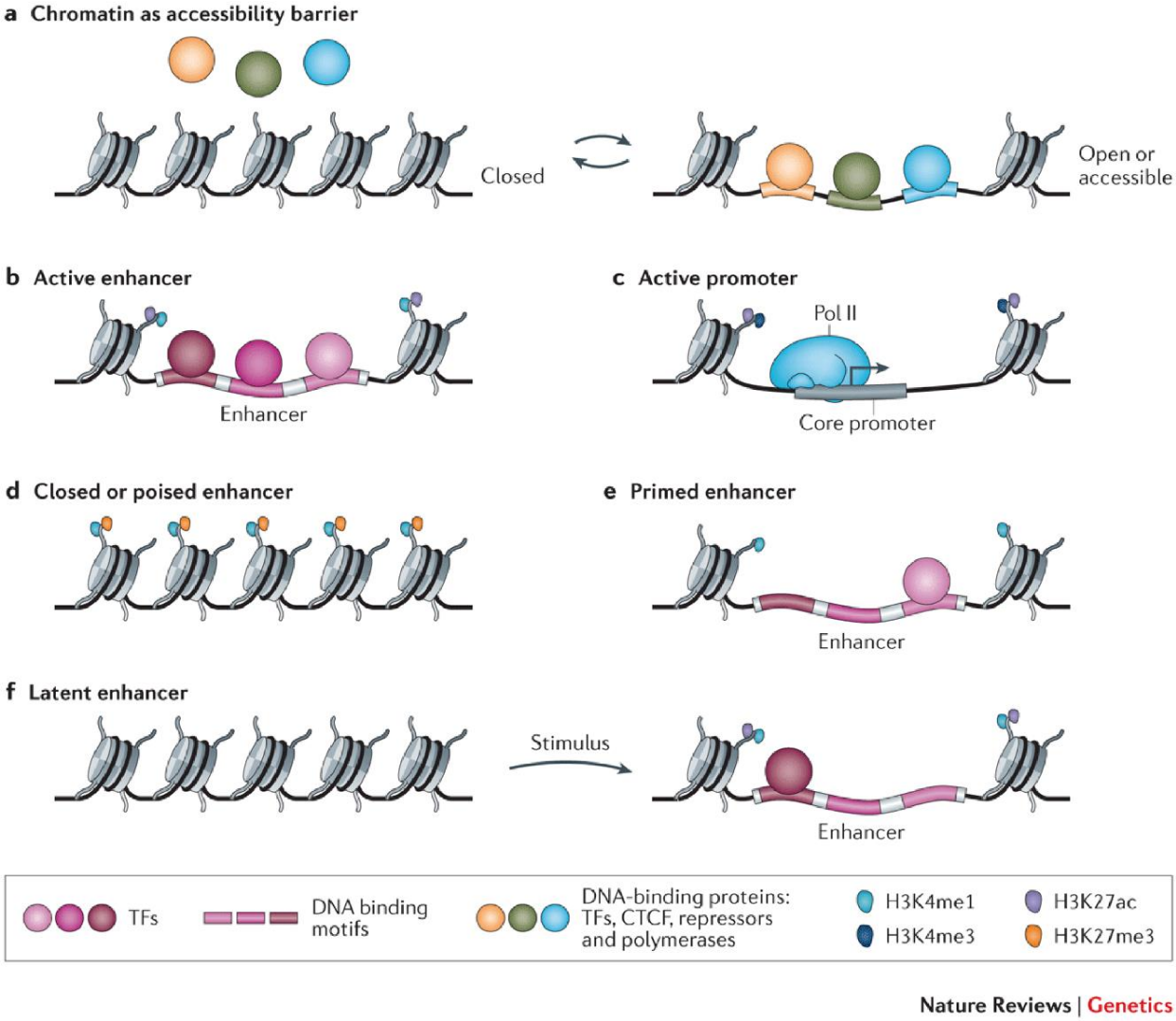


Figure 1-2: Histone marks affect chromatin architecture.¶ (A) Closed chromatin with regular nucleosomes is inaccessible to DNA binding proteins whereas open chromatin has no nucleosomes and can thus be bound by transcription factors, repressors, polymerases and CCCCTC-binding factor. (B) An active enhancer is bound by transcription factors and adjacent nucleosomes show H3K4me1 and H3K27ac. (C) Nucleosomes flanking an active promoter are bound by H3K27ac and H3k4me3 modified nucleosomes. (D) A closed/poised enhancer shows a mix of activating H3K4me1 and repressing H3K4me3 modified nucleosomes. (E) Enhancers that need to be ready to respond to a developmental stage or external stimuli are 'primed' with H3K4me1, and similarly, latent enhancers (F) have closed chromatin wrapped round unmodified nucleosomes which can be easily modified with activating marks as required (reproduced with permission from Nature Reviews Genetics (Shlyueva et al., 2014)).

Enhancers identified in human embryonic stem cells have also been found in alternative inactive states. These are termed poised enhancers (Figure 1-2d) and have open chromatin and both H3K4me1 and H3K27me3 modifications, but lack H3K27ac. Poised enhancers acquire the full range of active enhancer modifications during the appropriate developmental stage of embryogenesis (Rada-Iglesias et al., 2011). Similarly, primed enhancers (Figure 1-2e) can be distinguished from their active equivalents due to a lack of H3K27ac, even though they may have open and accessible chromatin (Creyghton et al., 2010). Both of these inactive enhancer states allow rapid and flexible response of enhancer elements (Creyghton et al., 2010). It also suggests that the number of functional enhancers is much lower than was predicted by analysis of H3Kme1 modifications; conversely, it is recognised that transcription can occur in the absence of H3Kme1, which would increase the number of enhancers (Hödl and Basler, 2012).

Latent enhancers (Figure 1-2f) were described for the first time in 2013 and, unusually, lack the classic enhancer histone modifications in terminally differentiated cells but acquire them following stimulation. When stimulation subsides, the enhancer does not return to the inactive state but rather persists and is thus able to respond to the same stimulation even faster subsequently (Ostuni et al., 2013).

Most active mammalian enhancers are transcribed bidirectionally to produce enhancer RNAs (eRNAs), a novel category of non-coding transcription. Recognised as early as 1979 (Boseley et al., 1979) but not really understood until the post-genomic era (Natoli and Andrau, 2012), eRNA transcription is now understood to be a regulated process and not simply transcriptional noise. Remarkably, some enhancers can recruit PolII when inserted on to a plasmid in the absence of a promoter (Moreau et al., 1981). A minority of eRNAs have a defined function: one example is the eRNA capable of recruiting (or 'trapping') the Yin-Yang transcription factor (Sigova et al., 2015). Others are presumed to have *cis* acting effect by opening chromatin caused by the action of transcription itself. The study of eRNAs is complicated by their rapid turnover (Pefanis et al., 2014).

1.4.5 Super enhancers

It was established in *Drosophila* that housekeeping genes (whose function is required across lineages and throughout development) have simple promoters that require minimal enhancement to be transcribed. Genes that must be transcribed in a more reflexive way, such as those involved in development or in response to a signal, have promoters with inactive resting states, but which are responsive to multiple enhancer elements (Zabidi et al., 2015). This mechanism has redundancy which therefore offers flexibility – each enhancer can respond to different transcription factors with the same end result (Goto et al., 1989).

In certain cell types, enhancers can also function in concert – known collectively as a 'super-enhancer' – to drive very high expression of a gene necessary for lineage differentiation, and are responsible for the lineage plasticity necessary for stem cells to retain their 'stemness' (Adam et al., 2015).

1.4.6 Cancer-specific enhancers

Typical cancer mutations often deregulate enhancers controlling growth-related genes, freeing them from the constraints of signal dependency. This can happen in one of two ways; first, trans-acting factors such as global epigenetic changes affecting accessibility of chromatin or the aberrant activation of transcription factors can lead to increased enhancer activity – for example, the *KLF5* transcription factor is frequently found to be amplified in lung cancer and promotes tumorigenesis. In a small number of lung cancers without focal amplification of the gene, point mutations in the *KLF5* DNA binding domain increase the DNA binding specificity allowing it to bind to novel binding sites and activate new genes (Zhang et al., 2018). Secondly are cis-acting factors which affect an insulator or an enhancer itself, such as the recurrent mutations which introduce transcription factor binding sites in the intronic region of *LMO* in T-ALL (Rahman et al., 2017).

Enhancers, just like the gene expression they control, are frequently lineage specific. Transcriptional deregulation underlies malignancy and enhancers are therefore vulnerable to 'hijacking' (Northcott et al., 2014). The various *MYC* enhancers found in cancers of different lineages illustrate this point well – an enhancer 1.47Mb downstream of *MYC* is found only in T-ALL (Herranz et al., 2014); another 90Kb upstream is specific to medulloblastoma (Lin et al., 2016); another 1.7Mb downstream is specific to B-ALL (Katerndahl et al., 2017); another 450Kb downstream is specific to lung adenocarcinoma and finally, 800Kb downstream is another specific to uterine corpus endometrial carcinoma (Zhang et al., 2016). Many of these are subject to focal amplification in the corresponding malignancies.

A recent study used H3K27Ac and PolII ChIP-seq to undertake enhancer profiling of primary adult T-cell leukaemia/lymphoma (ATLL) samples. They found enrichment of super enhancers responsible for T-cell activation, and at several recognised cancer genes in the ATLL samples (but not mature T-cells). Most interestingly, however, they found an enhancer associated with *TIAM2*, which has not previously been reported in ATLL, and promotes growth and survival of ATLL cells. Knockdown of the gene has an inhibitory effect on growth primarily due to apoptosis (Wong et al., 2017). This example illustrates that there are still genes involved in leukemogenesis which have not yet been identified by traditional genetic or genomic methods.

1.4.7 Technical approaches to identifying enhancers

The varying properties of an enhancer described above also means there are various different

techniques which can be used to locate enhancers but unfortunately no single technique is capable of identifying all.

The simplest is *in vivo* identification of transcription factor binding sites using ChIP-seq (Hallikas et al., 2006). Another, as described in section 2.4.4, is ChIP-seq using antibodies to modified histones. Bisulphite sequencing (which exploits the methylation of cytosine to protect against conversion to uracil) can be used to investigate the methylation status of regulatory sequences (Fraga and Esteller, 2002).

A more general approach is to identify the accessibility of chromatin. This can be undertaken using DNase-seq – exposed chromatin is hypersensitive to digestion whereas nucleosome compaction is protective – or variants such as FAIRE-seq (Song et al., 2011) and ATAC-seq. These techniques rely on the preferential extraction of exposed DNA and transposase-accessibility, respectively. It is also possible to predict enhancer sequences using computational methods relying on conservation of non-coding sequences (Pennacchio et al., 2006).

Evidence to date, compiled using all of the above methods would suggest that active enhancers have loose, accessible chromatin which is free from nucleosomes, and has hypomethylated CpG dinucleotides (Stadler et al., 2011). The enhancer regions are flanked by histones which have H3K27 acetylation and H3K4 monomethylation but not H3K4 trimethylation (Yukawa et al., 2014).

Understanding the distribution and composition of enhancers does not end with their identification, however. Not all enhancers regulate neighbouring gene promoters; some mammalian promoter-enhancer interactions are known to span over a megabase, such as the *MYC* super enhancer which is 1.7Mb away from its promoter. To define these interactions, it may be necessary to use chromatin interaction analysis by paired-end tag sequencing (ChIA-PET; (Fullwood et al., 2009)) or chromatin conformation capture sequencing (3C; (Dekker et al., 2002)) and variants. Establishing a model by which novel enhancers are formed during transformation is therefore of crucial importance in our understanding of leukemogenesis.

1.4.8 DNA methylation

Methylation of the C5 cytosine in CpG dincucleotides is established and maintained by DNA methyltransferases (Bird, 2002), and once in place, read by the methyl CpG binding proteins (Robertson, 2005). The transcription of methylated DNA is generally repressed, due to inhibition of transcription factor binding and due to the binding of the methyl CPG binding proteins, hence promoter regions are often unmethylated. Methylated DNA is frequently found in repetitive regions of satellite DNA and interspersed repeats (Song et al., 2005).

In 1983, it was discovered that cancer genomes are hypomethylated compared to their non-malignant counterparts (Feinberg and Vogelstein, 1983a). This hypomethylation can lead to activation of gene expression (Feinberg and Vogelstein, 1983b), to a mutator phenotype due to deficient mismatch repair, and to chromosomal instability as a consequence of unstable heterochromatin, particularly of the pericentromeric regions (Qu et al., 1999).
In CML, for example, there is a progressive methylation of the non-translocated *ABL1* allele from undetectable at diagnosis to almost inevitable at progression to blast crisis (Zion et al., 1994), and this phenomenon is unique to the p210 form and not the p190 (Shteper et al., 2001).

The study of epigenetic abnormalities in ALL has so far been mostly limited to DNA methylation; early studies were limited to a handful of genes already known to be of interest, such as *MDR1*, *ABL1*, *p15*, *p16, p73* and CD10 and were mostly undertaken by one group (Garcia-Manero et al., 2002; Garcia-Manero et al., 2003). Later studies used genome wide techniques which uncovered aberrant methylation of multiple CpG was associated with a poorer overall survival (Kuang et al., 2008).

1.5 Novel oncogenic super enhancers

1.5.1 The TAL Super Enhancer in T-cell acute lymphoblastic lymphoma

In 2014, Mansour and colleagues discovered a *TAL1* super enhancer in T-cell acute lymphoblastic leukaemia (Mansour et al., 2014); this was the first time the genetic mechanism responsible for the generation of oncogenic super enhancers had been identified. In around one quarter of patients with T-cell acute lymphoblastic leukaemia (T-ALL) a *STIL-TAL1* fusion gene is formed by a cryptic deletion of approximately 80Kb from the intervening sequence between the ubiquitously expressed *STIL* gene (*SCL/TAL1 Interrupting Locus*) and the *TAL1* transcription factor (*T-cell Acute Lymphocytic leukaemia 1*) genes. The fusion results in overexpression of the latter gene's coding sequence due to proximity of the former's regulatory sequences (Aplan et al., 1990; Breit et al., 1993; Brown et al., 1990).

Interestingly, however, further subgroups of patients have no detectable *STIL-TAL1* fusion but do have monoallelic overexpression of *TAL1* (Ferrando et al., 2004; Leroyviard et al., 1994). The authors used H3K27Ac ChIP-seq to analyse the Jurkat T-ALL cell line and uncovered a broad and strong H3K27Ac mark 7.5kb upstream of the *TAL1* gene, which was not present in various other cell types and lines, including other T-ALL lines which had no *TAL1* expression (Sanda et al., 2012). This followed the finding of a chromatin interaction by chromosome conformation capture (3C) sequencing which was reported by a second group (Zhou et al., 2013b). Sequencing of the region in the Jurkat cell line and subsequently in T-ALL patients revealed a variable, heterozygous insertion which introduced a de novo binding motif for the MYB transcription factor. To confirm the de novo binding site was responsible for the *TAL1* overexpression, the authors then cloned 400bp fragments from both the mutant and wild type alleles and tested them in a reporter assay (Mansour et al., 2014). The findings of this group showed that de novo acquisition of mutations in intergenic sequence can result in overexpression of oncogenes in cytogenetically normal leukaemias.

1.6 A candidate for novel oncogenic enhancement

EVI1 is a nuclear transcription factor located at 3q26.2. Its aberrant expression in myeloid leukaemias and an emerging number of solid tumours confers a universally poor prognosis but – akin to expression of *TAL1* in the example above - the mechanism by which this aggressive oncogene is aberrantly expressed is only understood in a small proportion of cases.

The first reported cytogenetic abnormality of 3q (i.e. the long arm of chromosome 3) in a human leukaemia was published in 1976 (Rowley and Potter, 1976), however, the affected genes were not known at this time. In 1988, Mucenski et al. found a common site of provirus integration in AKXD-23 mice developing exclusively myeloid tumours and the locus was thus termed the *Ecotropic Viral Integration site 1* or *EVI1* (Mucenski et al., 1988). In 1992 it was discovered that *EVI1* expression blocks differentiation (Morishita et al., 1992a) and that the previously observed 3q rearrangements in human leukaemia also resulted in activation of *EVI1* (Morishita et al., 1992b). Later, in 1996, it was established that the locus actually produced three distinct transcripts in normal tissues: the *EVI1* transcript, the Myelodysplastic *Syndrome 1* (*MDS1*) transcript, and a normal fusion transcript termed the *MDS1 and EVI1 COMplex* (*MECOM*) (Fears et al., 1996; Soderholm et al., 1997). The whole locus, encoding all of the variants above, was hence renamed *MECOM*.

Translation of *EVI1* starts in exon 3 but interestingly exon 2 is highly conserved (>90% between human and mouse) whilst exon 1 is variable. Exon 2 is only translated when part of

the longer *MECOM* transcript. Excluding exon 1, there is 91% nucleotide and 94% amino acid conservation between human and mouse (Morishita et al., 1990b).

As shown in Figure 1-3, the EVI1 protein contains two zinc finger binding domains; the first (domains 1-7) binds to TGACAAGATAA and the second (domains 8-10) to GAAGATGAG (Funabiki et al., 1994; Perkins et al., 1991a). Interestingly, the first of these consensus sequences is also bound by the GATA1 transcription factor (Kreider et al., 1993) which plays a crucial role in the normal development of haematopoiesis. It was initially thought that aberrantly expressed EVI1 outcompeted GATA1 for these binding sites, resulting in repression of GATA1 targets and blocking of differentiation; however, more recent studies showed that EVI1 actually interacts with the GATA1 protein directly (Laricchia-Robbio et al., 2006). The locus codes for three different proteins – the standard EVI1 protein (145kDa), a truncated version with deletion of zinc fingers 6 and 7 from the first zinc finger domain (EVI1-Δ324) and the full length MDS1-EVI1 protein with an additional positive regulatory domain (Maicas et al., 2012). Intriguingly, a Spanish group showed that EVI1 binds to its own promoter, and that the EVI1 isoform activates transcription but the Δ324 and MECOM isoforms repress it (Maicas et al., 2017).

Figure 1-3: The MECOM protein has an additional PR domain that is not present in the EVI1 protein. ¶Summary of the protein domains present in MECOM and EVI1. The MDS1 exons (shown in red) contribute a positive regulatory (PR) domain to the MECOM protein only; EVI1 exons (shown in blue) contribute 10 zinc finger domains (numbered; zinc fingers 6 and 7 are not present in the Δ324 isoform), a repressor domain (RD) and an acidic domain (AD). Adapted from (Maicas et al., 2012).

There is currently insufficient evidence from mouse models to confirm whether expression of *EVI1* alone is capable of inducing leukaemia, since the data is contradictory. Multiple groups have forced expression via retroviral transduction and transplanted the transduced cells back into irradiated mice, but the results are inconsistent, with mice in some studies developing MDS alone (Buonamici et al., 2004), some developing AML (Yoshimi et al., 2011), and some requiring co-expression of other genes such as *Hox9/Meis1* (Jin et al., 2007) or *RUNX1* (Cuenco and Ren, 2004) for frank AML.

Interestingly, a polymorphism downstream of *MECOM*, rs2201862, has been associated with a predisposition to myeloproliferative disorders in two recent genome wide association

studies (Tapper et al., 2015; Trifa et al., 2018). The T/C polymorphism is 153Kb downstream of *MECOM* and overlaps with two long intergenic non-coding RNAs and resulted in an overall risk of 1.4 (CI 1.1-1.8; p=0.0005) in the latter study.

1.6.1 *EVI1* is a developmental regulator and its expression is tightly controlled during development

Notably, *EVI1* is only expressed at very low levels in most normal adult tissues. Studies of murine development using in situ hybridisation and Northern blotting have shown that *EVI1* expression is at its highest between 9.5 and 12.5 days (Perkins et al., 1991b).

Knock out studies in mice have yielded results in agreement with this, with homozygous knockouts having largely normal development until 8.5 days post coitum, but soon after various malformations became apparent. Death occurred by 11.5 days post coitum (Hoyt et al., 1997). There are just two reports of monoallelic deletions of *EVI1* found in humans; both patients had congenital thrombocytopenia and dysmorphism (Bouman et al., 2016; Nielsen et al., 2012). Inappropriate expression in the bone marrow inhibits correct differentiation of the erythroid lineage (Kreider et al., 1993) and causes anaemia, dyserythropoiesis and dysmegakaryopoieses (Laricchia-Robbio et al., 2006).

Gerhardt et al established that *EVI1* expression is detectable in both CD34- and CD34+ cells in samples from patients with juvenile myelomonocytic leukaemia and in healthy donors, but is primarily expressed in the CD34+ cells (Gerhardt et al., 1997).

There is some evidence that *EVI1* and the full length *MECOM* have an antagonistic relationship and that *MECOM* expression is repressed by that of *EVI1* (Soderholm et al., *1997)*. However, not at all studies have agreed with this finding. Haas et al found that expression of any of the transcripts – including *MECOM* – conferred a poor prognosis (Haas et al., 2008).

1.6.2 *EVI1* is overexpressed in myeloid leukaemia due to cytogenetic rearrangement of chromosome 3

It has long been recognised that cytogenetic rearrangements can cause haematological malignancy (Mitelman et al., 2007). Cytogenetic rearrangements can result in two types of activated oncogene. The first of these are fusion genes brought about by translocations occurring within gene bodies that generate novel fusion proteins. These are chimeric, with both translocation partners making a contribution to the composition and function of the protein. The classic example of a fusion gene is that of *BCR-ABL1* in CML and ALL, as

described above. In the second scenario, translocations can occur that bring a proto-oncogene under the control of an ectopic promoter or regulatory sequence, without altering the coding sequence of the gene, such as the various IGH translocations in B-cell malignancy.

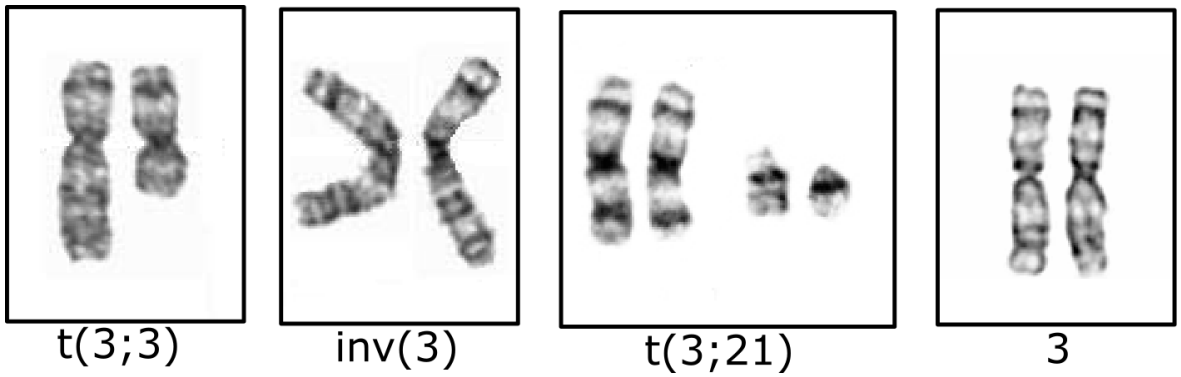


Figure 1-4: Cytogenetic rearrangements can result in overexpression of *EVI1*.¶The t(3;3) and inv(3) both result in *EVI1* overexpression and GATA2 haploinsufficiency as a consequence of ectopic enhancement by the distal GATA2 haematopoietic enhancer. The t(3;21) results in a *RUNX1-EVI1* fusion protein. However, some leukaemia patients have high *EVI1* expression in the absence of any cytogenetic rearrangements (far right).

MECOM, and its component genes, are unusual because they participate in both types of translocation and these can be found in AML, MDS and CML. It was recognised as early as 1992 that cytogenetically visible rearrangements (examples are shown in Figure 1-4) of chromosome 3 were responsible for the overexpression of *EVI1* (Morishita et al., 1992b), and the fusion protein resulting from the t(3;21) was identified in 1994 (Mitani et al., 1994). Mitani et al (Mitani et al., 1994) screened a cDNA library of a cell line with a t(3;21) and found a 180kDa fusion protein comprising the N-terminus of the *RUNX1* gene and the C-terminus of the *MECOM* gene. It was later shown that the *RUNX1-EVI1* fusion protein represses the growth-inhibitive effects of TGFβ (Kurokawa et al., 1998).

However, the mechanism of overexpression in inv(3) and t(3;3) leukaemias had not been uncovered until relatively recently. In 2014, two papers were published almost simultaneously which established that, contrary to the prevailing hypothesis, a single oncogenic enhancer was responsible for *EVI1* overexpression in leukaemias with both the inv(3) and t(3;3) (Groeschel et al., 2014; Yamazaki et al., 2014). Using targeted sequencing of the long arm of chromosome 3, the Delwel group established that the 3q21 breakpoints clustered in a 130kb region between *GATA2* and *RPN1*. The prevailing hypothesis at the time – which had been incorporated into the World Health Organisation guidelines (Swerdlow et al., 2008) - stated that *EVI1* overexpression was a consequence of ectopic enhancement by *RPN1*. However, the group used chromosome conformation capture sequencing to establish

an interaction that was actually between the *EVI1* promoter and the 18kb region 3' of *RPN1* which was commonly translocated. Translocation and inversion of the region separates the *GATA2* gene body from its enhancer – the *GATA2* distal haematopoietic enhancer or G2DHE which was present in the 18kb region – and the enhancer instead exerted its effect on the *EVI1* promoter. Remarkably, the effect of the enhancer was unrelated to the orientation of the *EVI1* gene which differs depending on whether the region is translocated or inverted. The juxtaposition of the G2DHE to *EVI1* also has a concomitant effect on the expression of *GATA2* gene, from which it was separated, and it was established that haploinsufficiency of *GATA2* also contributed to the disease (Katayama et al., 2017). The interaction was unique to cell lines with a cytogenetic rearrangement of 3q and was then refined to the enhancer element using H3K27Ac, H3K4me3 and H3Kme1 ChIP-seq to two p300 binding regions. Finally, a genome editing approach to silence the enhancer and confirm that targeting the enhancer efficiently attenuated *EVI1* expression (Groeschel et al., 2014).

Other, rarer rearrangements resulting in *EVI1* expression have been reported. For example the t(3;12)(q26;p13) results in a fusion gene comprising the first two exons of *ETV6* (previously known as *TEL*) to the second exon of *EVI1* (Peeters et al., 1997). Overexpression of *EVI1* has also been reported following cryptic amplification of the *EVI1* locus via double minutes (Volkert et al., 2014).

1.6.3 EVI1 is overexpressed in solid tumours

Expression of *EVI1* is reported in a growing number of solid tumours, such as breast (Wang et al., 2017), prostate (Wang et al., 2016), colon (Gao et al., 2011) and ovarian cancer (Brooks et al., 1996) where it is invariably associated with a poor prognosis.

In glioblastoma multiforme, EVI1 expression was detected by immunohistochemistry (IHC) in all 86 of their cohort, and classified as highly expressed in 41.9%. High expression was associated with a significantly poorer prognosis ($p=0.033$) with a 1-year survival rate of 20.6% compared to 64.7% in low expression (Hou et al., 2016).

1.6.4 EVI1 is also overexpressed in patients with no detectable cytogenetic rearrangement of 3q

In addition to the rearrangements described in section 1.6.2, *EVI1* expression can occur in the absence of cytogenetic rearrangements of 3q (Russell et al., 1994). Despite the advances in our understanding of *EVI1* expression in the 3q-rearranged leukaemia, there is still very little progress in understanding what causes *EVI1* expression in leukaemia with no 3q rearrangements.

Gröschel et al. tested a cohort of 1,382 newly diagnosed AML patients below the age of 60 and found *EVI1* expression in 10.7% based on expression of 0.1x compared to the SKOV3 cell line (which has amplification of the MECOM locus and thus high expression; (Nanjundan et al., 2007)) and normalised to *PBGD*. They found that *EVI1* expression was an independent predictor of poor prognosis as measured by complete remission rates (odds ratio 0.54, p=0.002) and event free survival (hazard ration 1.46, p<0.001). Notably, they recommend bone marrow transplantation for patients who achieve complete remission (Groeschel et al., 2010).

One group found that *EVI1* overexpression is particularly prevalent in AML patients with *MLL* translocations – 43% of their cohort expressed *EVI1*. They showed that knockdown of *EVI1* in *EVI1*-expressing *MLL* rearranged cells inhibited leukemic growth, both *in vivo* and *in vitro,* and thus concluded that *EVI1* represents a critical factor in the pathogenesis of leukaemia (Bindels et al., 2012). *EVI1* is also overexpressed in 10-15% of myelodysplastic syndrome (MDS) where it is again associated with a poor prognosis (Nimer, 2008).

Chromosomal abnormalities of 3q are rare in paediatric leukaemia (Raimondi et al., 1999), however, *EVI1* expression is detected. A Dutch group studied a cohort of 228 paediatric AML patients and found *EVI1* expression in 9% of *de novo* cases and 50% (albeit in a small number – just 4 of 8 patients) in secondary AML. Notably, the group detected no 3q abnormalities, and even excluded the presence of cryptic abnormalities using a FISH strategy. Echoing that of adults, they found *EVI1* expression coincided with *MLL* translocations and monosomy 7, both of which confer a poor prognosis (Balgobind et al., 2010).

Given the proposed antagonistic relationship between *EVI1* and *MECOM, o*ne group studied in detail the different transcripts expressed from the *EVI1* locus and classified their cohort of 319 AML patients by the combination of *EVI1* versus *MDS1-EVI1* (i.e. *MECOM*) yielding four groups: $EVI1^{+}MECOM^{-}$, $EVI1^{+}MECOM^{+}$, $EVI1^{-}MECOM^{+}$ *and* $EVI1^{-}MECOM^{-}$. They found that only those patients expressing *EVI1* transcripts had a poor prognosis and this held true regardless of their *MECOM* expression level (Barjesteh van Waalwijk van Doorn-Khosrovani et al., 2003).

EVI1 expression in CML has been well established in BC but less well studied in CP. One small study of 14 CML-BC patients detected *EVI1* expression in 10 (71%), none of whom had a cytogenetic rearrangement (Ogawa et al., 1996a). Expression is found in CML patients with and without 3q rearrangements (Morishita et al., 1990a; Ogawa et al., 1996b). Crucially,

EVI1 expression is detectable in CD34+ myeloid precursor cells of CML patients and of healthy individuals. In *BCR-ABL1* positive CD34+ cells *EVI1* expression is repressed by imatinib treatment but imatinib treatment in *BCR-ABL1* negative CD34- cells had no effect on *EVI1* expression, suggesting that the regulation of *EVI1* is linked to the activity of the *BCR-ABL1* protein (Roy et al., 2012). A Japanese group found that the high *EVI1* expression in CML-CP stem cells conveyed superior LSC survival, contributing to the leukaemia initiating potential and resistance to tyrosine kinase inhibition (Sato et al., 2014).

Daghistani et al found *EVI1* expression at the time of imatinib failure correlated with a lack of response to the second generation TKIs and thus predicted patients who may benefit from early transplant (Daghistani et al., 2010). Manachai et al recently reported that patients in CML-BC with high *EVI1* expression also had activated nuclear β-catenin. The authors showed that increased *BCR-ABL1* expression in myeloid leukaemia cell lines enhanced the expression of *EVI1* but that this did not occur in lymphoid cell lines (Manachai et al., 2017).

Maicas and colleagues used ChIP and site directed mutagenesis to characterise the *EVI1* promoter and found that both RUNX1 and ELK bind to the 318bp region identified (Maicas et al., 2013). A Spanish group identified a direct interaction between EVI1 protein and the promoters of two miRNA, miR-1-2 and miR-133-a-1, although only the former was found to be involved in the aberrant expression in cell lines (Gomez-Benito et al., 2010). This group and others also showed that the presence of EVI1 protein was associated with promoter methylation and down regulation of the hsa-miR-124 (Dickstein et al., 2010; Vázquez et al., 2010).

1.7 Aims

The aim of this thesis is to investigate the genetics and epigenetics of leukaemia by asking three questions:

1. Does oncogenic stress promote genetic instability associated with leukaemia progression?
2. Does stable transformation of B cell precursors depend on the adaptation of epigenetic changes?
3. Are epigenetic changes responsible for the expression of *MECOM* in poor prognosis myeloid leukaemia?

2 Transcription sensitizes lineage-specific genes to DNA damage in leukaemia cells

2.1 Contribution of others to this chapter

All bioinformatic analysis was undertaken by Dr Mark Robinson. All mouse handling and the majority of cell culture and ChIP-seq was undertaken by Dr Bryant Boulianne. Dr Jenny Thomas performed the DsRed experiment. I performed all DNA-FISH, all RNA-seq, all ChIP-qPCR and some of the cell culture and ChIP-seq processing. This chapter was published in Cell Reports (see Appendix) and thus figures were prepared by Dr Niklas Feldhahn, Dr Mark Robinson, Dr Bryant Boulianne and by me (Boulianne et al., 2017).

2.2 Hypothesis

Cellular stress in leukaemia cells resulting from oncogene expression (i.e. oncogenic stress) promotes genetic instability associated with leukaemia progression.

2.3 Background

It has long been established that B-lineage genes are frequently mutated in B-lineage leukaemia, however, the mechanistic basis of this observation is not understood. One possibility is that mutations arise randomly in genes of all lineages but those in B-lineage genes are selected for in B-lineage leukaemia. A second possibility is that the B-lineage genes are natural targets for mutation in B-lineage cells, and thus are present in B-lineage leukaemia. In order to investigate these possibilities, we used γH2AX and H3K27Ac chromatin immunoprecipitation sequencing (ChIP-seq), RNA-seq and DNA-FISH to investigate the relationship between DNA damage and active promoters in B-cells undergoing oncogenic crisis.

2.4 Results

2.4.1 Oncogene expressing B-cell precursors suffer genome-wide DNA damage

In order to investigate potential DNA damage resulting from oncogene expression during the onset of leukaemia (analogous to *BCR-ABL1*-positive B-ALL) we aimed to replicate the *in vivo* leukaemia-initiating events *in vitro* by forced oncogene expression in primary B-cell precursors derived from murine bone marrow. In brief, bone marrow cells were cultured *ex vivo* with murine IL7, which promotes the selective survival and proliferation of B-cell precursors, and transduced on day two and day three with retroviral MIGR1 vectors expressing the *BCR-ABL1*, *MYC* or the empty vector (EV) as control as outlined in Figure 2-1. Flow cytometric analysis on day 10 (see Figure 2-2) confirmed selective survival of normal or oncogene expressing B-cell precursors. In order to increase the sensitivity of the

DNA damage detection by ChIP-seq in consecutive experiments, bone marrow from P53-binding protein 1 (53BP1)-deficient mice (*53BP1*$^{-/-}$) was used. 53BP1 is a DNA damage response (DDR) protein, which localises to damaged DNA in order to recruit additional DDR proteins. Absence of 53BP1 results in increased persistence of induced DNA damage and has been used by others to facilitate DNA damage detection by ChIP-Seq (Barlow et al., 2013; Hakim et al., 2012; Yamane et al., 2011; Yamane et al., 2013).

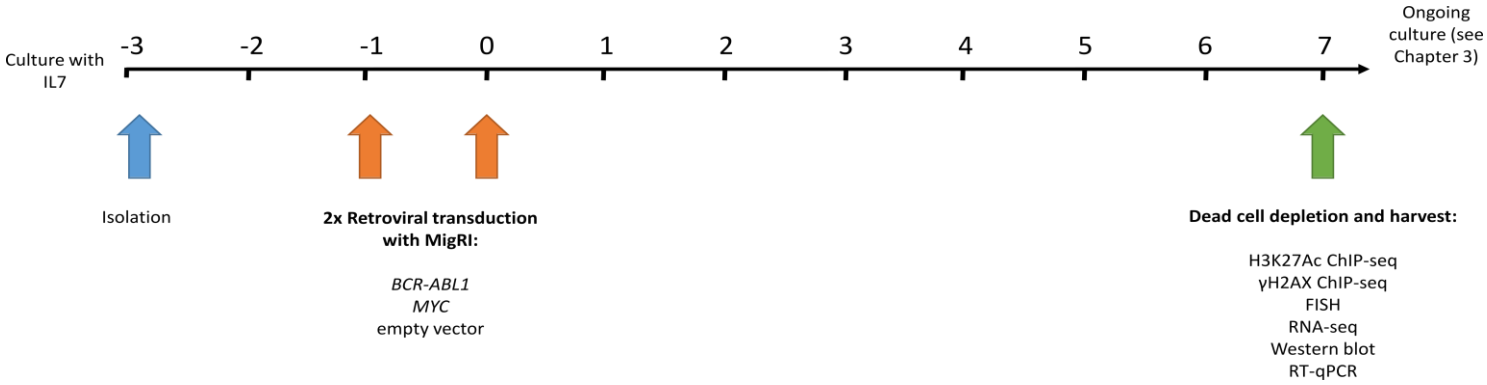


Figure 2-1: Outline of the oncogene induction experimental approach. ¶Whole murine bone marrow was cultured with IL7 and transduced with retroviral vectors expressing an oncogene. On day seven post infection when they were suffering from oncogenic stress, cells were harvested for γH2AX ChIP-seq (n=2 for both oncogenes), H3K27Ac ChIP-seq (n=1 for both oncogenes), FISH (n=5 for all conditions), RNA-seq (n=2 for all conditions), Western blot (n=1 for all conditions) and RT-qPCR (n=3). The sample size (n) refers to independent transduction and library preparation.

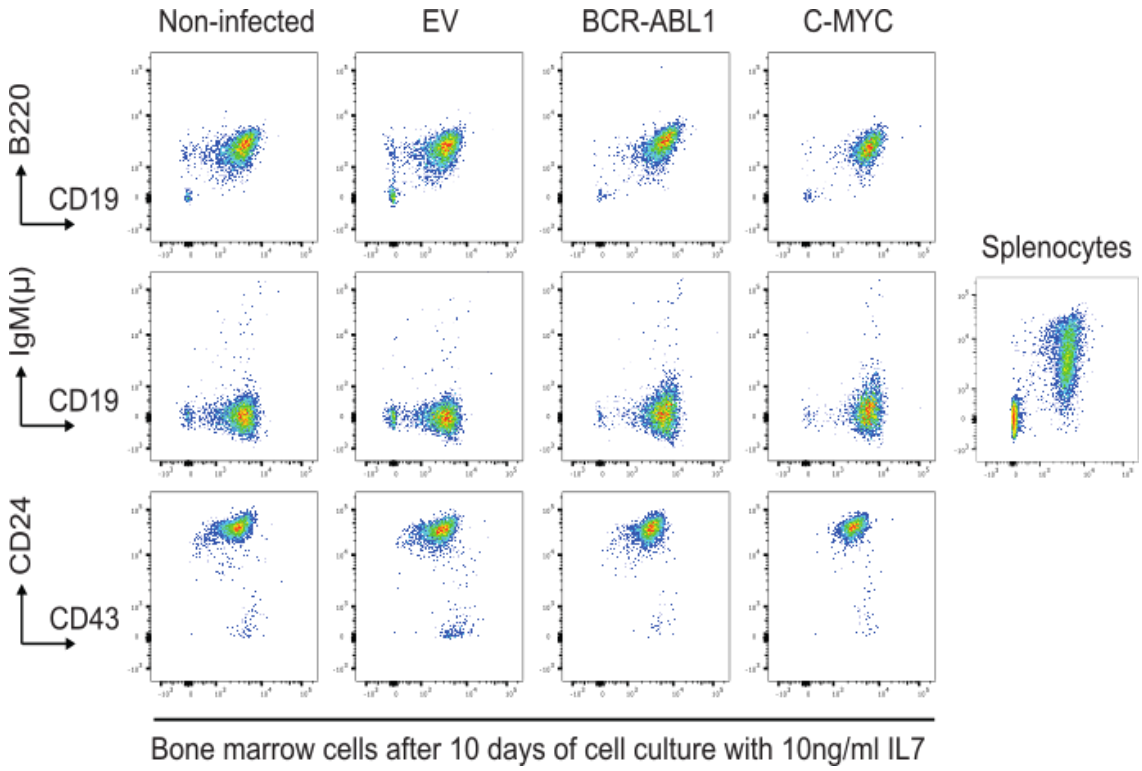


Figure 2-2: Whole murine bone marrow cultured with IL7 for ten days enriched for B-cell precursors. ¶Flow cytometric analysis a representative experiment (n=1) of non-infected and cells infected with the empty vector, BCR-ABL1 and MYC using antibodies for B-cell antigens. Staining of splenocytes with IgM is included as a positive control. Analysis and figure by Dr Bryant Boulianne.

Acute, unscheduled oncogene induction is known to induce apoptosis or senescence in primary cells, and apoptotic responses to *BCR-ABL1* and *NRAS* have been reported by others (Dengler et al., 2011; Shojaee et al., 2015). In agreement, following the induction of *BCR-ABL1* or *MYC* oncogene, oncogene-expressing cells entered a phase of increased proliferation but suffered an extended period of decreased viability, as depicted in Figure 2-3.

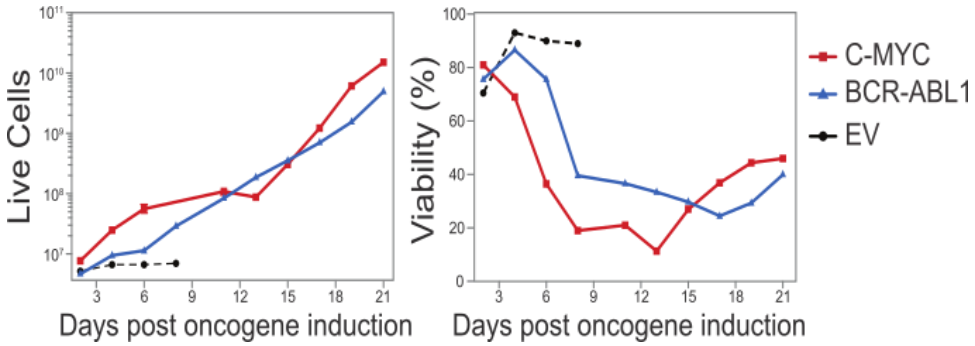


Figure 2-3: Oncogene induced cells had decreased viability and increased proliferation. ¶Plots show the proliferation and viability of the B-cell precursors in the 21 days following their transduction with *BCR-ABL1*, *MYC* or empty vector – a representative experiment (n=1) of five is shown. Analysis and figure by Dr Bryant Boulianne.

Oncogene-expressing cells showed increased levels of the DNA damage associated protein biomarkers, i.e. phosphorylated p53 and γH2AX, compared to the empty vector or uninfected cells by Western blot analysis (Figure 2-4).

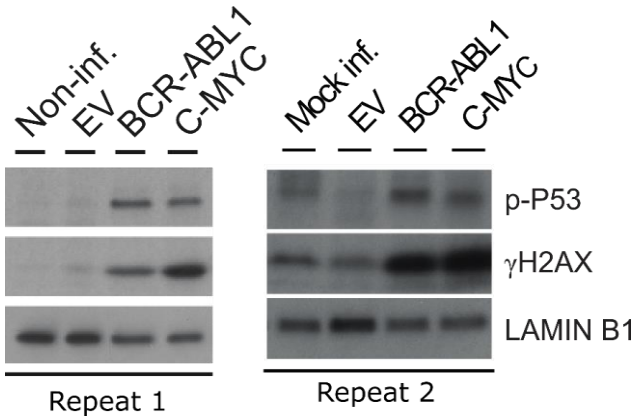


Figure 2-4: Oncogene-induction triggered a DNA damage response. ¶Western blots showing DNA-damage response proteins (phosphorylated P53 and γH2AX) at D7 post-infection in two representative experiments. Lamin B1 was used as a loading control. Blot and figure by Dr Bryant Boulianne.

Of note, oncogene expressing cells exhibit reduced viability. DNA damage pathway activation has been associated with apoptosis, so in order to distinguish oncogenic stress-related DNA damage from apoptosis-induced DDR activation it was necessary to ensure the population of cells studied contained live cells only. Column-based magnetic activated cell sorting (MACS) was therefore used to positively select for apoptotic and dead cells using magnetic beads which recognise an antigen in the plasma membrane of dead and apoptotic cells only. Following positive selection for the apoptotic/dead cell antigen, the flow-through from the column contained live cells only. Depletion of dead cells from this flow through was

confirmed by Trypan blue (which is excluded from live cells) staining and by flow cytometry (Figure 2-5).

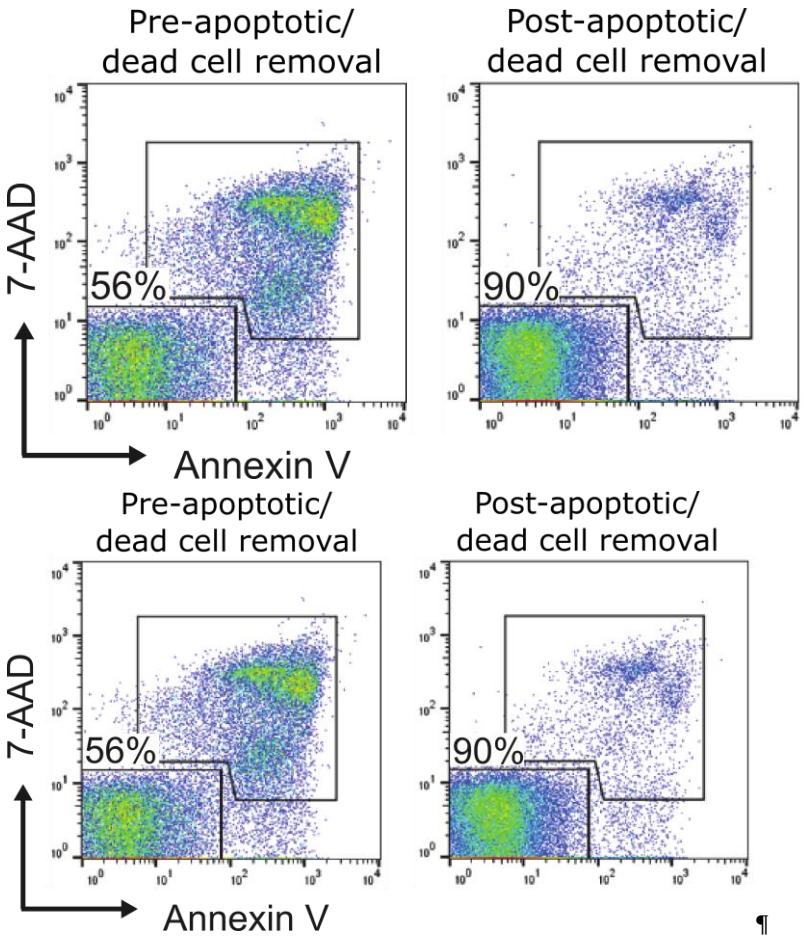


Figure 2-5: Flow cytometry confirmed that the apoptotic population was removed by dead cell depletion¶7 aminoactinomycin D (7-AAD) is excluded from live cells and Annexin V is a biomarker of apoptosis. One representative experiment is shown. Analysis and figure by Dr Bryant Boulianne.

Western blot analysis of dead/apoptotic cells depleted cells confirmed that the observed DNA damage pathway activation in oncogene expressing cells was unrelated to apoptosis (Figure 2-6).

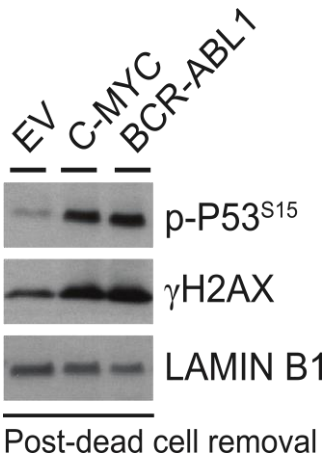


Figure 2-6: DNA damage response was not caused by apoptosis.¶Western blot showing DNA-damage response proteins are expressed in the dead-cell depleted population. One representative experiment is shown. Blot and figure by Dr Bryant Boulianne.

As mentioned in section 1.3.2, genomic alterations in leukaemia have been repeatedly associated with AID and/or RAG1 activity. To investigate to what extent the observed DNA damage in oncogene expressing cells relates to AID or RAG1 function, AID and RAG1 expression was quantified by RT-qPCR (Figure 2-7) and Western blot analysis (Figure 2-8). In the experimental model (oncogene expression in 53BP1-/- B-cell precursors) oncogene expressing cells were not characterized by elevated AID or RAG1 expression Hence, RAG and AID are unlikely to represent the source of the endogenous DNA damage observed.

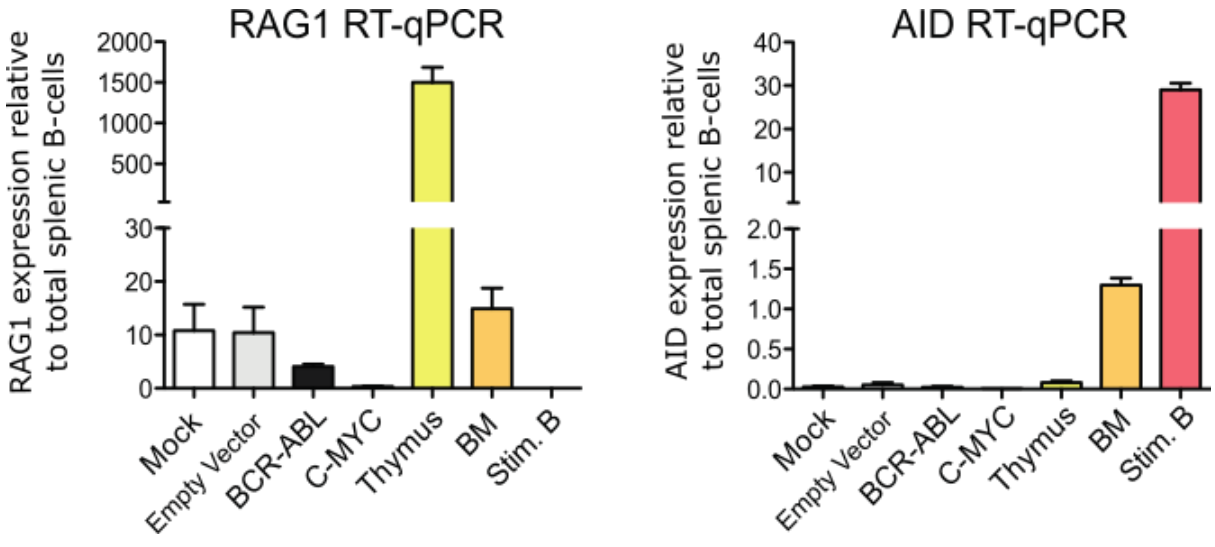


Figure 2-7: DNA damage was not linked to RAG1 or AID expression.¶RT-qPCR analysis (n=3) of *RAG1* and *AID* expression in the untransformed and transformed B-cell precursors at D7 post infection relative to total splenic B-cells, normalised to HPRT expression. Total thymus was used as a positive control for RAG1 and lipopolysaccharide/IL4 stimulated mature B-cells as a positive control for AID. RT-qPCR and figure by Dr Bryant Boulianne.

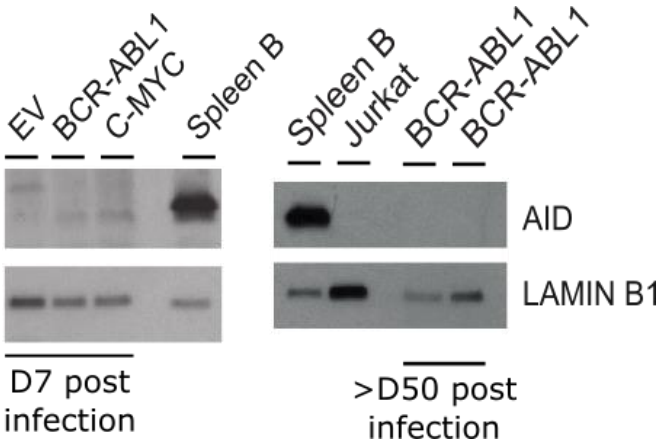


Figure 2-8: DNA damage was not linked to AID protein.¶Western blot analysis of AID protein in the transformed and untransformed B-cell precursors at D7 and >D50 post infection. Lamin B1 was used as a loading control. One representative experiment is shown. Blot and figure by Dr Bryant Boulianne.

In order to identify the exact location of DNA damage induced by the oncogene activation, γH2AX ChIP-seq (n=2 for each oncogene) was undertaken in duplicate on the B-cell precursor populations following dead-cell depletion. Bioinformatic analysis of the DNA pulled down was undertaken as reported by others (Barlow et al., 2013; Rodriguez et al., 2012; Seo et al., 2012) – for more detail, see section 6.11. The number of uniquely aligned reads and the quality scores for these libraries are shown in Figure 2-10 and Figure 2-11. Note, all experiments were conducted using varying proportions of male and female mice therefore the sex chromosomes were excluded for all bioinformatic analysis.

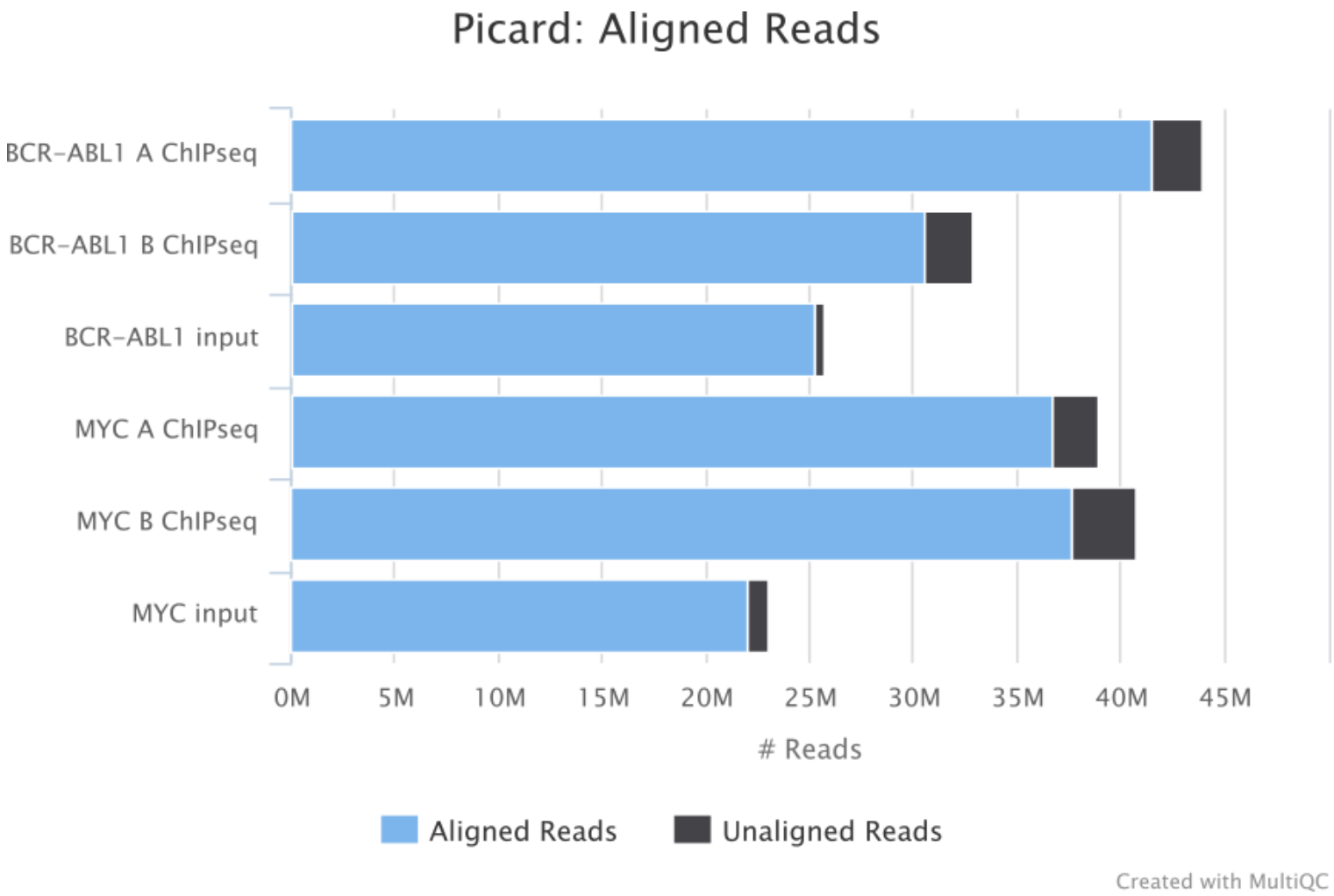


Figure 2-9: Picard alignments show that each of the four γH2AX ChIP-seq library yielded a minimum of 31 million uniquely mapped reads.¶Two corresponding input libraries had a minimum of 22 million uniquely mapped reads. Figure generated using (Ewels et al., 2016)

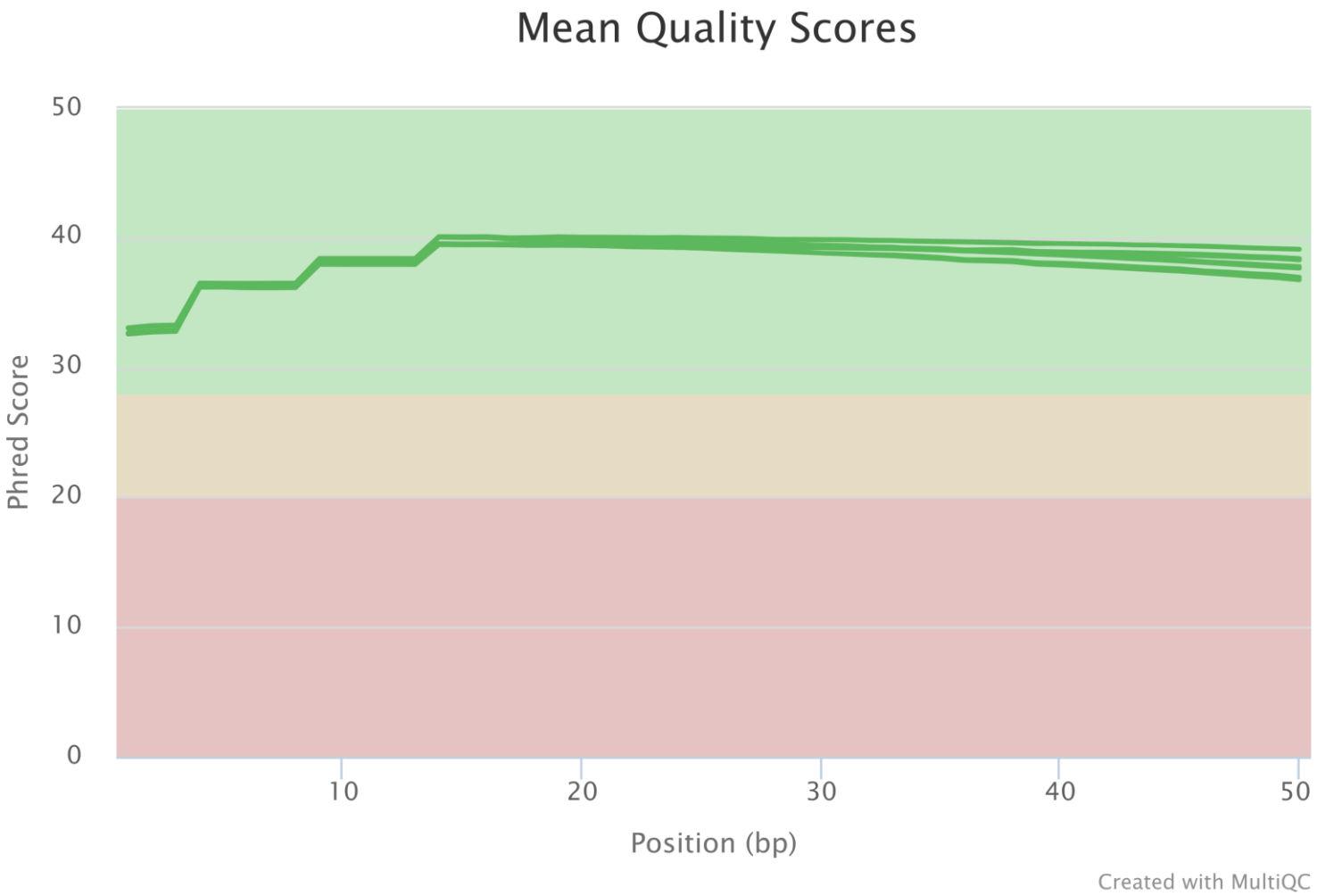


Figure 2-10: Each of the four γH2AX ChIP-seq libraries and both input libraries were all of very good quality with Phred scores >30.¶Figure generated using (Ewels et al., 2016).

The signals were highly reproducible between the two biological repeats (as summarised in Figure 2-11 and depicted in Figure 2-12). Of the peaks with a signal above background, 99% and 93% were in common between the first and second replicate for BCR-ABL1 and MYC expressing populations, respectively, and the signal intensity was highly correlated (r=0.90 and 0.86; p=<1x10^{-15} for both).

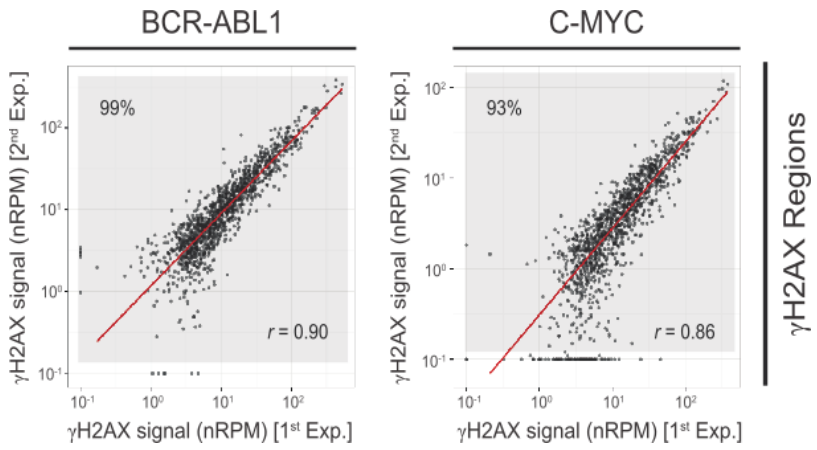


Figure 2-11: Overall γH2AX-signals were highly reproducible between experiments ¶Scatterplot showing Reads Per Million normalised to input/isotype control (nRPM) of the γH2AX regions for BCR-ABL1 and MYC expressing B-cell precursors are shown for two biological repeats. The red line indicates linear regression; p=<1x10^{-15}). For more information see section 6.11.6. Bioinformatic analysis and figure by Dr Mark Robinson.

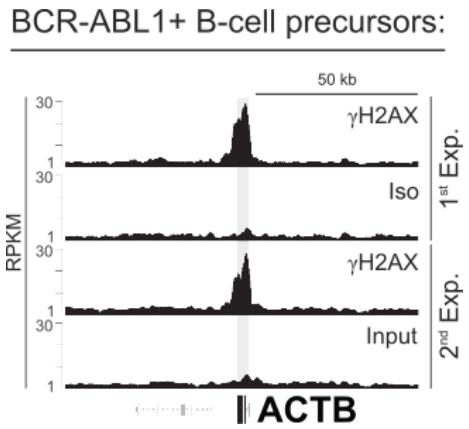


Figure 2-12: Individual γH2AX peaks were highly reproducible between experiments.¶ Custom track showing a representative γH2AX peak at actin beta from two biological repeats. The isotype control library shows the background signal. Gene bodies and their orientation are indicated below in black. Figure by Dr Niklas Feldhahn.

The analysis revealed 889 γH2AX regions that were induced by *BCR-ABL1* and 960 that were induced by *MYC*. As shown in Figure 2-13, the signal intensity of the γH2AX-enriched regions was also remarkably consistent for both BCR-ABL1 and MYC-induced experiments; of the peaks with a signal above background, 99% were detected in both oncogene expressing

populations with a correlation of 0.89 ($p=<2.2x10^{-16}$). Expression of *BCR-ABL1* is known to increase expression of *MYC* so the observed overlap may be due to the common effect of *MYC* activation, or more simply, may be the common result of oncogene expression, such as replicative stress (Cleveland et al., 1989; Sawyers, 1993).

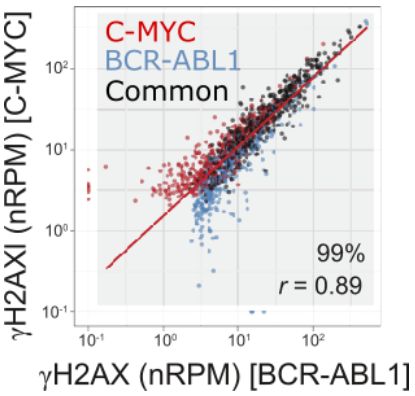


Figure 2-13: γH2AX signal intensity was comparable for both BCR-ABL1 and MYC oncogene induction.¶Scatterplot showing the reads per million normalised to input/isotype control (nRPM) of the γH2AX regions (i.e. signal intensity) is comparable for both BCR-ABL1 (blue) and MYC (red) expressing B-cell precursors (n=2 for each oncogene). The red line indicates linear regression. Bioinformatic analysis and figure by Dr Mark Robinson.

The identified γH2AX regions were therefore merged for further analysis, resulting in a total of 1289 regions for further analysis. These were distributed over all the murine chromosomes (Figure 2-14). Further, the total coverage across all the chromosomes was highly correlated (r=0.87) with the exon coverage per chromosome (i.e the percentage of all exons in the murine genome present on each chromosome; shown in Figure 2-15), which varies from 2.1% on chromosome 11 to just 1.0% on chromosome 12, indicating the distribution was not simply at random across the chromosomes.

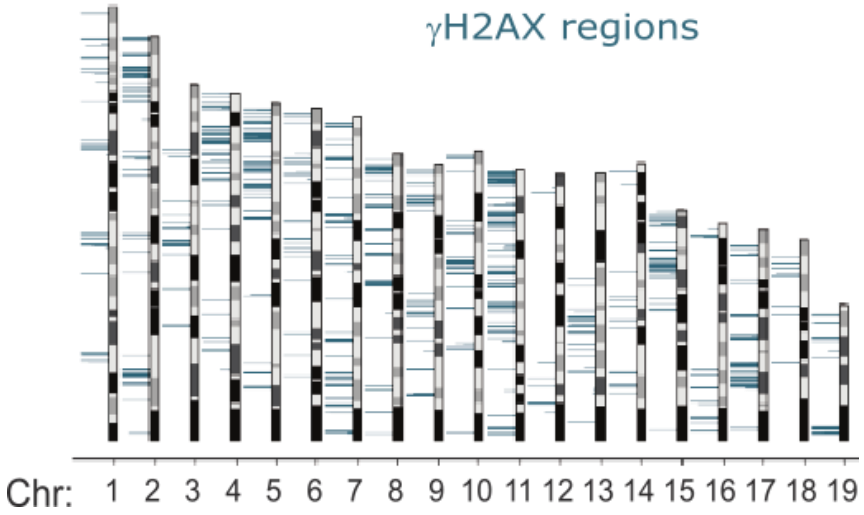


Figure 2-14: γH2AX regions were distributed across all chromosomes.¶Karyogram showing the γH2AX region distribution from two biological repeats over the mouse genome in blue. Length of bars indicates the intensity of the γH2AX signal (RPM of the BCR-ABL1 library; range 0-100RPM) for each. See section 6.11.5 for details. Bioinformatic analysis and figure by Dr Mark Robinson.

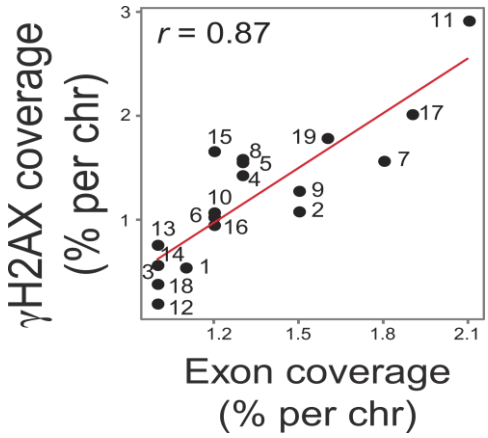


Figure 2-15: γH2AX coverage across chromosomes was highly correlated with exon coverage.¶Linear regression fit is indicated by the red line; data shown is for combined BCR-ABL1 and MYC regions from two biological repeats. Bioinformatic analysis and figure by Dr Mark Robinson.

2.4.2 γH2AX-marked genomic regions occur at B-cell precursor-specific genes and relate to endogenous DNA damage

Genes that overlapped with γH2AX regions were then examined in more depth. Of the 1,289 γH2AX regions identified, 788 overlapped with more than one gene, resulting in a total of 2,300 genes marked by γH2AX. Notably, many of the most prominent γH2AX regions located to genes associated with B-cell precursor function (Figure 2-16) including those genes involved in the pre-B cell receptor complex (Figure 2-17, left), pre-B cell receptor signalling (middle), and B-cell identity related transcriptional pathways (right).

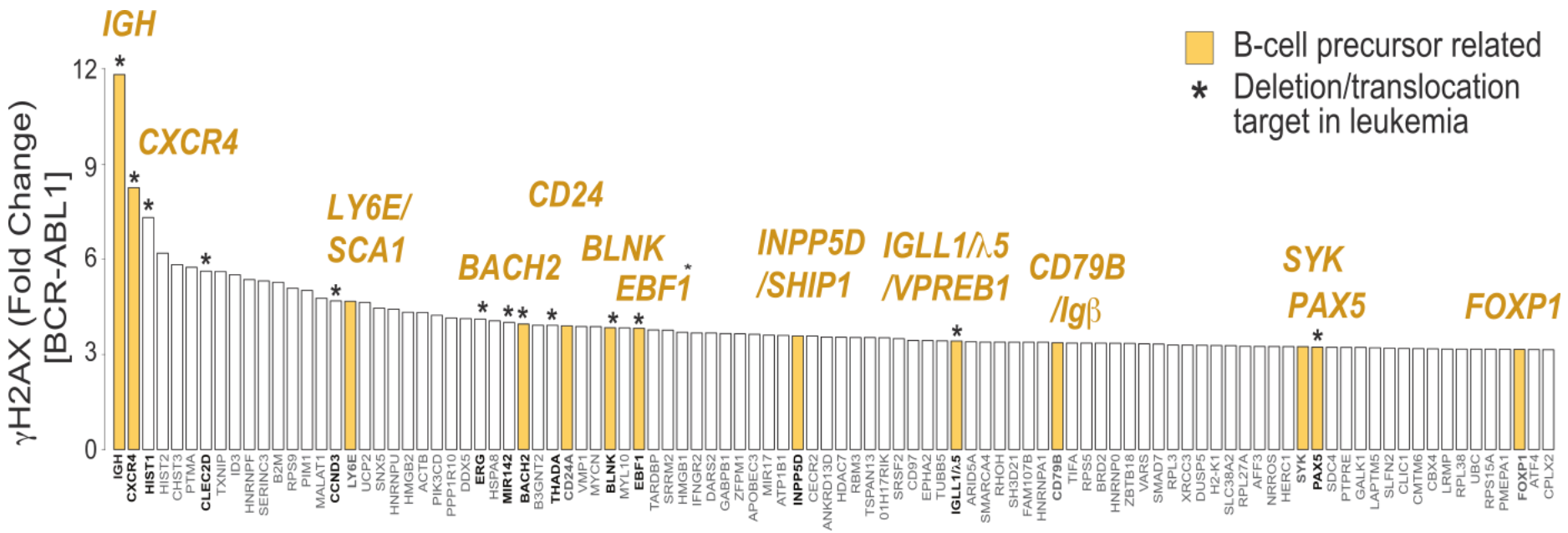
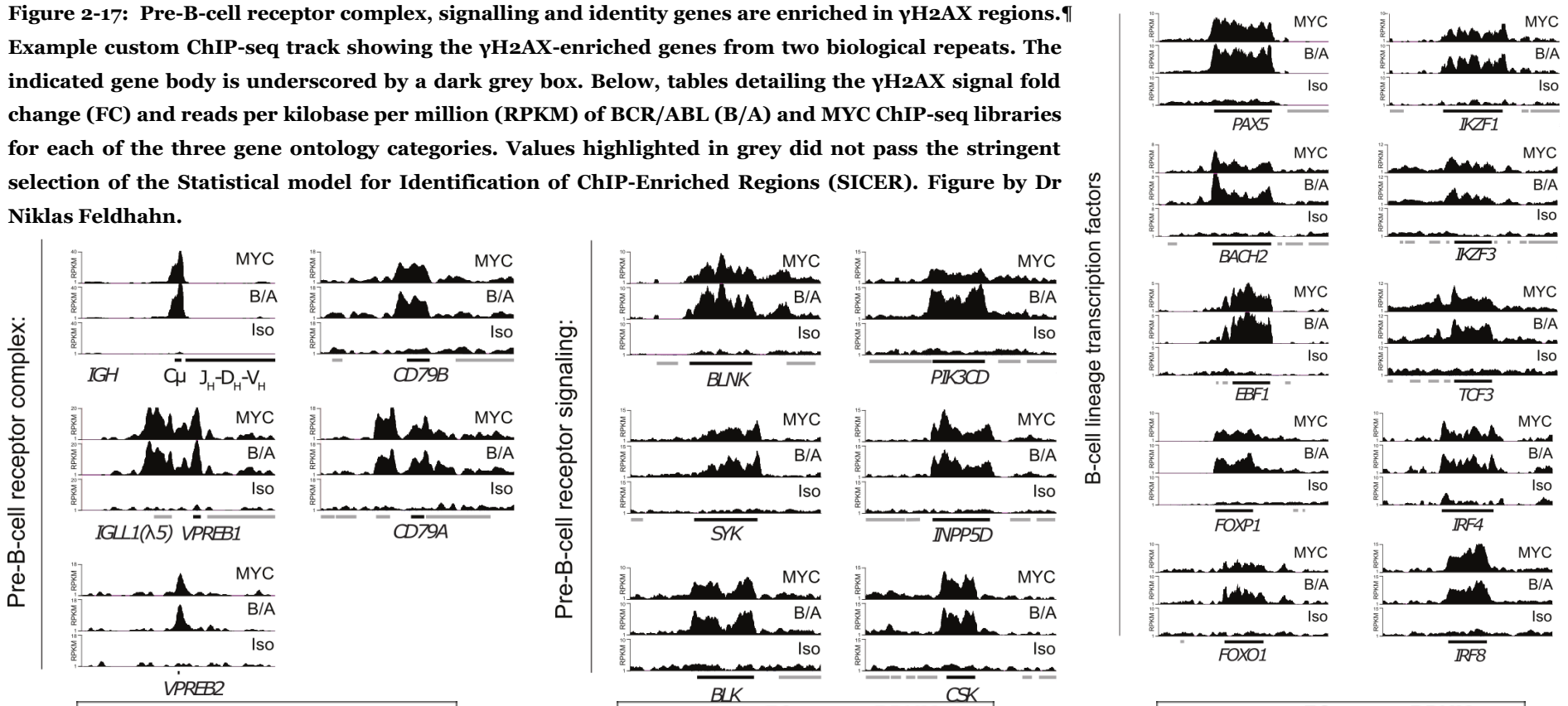


Figure 2-16: Prominent γH2AX regions affected lineage-specific genes¶ Bar chart of the top ranking γH2AX regions identified in *BCR-ABL1* transformed B-cell precursors from two biological repeats and their average fold change values. Genes that overlap with the γH2AX regions are indicated on the X-axis; those involved in B-cell precursor related pathways are highlighted in yellow, and those which are the target of deletions and/or translocations in B-cell leukaemias are indicated with an asterisk. Regions were combined if the cover the same gene.

Figure 2-17: Pre-B-cell receptor complex, signalling and identity genes are enriched in γH2AX regions.¶ Example custom ChIP-seq track showing the γH2AX-enriched genes from two biological repeats. The indicated gene body is underscored by a dark grey box. Below, tables detailing the γH2AX signal fold change (FC) and reads per kilobase per million (RPKM) of BCR/ABL (B/A) and MYC ChIP-seq libraries for each of the three gene ontology categories. Values highlighted in grey did not pass the stringent selection of the Statistical model for Identification of ChIP-Enriched Regions (SICER). Figure by Dr Niklas Feldhahn.



Gene	FC B/A	FC MYC	RPKM B/A	RPKM MYC
IGH	11.8	11.8	3.3	3.9
IGLL1	3.4	3.6	2.2	2.2
VPREB1	3.4	3.6	2.2	2.2
VPREB2	2.9	2.3	2.2	1.7
CD79A	2.6	3.3	2.1	3.1
CD79B	3.4	2.5	2.4	2.3

Gene	FC B/A	FC MYC	RPKM B/A	RPKM MYC
SYK	3.2	2.7	2.2	2.1
BLNK	3.8	3.0	2.4	1.8
CSK	2.9	3.0	3.2	3.7
INPP5D	3.6	3.4	3.0	2.8
BLK	2.5	2.0	2.1	1.7
PIK3CD	4.2	2.1	4.2	2.1

Gene*	FC B/A	FC MYC	RPKM B/A	RPKM MYC
PAX5	3.2	3.2	2.7	2.6
EBF1	3.8	3.7	1.6	1.5
BACH2	4.0	2.9	2.4	1.0
TCF3	2.2	1.9	2.1	2.2
IKZF1	3.0	2.5	1.9	1.5
IRF8	2.8	3.1	3.5	2.9

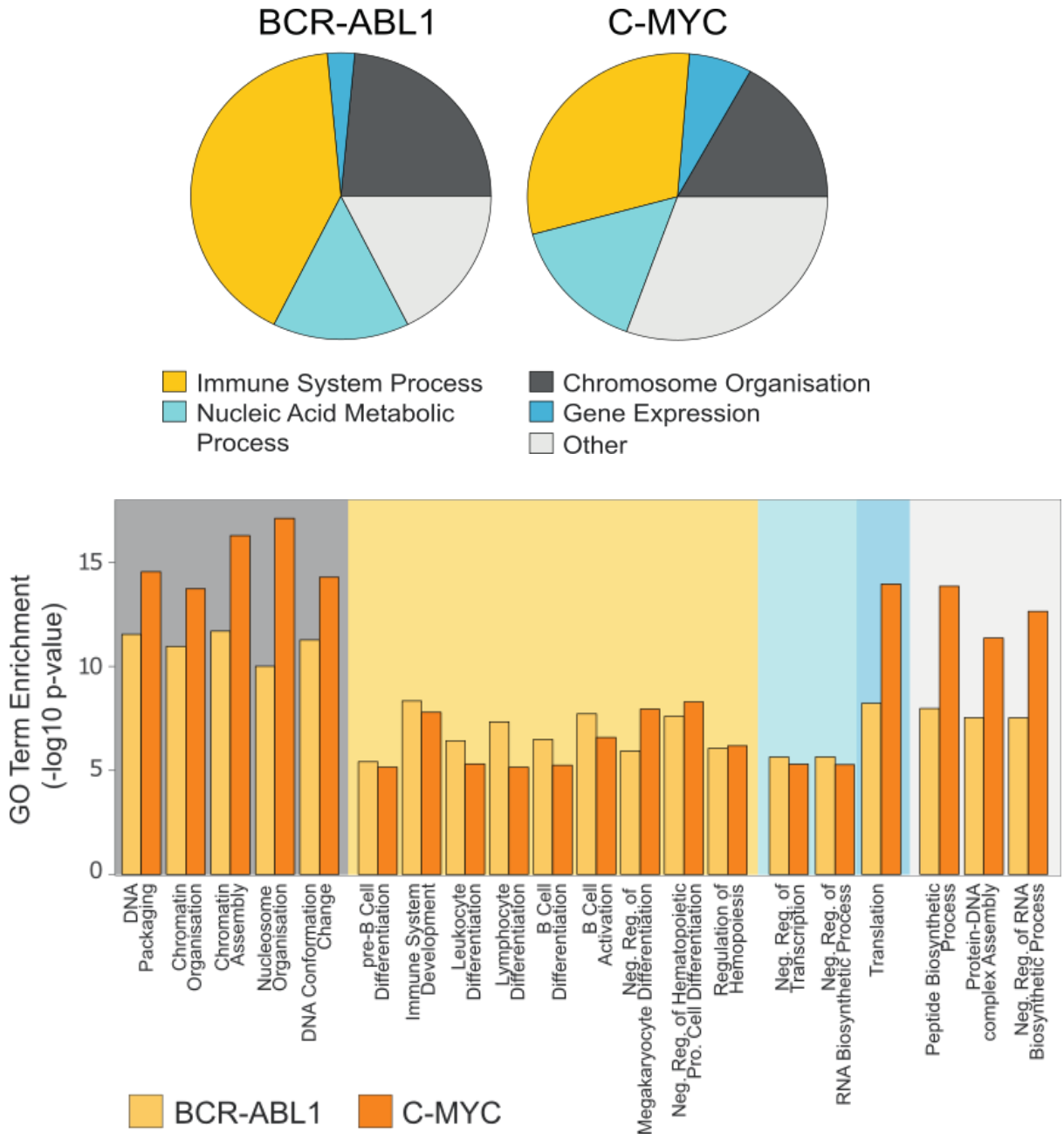


Figure 2-18: Gene ontology pathway analysis showed γH2AX regions frequently affected immune-system related processes.¶(Top) Pie charts show all pathways identified by gene ontology enrichment analysis for *BCR-ABL1* and *MYC* expressing B-cell precursors. (Bottom) Bar chart with further detail of the gene ontology pathways highlighted in the same colours. For more information, see section 6.11.7. Bioinformatic analysis and figure by Dr Mark Robinson.

In order to analyse all γH2AX-regions for associated genes, gene ontology (GO) analysis was performed on the complete set of regions. GO analysis confirmed the analysis of the top-ranked γH2AX-regions as it revealed that the major pathway affected in each transformed B-cell precursor population related to immune system processes (Figure 2-18) including lymphocyte and leukocyte differentiation.

In order to confirm DNA damage at identified γH2AX regions with an independent method, interphase fluorescence in situ hybridisation (FISH) was undertaken using dual colour break apart probes. Each assay comprised a pair of fluorescently labelled probes which hybridised to the target gene and its flanking regions – an intact gene with both probe components in close proximity therefore yielded a yellow fusion signal, whilst any damage which affected the target gene and resulted in physical disruption of the chromatin resulted in separation of one of the labelled probes. Cells were harvested for FISH at the same time points (day 7) as used for the previous experiments (e.g. ChIP-Seq analysis). *PAX5* and *BLNK*, two prominent γH2AX hotspots identified by ChIP-seq were selected for analysis. In addition, a γH2AX-negative gene (*FOXP2*) was used as a control probe (herein referred to as the 'coldspot').

As shown in **Error! Reference source not found.**, the automated analysis detected physical separation of the two probe halves significantly more frequently for the hotspot probes in transformed B-cells (*BCR-ABL1* [B/A] or *MYC*) than in the control (empty vector; EV) cells (*BLNK* empty vector versus *BCR-ABL1* p=0.0011, empty vector versus *MYC* p=0.0039; *PAX5* empty vector versus *BCR-ABL1* p=0.0040, empty vector versus *MYC* p=0.0086) and in neither population for the cold spot control (*FOXP2* empty vector versus *BCR-ABL1* or *MYC* not significant).

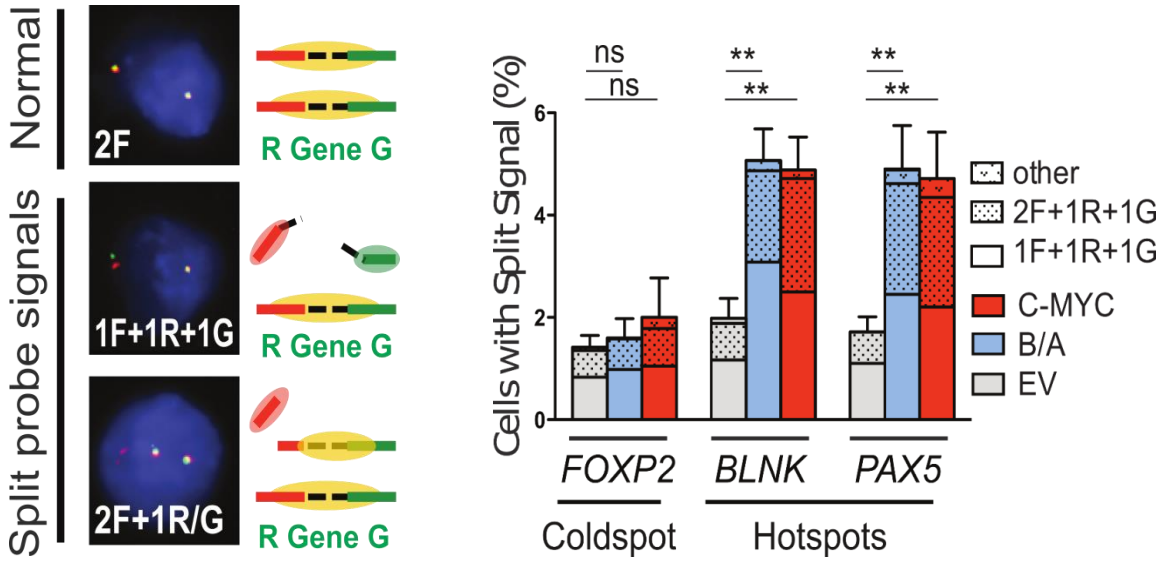


Figure 2-19: γH2AX hotspots correlate with sites of active DNA damage.¶(L) An overview of the FISH signal patterns. (R) Bar chart showing results from dual colour break apart FISH probe hybridisations undertaken on day seven (n=5 hybridisation experiments, with a minimum of 1000 interphases analysed per hybridisation) verifying significantly increased DNA damage in the oncogene induced populations for using at the γH2AX hotspots (PAX5 and BLNK) with no significant increase for the cold spot control (FOXP2) or the empty vector populations. Bars indicate mean with error bars showing ± standard error of the mean (SEM).

In a separate pilot experiment, γH2AX signals in early (day 7 post infection) transduced cells were compared to those from long term (more than 50 days post infection) transformed cells by ChIP-Seq. While almost all γH2AX regions could be detected at both stages, we observed that the most prominent γH2AX region identified in early transformed cells (at *IGH*) disappeared in the >D50 cells (Figure 2-20, left). Targeted investigation of this locus in a larger panel of >D50 cells derived from independent experiments by endpoint PCR revealed a biallelic internal deletion in two of the seven oncogene-expressing populations (Figure 2-20, right). Notably, *IGH* deletions were only found in *BCR-ABL1*-transformed cells, though the limited number of samples analysed may further limit the interpretation of this finding.

Collectively, these results show that oncogene activation in B-cell precursors results in DNA damage affecting B-lineage genes, and that this damage could predispose to genomic deletion during successful transformation.

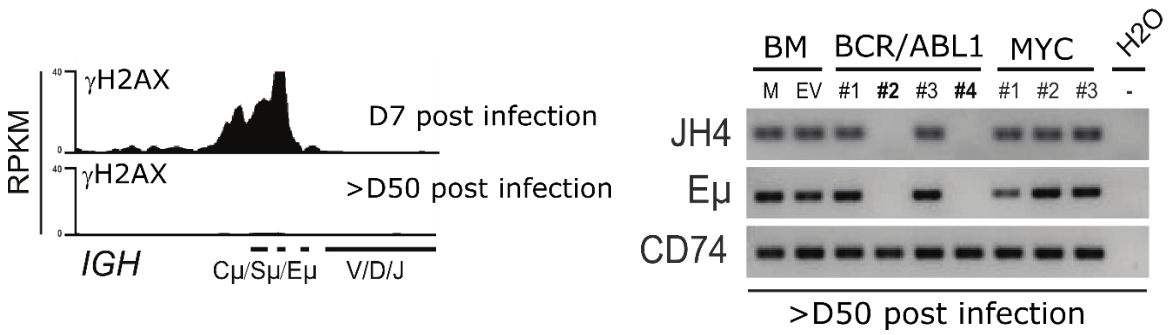


Figure 2-20: DNA damage resulted in loss of the IGH locus.¶(L) Custom ChIP-seq track showing complete loss of the D7 IGH signal in the BCR-ABL1 transformed B-cell precursors by D50+. (R) Agarose gel image showing PCR amplification of genomic DNA from selected D50+ transformed B-cell precursors. Loss of the JH4 and Eμ regions compared to control CD74 region can be seen in two of the three *BCR-ABL1* experiments shown. This was not detected in three MYC experiments. PCR and figure by Dr Niklas Feldhahn.

2.4.3 Genes known to be mutated in B-ALL frequently overlap with γH2AX regions

In order to evaluate the potential relevance of oncogenic stress associated DNA damage identified for genomic alterations reported for B-ALL patients, the combined γH2AX region list was compared to three publicly available datasets detailing gene mutations found in human leukaemia. Genes were excluded from the comparison if they were not present in B cell leukaemia, if they related to abnormalities other than deletions or translocations, or if there was no murine equivalent. Genes that overlapped with γH2AX regions in the analysis were significantly enriched for genes identified by these datasets as recurrently mutated in human B-ALL – 41% with B-ALL (Mullighan et al., 2007), 26% with high risk B-ALL (Mullighan et al., 2009b) and 19% with Philadelphia positive-like ALL (Roberts et al., 2012) – see Table 2-1.

Table 2-1: γH2AX enriched murine genes overlap with genes found to be abnormal in human B-ALL.¶Genes were included if they were present in B-ALL, present in the mouse genome and were annotated. Two-by-two contingency table analysis confirmed significant enrichment of γH2AX regions (Swaminathan et al., 2015)

Data for comparison	Type	Overlapping loci	Genes	Significance
(Mullighan et al., 2007) GSE5511	B-ALL	16/39 (41%)	*BLNK, BTG1, C19ORF7, DLEU2, FKB2, EBF1, ERG, FAM65A/RLTPR, HIST1, IKZF1. IKZF3,*	P<0.0001

			LEF1, PAX5, RUNX1, TCF3, THADA	
(Mullighan et al., 2009a; Mullighan et al., 2009b) GSE11877 GSE12995	High-risk B-ALL	11/42 (26%)	*ADAR, ARID1B, BTG, DLEU2, EBF1, ERG, HIST1, IKZF1, PAX5, TCF3, VPREB1*	P<0.002
(Roberts et al., 2012) GSE11877	Philadelphia-like B-ALL	8/42 (19%)	*BCR, EBF1, IGH, KLF2, PAX5, RCSD1, SH2B3, UHRF1*	P<0.034

A total of 27 genes were identified in the combined region list, including *IGH*, *IKZF1* and *TCF3* – see Figure 2-21 for custom ChIP-seq tracks of all 27 genes. A further 8 genes (shown in Figure 2-22) which have been reported to be mutated in B-ALL showed γH2AX signals were also present but the γH2AX signal was insufficiently strong when normalised to input to meet the region-calling criteria.

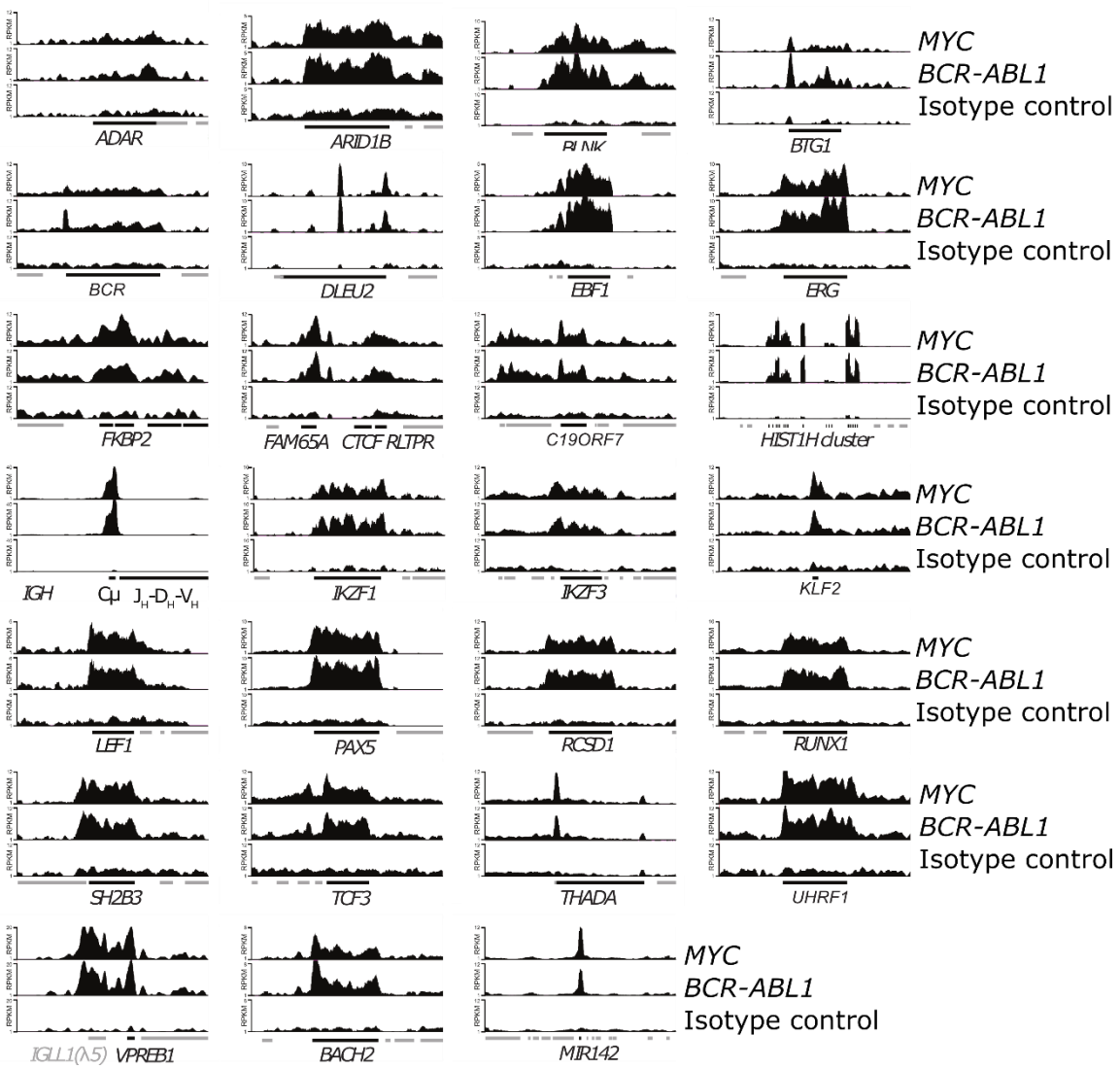


Figure 2-21: γH2AX enriched murine genes overlap with genes found to be abnormal in human B-ALL. ¶Custom ChIP-seq track images for the 27 genes listed in Table 2-1, showing reads from two biological repeats for each oncogene. The scale between genes varies but is constant for the *MYC, BCR-ABL1* and isotype control tracks within each gene. Figure by Dr Niklas Feldhahn.

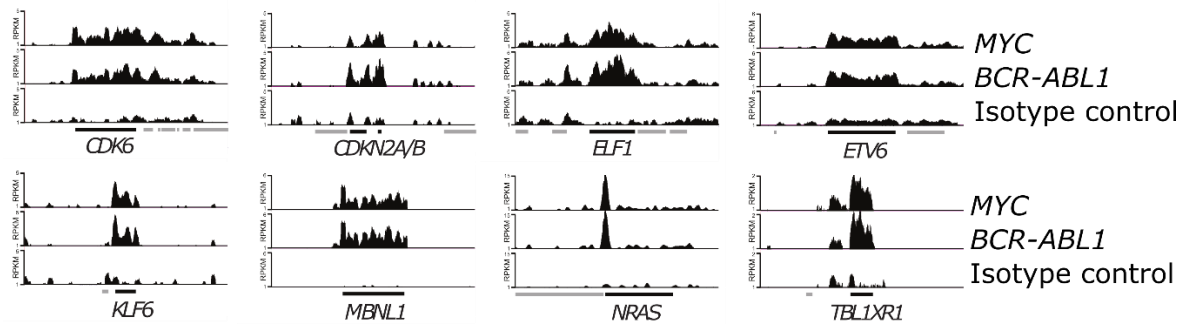


Figure 2-22: Further γH2AX marked murine genes overlap with genes found to be abnormal in human B-ALL.¶Custom ChIP-seq track showing an additional 8 genes reported to become defective in human B-ALL and which have γH2AX signals in B-cell precursors transformed by BCR-ABl1 and MYC but which did not pass the threshold for significance. The scale between genes varies but is constant for the *MYC, BCR-ABL1* and isotype control tracks within each gene. Figure by Dr Niklas Feldhahn.

2.4.4 γH2AX-marked regions coincide with expressed genes

Whilst analysing the function of the genes that overlapped with γH2AX, it became apparent that they were typically highly expressed in B-cell precursors (shown in Figure 2-16). In order to determine if DNA damage generally coincided with expressed genes, the exact location of the γH2AX signals was thus analysed by meta-gene analysis. This analysis showed that the γH2AX signal was infrequently detected in intergenic regions but frequently detected over gene bodies (as illustrated in Figure 2-23).

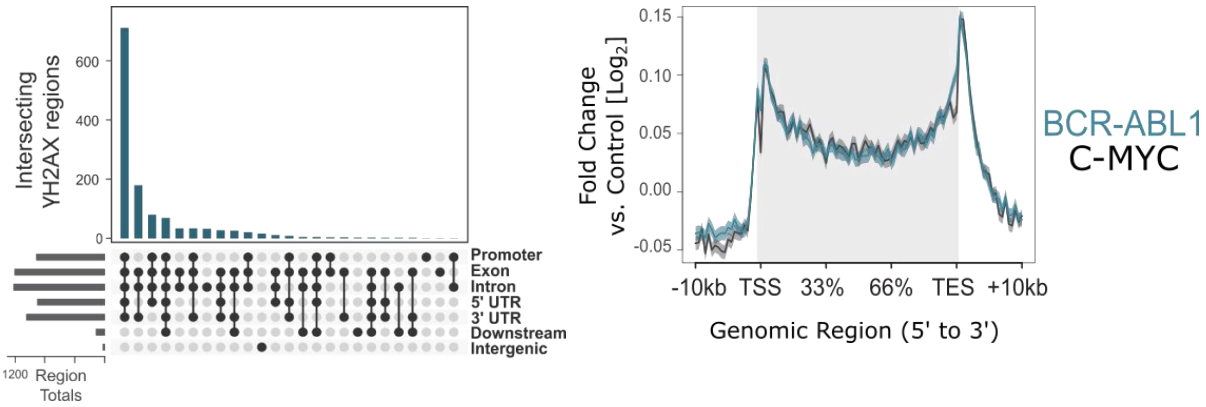


Figure 2-23: γH2AX regions often specifically covered the gene body.¶Meta-gene analysis of the distribution of γH2AX reads over the gene body; signals are normalised to the isotype control. (L) Overlap of the γH2AX regions with specified gene features. (R) Distribution of γH2AX reads in the gene bodies versus upstream or downstream regions. Bioinformatic analysis and figure by Dr Mark Robinson.

The gene body specific pattern observed suggested transcription as a potential cause for γH2AX region formation. We hence investigated the conjunction of γH2AX region formation

and transcription. In order to do so, H3K27Ac signals in transformed cells were analysed by ChIP-seq (n=1 for each oncogene). The number of uniquely aligned reads and the quality scores for these libraries are shown in Figure 2-24 and Figure 2-25. H3K27Ac is a histone modification that is primarily used for the identification of active enhancers but is also present at promoters. Akin to γH2AX signals, the H3K27Ac-ChIP-seq signals in transformed B-cell precursors located to transcriptional start sites in both of the transformed B-cell precursor populations, as shown in Figure 2-26.

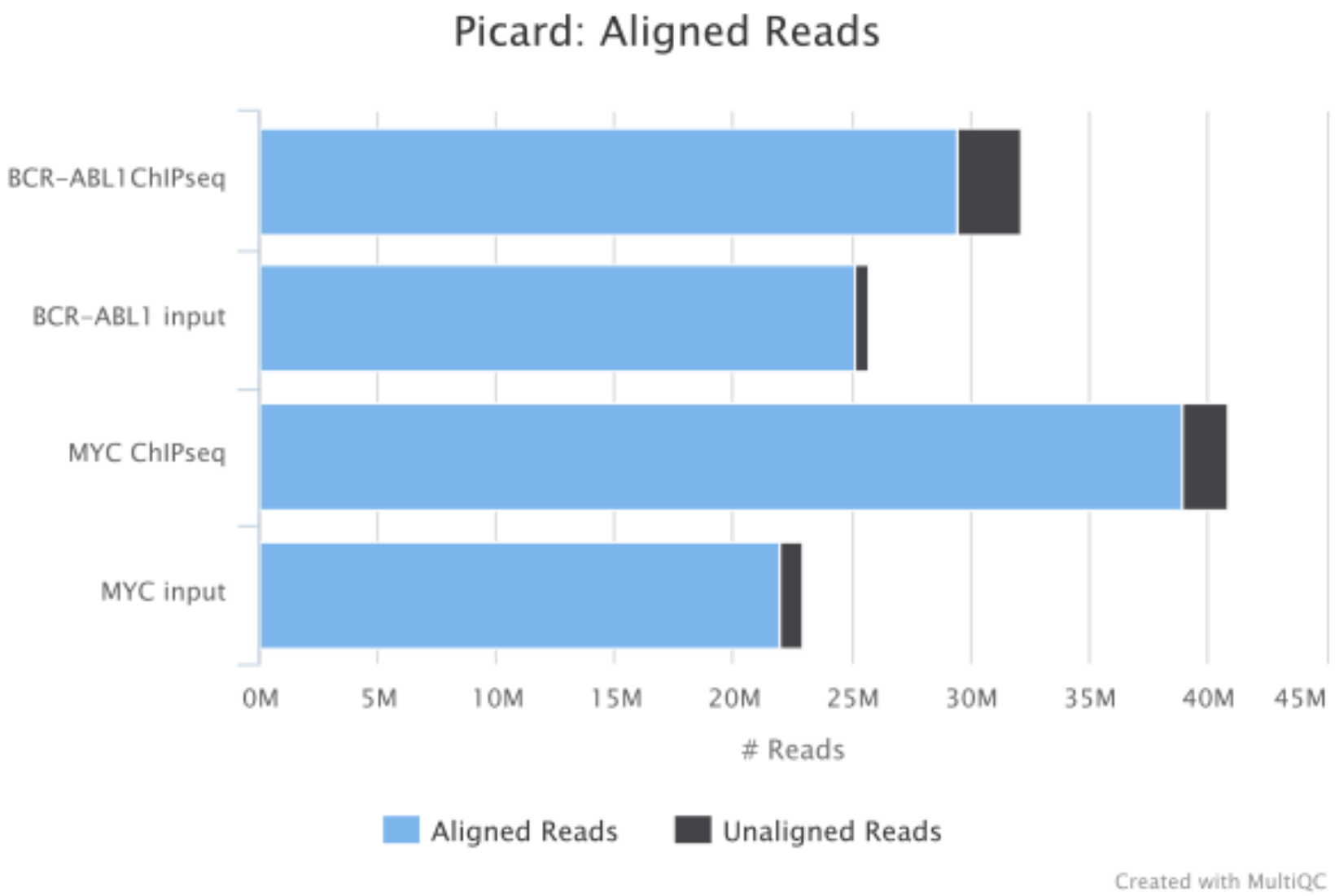


Figure 2-24: Picard alignments show that each of the day 7 H3K27Ac ChIP-seq libraries yielded a minimum of 29 million uniquely mapped reads.¶Two corresponding input libraries had a minimum of 22 million uniquely mapped reads. Figure generated using (Ewels et al., 2016)

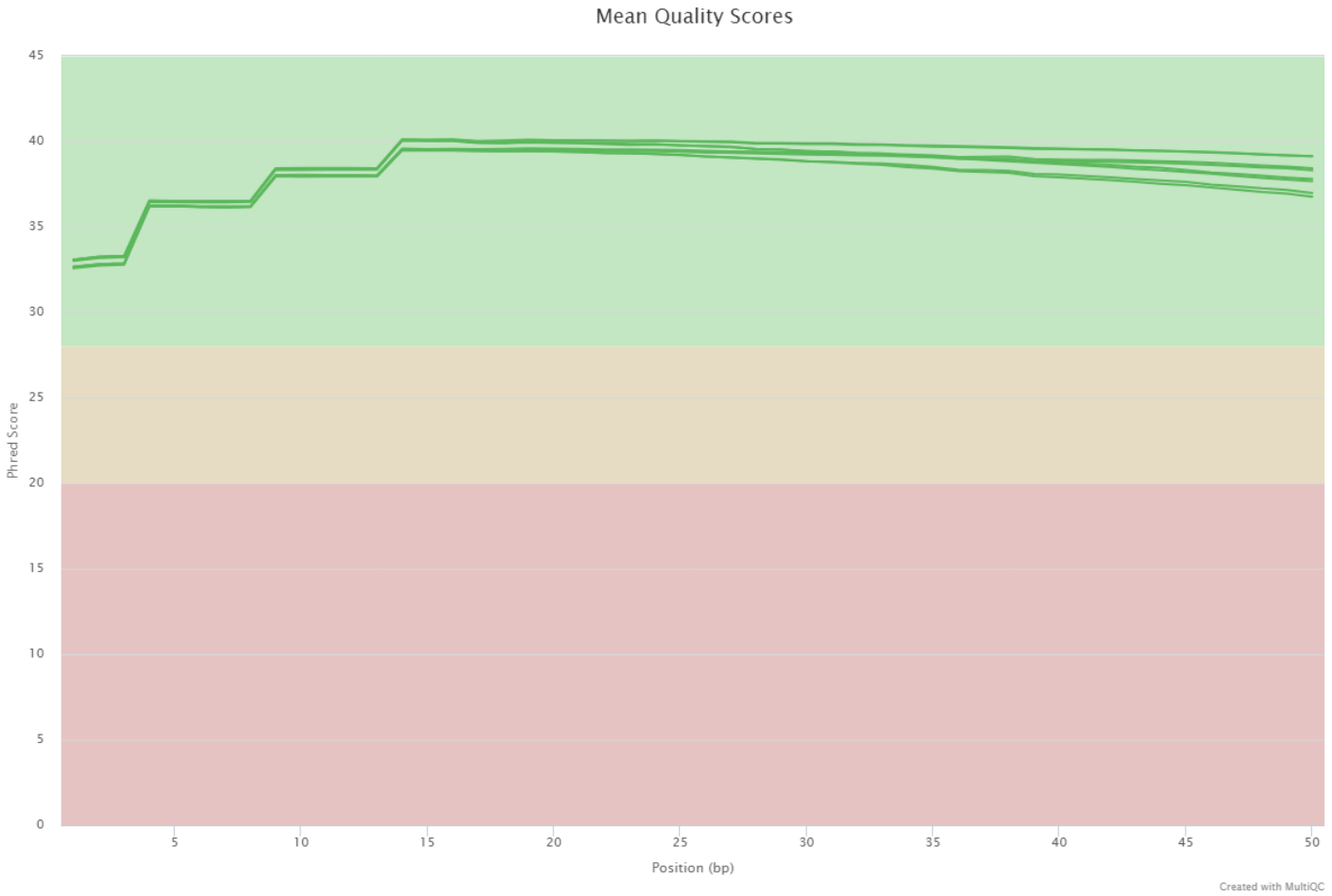


Figure 2-25: Both of the H3K27Ac ChIP-seq libraries and both of the input libraries were of very good quality with Phred scores >30.¶Figure generated using (Ewels et al., 2016).

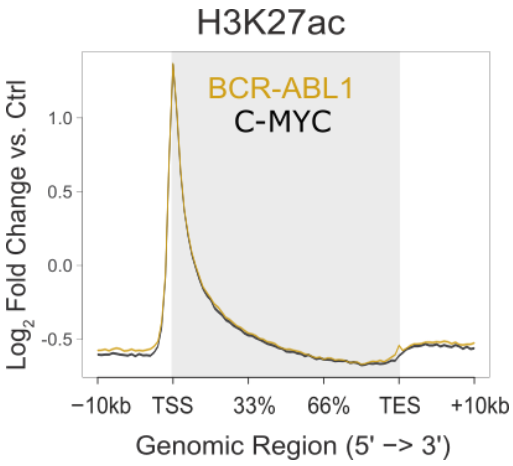


Figure 2-26: H3K27Ac reads frequently occur at the promoter.¶ Meta-gene analysis of the distribution of H3K27Ac reads over the gene body. Bioinformatic analysis and figure by Dr Mark Robinson.

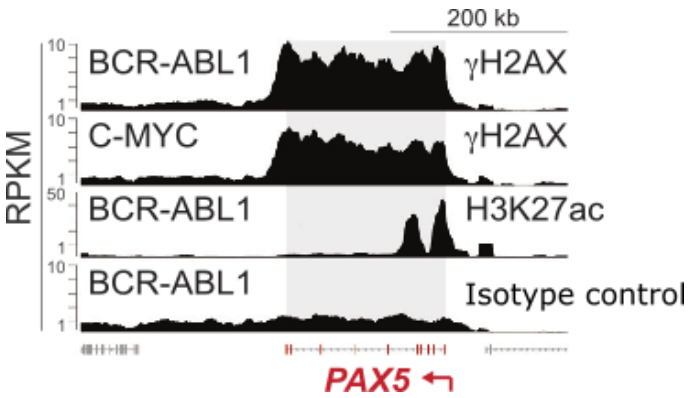


Figure 2-27: H3K27Ac and γH2AX regions frequently overlap.¶ Custom ChIP-seq track showing the comparative density of γH2AX, H3K27Ac and isotype control reads for *PAX5* in *BCR-ABL1* and *MYC* expressing B-cell precursors. Figure by Dr Niklas Feldhahn.

As depicted in Figure 2-27, visual analysis showed that γH2AX-enriched genes displayed a distinctive H3K27Ac peak at the promoter region. In order to better quantify the overlap of the two signals the association was analysed using bioinformatics. In order to do so, normalised γH2AX signals were further processed to yield 10kb γH2AX regions which either overlapped, or did not overlap, H3K27Ac signals; 85% of *BCR-ABL1* regions and 86% of *MYC* reads were found to overlap.

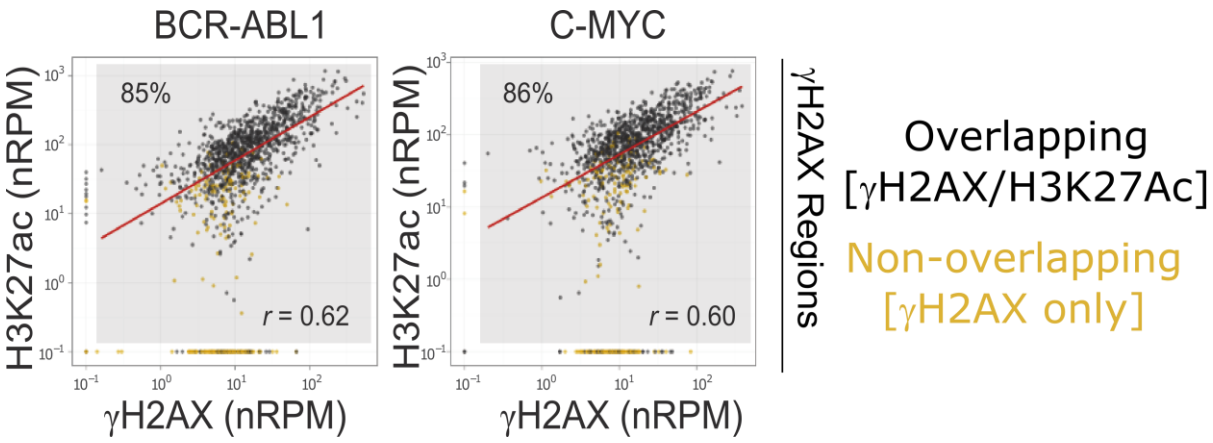


Figure 2-28: Overlapping H3K27Ac and γH2AX regions have comparable signal intensity. ¶Each black dot indicates a 10Kb γH2AX region which overlaps an H3K27Ac signal; yellow dots indicate a 10Kb γH2AX region which does not overlap with an H3K27Ac signal. The scatterplots show γH2AX signal intensity on the x-axis and H3K27Ac signal intensity for γH2AX regions on the y-axis, for BCR-ABL1 (left plot, Pearson's correlation *r*=0.62) and MYC (right plot, *r*=0.60) expressing populations. The red line indicates linear regression. (The dots within the grey are those above background, 85% and 86%, respectively). Bioinformatic analysis and figure by Dr Mark Robinson.

There was a significant correlation (r=0.62 and 0.60; p<02.2x10^{-16} for both) between the intensity of the normalised signals from γH2AX signals compared to H3K27Ac signals for the γH2AX regions in BCR-ABL1 and MYC expressing populations, respectively, as shown in Figure 2-28.

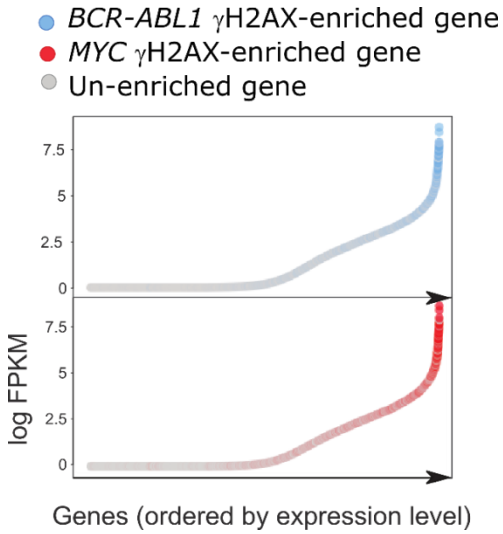


Figure 2-29: γH2AX-enriched genes were highly expressed. ¶The correlation between gene expression level (ranked by increasing fragments per kilobase of transcript per million reads, FPKM). Genes enriched by γH2AX are indicated in blue (BCR-ABL1, n=2) and red (MYC, n=2). Bioinformatic analysis and figure by Dr Mark Robinson.

In an independent approach, γH2AX region formation and transcription was compared using RNA-Seq (n=2 for EV and for each oncogene). The number of uniquely mapped reads and the quality scores for these libraries are shown in
Figure 2-30 and Figure 2-31.

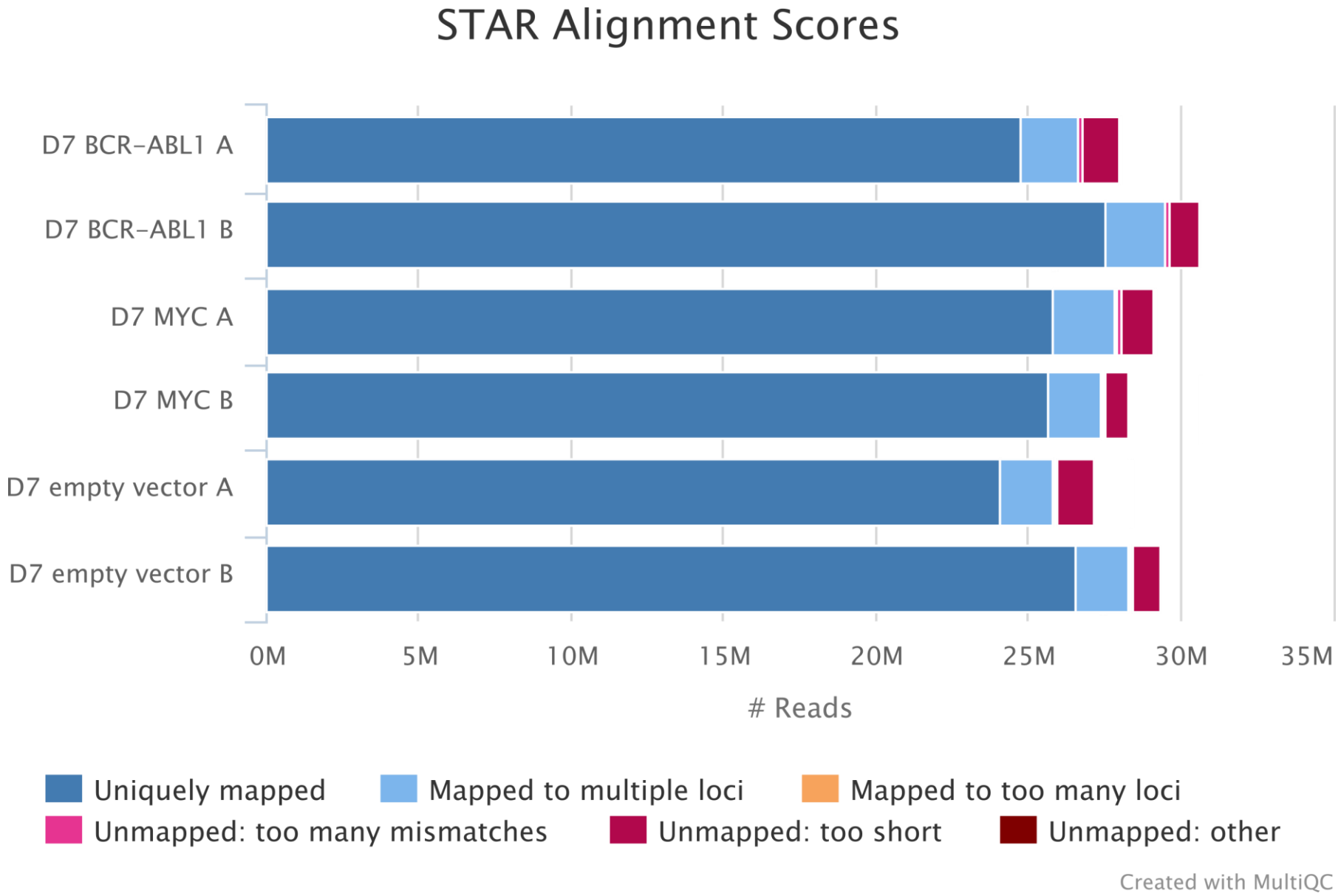


Figure 2-30: Spliced Transcripts Alignment to a Reference (STAR) alignment scores show that each RNA-seq library yielded a minimum of 24 million uniquely mapped reads. ¶Figure generated using (Ewels et al., 2016)

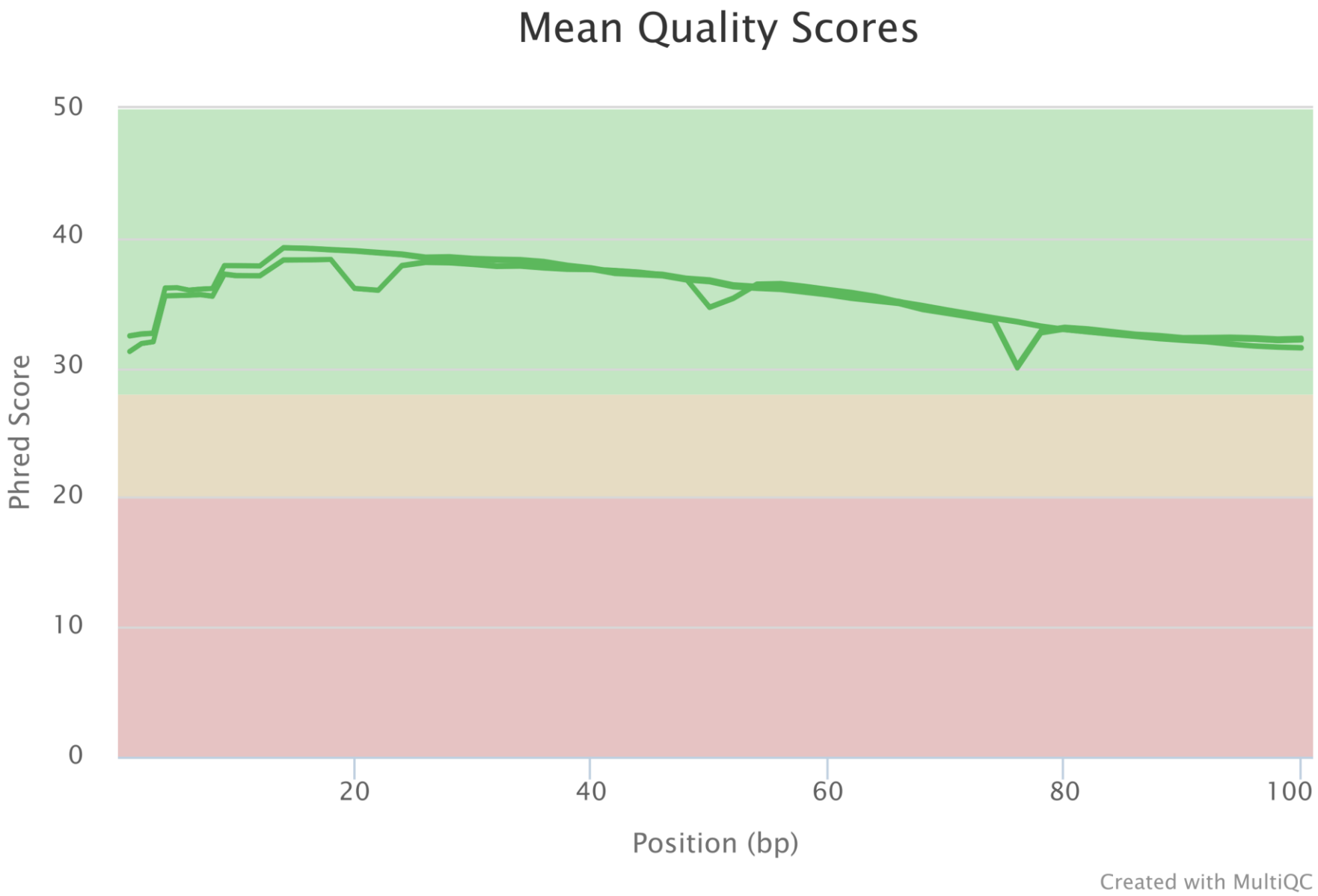


Figure 2-31: RNA-seq libraries were all of very good quality with Phred scores >30. ¶Figure generated using (Ewels et al., 2016)

As in all other experiments, messenger RNA (mRNA) used for RNA-seq analysis was taken from apoptotic/dead cell depleted BCR-ABL1- or MYC-expressing B-cell precursors on day 7 post oncogene induction. RNA-Seq analysis confirmed that γH2AX regions were enriched amongst the highly expressed genes (Figure 2-29).

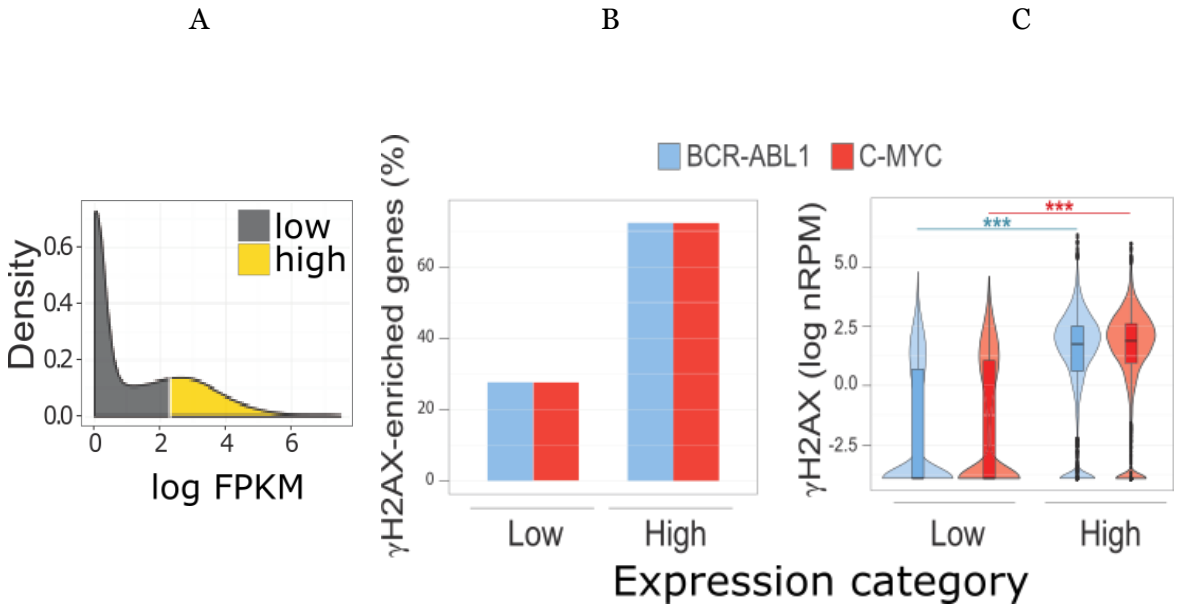


Figure 2-32: Highly expressed genes are enriched for γH2AX.¶(A) Kernel density plot of FPKM values showing high and low expression categories split based on median expression. (B) Bar chart showing the relative distribution of γH2AX enriched genes in these low and high expression categories. (C) Violin plot showing the majority of the highly expressed genes had a high γH2AX signal whereas the γH2AX signal in the lowly expressed genes showed a wider distribution ($p<1x10^{-15}$ using the Mann-Whitney *U* test). Bioinformatic analysis and figure by Dr Mark Robinson.

Genes were then categorised as high or low expression based on their median expression (Figure 2-32A). The bar chart in Figure 2-32B shows that the 73% of γH2AX enriched genes were highly expressed and similarly in Figure 2-32C, the γH2AX signal intensity was increased with higher expression.

Interestingly, the small proportion of genes that showed differential expression in the separate *BCR-ABL1* (signals highlighted in blue) and *MYC* (signals highlighted in red) transformed populations also showed a corresponding difference in signal intensity (shown in Figure 2-33).

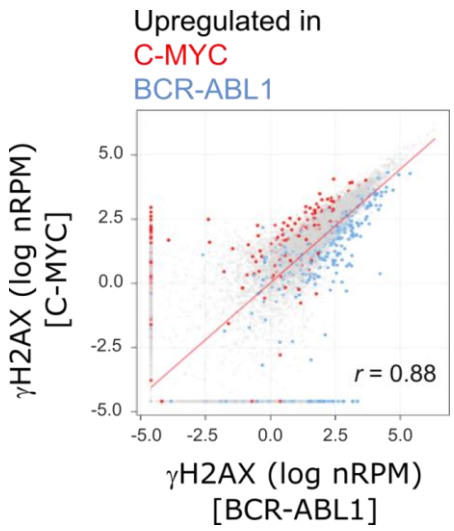


Figure 2-33: Genes which were expressed differentially in *BCR-ABL1* and *MYC* also had an associated change in γH2AX signal intensity. ¶Scatterplot showing the difference in nRPM γH2AX signal across the extended gene regions between MYC and BCR-ABL1 populations. Genes upregulated in MYC are shown in red, those upregulated in BCR-ABL1 are shown in blue and the remainder in grey. The red line indicates linear regression. Bioinformatic analysis and figure by Dr Mark Robinson.

In order to functionally validate these findings, expression of a DsRed transgene was induced using doxycycline in *MYC* transformed cells (see Figure 2-34 for experimental overview). The expression of the transgenic red fluorescence protein (see Figure 2-36) triggered an accumulation (normalised to a rabbit IgG isotype control) of γH2AX at the expression cassette of doxycycline-treated cells that was 2.1 times greater than that of untreated control cells, which was detectable by γH2AX ChIP-qPCR (Figure 2-35).

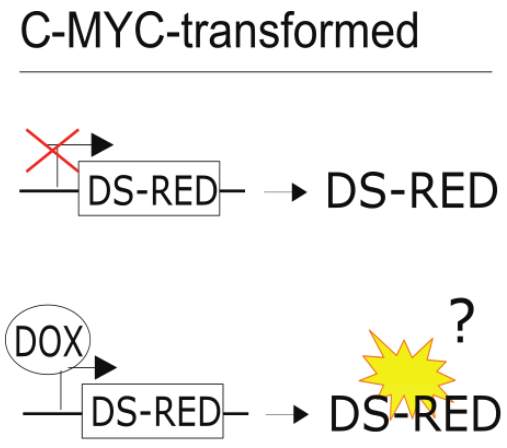


Figure 2-34: Outline of the induced DsRed expression experiment

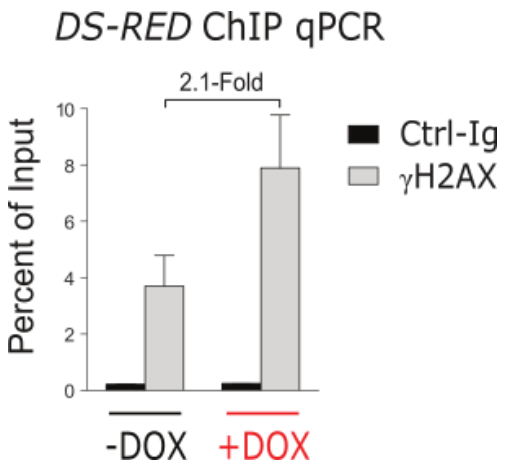


Figure 2-35: Induced expression of a gene results in increased DNA damage. ¶Bar chart showing the increase of DsRed sequence in doxycycline treated (labelled in red) and untreated (labelled in black) cells as determined by γH2AX ChIP-qPCR and normalised to corresponding input DNA. PCR and figure by Dr Jenny Thomas.

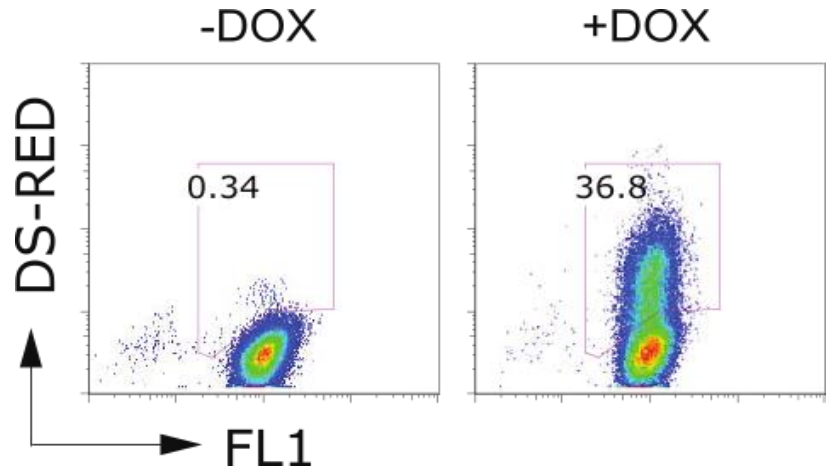


Figure 2-36: Doxycycline treatment resulted in expression of DsRed. ¶ Flow cytometric confirmation of doxycycline-induced expression of DsRed in MYC-transformed B-cell precursors. Analysis and figure by Dr Jenny Thomas.

2.4.5 Transcription-related fragility contributes to the DNA damage

As described in section 1.3, there are several established causes of transcription-related fragility; the γH2AX regions were therefore analysed to determine the effect of transcription-related fragility on the lineage-specific damage.

First, in light of recent evidence which shows that the torsional stress of transcription varies according to the proximity and direction of transcription neighbouring genes, an analysis of orientation and proximity was undertaken (Hamperl et al., 2017). Genes which occurred within 10kb were considered to be gene pairs and each gene pair was then categorised as transcribed in tandem, convergently or divergently; those with no additional gene within 10kb were categorised as isolated, as depicted in Figure 2-37. The number of γH2AX signals was indeed increased in paired genes compared to those in isolation consistent with published reports (Meng et al., 2014; Pannunzio and Lieber, 2016). However, the orientation of the gene pairs did not further increase the γH2AX signals.

Consistent with the recognised increased fragility of gene pairs, however, the *intensity* of the γH2AX signals was increased at the convergent or divergent gene pairs compared to the tandemly arranged pairs (see Figure 2-38 ; $p=<1 \times 10^{-6}$ for convergent versus tandem and divergent vs tandem using the Mann-Whitney test).

	γH2AX regions:	Randomly expected:
Tandem	24%	16%
Convergent	16%	10%
Divergent	11%	8%
Isolated	30%	54%
Overlapping	19%	12%

Figure 2-37: γH2AX regions were enriched for genes transcribed in close proximity.¶ Gene pairs were considered to be within 10kb. Bioinformatic analysis and figure by Dr Mark Robinson.

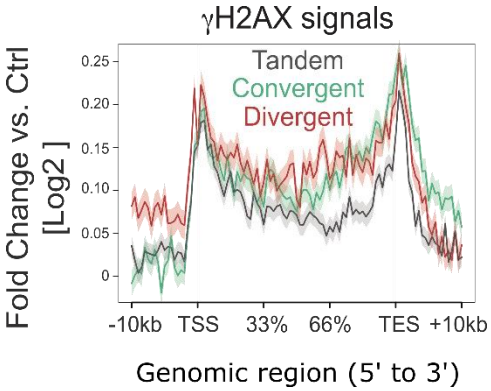


Figure 2-38: The intensity of γH2AX signals was increased for convergent and divergent orientated genes compared to tandem orientated.¶The mean fold change in BCR-ABL1 versus isotype control is shown, normalised for gene length. Bioinformatic analysis and figure by Dr Mark Robinson.

Second, given that the formation of R-loops aggravates the conflict between transcription and replication and can increase the risk of DNA damage, the propensity of the sequence within the γH2AX regions to form R-loops was assessed *in silico*. R-loop forming sequences (RLFSs) were predicted bioinformatically as previously described and γH2AX regions were analysed for enrichment of these RLFS. As shown in Figure 2-39 there was a significant enrichment of RLFSs in the γH2AX regions compared to all genes and compared to the highly expressed genes alone (both p=0.0001).

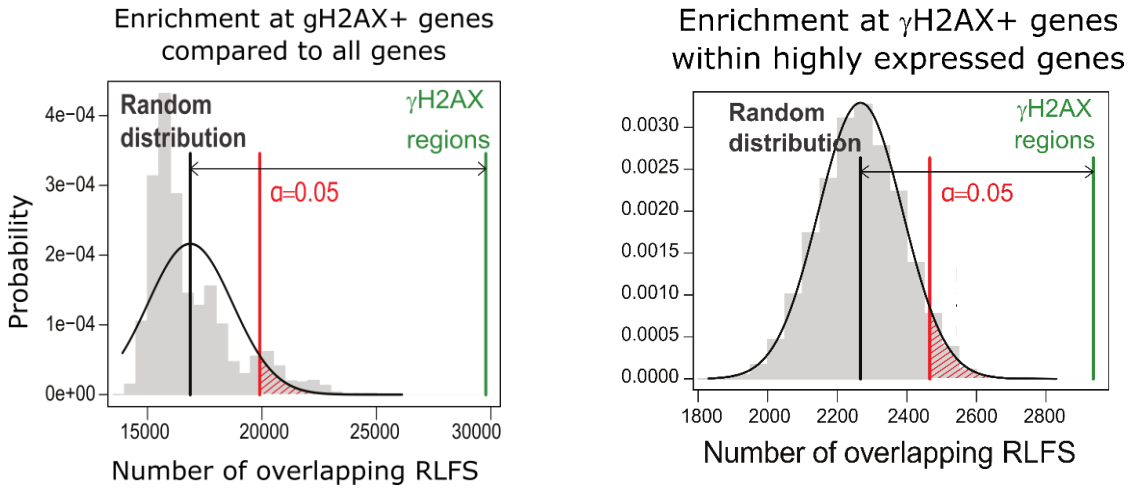


Figure 2-39: R-loop forming sequences were significantly enriched at γH2AX regions compared to all genes (left) and compared to the top 1,000 highly expressed genes only (right). The distribution of all predicted RLFSs is shown by the grey histogram by random sampling of 10,000 permutations. The vertical lines indicate the number of RLFS which overlap with a γH2AX region – the black line indicates the mean number of overlapping RLFS, the red line indicates the significance threshold (α=0.05) and the green line indicates the actual number of RLFSs overlapping with γH2AX - associated genes (p=0.0001). Bioinformatic analysis and figure by Dr Mark Robinson.

Since R-loops primarily cause DNA damage by promoting transcription-replication conflicts, the identified γH2AX regions were further analysed for overlap with early replicating fragile sites (ERFS). Breakage of ERFS has been linked with transcription-replication conflicts upon induction of replicative stress (Barlow et al., 2013). There was a significant overlap - with 55% of the ERFSs overlapping with 26% of γH2AX regions, as shown in Figure 2-40.

Hydroxyurea, which inhibits the production of deoxyribonucleotides and thus induces a form of replicative stress caused by stalled replication forks, was used by Barlow et al to identify ERFS. Interestingly, when comparing the actual γH2AX signal distribution on individual gene loci between the Barlow dataset and the current dataset, substantial differences were observed between hydroxyurea (HU)-treated and oncogene-expressing cells, potentially indicating a different mechanism of DNA damage, shown in Figure 2-41. However, as the γH2AX-ChIP-Seq data on HU-treated cells used for comparison was retrieved from proliferating mature B-cells, the comparison of the two datasets and its meaningful interpretation was limited.

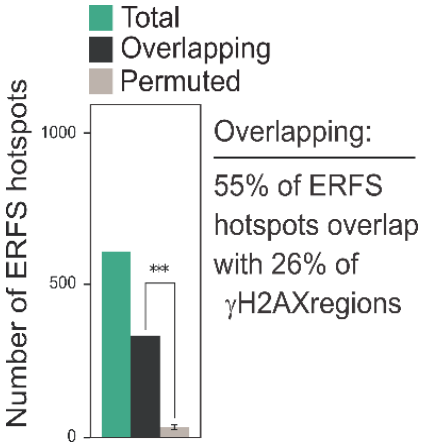


Figure 2-40: γH2AX regions overlapped with early replicating fragile sites.¶Bar diagram showing the ERFS hotspots (green) against the number which overlap with γH2AX regions (black). Significance derived from the permuted number from random re-sampling (grey). Bioinformatic analysis and figure by Dr Mark Robinson.

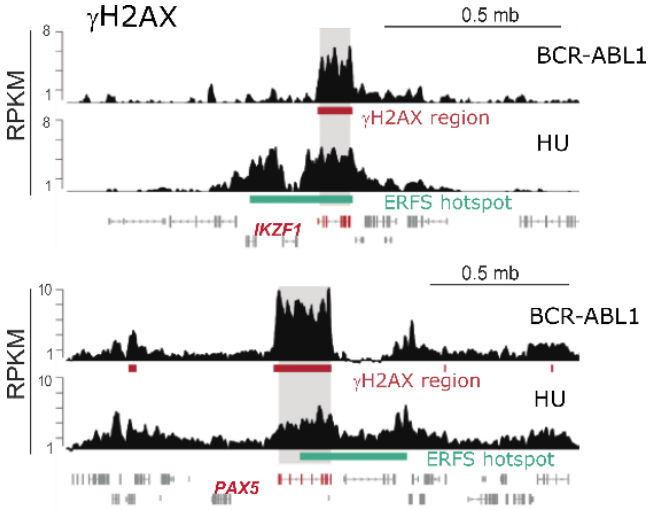


Figure 2-41: Hydroxyurea-induced γH2AX signals were different from overlapping oncogene-induced γH2AX signals.¶Custom ChIP-seq γH2AX track showing the overlap between γH2AX reads from hydroxyurea (green) and *BCR-ABL1* (red) induced early replicating fragile sites (ERFS) at two representative loci (*IKZF1* and *PAX5*). Figure by Dr Niklas Feldhahn.

2.4.6 Active transcription predisposes a gene to sustaining DNA damage

Results so far suggested that oncogene expressing B-cell precursors experience transcription-associated DNA damage. As a logical consequence, B-lineage specific genes that are induced by B-cell specific differentiation programs in B-cell precursors become prone to DNA damage. This raised the question as to whether a change in lineage identity, for example towards the myeloid lineage, would cause a change in H2AX hotspot formation. A *BCR-ABL1*-transformed murine B-cell precursor cell line generated by the Müschen Laboratory (Chen et al., 2015) that carried a doxycycline-inducible *CEBP*A transgene was used to functionally investigate the proposed link between lineage-specific transcription and DNA damage (see Figure 2-42 for an overview of the experiment). CEBPA is a myeloid transcription factor, the expression of which has been shown to efficiently convert B-lymphoid precursors towards the myeloid lineage (Cammenga et al., 2003; Di Tullio et al., 2011; Zhang et al., 2004). Treatment with doxycycline induced the expression of *CEBPA* and thus triggered a switch from CD11B- lymphoid phenotype to a CD11B+ myeloid phenotype. Doxycycline treated cells were further enriched using anti-CD11B magnetic bead cell sorting to reach a purity of >90% CD11B+ cells. Sorted myeloid and lymphoid fractions of BCR-ABL1-expressing cells were further depleted of dead/apoptotic cells and used for γH2AX-ChIP-seq analysis as before (see Figure 2-43).

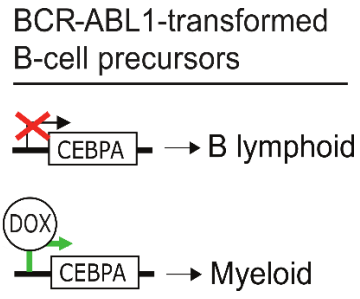


Figure 2-42: Outline of the myeloid lineage induction experimental approach.

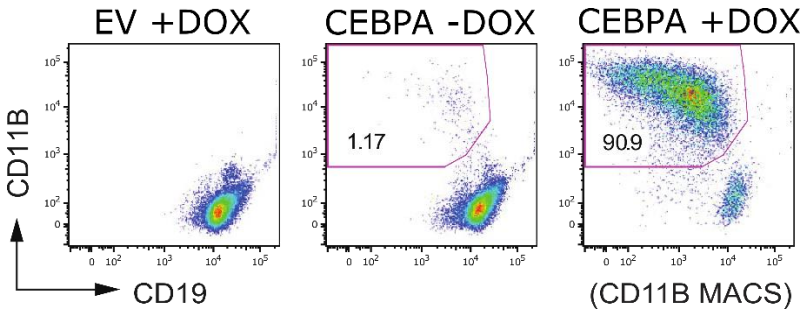


Figure 2-43: Flow cytometric analysis of the untreated and treated populations¶, verifying that doxycycline treatment of B-cell precursors with the CEBPA transgene induces a myeloid phenotype. Treated cells were purified (>90%) with anti-CD11B magnetic beads prior to ChIP-seq. Figure by Dr Niklas Feldhahn.

It was therefore possible to examine the de novo DNA-damage sites, again using γH2AX-ChIP-seq and DNA-FISH, which potentially arose as a consequence of the induced lineage switch. The number of uniquely aligned reads and the quality scores for these libraries are shown in Figure 2-44 and Figure 2-45.

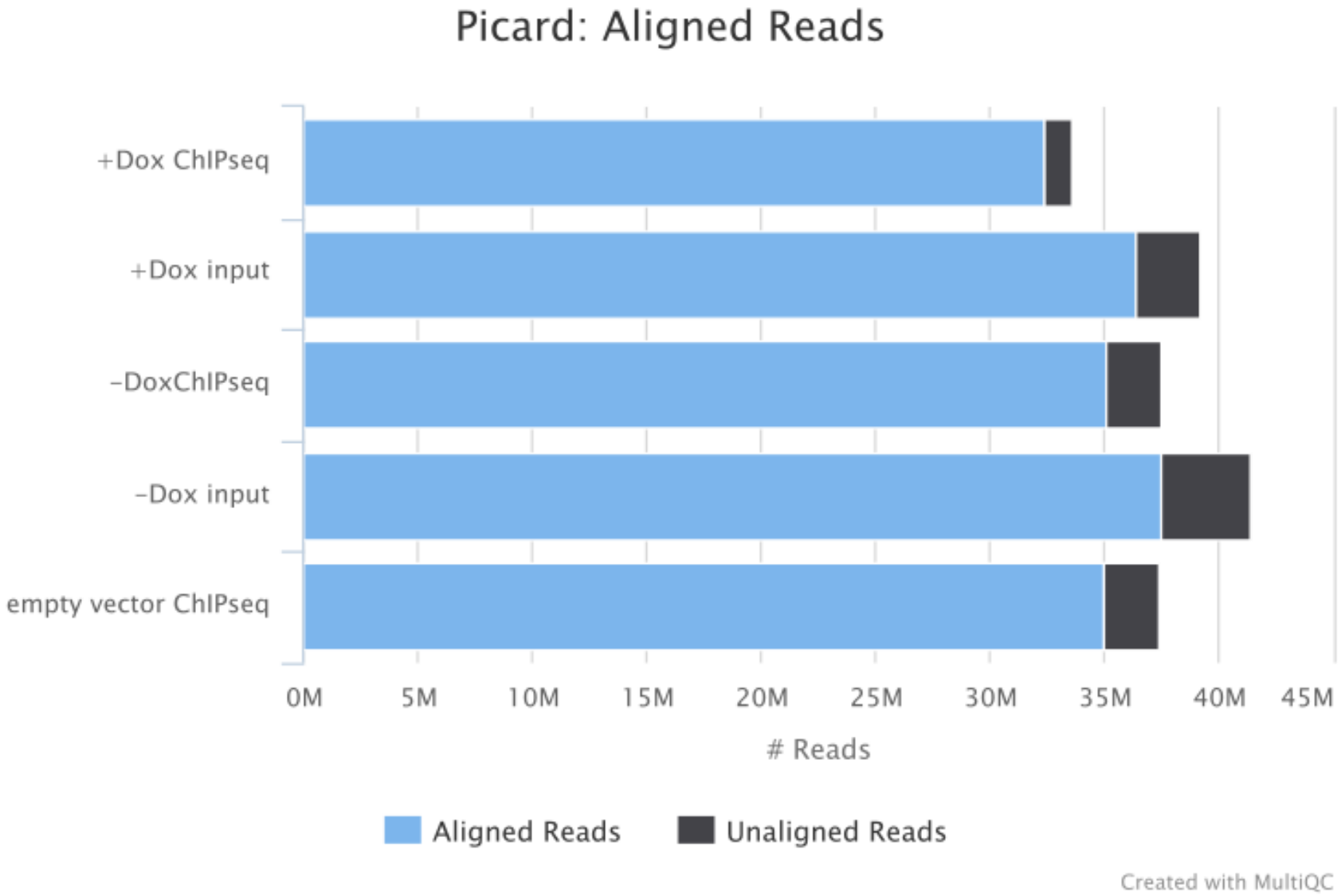


Figure 2-44: Picard alignments show that each of the three γH2AX ChIP-seq libraries yielded a minimum of 32 million uniquely mapped reads.¶Two corresponding input libraries had a minimum of 36 million uniquely mapped reads. Figure generated using (Ewels et al., 2016)

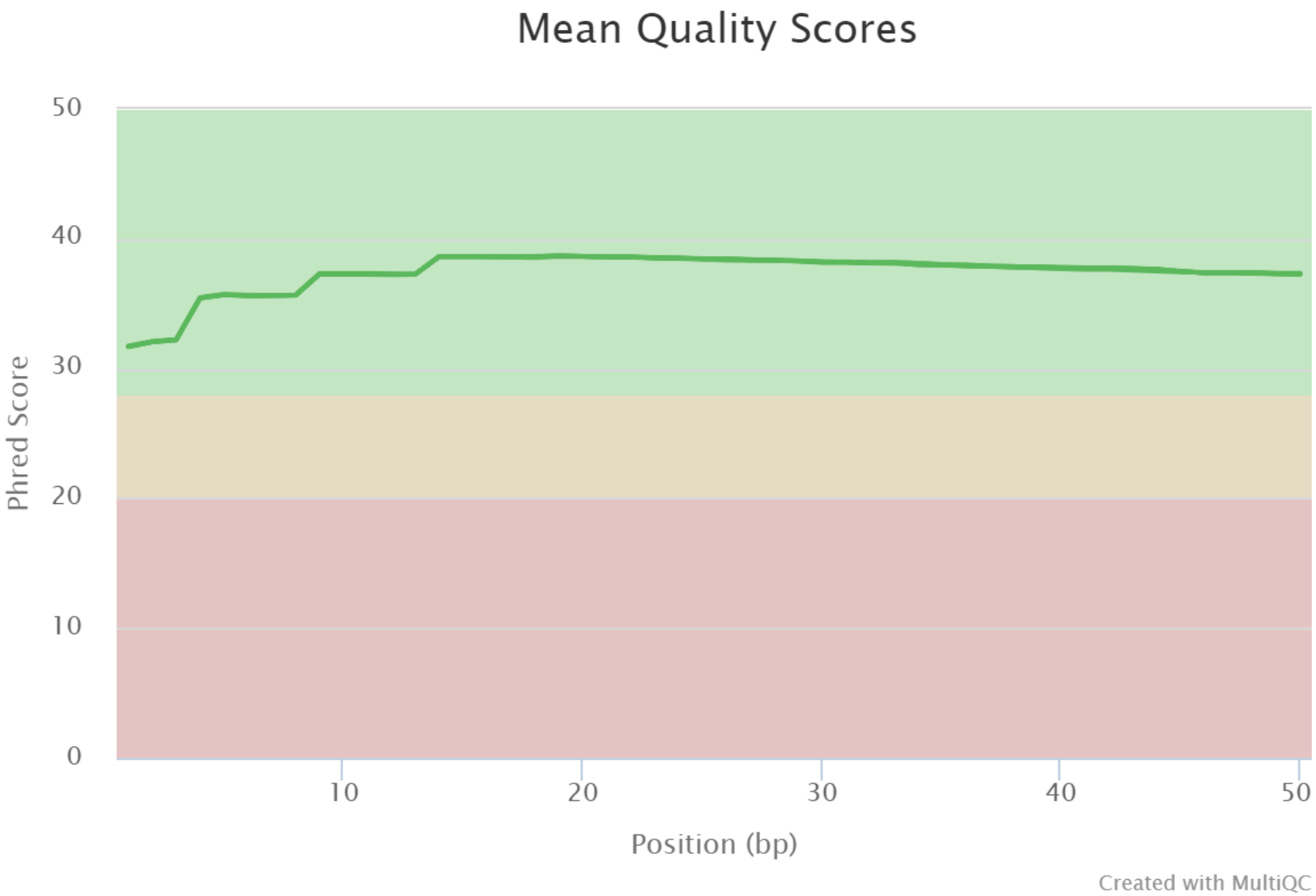


Figure 2-45: All three γH2AX ChIP-seq libraries and both input libraries were all of very good quality with Phred scores >30.¶Figure generated using (Ewels et al., 2016).

As predicted, the myeloid-converted fraction of BCR-ABL1+ cells (CD11B+, CEBPA+DOX) contained de novo γH2AX-ChIP-seq hotspots at myeloid-specific genes (a representative ChIP-seq track for *LYZ2* is shown on the right of Figure 2-46) that were not present in the lymphoid fractions analysed as controls. In contrast, γH2AX hotspots at non-lineage specific genes were present in all fractions (the *HIST1* cluster is shown on the left), γH2AX hotspots at lymphoid lineage specific genes showed only minor reductions of γH2AX signal in the myeloid-converted cells (*IGH* is shown in the middle) which is likely to be due to retention of some lymphoid phenotype (cells analysed are CD11B+ but CD19^{low}).

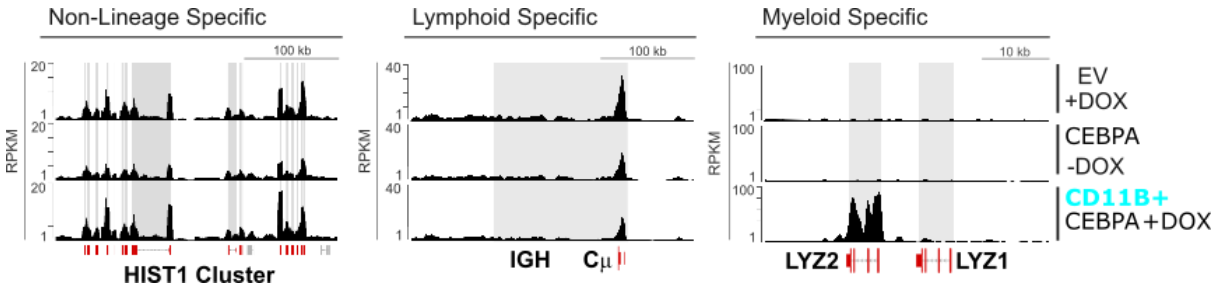


Figure 2-46: Induction of a myeloid phenotype caused DNA damage at myeloid-specific genes.¶ Custom ChIP-seq track showing γH2AX reads for the *HIST1* cluster (non-lineage specific, left), *IGH* (lymphoid specific, middle) and *LYZ2* (myeloid specific) in populations with lymphoid and induced myeloid (highlighted in blue, bottom track) phenotypes. The gene bodies are highlighted in grey. Figure by Dr Niklas Feldhahn.

DNA-FISH was performed, as described in section 2.4.1, to confirm the γH2AX signals seen by ChIP-seq represented DNA damage but instead using two myeloid hot spots (*LYZ2* and *THBS1*) and the same coldspot (*FOXP2*). Again, a significantly increased number of split signals were scored in the two myeloid hotspots in the myeloid-switched populations compared to the lymphoid populations. No significant increase was detected in the coldspot (shown in Figure 2-47).

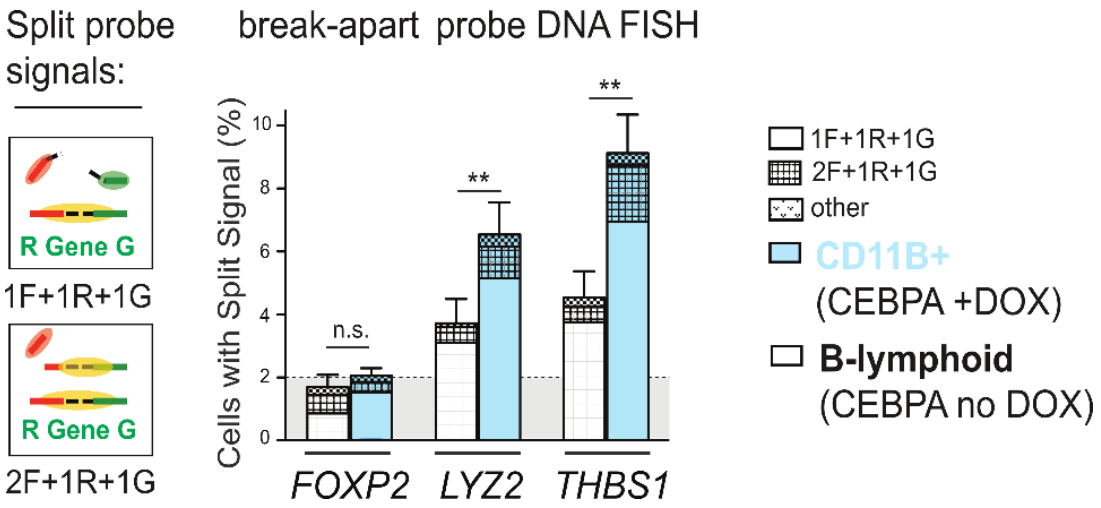


Figure 2-47: Myeloid hotspots correlate with sites of active DNA damage in myeloid switched cells. Bar chart showing results of FISH experiments (n=5) using dual colour break apart probes for one cold spot (*FOXP2*) and two hotspots (*LYZ2* and *THBS1*) on untreated lymphoid (white) and induced myeloid (blue) populations.

Analysis of the ten genes with the highest de novo γH2AX signals in myeloid-converted cells showed that all of these genes associated with functions in myeloid cells (Table 2-2).

Table 2-2: The highest ranked H2AX regions in switched myeloid populations were enriched for myeloid specific function

Gene	**Fold change**	**Function**	Reference
THBS1	27.7	Crucial for IL10 production; expressed by macrophages	(Zhao et al., 2014)
LYZ2	11.6	Phagocytosis; LYSM promoter used for myeloid transgenes	(Clausen et al., 1999)
PSAP	6.0	Lysosomal trafficking; expressed by myeloid cells	(Zhou et al., 2015)
TXNIP	5.8	Inflammasome; ROS production; expressed by macrophages	(Park et al., 2013)
SIRPA	5.8	Inhibitory receptor; expressed on macrophages	(Veillette et al., 1998)
PLIN2	5.5	Expressed by macrophages	(Feingold et al., 2010)
CD33	5.3	Highly expressed on myeloid progenitor cells	(Paul et al., 2000)
SLFN2	5.1	Mutation of SLFN2 caused myeloid immunodeficiency	(Berger et al., 2010)
TRIB1	5.1	Overexpressed in AML; causally linked to leukemogenesis	(Yokoyama et al., 2010)
MMP8	5.0	Macrophage differentiation; mutated in AML	(Kim et al., 2014)

Gene ontology pathway analysis of the complete set of de novo γH2AX hotspots in myeloid-converted cells further confirmed that genes overlapping with de novo γH2AX regions were strongly enriched for myeloid-specific pathways.

3 Epigenetic reprogramming during leukemic transformation of B-cell precursors

3.1 Contribution of others to this chapter

All bioinformatics analysis (including all figures except ChIP-seq tracks) was undertaken by Dr Mark Robinson. I performed all cell culture, ChIP-seq and RNA-seq experiments.

3.2 Hypothesis

BCR-ABL1 expression in murine B-cell precursors induces histone modification and novel or aberrant enhancers during stable transformation, via activity of the kinase and/or due to the selection pressure during clonal outgrowth of the successfully transformed cells; these enhancers are conserved in human Philadelphia-positive lymphoid leukaemia.

3.3 Background

There is increasing evidence that changes in epigenetic regulation play an important role in leukaemia development. H3K27Ac ChIP-seq and RNA-seq were therefore used to explore the contribution of epigenetic alterations to the stable transformation of the B-cell precursors, with special interest on the activation of novel enhancers following stable transformation.

3.4 Results

3.4.1 Oncogene expressing populations had similar H3K27Ac peaks during oncogenic crisis but remarkably different peaks once stably transformed

In order to screen for novel epigenetic alterations that contribute to successful transformation separate populations of the oncogene-expressing cells analysed in Chapter 2 were maintained in culture until they were stably transformed to a cell line (referred to from now on as day 60+) – this is summarised in Figure 3-1.

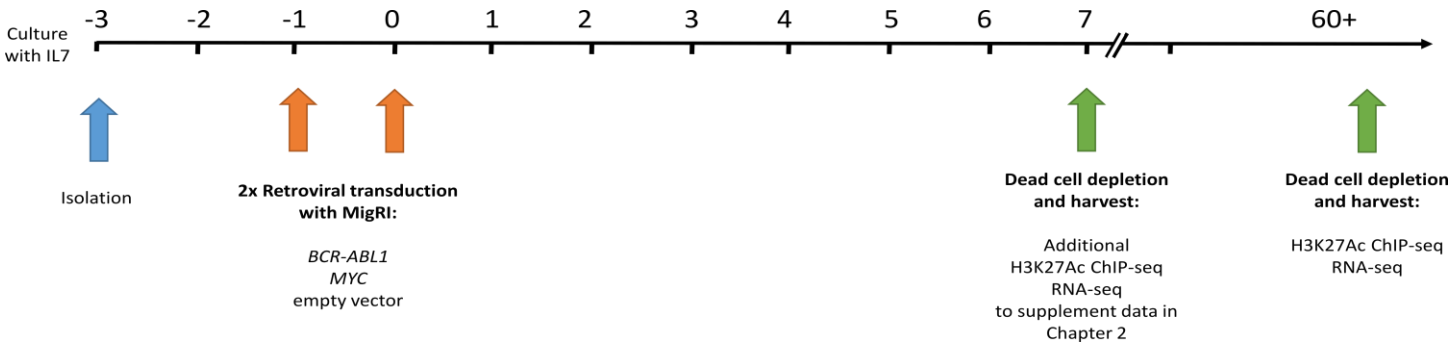


Figure 3-1: Outline of the stable transformation experimental approach. ¶Whole murine bone marrow was cultured with IL7 and transduced with retroviral vectors expressing an oncogene. On day seven post infection when they were suffering from oncogenic stress, in order to increase the datasets from Chapter 2, cells were harvested for: H3K27Ac ChIP-seq (n=1 for EV) and RNA-seq (total n=4 for EV and n=5 for both oncogenes). On day 60+ when cells had outgrown from the period of oncogenic crisis and were successfully transformed, cells were harvested for: H3K27Ac ChIP-seq (n=2 for both oncogenes) and RNA-seq (n=3 for both oncogenes).

At this point, H3K27Ac ChIP-seq was undertaken, as before. This enabled a comparison between empty vector infected B-cell precursors (cells infected with empty vector as control and harvested 7 days after transduction; n=1), oncogene-expressing cells which were in a period of acute oncogenic stress (*BCR-ABL1* or *MYC* transduced cells harvested 7 days after transduction and experiencing a temporary phase of stress described in chapter 2; n=1 for each oncogene), and oncogene-expressing cells which had survived the oncogenic stress to reach stable transformation (*BCR-ABL1* or *MYC* transduced cells harvested ~60 days after transduction and exhibiting >90% viability; n=2 for each oncogene). The number of uniquely aligned reads and quality scores for the new ChIP-seq libraries and input libraries is shown in
Figure 3-2 and Figure 3-3.

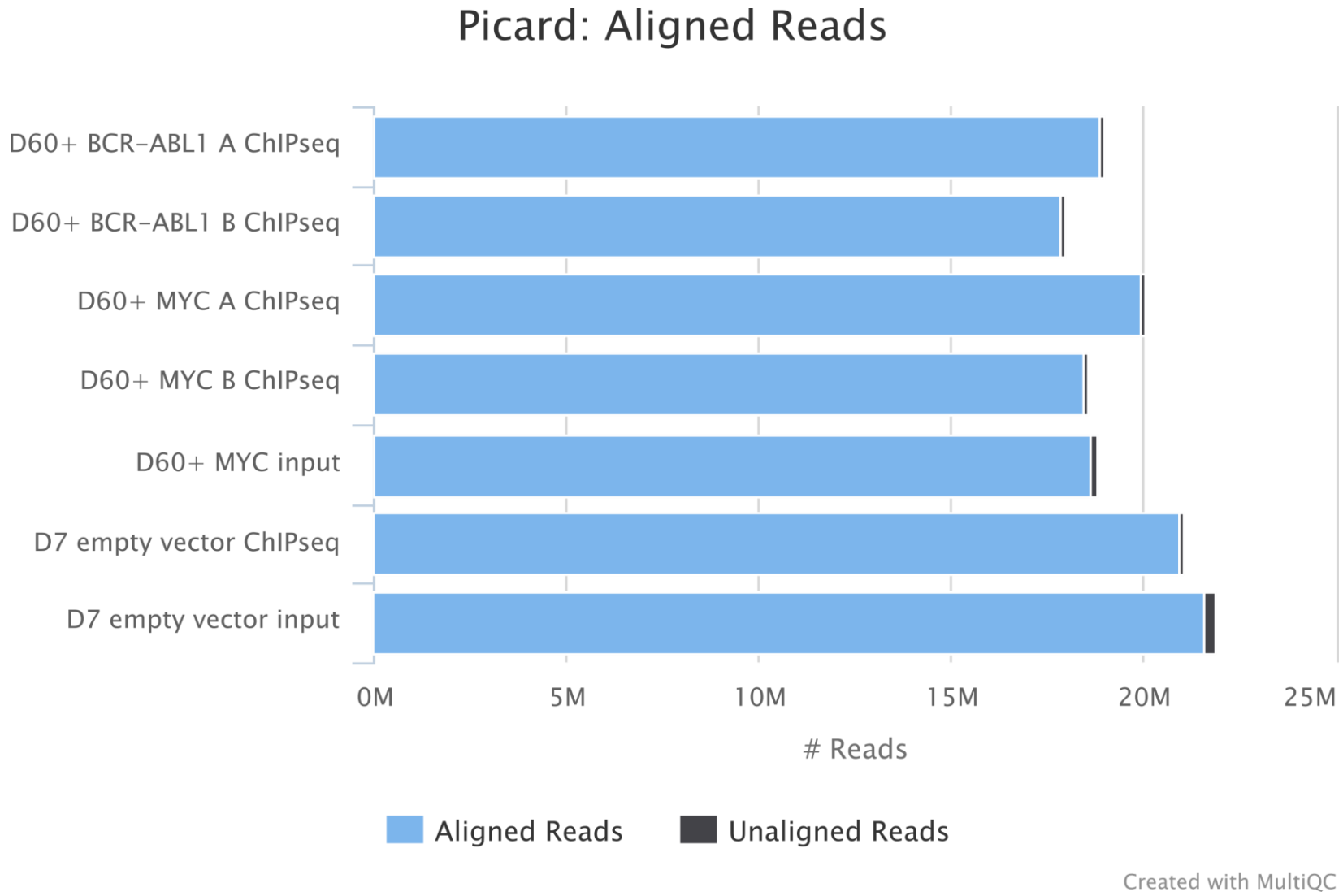


Figure 3-2: Picard alignments show that each of the five H3K27Ac ChIP-seq library yielded a minimum of 34 million uniquely mapped reads.¶Two corresponding input libraries had a minimum of 36 million uniquely mapped reads. Figure generated using (Ewels et al., 2016)

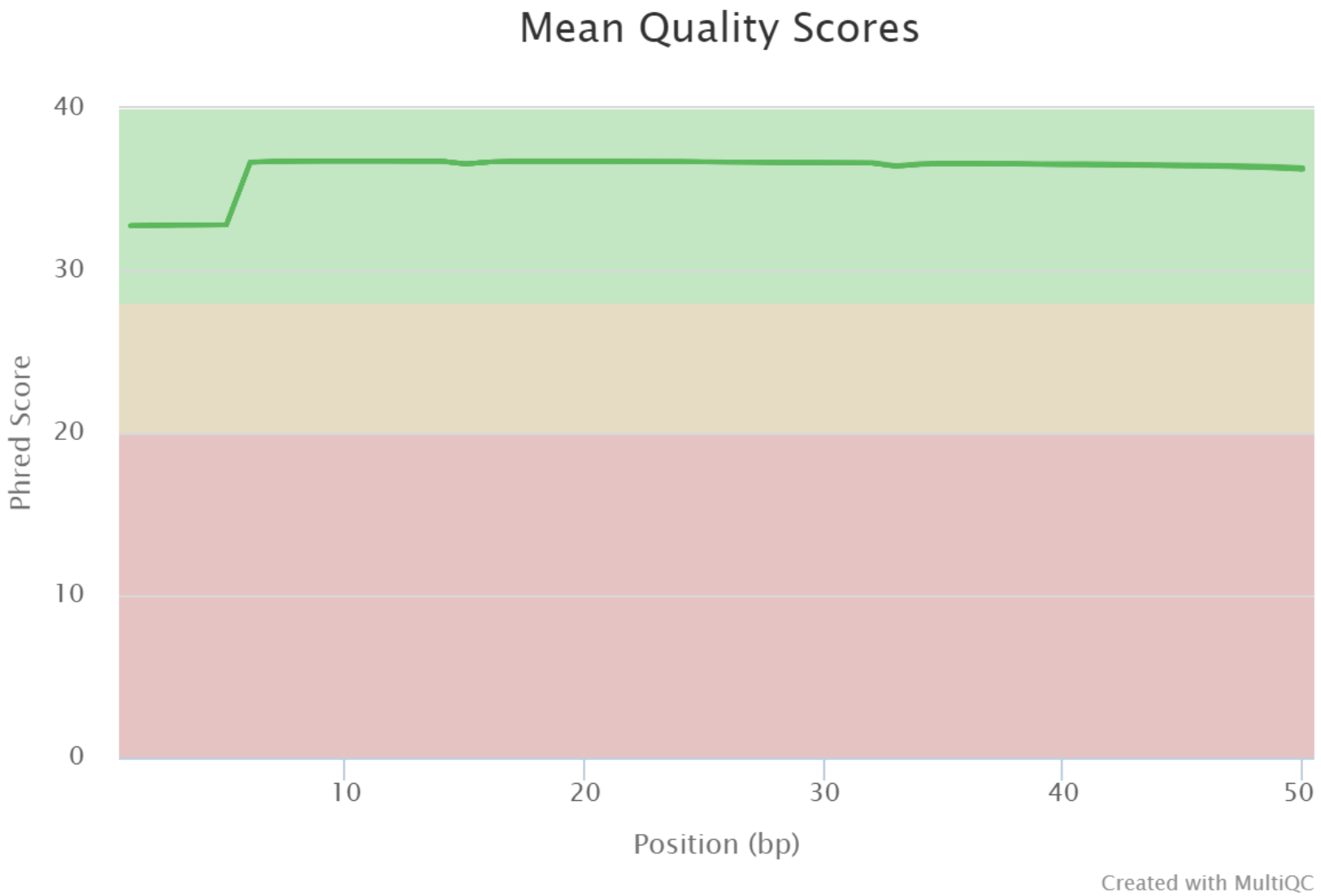


Figure 3-3: All five H3K27Ac ChIP-seq libraries and both input libraries were all of very good quality with Phred scores >30.¶Figure generated using (Ewels et al., 2016)

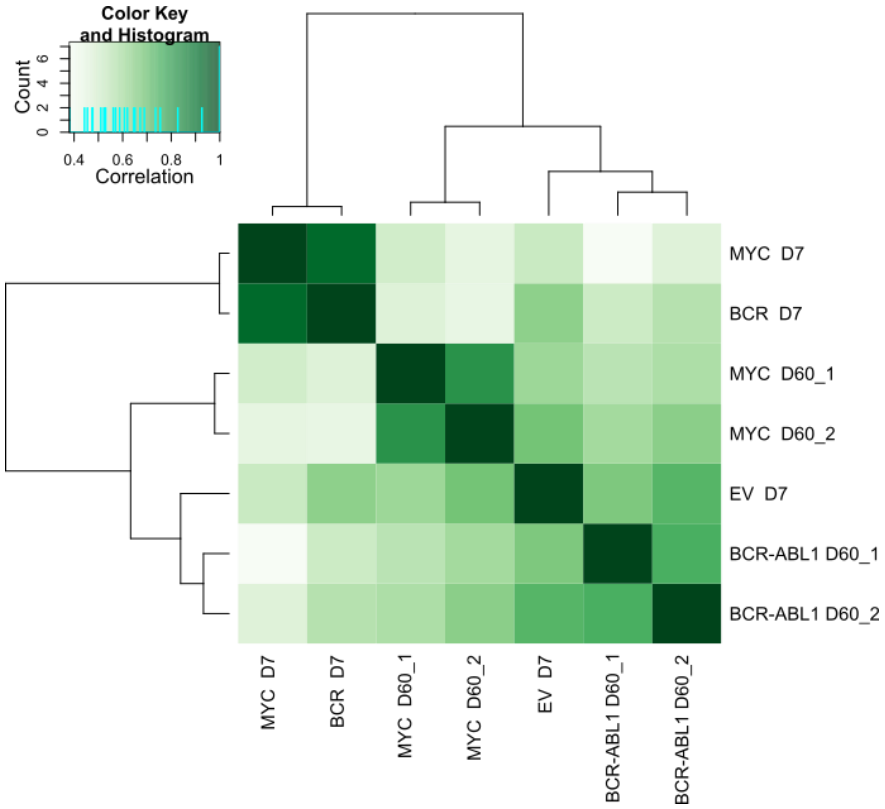


Figure 3-4: Stably transformed cells expressing *BCR-ABL1* or *MYC* were more dissimilar to each other than to the empty vector. Heatmap representation of the log10 Pearson's correlation values of H3K27Ac peak intensity in all seven conditions. Peaks are ordered by similarity (peak intensity) as shown by the dendrogram. The day 7 empty vector population had more H3K27Ac peaks in common with the day 60+ stably transformed cells than the day 7 cells in oncogenic crisis. Bioinformatic analysis and figure by Dr Mark Robinson.

For a general overview of the different conditions, a heatmap of all H3K27Ac peaks was constructed – shown in Figure 3-4. Each of the five conditions (including two biological replicates of the day 60+ stably transformed populations) was ordered by the similarity of the peak intensity as shown in the dendogram on the axes. As expected, the two biological replicates of the day 60+ stably transformed cells with *BCR-ABL1* or *MYC* were highly correlated (0.73 and 0.82, respectively) with each other. Unexpectedly, the H3K27Ac peaks in the day 60+ stably transformed cells expressing *BCR-ABL1* were more dissimilar to the day 60+ stably transformed cells expressing *MYC* (0.54 and 0.58) than they were to the day 7 empty vector cells (0.62 and 0.70). This is particularly interesting given that the two day 7 oncogenic crisis populations expressing *BCR-ABL1* or *MYC* were the most highly correlated of all seven conditions (0.93). This would indicate that oncogenic crisis induced by either oncogene results in a reproducibly similar pattern of H3K27Ac but that the same two oncogenes induced reproducibly different routes to stable transformation. This may be

because *MYC*, a transcription factor, induces more dramatic changes to the epigenome than *BCR-ABL1*, a tyrosine kinase.

3.4.2 Stable transformation induced broad, intense H3K27Ac peaks outside of promoter regions consistent with novel enhancers

In order to examine the differences between the conditions, an analysis of only those peaks which were different between the early and late conditions was undertaken. A total of 1,452 peaks differed significantly in the day 7 conditions versus the day 60 conditions ($p<0.05$). As shown in Figure 3-5, analysis of the H3K27Ac peaks showed that the overall distribution of H3K27Ac throughout the different functional locations of the genome shifted depending on the time point in the experiment. In general, oncogene–expressing cells at crisis stage (day 7) had peaks present at promoters or overlapping with a gene (highlighted with a black frame) that resembled narrow and intense signals. The remaining peak locations showed low signal intensity.

Notably, in the day 60 stably transformed populations, all categories of peak location *except* the promoter peaks showed a notable increase in broad, intense peaks (these are highlighted with a blue frame). The non-promoter peaks are consistent with the profile of newly activated enhancers. In addition, the intensity of the remaining peaks has returned to at least that of the day 7 empty vector, consistent with resumption of routine (albeit malignant) cellular function as opposed to a response to oncogenic crisis.

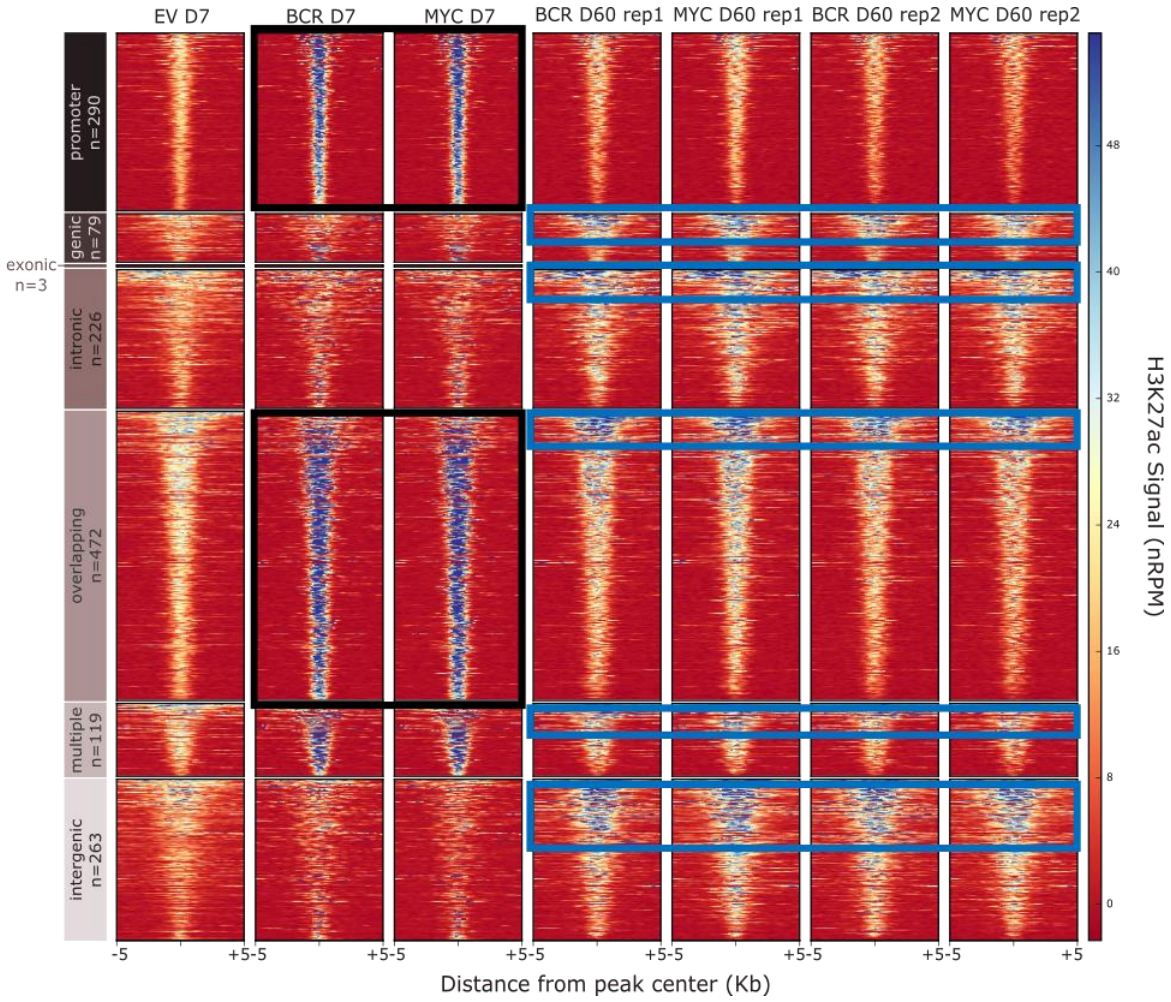


Figure 3-5: H3K27Ac ChIP-seq revealed peaks of different intensity in stably transformed populations.¶Heatmap image showing H3K27Ac ChIP-seq peaks for differential signals in day 7 cells infected with empty vector (EV; n=1)) only versus day 7 cells expressing *BCR-ABL1* (BCR) or *MYC* during a period of oncogenic stress (n=1 each) versus two biological replicates of the stably transformed cells which grew out of each oncogene-expressing population at day 60+ (n=2 per oncogene). The first column indicates the location of the differential peaks – promoter only, genic, exonic, intronic, overlapping (genic and intergenic), multiple (overlapping two or more genes) and lastly intergenic only. The last column indicates the H3K27Ac normalised read counts per million (signal intensity). Peaks shown (n=1,452) were significantly ($p<0.05$) different in day 7 (both conditions) versus day 60+ stably transformed. Bioinformatic analysis and figure by Dr Mark Robinson.

3.4.3 Stable transformation induced H3K27Ac changes in cell activation and proliferation pathways

As shown in Figure 3-6, semantic comparison of the gene ontology annotation was undertaken to detect similarities between the genes associated with H3K27Ac signals that were significantly different in the day 60+ stably transformed populations compared to the day 7 empty vector and oncogenic crisis populations (Yu et al., 2010).

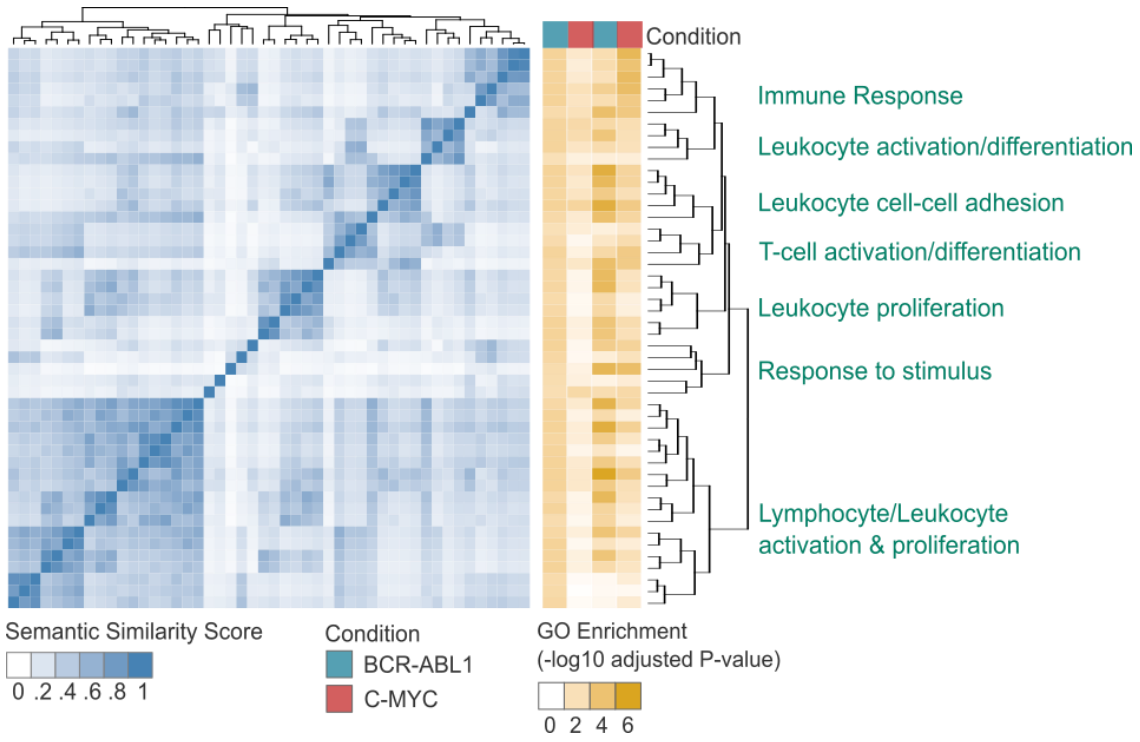


Figure 3-6: Enriched gene signatures related to cell activation and proliferation. ¶Heatmap showing gene ontology semantic similarity analysis (GOSemSim) of genes associated with significantly different H3K27Ac ChIP-seq peaks in day 60+ *BCR-ABL1* or *MYC* stably transformed populations versus day 7 empty vector or oncogenic crisis populations. The day 60+ populations showed enrichment for various gene ontology signatures relating to cell activation and proliferation. Heatmap colours represent similarity index (blue) and enrichment (brown). Bioinformatic analysis and figure by Dr Mark Robinson.

This shows that, whilst the H3K27Ac peaks that were significantly different between *BCR-ABL1* and *MYC* stably transformed day 60+ populations correlated less well with each other than with the empty vector, the enriched gene signatures representing the overall cellular outcome, such as cell activation and proliferation, were similar.

3.4.4 RNA-seq analysis of gene expression showed stable transformation was associated with activation of pathways necessary for increased proliferation.

In order to understand the pathways inhibited and/or activated during oncogenic stress and stable transformation, further RNA-seq analysis was undertaken on the day 7 empty vector populations (n=2+2), the day 7 BCR-ABL1-expressing populations in oncogenic crisis (n=2+3 for each oncogene) and the day 60+ stably transformed populations (n=3 for each oncogene). The number of uniquely mapped reads and the quality scores for these libraries are shown in Figure 3-7 and Figure 3-9.

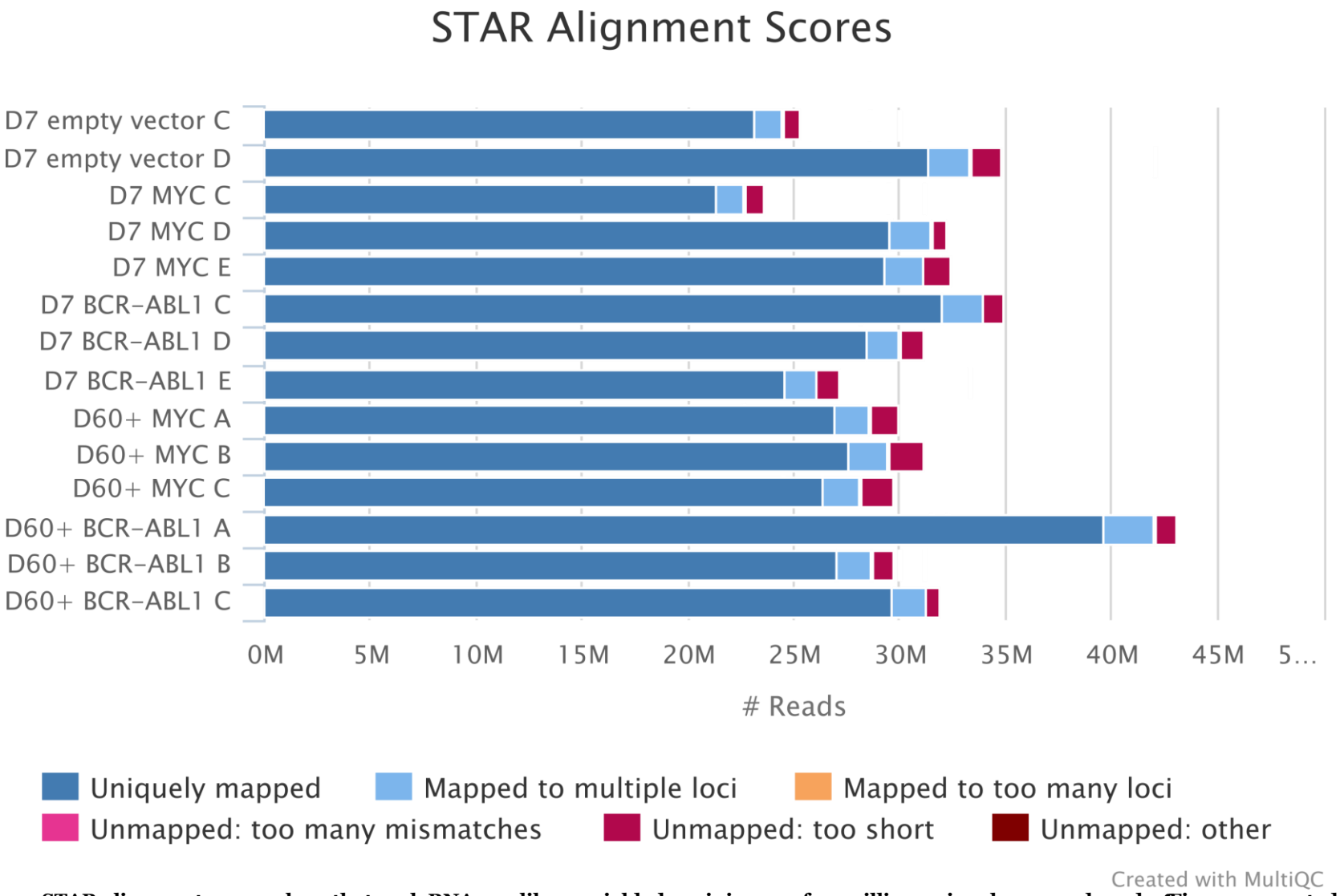


Figure 3-7: STAR alignment scores show that each RNA-seq library yielded a minimum of 22 million uniquely mapped reads. ¶Figure generated using (Ewels et al., 2016).

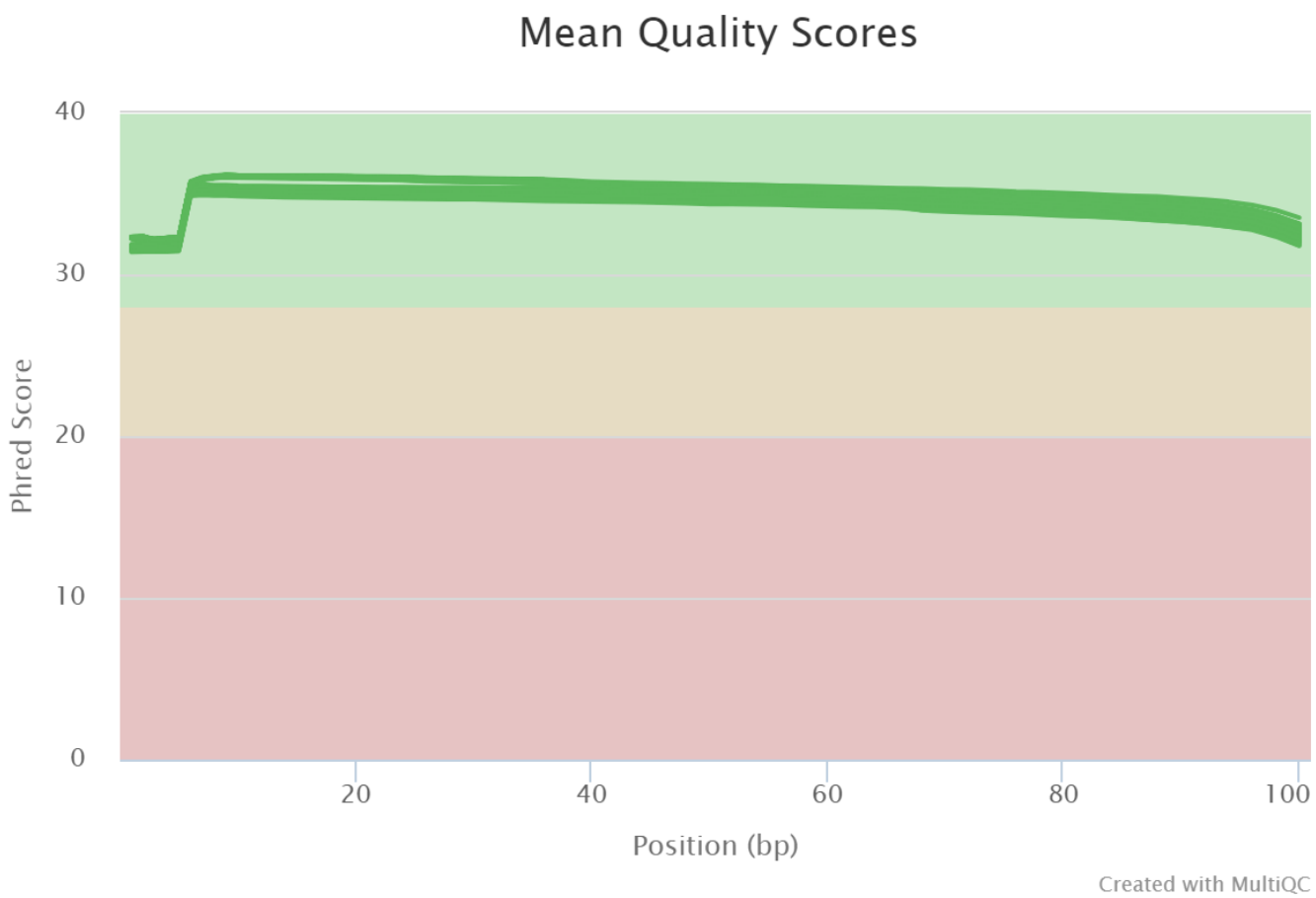


Figure 3-8: Each of the RNA-seq libraries libraries were of very good quality with Phred scores >30.¶Figure generated using (Ewels et al., 2016).

As shown in Figure 3-9, each comparison revealed significantly inhibited and activated pathways (as defined by the Kyoto Encyclopedia of Genes and Genomes (KEGG)).

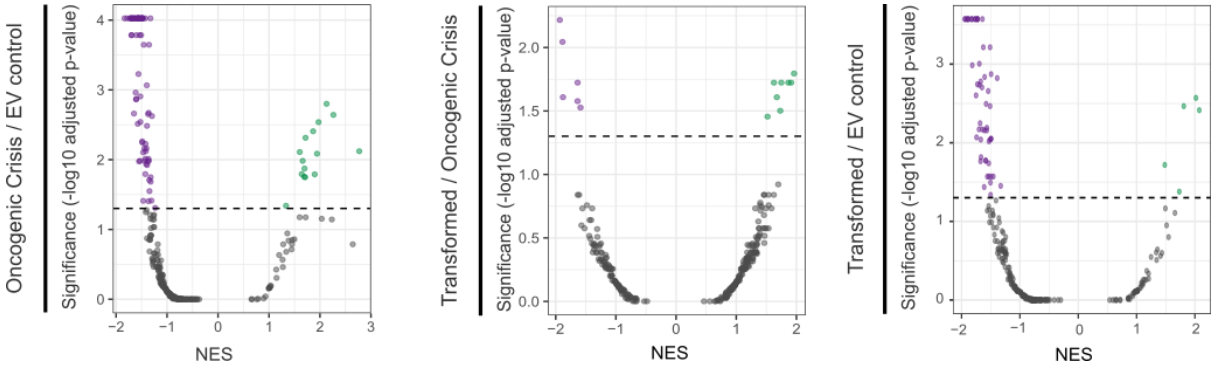


Figure 3-9: Oncogenic crisis and stable transformation resulted in significant activation and inhibition of KEGG pathways. ¶Differentially expressed pathways were identified in day 7 empty vector, day 7 *BCR-ABL1*-expressing oncogenic crisis and day 60+ *BCR-ABL1*-expressing stably transformed populations. Volcano plots showing KEGG pathway enrichment using normalised expression scores (NES). The dashed line in each indicates significance (p=0.05) - note the scale differs in each. Those genes above the dashed line are significantly inhibited (purple) or significantly activated (green). Bioinformatic analysis and figure by Dr Mark Robinson.

The top 20 KEGG pathways in Figure 3-9 that were inhibited and activated in each comparison are listed in Figure 3-10. Those which were significantly ($p<0.05$) inhibited are coloured in purple and those significantly activated in green whilst those which remain in grey did not reach significance. This shows that oncogenic crisis and stable transformation induced by *BCR-ABL1* results in activation of pathways necessary for increased proliferation. This includes DNA replication, transcription, translation, replication and repair, for example.

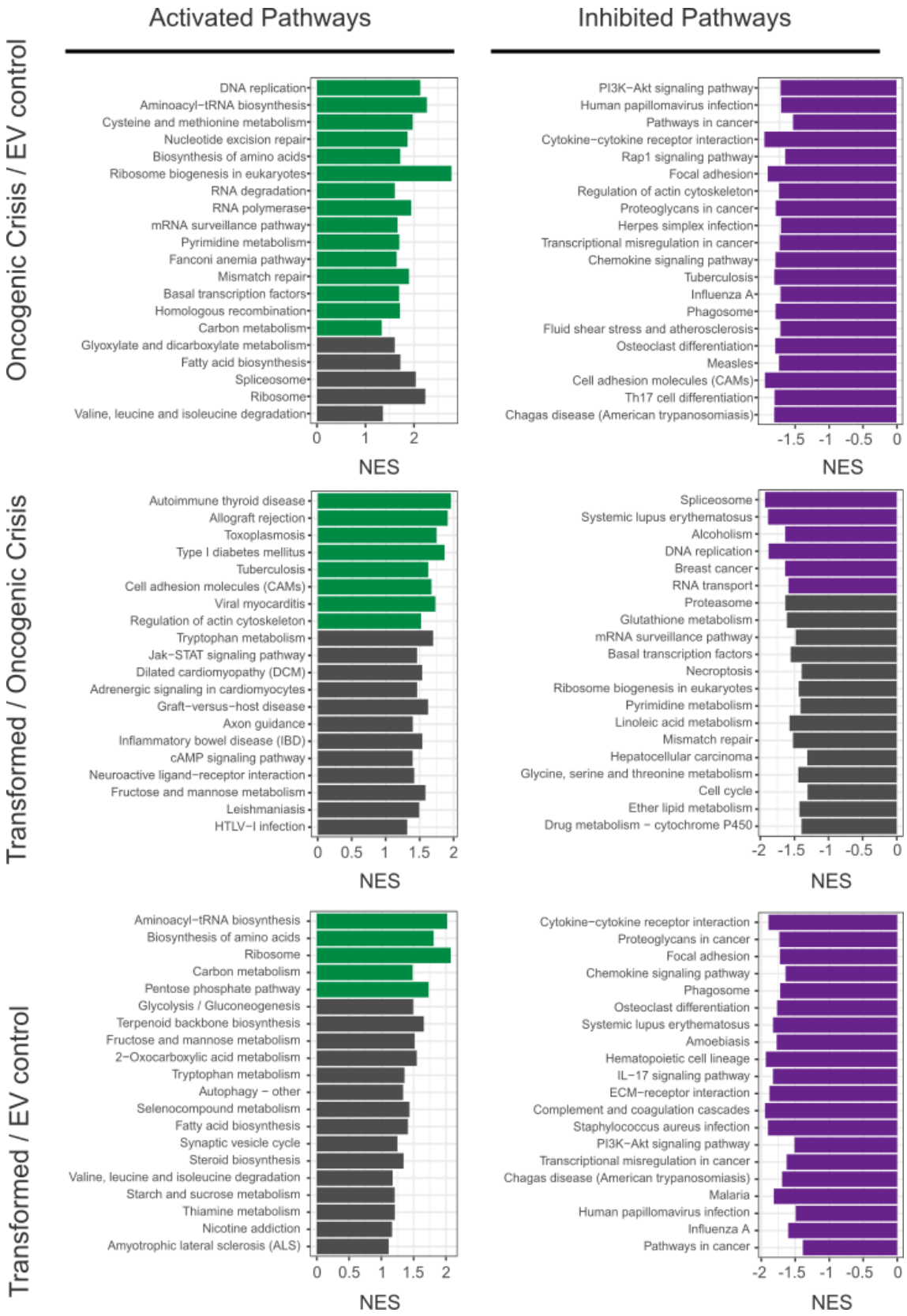


Figure 3-10: Oncogenic crisis and stable transformation resulted in activation of transcription, replication and repair pathways and inhibition of immune pathways. ¶The top 20 activated (green) and inhibited (purple) KEGG pathways for each comparison made in Figure 3-9 are shown. Bioinformatic analysis and figure by Dr Mark Robinson.

3.4.5 Combined H3K27Ac ChIP-seq and RNA-seq of stably transformed *BCR-ABL1*-expressing populations revealed epigenetic and genetic differences relevant to human Ph+ leukaemia

In light of the reproducible differences between the day 60+ stably transformed populations expressing either *BCR-ABL1* or *MYC,* subsequent analyses focused on the co-operating events that result in stable transformation of *BCR-ABL1*-expressing cells only. The differential peaks in the day 7 empty vector and *BCR-ABL1*-expressing oncogenic crisis populations versus day 60+ stably transformed populations were compared; this time only those which were highly significantly different ($p=1x10^{-4}$) were included, reducing the number of peaks from 1,452 to just 84, as shown above the red line in in Figure 3-11.

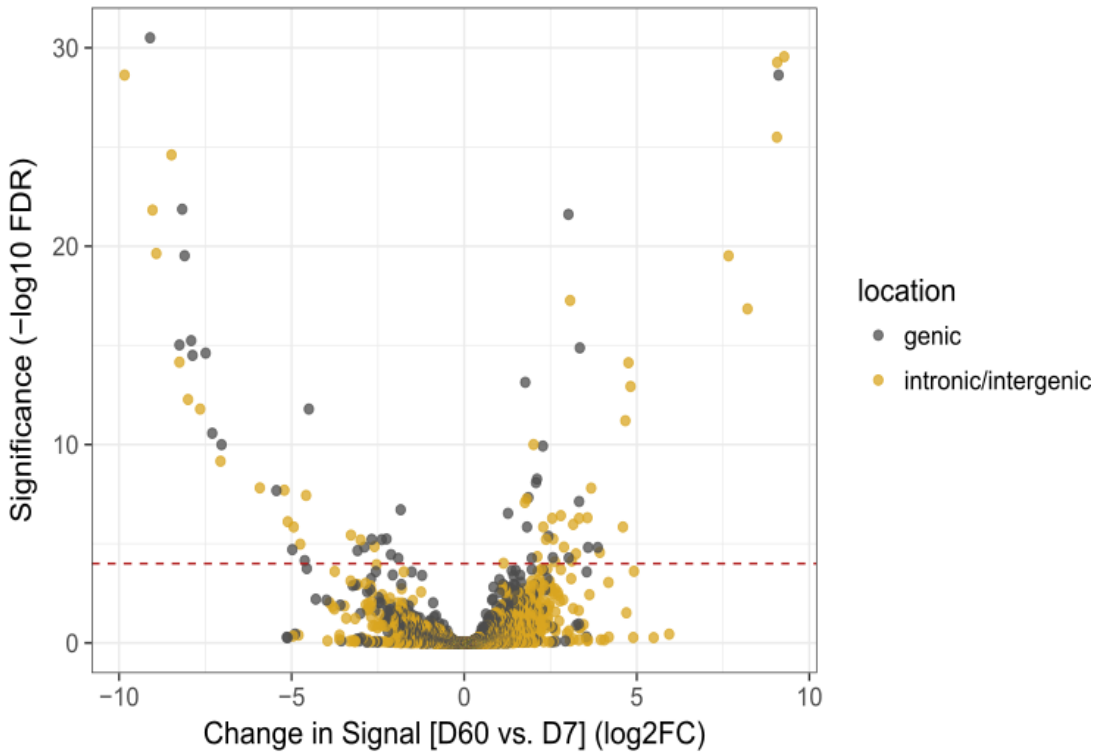


Figure 3-11: Volcano plot showing all 11,308 H3K27Ac *BCR-ABL1* peaks. ¶The log2 fold change in the day 7 empty vector and oncogenic crisis populations versus the day 60+ stably transformed populations is indicated on the Y-axis and the –log10 significance of the change on the X-axis. The dashed red line indicates highly significant (p=0.0001) differences (n=84). Intergenic and intronic signals are highlighted with yellow circles; grey circles represent all other genic locations. Bioinformatic analysis and figure by Dr Mark Robinson.

Next, an investigation of the correlation between the 84 highly significantly different H3K27Ac ChIP-seq signals and gene expression (as quantified by RNA-seq) was undertaken. For this analysis, the expression of the most proximal gene identified by RNA-Seq was

assigned to the respective intergenic peaks. The scatter plots in Figure 3-12 show that there is indeed a strong correlation between the expression of each gene and the intensity of the H3K27Ac signal in day 60+ stably transformed populations versus the day 7 empty vector control population (r=0.71; left) or the day 7 oncogenic crisis populations (r=0.69; right); both correlations are highly significant ($p<1x10^{-9}$ and $p<1x10^{-8}$, respectively). Intergenic and intronic peaks are highlighted with a yellow dot to distinguish them from all other peak locations (shown with black dots). Selected genes are indicated.

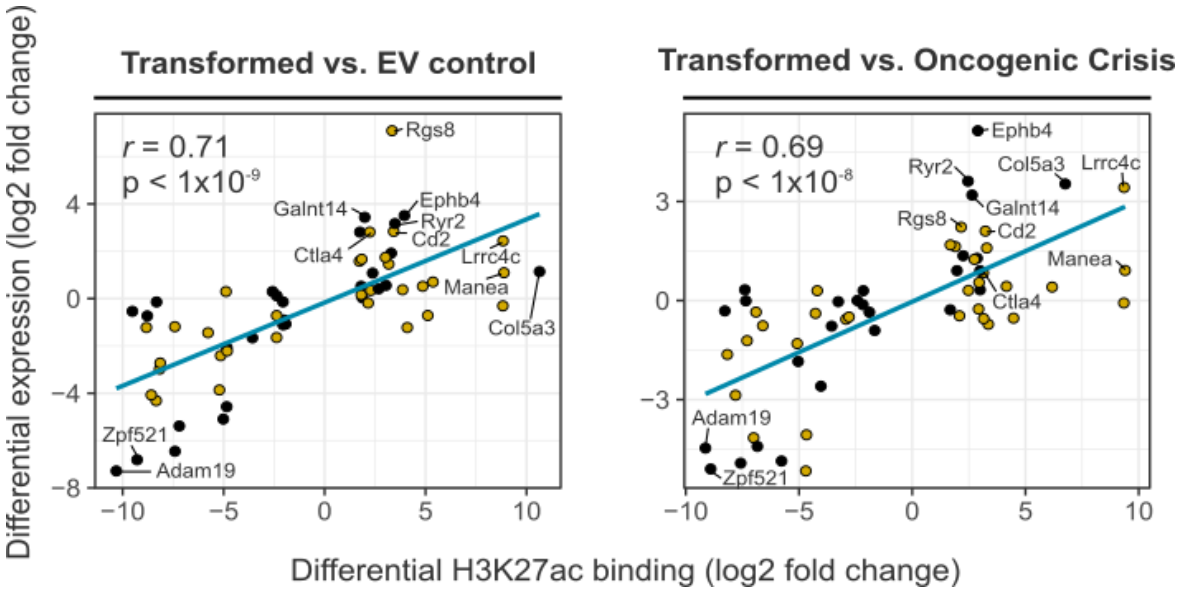


Figure 3-12: Differential H3K27Ac signals correlated with differential expression in *BCR-ABL1* stably transformed populations. ¶Scatterplots showing the H3K27Ac signal intensity in day 60+ stably transformed cells (transformed) compared to day 7 empty vector (EV; left) or day 7 *BCR-ABL1*-expressing cells in oncogenic crisis (oncogenic crisis; right). The expression of the most proximal gene was associated with all intergenic signals. Intergenic and intronic signals are highlighted with yellow circles; black circles represent all other genic locations. The blue line represents linear regression fit. The *r* value and the p-value of the Pearson's correlation is indicated. Genes shown (n=84) associated with highly significantly ($p=1x10^{-4}$) differential peaks in day 7 (both conditions) versus day 60+ stably transformed. Bioinformatic analysis and figure by Dr Mark Robinson.

The day 60+ stably transformed putative enhancer-associated genes were then analysed in more detail. One interesting example from this list is *Ctla4*, which featured amongst the highly significantly different genes twice: once because of an intergenic peak and once because of a different overlapping peak – this is shown in Figure 3-13. *Ctla4* is a negative regulator of T-cell activation and there is evidence that its activation contributes to avoidance of tumour immune surveillance (Laurent et al., 2007). The empty vector and the day 7 *BCR-ABL1* expressing cells show a shallow signal which is most intense at the *Ctla4* gene promoter but which extends to cover the entire length of the gene body. This has increased significantly in the day 60+ stably transformed *BCR-ABL1* expressing cells that show strong peaks in the location of the previously shallow ones, and in addition the emergence of intergenic peaks both upstream and downstream in one of the two replicates. These stronger peaks in day 60+ stably transformed populations corresponded with an increase in *Ctla4* expression of 6.94 fold compared to the empty vector ($p=8.07x10^{-7}$).

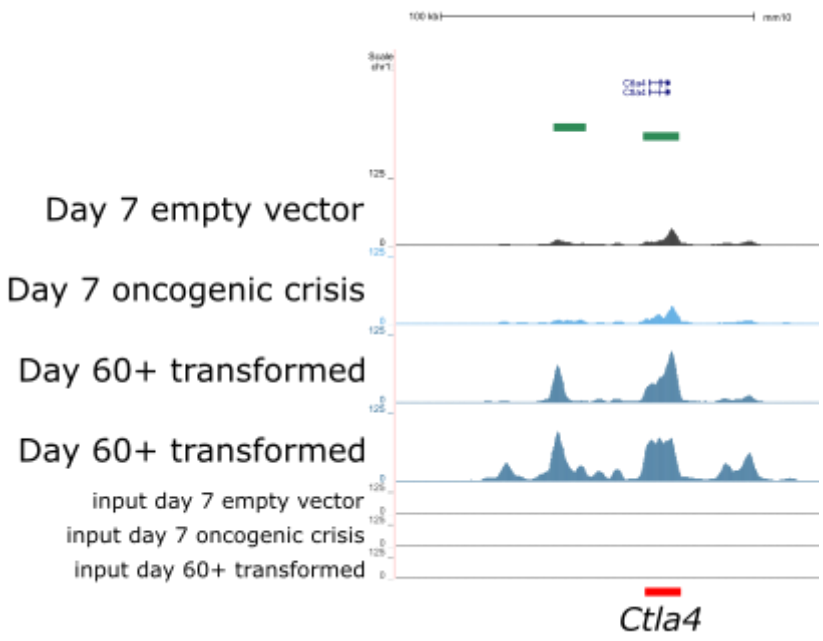


Figure 3-13: Increased expression of *Ctla4* was associated with an increased H3K27Ac peak in day 60+ *BCR-ABL1*–expressing stably transformed cells. ¶ Custom ChIP-seq track showing H3K27Ac ChIP-seq peaks at *Ctla4* at day 7 and day 60 for *BCR-ABL1*-expressing cells. The gene body is indicated in red and the significantly different peaks are indicated in green

Another interesting example of a putative enhancer regulated gene in stably transformed D60 cells is *Galnt14*. *Galnt14* is an peptidyl-O-gylcosyltransferase whose expression is necessary for sensitivity to apoptosis-inducing therapy (Wagner et al., 2007). As shown in Figure 3-14, an H3K27Ac ChIP-seq peak overlapping *Galnt14* was significantly increased in both day 60+ stably transformed replicates, along with the emergence of an intronic peak.

This alteration was associated with a 10.86x increase in expression compared to day 7 empty vector ($p=1.18x10^{-55}$) and with a 9.13x increase in expression compared to the day 7 *BCR-ABL1* expressing cells in oncogenic crisis ($p=5.49x10^{-54}$).

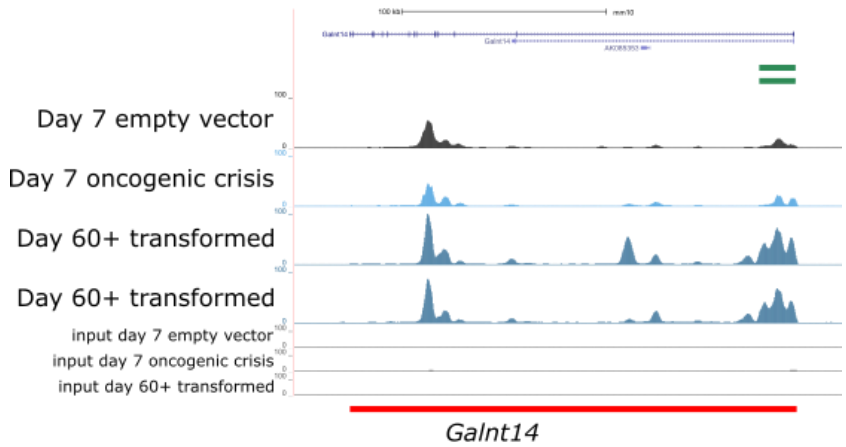


Figure 3-14: Increased expression of *Galnt14* was associated with an increased H3K27Ac peak in day 60+ stably transformed cells. ¶ Custom ChIP-seq track showing H3K27Ac ChIP-seq peaks at *Galnt14* at day 7 and day 60 for *BCR-ABL1*-expressing cells. The gene body is indicated in red and the significantly different peaks are indicated in green.

Whilst this approach to analysing the differential peaks has clearly identified some interesting interactions the nature of enhancers and their ability to interact over long distances mean this approach is not without its flaws. The well characterised *MYC* super enhancer, approximately 1.7Mb upstream (Shi et al., 2013), was correctly identified as having a highly significantly differential peak intensity in the day 7 empty vector and day 7 oncogenic crisis versus day 60+ stably transformed *BCR-ABL1*-expressing populations – this is shown in Figure 3-15. However, the differential peak intensity was incorrectly assigned to the most proximal gene. This example is not likely to be isolated so it will be important to clarify the enhancer-promoter interactions for a meaningful analysis, and not just rely on identification of putative enhancers by proximity.

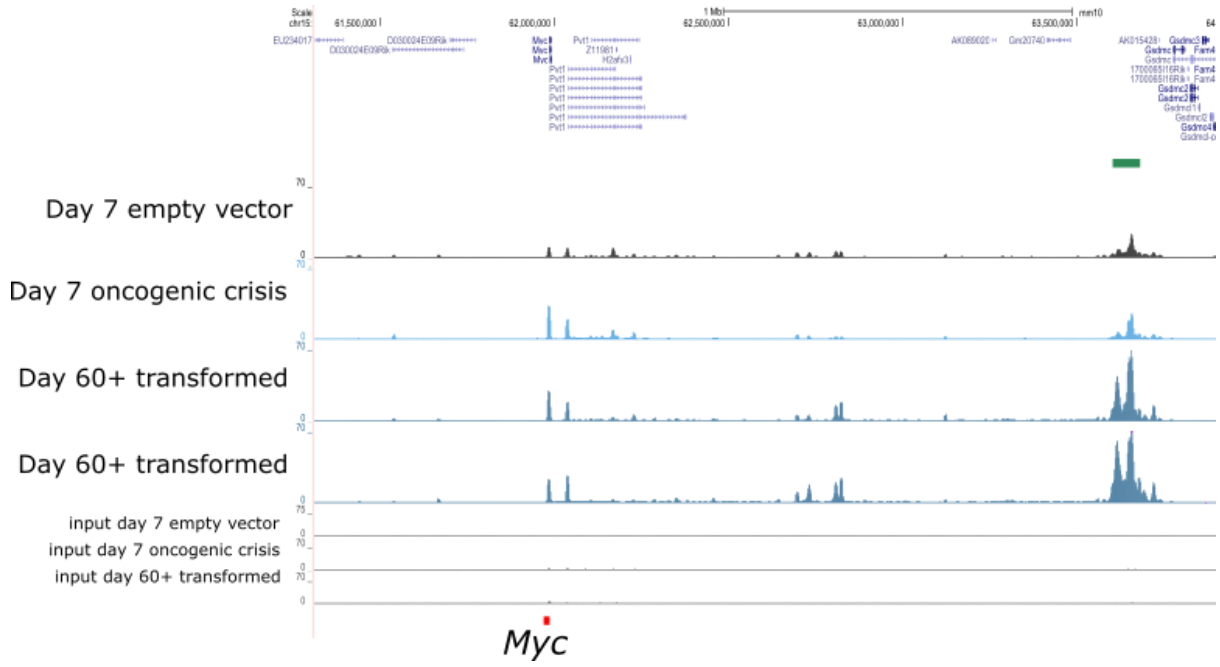


Figure 3-15: The well characterised *MYC* super enhancer was detected in the *BCR-ABL1*-expressing stably transformed day 60+ model but was wrongly assigned to a more proximal gene in this analysis.¶Custom ChIP-seq track showing H3K27Ac ChIP-seq peaks at day 7 and day 60 for *BCR-ABL1*-expressing cells. The gene body is indicated in red and the significantly different peak (which coincides with the *MYC* super enhancer) is indicated in green.

Putative enhancers were then compared to three Ph+ (TOM1, SD1 and BV173) and three Ph- (Reh, SMS-SB and Kasumi-2) human B-ALL cell lines, in order to verify that enhancers identified in the murine model were also present in human leukaemia. Several established enhancers, such as the *MYC* super enhancer (described above) and the *PAX5* super enhancer were indeed present in a proportion of the B-ALL cell lines, confirming the validity of this model to identify relevant human enhancers. One novel H3k27Ac peak called in *BCR-ABL1* oncogenic crisis and stably transformed populations was located within the intron of *BCL2*. Interestingly, this was also present in human B-ALL cell lines but was restricted to the Ph+ B-ALL cell lines only suggesting this modification may be *BCR-ABL1* dependent (shown in Figure 3-16).

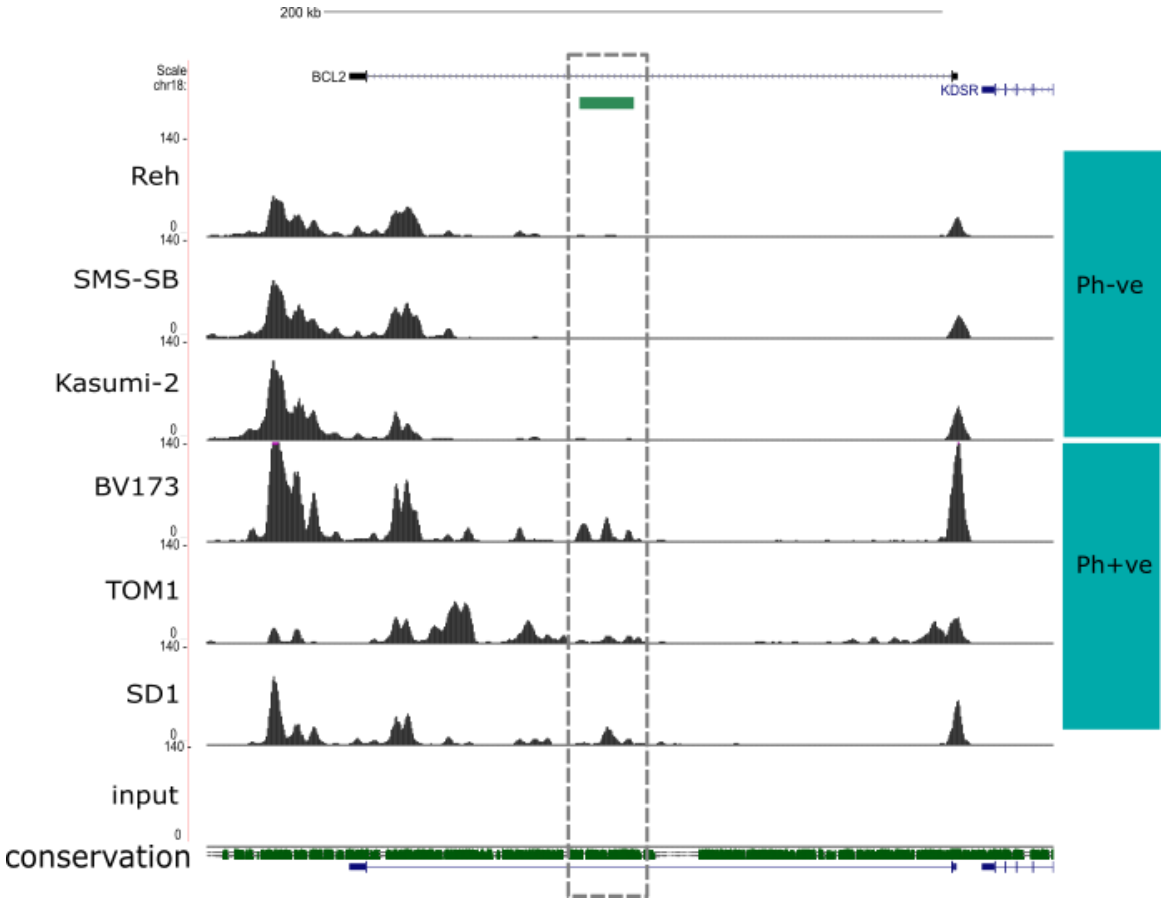


Figure 3-16: An intronic putative enhancer identified in the mouse model was present only in the Ph+ human B-ALL cell lines.¶Custom H3K27Ac ChIP-seq track showing the *BCL2* region in six different human B-ALL cell lines; Reh, SMS-SB and Kasumi-2 (top) are Ph- whilst BV173, TOM1 and SD1 (bottom) are Ph+. The location of the differential murine H3K27Ac peak is indicated with a green bar above the tracks and highlighted with a grey dashed box. The sequence conservation track at the bottom shows alignment with mouse.

4 Genomic drivers of *EVI1* expression in myeloid leukaemia

4.1 Contribution of others to this chapter

All bioinformatics analysis (including Figure 4-24) was undertaken by Dr Mark Robinson. Immunohistochemistry was performed by the Imperial College Healthcare Molecular Pathology service and sequencing was performed by the Medical Research Centre's Genomics Facility. I performed all sample processing, tissue biobanking and all other experiments.

4.2 Hypothesis

EVI1 expression in poor prognosis leukaemia with no cytogenetic rearrangement of *EVI1* is caused by aberrant activation of a novel oncogenic enhancer.

4.3 Background

As detailed in section 1.6, *EVI1* is a tightly controlled transcription factor which is essential for normal haematopoietic development and whose overexpression causes differentiation arrest. It is aberrantly expressed in a proportion of myeloid leukaemia. This aberrant expression is understood in some cases - cytogenetic rearrangements result in fusion proteins transcribed from ectopic promoters, or in the normal protein transcribed from ectopic enhancers – but remains idiopathic in others. The similarity between the aetiology of *EVI1* expression in myeloid leukaemia and that of *TAL1* expression in T-cell leukaemia – and the recent discovery of a novel oncogenic super enhancer in the latter – triggered the investigation of the non-coding genome in CML to explore whether a similar mechanism may underpin aberrant *EVI1* expression as described in the following investigation.

4.4 Results

4.4.1 EVI1 expression was detected in chronic and blast phase samples

EVI1 overexpression is detectable in various myeloid leukaemias including myelodysplastic syndrome, chronic myeloid leukaemia and acute myeloid leukaemia. As discussed above, however, CML is defined by a single gene fusion, *BCR-ABL1*, and even in the later stages of disease has a relatively simple genome compared to the heterogeneous nature of MDS and AML. For this reason, CML was initially chosen as a model disease for this investigation. In order to investigate the possibility of a novel oncogenic enhancer of *EVI1*, a sub collection of the Imperial College Healthcare Tissue Bank was initiated (reference number MEC_AR_16_030) and access to this sub collection for this project was granted in October 2016 and extended in November 2017 (R16065). See appendix for the Imperial College Healthcare Tissue Bank Approval.

When samples of a sufficient size were received, RNA and genomic DNA were extracted and cells were harvested for FISH, cryopreserved and cross-linked. Over 18 months, 31 suitable patient samples were received for the study.

First, RT-qPCR of cDNA derived from the samples was undertaken to establish the fold change of *EVI1* expression compared to the *BCR-ABL1* positive myeloid cell line, LAMA84, normalised to a stably expressed control gene, *GUSB* (Wang et al., 2006) using the delta delta CT (cycle threshold) method, as shown in Figure 4-1. A primer pair targeting exon 14, which is common to all *EVI1* transcripts, including the *MECOM* transcript, was used.

Samples from healthy volunteers (n=5) were negative for *EVI1* expression. The CML-BC (n=7) samples showed a wide variation in expression levels, with negative, low and high expression. The cell lines tested (n=3) were also variable, with KCL22 classified as low expression and K562 and LAMA84 (to which the expression of all samples was compared) as high.

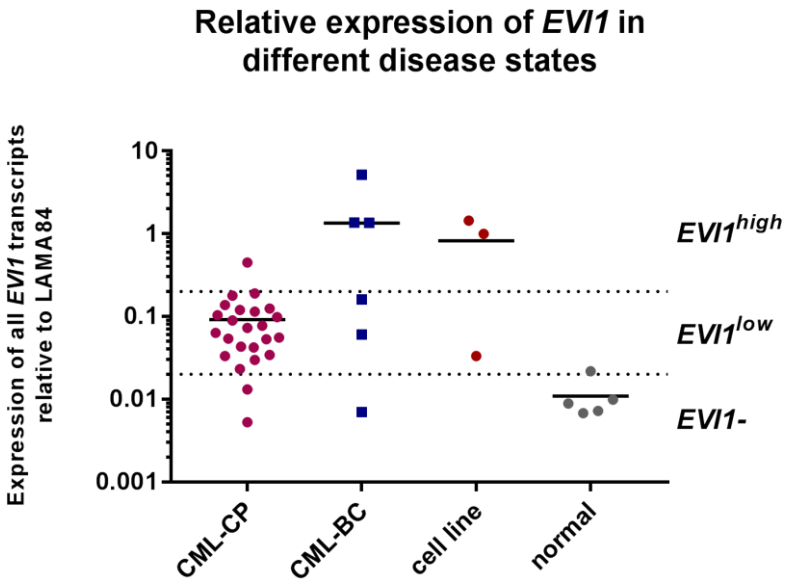


Figure 4-1: *EVI1* expression was highest in the CML-BC samples.¶Scatter plot showing the log_{10} expression of total *EVI1* transcripts expressed relative to that of LAMA84 normalised to GUSB using ΔΔCt method.

Samples were categorised as high expression ($EVI1^{high}$ if 0.2x or above), low expression ($EVI1^{low}$ if 0.02-0.2x) and negative ($EVI1^{-}$ if less than 0.02x). As shown in Figure 4-2, there was little variation in the range of expression for the CML-CP samples with most expressing between 0.02x and 0.2x the *EVI1* of LAMA84. The majority of CML-CP samples were thus categorised as $EVI1^{low}$, with two samples categorised (MEC122 and MEC124) as $EVI1^{-}$ and a single sample (MEC140) as $EVI1^{high}$.

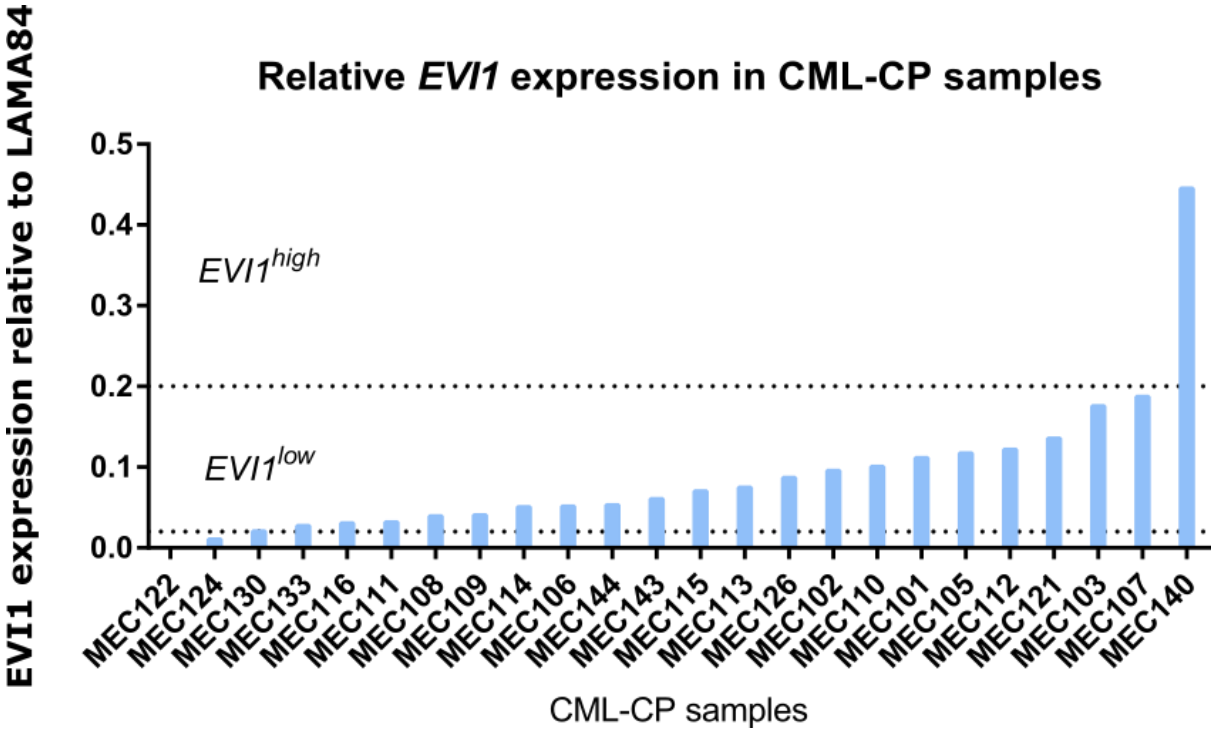


Figure 4-2: The majority of CML-CP samples had low expression of *EVI1*. Bar chart showing the expression classifications of *EVI1* for all 24 CML-CP samples (total *EVI1* transcripts expressed relative to that of LAMA84 normalised to GUSB using ΔΔCt method). A single sample (MEC140) was classified as having high expression, and two (MEC122 and MEC124) as being negative for expression.

Samples with *EVI1*^{high} from any disease category (n=7) plus all three cell lines were then tested by fluorescence in situ hybridisation (FISH) using a triple colour break apart probe (Kreatech Diagnostics, schematic shown in Figure 4-3) for the EVI1 locus, to determine the presence or absence of a cytogenetic rearrangement. A minimum of 100 interphases were examined for each sample, with up to five metaphases when available.

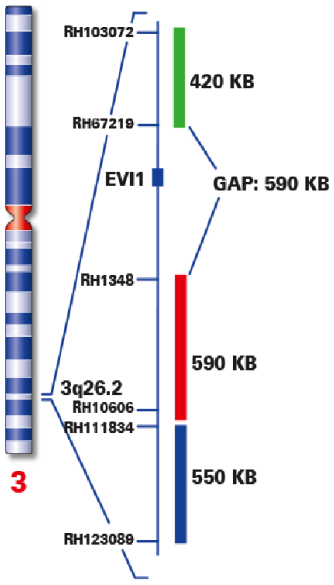


Figure 4-3: The triple colour break apart FISH probe detects rearrangements of the *MECOM* locus.¶The three sections of the probe co-localise in a single fusion to the *MECOM* locus on a normal homologue for chromosome 3; if translocated or inverted, the fusion splits. Figure taken from Kreatech Diagnostics.

One CML-BC sample (with the highest expression shown in Figure 4-1) was excluded as the patient had a t(3;3) as shown in Figure 4-4; the fusion signal on the larger derivative chromosome 3 has split, leaving the green component in the original 3q26.2 locus and translocating the red and aqua components to 3q21 on other homologue. The remaining intact fusion signal moves with the reciprocally translocated segment. All other patient samples and cell lines showed two intact fusion signals only, consistent with no rearrangement of the EVI1 locus, and indicating an unexplained mechanism for *EVI1* expression.

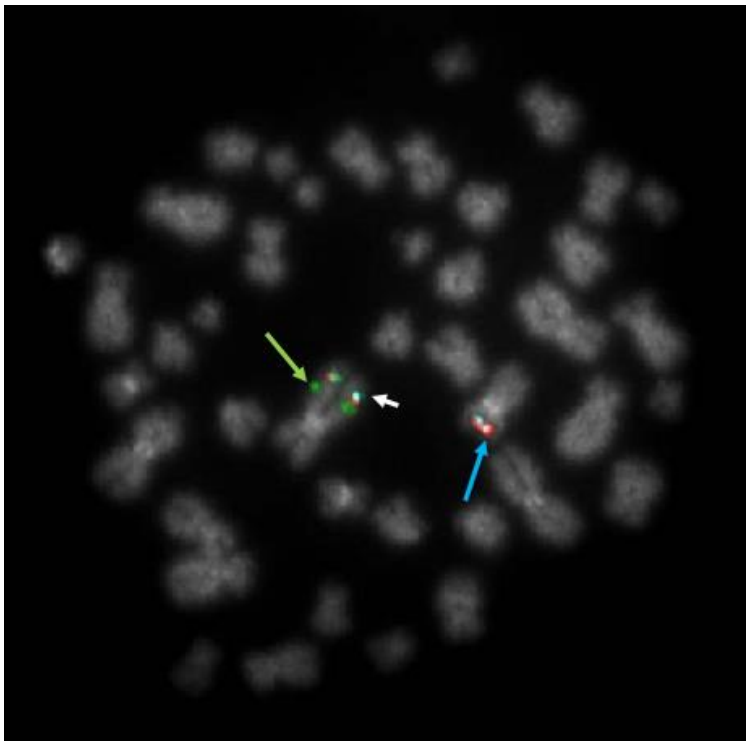


Figure 4-4: One CML-BC sample was excluded from the study following detection of a t(3;3) translocation. ¶ Metaphase image showing the results of the same triple colour break apart FISH assay. The *MECOM* locus which originated on the larger der(3) has split leaving a green signal only [green arrow]. The red and aqua signals have translocated to the smaller der(3) [blue arrow]. One intact *MECOM* locus and fusion signal remains, distal to the breakpoint on the larger der(3) [white arrow].

A representative metaphase from a patient sample with no cytogenetic rearrangement of chromosome 3 is shown in Figure 4-5. A summary of the patients from which the remaining samples (n=30) was taken in shown in Table 4-1.

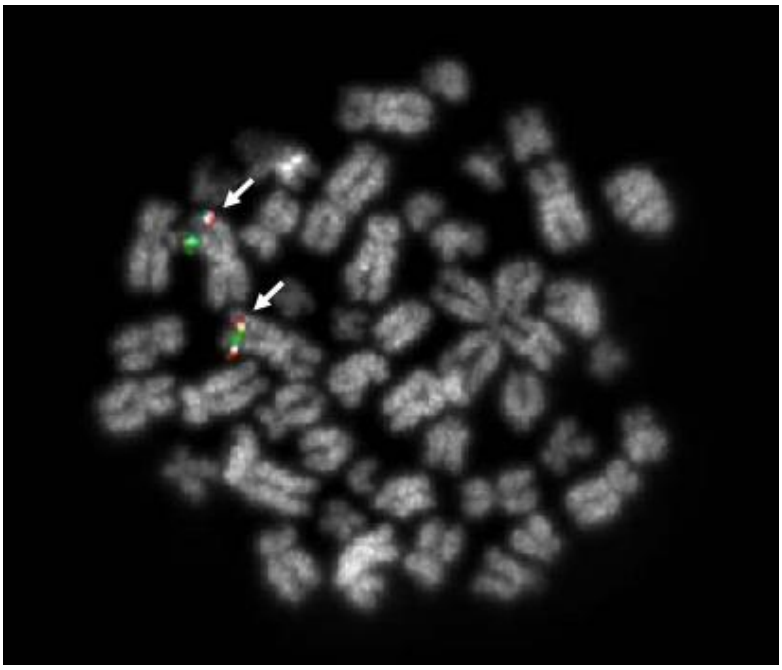


Figure 4-5: All other samples with *EVI1*^{high} expression had no detectable rearrangement of the *MECOM* locus. ¶ Metaphase image showing the results of the triple colour break part assay. Intact fusion signals, comprising green, red and aqua probe components, have hybridised to the long arm of both homologues of chromosome 3 [indicated by a white arrow]. This sample has no cytogenetic rearrangement of the *MECOM* locus.¶

Table 4-1: Patient characteristics for suitable samples received and bio banked.

	No. of EVI1^{-} patients	*No. of EVI1^{low} patients*	*No. of EVI1^{high} patients*	*Total*
Sex				
Male	1	13	2	16 (53.3%)
Female	2	10	2	14 (46.6%)
Age group				
<35 yrs.	1	6	1	8 (26.6%)
35-60 yrs.	1	13	2	16 (53.3%)
>60 yrs.	1	4	1	6 (20.0%)
Disease				
CML-CP	2	21	1	24 (80.0%)
CML-BC	1	2	3	6 (20.0%)
Total	3 (10.0%)	23 (76.6%)	4 (13.3%)	**30**

4.4.2 In chronic phase MECOM expression is derived from the CD34+ compartment

Consistent with other reports (Gerhardt et al., 1997), low level expression (less than 0.2 times that of LAMA84) was detected in the majority of CML-CP patients which was likely due to the proportion of CD34+ cells present in the sample. CML-CP samples have, by definition, fewer than 10% blasts in the peripheral blood or bone marrow whereas CML-BC samples have blasts comprising 20% or more of the peripheral blood white blood cells or nucleated bone marrow cells (Swerdlow et al., 2008). Where available, the CD34 content of the samples (as measured by flow cytometry undertaken by the Imperial Molecular Pathology department as part of routine diagnostic testing) was compared to the *EVI1* expression.

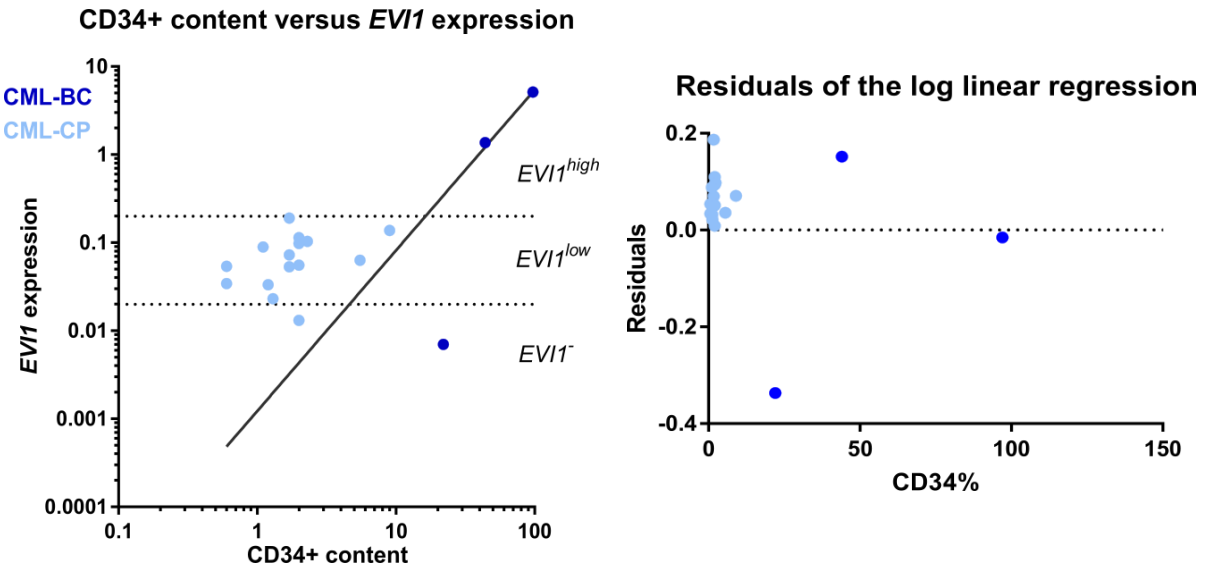


Figure 4-6: In CML-CP, CD34+ content correlates with *EVI1* expression. ¶Scatter plot (left) showing CD34+ content and *EVI1* expression for all samples for which this information was available. CML-BC are indicated in dark blue and CML-CP in light blue; the horizontal dotted lines indicate the thresholds for *EVI-*, *EVI1low* and *EVI1high* expression. Linear regression is shown with a black line. The residuals of the log linear regression (right) shows the CML-BC with 22% CD34+ but negative *EVI1* expression is an outlier.

Notably, the relationship between CD34+ content and *EVI1* expression differs in the CML-CP from the CML-BC samples. As shown in Figure 4-6 (left), the CD34+ content in CML-CP ranges from 0.6% to 9%. In CML-BC the CD34+ content is, as expected, far higher than that of CML-CP ranging from 20% to 97%. The *EVI1* expression in the latter samples, however, ranges from 0.007x to over 5x for CML-BC.

Whilst the pattern of expression in CML-CP samples suggests that the overall expression is simply reflective of the proportion of CD34+ haematopoietic progenitor cells, the varying level of expression in the CML-BC samples –notably, MEC125, which has 22% CD34+ but *EVI1* expression of just 0.007x and classified as negative – would indicate that a different mechanism is responsible. Residuals of the log linear regression (right) shows that this sample is indeed the outlier. It is therefore likely that the interactions which may be detected in CML-CP reflect the normal interactions of a haematopoietic progenitor cell whereas those detected in CML-BC reflect an abnormal, malignant process.

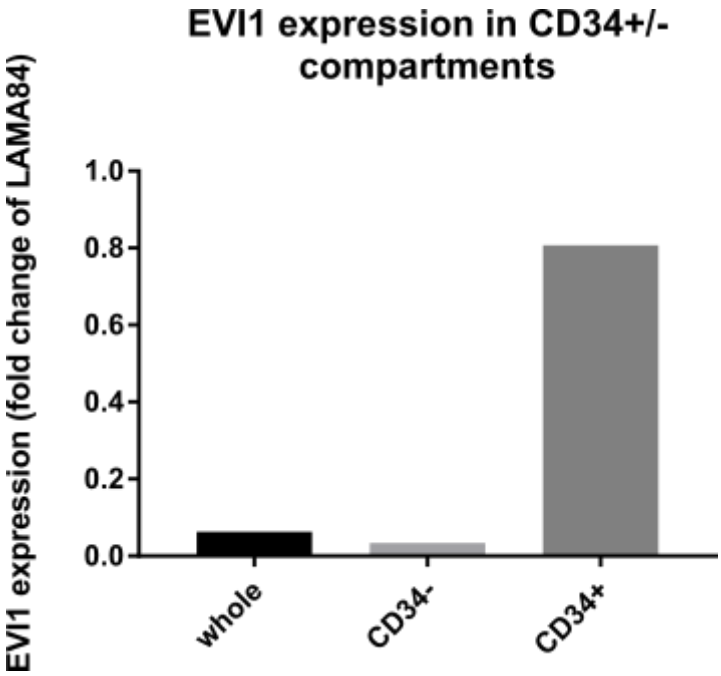


Figure 4-7: *EVI1* is expressed in the CD34+ compartment only during chronic phase. ¶Bar chart showing the *EVI1* expression (as a fold change relative to the LAMA84 cell line and normalised to GUSB) of a mononuclear cells from a CML-CP (MEC143): the whole leukophoresis sample (left), the positively selected CD34+ fraction (right) and the depleted CD34- fraction (middle).

In order to further investigate this correlation, it was necessary to examine the *EVI1* expression derived from the CD34+ and CD34- fractions separately. First, mononuclear cells were isolated by density centrifugation from a leukophoresis sample taken at diagnosis of CML-CP. CD34+ cells were then positively selected by magnetic activated cell sorting and the CD34- flow through was also retained. Analysis of *EVI1* expression was then undertaken using RT-qPCR, as before, of all three fractions (whole, CD34+ and CD34-). As shown in Figure 4-7, the low level of *EVI1* expressed by the total sample derived almost entirely from the CD34+ compartment (residual $EVI1^{low}$ expression in the CD34- fraction is likely due to CD34+ contamination – this fraction is depleted of CD34+ rather than negatively selected). This is in agreement with various murine studies (Goyama et al., 2008; Kataoka et al., 2011; Laricchia-Robbio and Nucifora, 2008; Yuasa et al., 2005) and a human study (Steinleitner et al., 2012) which have shown that *EVI1* is expressed in normal adult CD34+ haematopoietic progenitor cells. It is important to note that CD34 expression alone does not define a CML stem cell, but rather, an early progenitor (Graham et al., 2007). It is likely that the expression of *EVI1* could be further refined to a cell of origin with further selection for stem cell defining surface markers (Helgason et al., 2010).

In the above experiments it is shown that there was a correlation between CD34 content and *EVI1* expression in CP and that, in a single patient, the *EVI1* expression was derived from the CD34+ fraction. The majority of leukaemic cells in CP – which all carry the Philadelphia translocation and are thus clonally derived from a leukaemic progenitor – do not therefore

express *EVI1*. In BC, however, all patient samples showed a much higher CD34+ fraction (by definition over 20% and up to 97%) and yet one showed lower *EVI1* expression (down to just 0.007x), therefore, there must be a different mechanism responsible for the *EVI1* expression in CML-BC.

4.4.3 The full length *MECOM* transcript contributes little to the overall *EVI1* expression

Multiple RT-qPCR assays of the *EVI* ^{high} cell lines and the one available *EVI* ^{high} CML-BC patient sample were therefore undertaken using primer pairs designed to identify all described *EVI1* transcripts from the *MECOM* locus, including that of *MECOM*, as shown in Figure 4-8 (Lugthart et al., 2008).

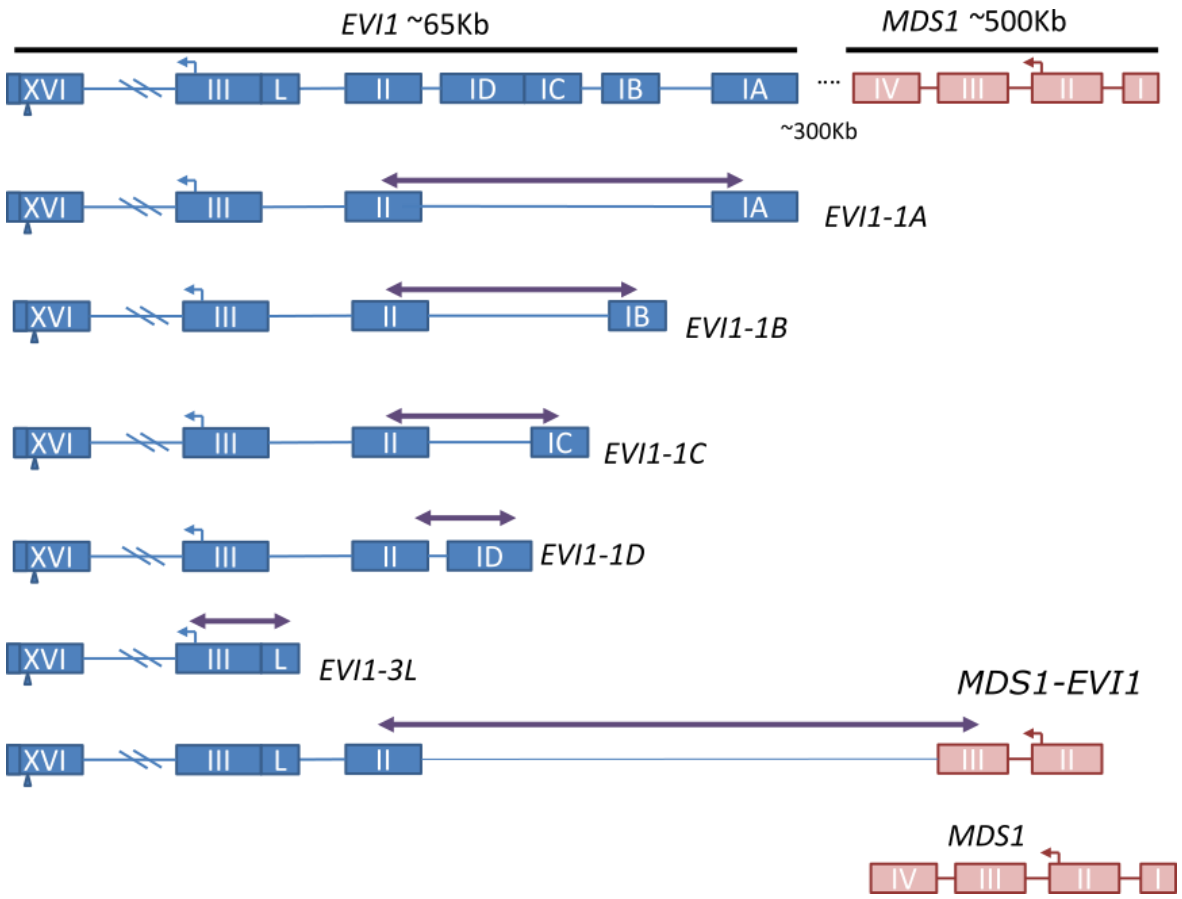


Figure 4-8: Location of primers used to detect all EVI1 transcripts¶*EVI1* exons are indicated in blue, *MDS1* in pink. The purple arrows indicate the different amplicons – not to scale. Adapted from (Lugthart et al., 2008).

The expression profile of the different transcripts expressed by each sample is shown in Figure 4-9; whilst K562 and LAMA84 have a broadly similar profile (albeit with a lower overall expression in the latter) with a majority of *EVI1-1A* transcripts. All three samples

expressed a small proportion of *EVI1-1D* and *-1B* transcripts, but none had detectable *EVI1-1C*. The significance of the alternative first exons is unclear, as all transcripts have the same translational start sites in exon 3 (Lugthart et al., 2008).

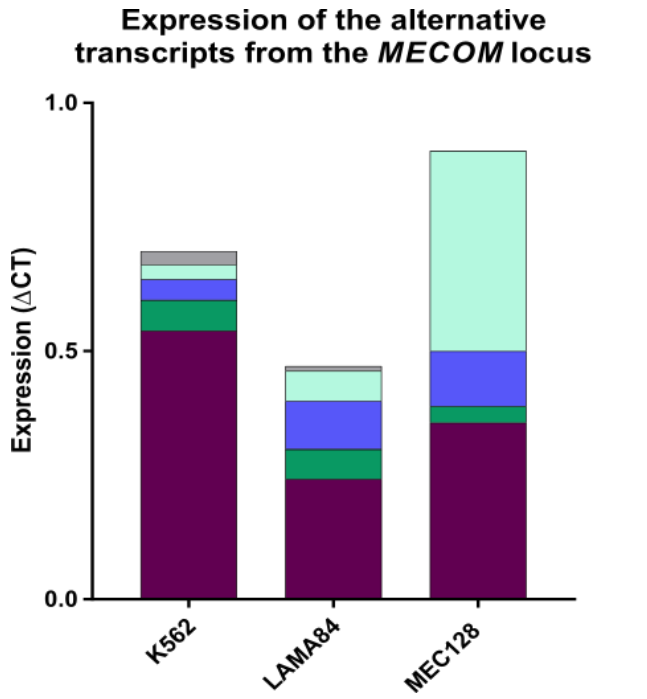


Figure 4-9: The CML-BC sample expressed a higher proportion of *EVI1-3L* transcripts than the CML cell lines. Bar chart showing expression of the six alternative transcripts expressed from the *MDS1-EVI1* locus in two high expressing cell lines and one CML-BC patient sample. Data shown is ΔCt of each transcript to *GUSB*.

The CML-BC patient sample (MEC128) has a noticeably larger proportion of *EVI1-3L*. This transcript has the usual exons 1 and 2 replaced by genomic sequence upstream of exon 3. Importantly, the *MECOM* transcript – which has a different *MDS1* promoter, approximately 500Kb upstream – had negligible (in LAMA84 and MEC128) or very low expression (in K562). This indicates that the relevant promoter for investigating promoter/enhancer interactions is the *EVI1* promoter which overlaps with exons 1A-1D.

4.4.4 Primary human blood and bone marrow samples from CML patients appear to lack H3K27Ac

Selected samples with *EVI1*^{high}, *EVI1*^{low} and *EVI1*$^{-}$ were then processed for H3K27Ac ChIP-seq. As used by Mansour et al. during their discovery of the *TAL1* super enhancer, this technique is widely used to identify enhancers. The first quality control step of the chromatin immunoprecipitation experiment, DNA quantification, showed protein-bound DNA was immunoprecipitated by the H3K27Ac antibody in all the samples (both cell line and primary CML samples) but not in the isotype control. The lack of DNA in the isotype control indicated that the ChIP experiment had not pulled down DNA indiscriminately and was thus successful (shown in Figure 4-10).

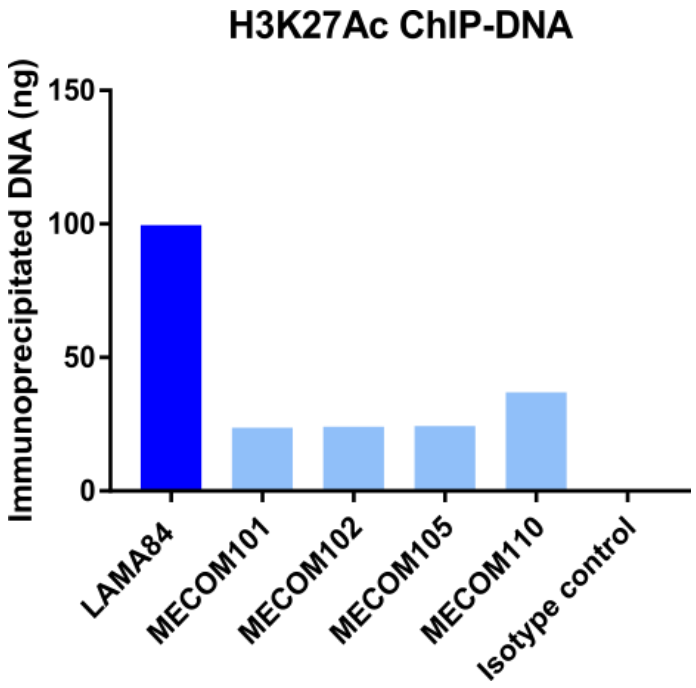


Figure 4-10: H3K27Ac ChIP yielded DNA from a human cell line and, to a lesser extent, primary CML cells ¶(all n=1)

The second quality control step, a Western blot to confirm that H3K27Ac was enriched in the immunoprecipitated lysate, unexpectedly showed a complete absence of H3K27Ac in the primary cells (see Figure 4-11).

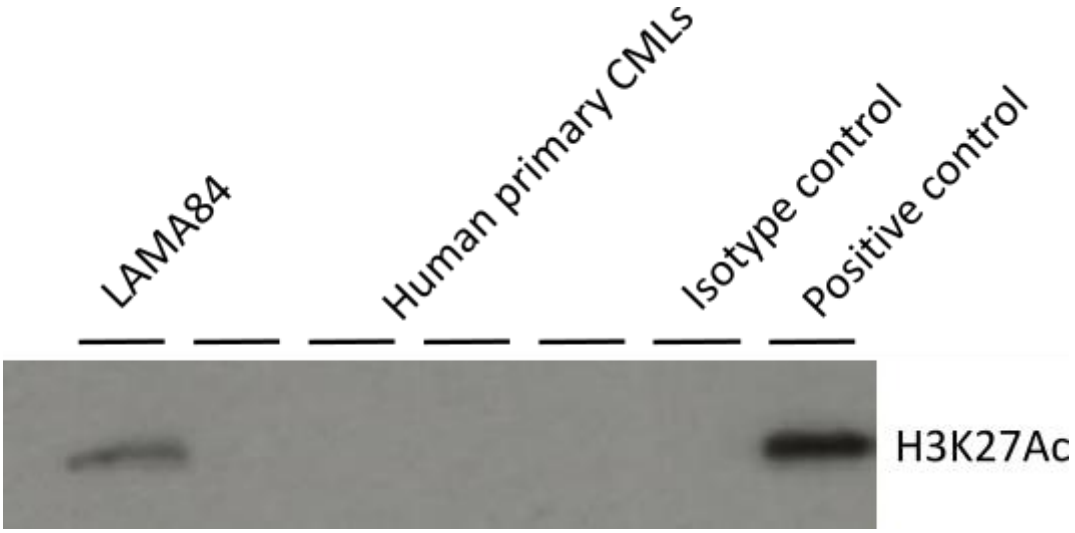


Figure 4-11: Western blot of H3K27Ac ChIP pull-down showed a complete lack of H3K27Ac primary CML.¶The human CML cell line (LAMA84) showed enrichment, indicating that this phenomenon was limited to primary cells – the same four as shown in Figure 4-10. Isotype and positive controls indicated the ChIP experiment and the Western blot were successful.

It has previously been reported that lysing of primary CML samples triggers a degradative activity – which cannot be inhibited by conventional protease inhibitors – which results in complete loss of the *BCR-ABL1* protein (Patel et al., 2006). A surrogate marker, such as the BCR-ABL1 substrate phospho-Crkl, is therefore required to assay for the kinase activity.

A further Western blot was therefore undertaken of primary cell lysate from a patient with CML-BC and a control patient (with chronic lymphocytic leukaemia) using antibodies for H3K27Ac and for total histone H3 (see Figure 4-12). The lysate was processed with and without histone deacetylase inhibitor (HDACi) according to the manufacturer's protocols. This showed both patients had histone H3 but confirmed that the CML patient had no detectable H3K27Ac as before. H3K27Ac was not increased in the presence of the histone deacetylase inhibitor indicating that the H3K27Ac was not present *in vivo* or that the deacetylation had occurred prior to the lysing of the cells.

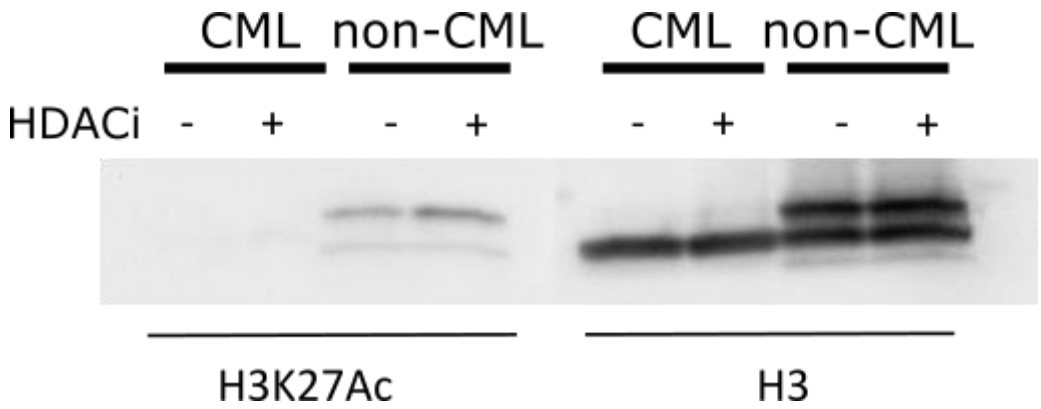


Figure 4-12 Western blot showed H3 - but not H3K27Ac - is present in primary CML lysate.¶Both H3 and H3K27Ac were present in a control non-CML protein lysate but only H3 was present in the protein lysate from the CML patient. Use of histone deacetylase made no detectable difference to the lack of H3K27Ac in the primary CML lysate.

In order to exclude the possibility that H3K27Ac was being degraded following lysis, immunohistochemical (IHC) staining for H3K27Ac was also undertaken using a bone marrow trephine biopsy taken from the same CML-BC patient. Bone marrow trephine biopsies are fixed directly in neutral buffered formalin and the cells are therefore not lysed.

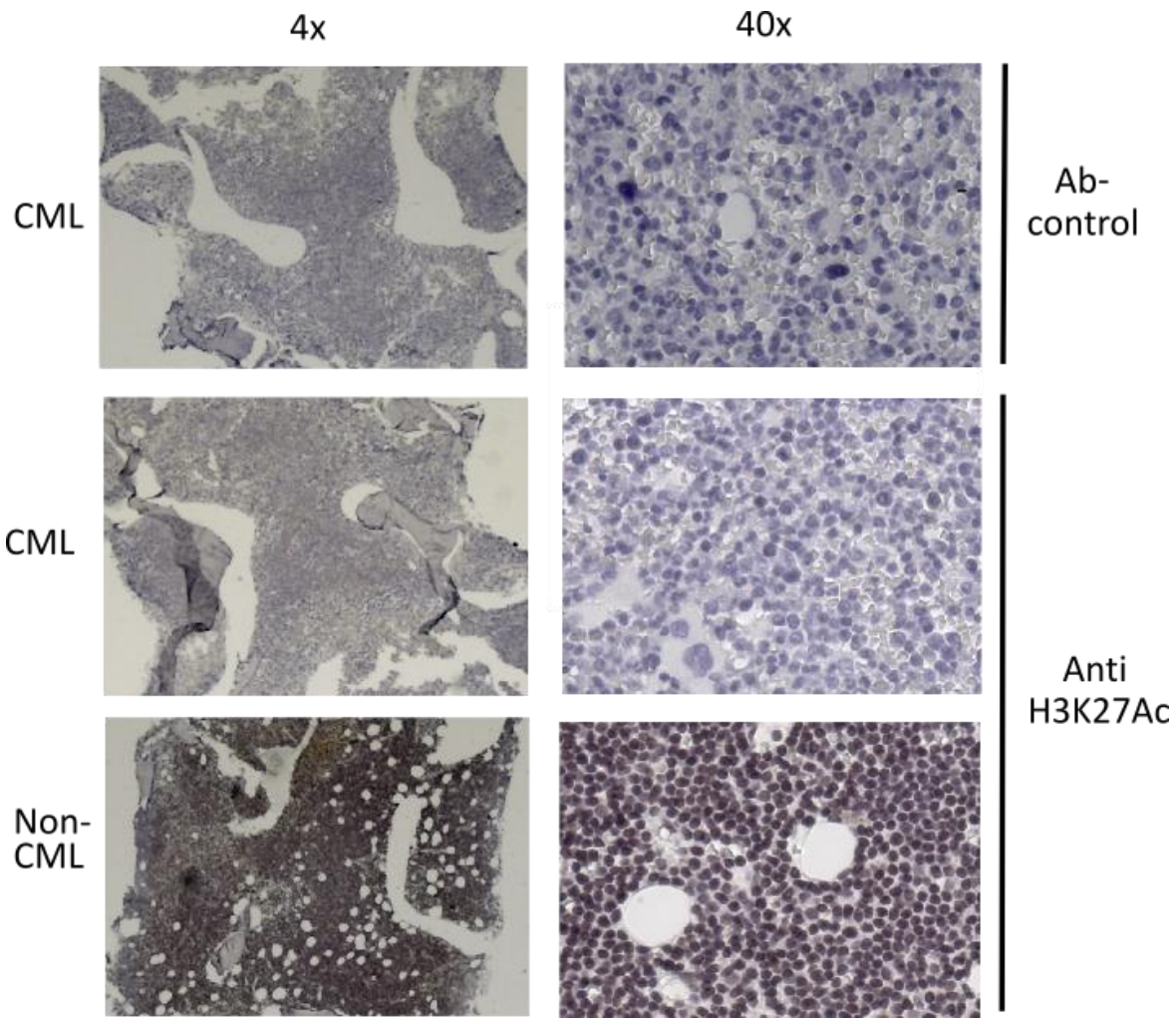


Figure 4-13: Immunohistochemical analysis showed no detectable H3K27Ac of a bone marrow trephine from a CML patient¶(middle). . Staining was comparable to the no-antibody control (top); a bone marrow trephine from a non-CML patient showed positive staining (bottom).

The IHC results (shown in Figure 4-13) showed strong H3K27Ac in the bone marrow trephine from the non-CML patient but no detectable H3K27Ac in the equivalent sample from the CML patient, again confirming the results from the Western blot. This result indicated that the deacetylation pre-existed *in vivo* and not as a consequence of sample handling or cell lysis.

This result is also in agreement with the work of Zhang et al, who demonstrated that CD34+ cells from newly diagnosed CML patients lacked acetylation of histones H3 and H4 and that these quiescent stem cells could be effectively targeted by TKI in combination with histone deacetylase inhibitors (Zhang et al., 2010).

4.4.5 H3K27Ac ChIP-seq is more sensitive than immunohistochemistry and Western blot

Despite the apparent absence of H3K27Ac by both IHC and Western blot, the immunoprecipitated DNA from all CML patients yielded sufficient ChIP-DNA for the successful preparation of ChIP-seq libraries. Therefore, H3K27Ac ChIP-seq analysis of CML-CP, CML-BC and CML cell lines was undertaken as planned (n=1 per patient or cell line) yielding successful alignments. A Venn diagram showing the peaks called in each category is shown in Figure 4-14.

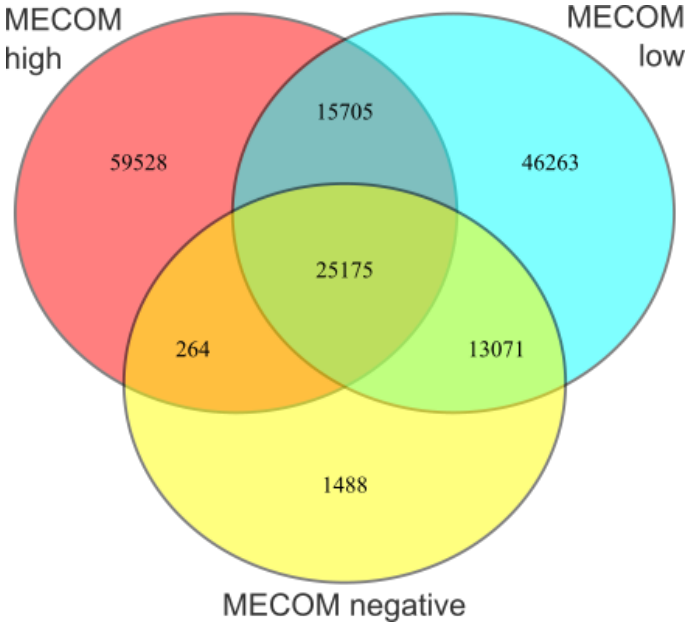


Figure 4-14: Venn diagram showing the number of peaks called in each category of MECOM expression.

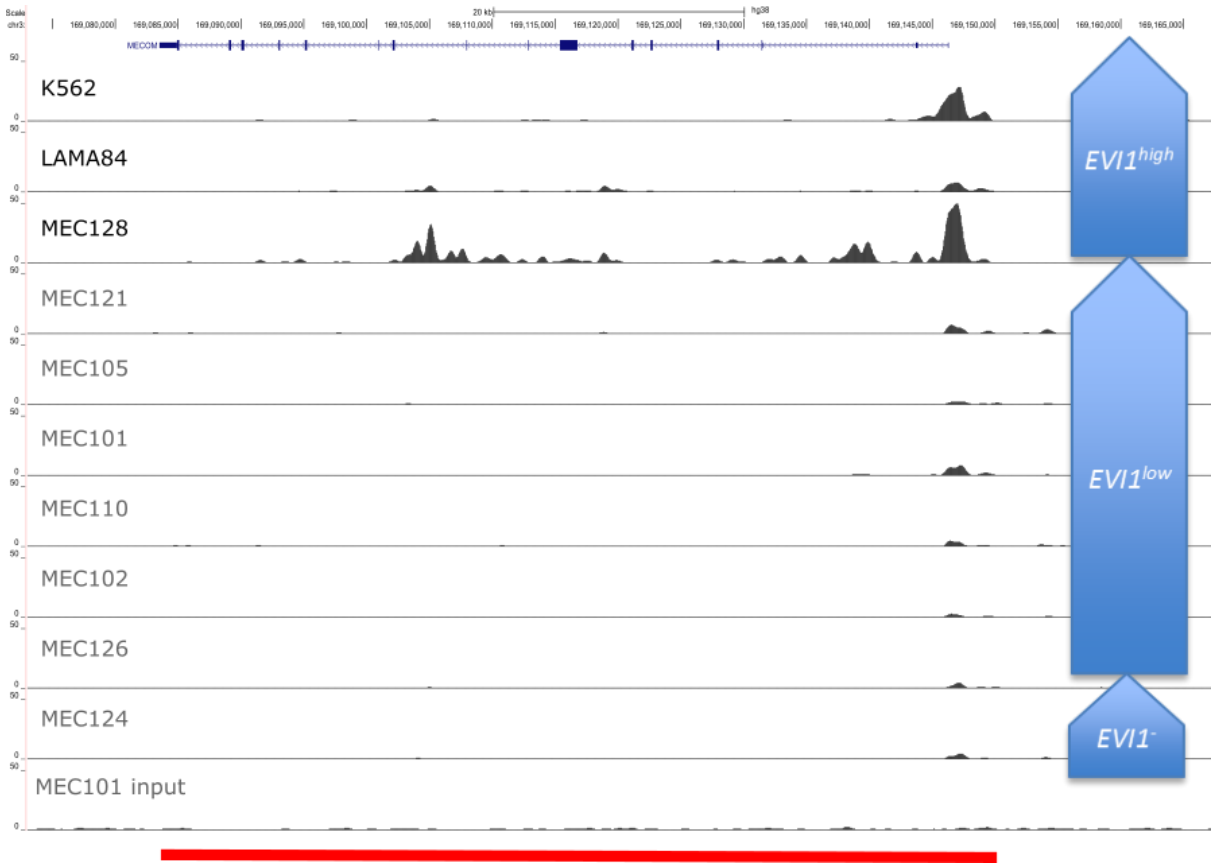


Figure 4-15: H3K27Ac ChIP-seq track showing the *EVI1* gene body ¶(highlighted in red). CML-CP samples are shown in grey and CML-BC samples are shown in black. A representative input track is shown at the bottom. *EVI1* expression classification is shown in blue on the right.

As shown in Figure 4-15 analysis of the H3K27Ac ChIP-seq showed a single shallow peak at the promoter in the CML-CP patient samples including the *EVI1*-negative CML-CP sample (MEC124). In contrast, a strong signal marked the promoter of both K562 and the CML-BC patient sample (MEC128). Notably, the CML-BC patient sample showed a series of H3K27Ac peaks throughout the gene body including a series of broad peaks which spanned a primarily intronic region of approximately 20Kb in the middle of the gene body. The strongest of these peaks were echoed in LAMA84 albeit at a lower intensity. H3K is not usually acetylated throughout gene bodies but is restricted to histones either side of the promoters/enhancers; this broad mark could therefore represent a novel oncogenic enhancer. Approximately half of all annotated enhancers are intergenic but they are less well studied (Shen et al., 2012). There is some evidence that intergenic enhancers attenuate the expression of the gene in which they are situated (Cinghu et al., 2017), however, this would be counterintuitive given the scenario above of activation of the *EVI1* oncogene.

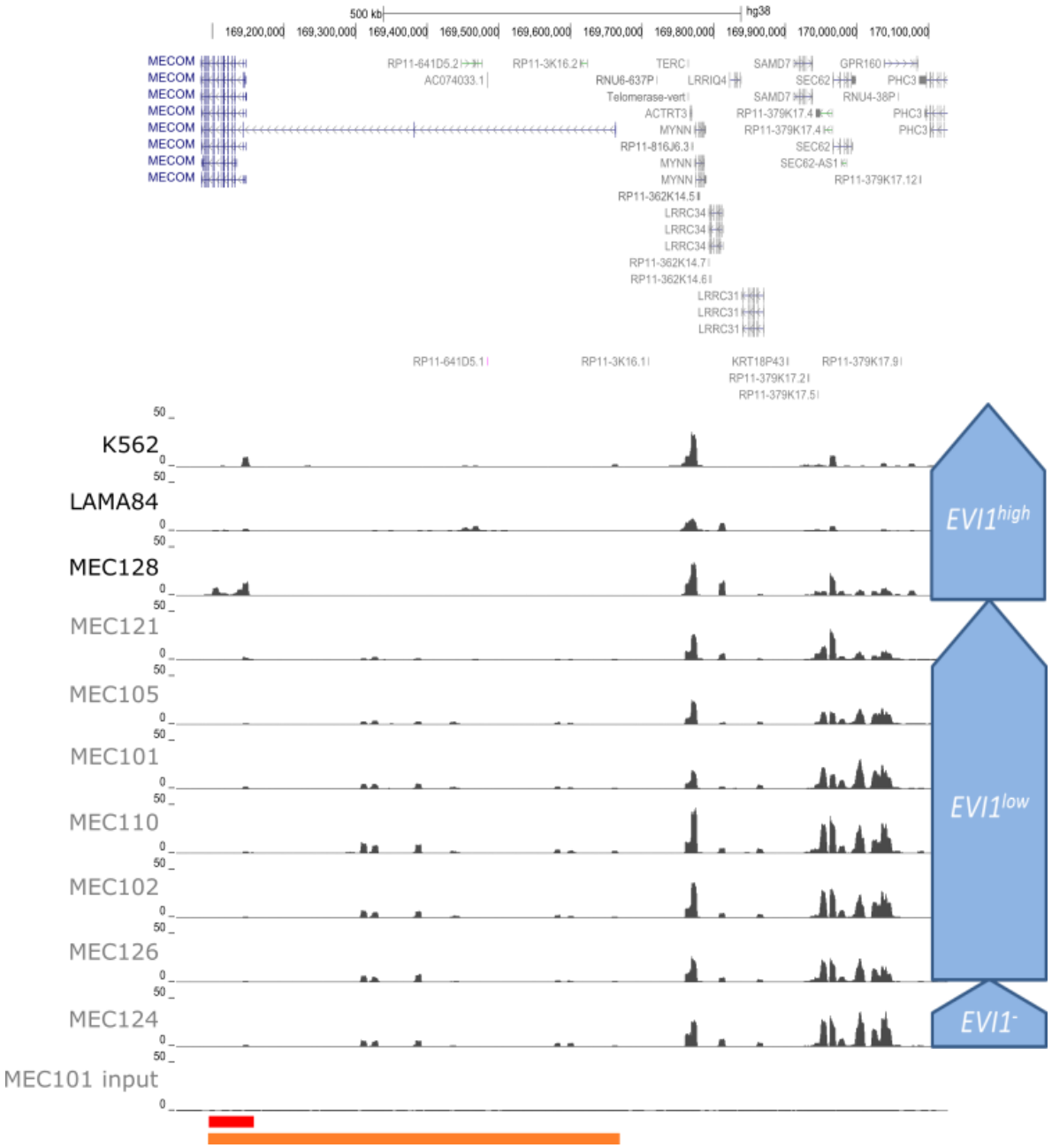


Figure 4-16: H3K27Ac ChIP-seq track showing the full *MECOM* gene body ¶(highlighted in orange) and the *EVI1* gene body (highlighted in red) and approximately 500Kb upstream. CML-CP samples are shown in grey and CML-BC samples are shown in black. *EVI1* expression classification shown in blue on the right.

As shown in Figure 4-16 the *MECOM* gene body and the region upstream of both, and indeed throughout the rest of the genome (data not shown), showed a huge amount of variation between the *EVI1*^{low} and *EVI1*$^{-}$ samples compared to the *EVI*^{high}; this was likely to be due to the effects of the EVI1 protein, which, as discussed in section 1.6 has a fundamental role in haematopoietic differentiation. Unfortunately, however, this made the identification of a candidate enhancer impossible. Hence, the H3K27Ac ChIP-seq could be used to validate a potential *EVI1* enhancer-promoter interaction but is insufficient to identify it.

4.4.6 Introduction of chromosome conformation capture to detect promoter-enhancer interactions

Chromatin conformation capture (3C) is a technique used to detect interactions within the genome which relies upon the likelihood of two interacting fragments being crosslinked and ligated together based on their proximity (Dekker et al., 2002). Whilst the original technique could only probe a suspected interaction between two known regions (one versus one), there have been various variant 3C techniques which have offered improvements on this, such as 3C-seq or 4C (one versus many (Stadhouders et al., 2013)) and Hi-C (all versus all (Belton et al., 2012). Of note, 3C was used to detect the interaction between the auto regulatory site of the *TAL1* complex and an intergenic site 8Kb upstream from the *TAL1* transcriptional start site (Zhou et al., 2013c)., which facilitated the finding of the super enhancer at a later date (Mansour et al., 2014).

The circularised chromatin conformation capture (4C) approach was therefore employed to directly interrogate chromatin-chromatin interactions involving the *EVI1* promoter region in the *EVI1*^{high} cell lines and CML-BC patient *EVI1*^{high} samples compared to *EVI1*^{low} and *EVI1*$^{-}$ patient samples. Chromosome conformation capture uses a region of interest to capture (hence, it is termed the 'bait') chromatin which physically interacts with it. It utilises restriction enzyme digest and a proximal ligation to capture the fragments of DNA that are bound by proteins in common – an overview of the technique is shown in Figure 4-17.

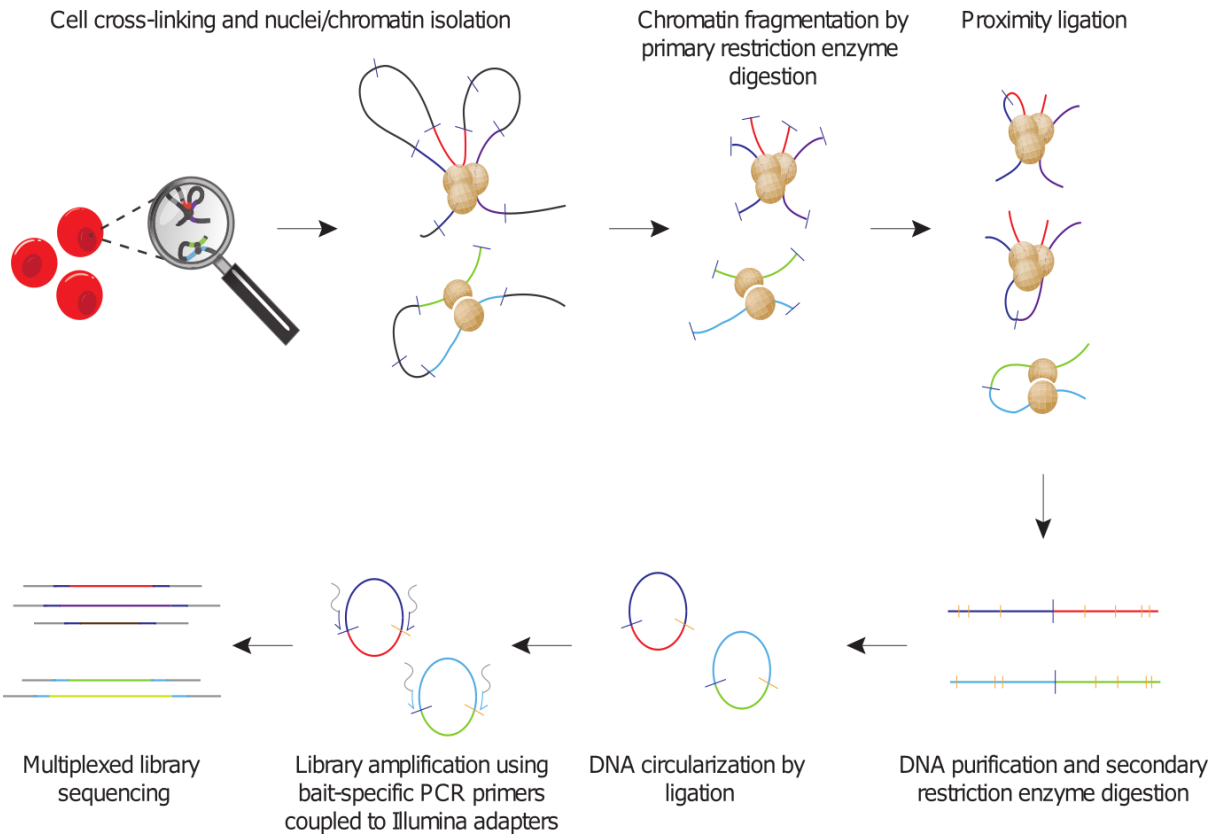
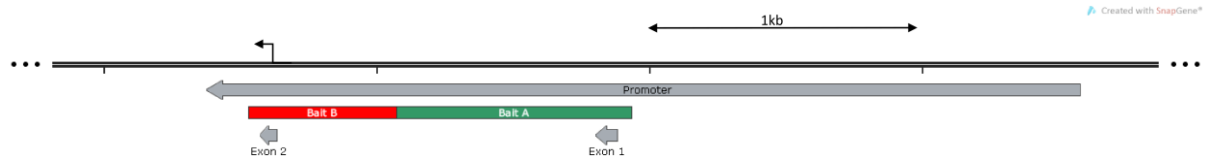


Figure 4-17: Overview of the 4C procedure.¶Cross-linked nuclei are digested with a primary restriction enzyme; the protein/DNA complexes are diluted to such an extent that when ligated, DNA fragments which were crosslinked into the same protein/DNA complex are likely to ligate to each other. DNA is then reverse-crosslinked to remove the protein and digested with a secondary restriction enzyme. A second ligation results in circularisation and finally, a 4C library is constructed by inverse PCR using primers specific to the region of interest (i.e. the *EVI1* promoter, or 'bait') with 3' overhang incorporating the sequences necessary for Illumina clustering, sequencing and multiplexing. Reproduced with permission from (Stadhouders et al., 2013).

Two baits – both entirely within the 3,200bp promoter region - were designed using EcoRI and DpnII as the primary and secondary restriction enzyme, respectively, as shown in Figure 4-18 (van de Werken et al., 2012).

Figure 4-18: Bait A and Bait B are located within the promoter region of *EVI1*.¶The translational start site is indicated with a black arrow, and the positions of exons 1 and 2 are indicated with filled grey arrows.



Various published protocols were adapted in order to use frozen, cross-linked pellets from the patient samples which had already been collected and stored according to the existing ChIP-seq protocol (see section 6.8.2) (Gondor et al., 2008; Stadhouders et al., 2013; van de Werken et al., 2012). Notably, these had been cross-linked with a final concentration of 1% formaldehyde (which was optimal for the ChIP-seq) rather than 2% most commonly used in the 3C protocols. Briefly, the frozen, cross-linked pellet was resuspended in water and homogenised on ice with a pellet pestle. This step – not required in ChIP because of the sonication – was necessary to allow access of the restriction enzyme to the chromatin. The sample was centrifuged and the supernatant removed; the pellet was resuspended in water and split into three separate digestion reactions. At this point, an aliquot was removed for Control 1 (the undigested sample). All reactions were digested with 500U of the primary restriction enzyme, EcoRI (except the Control 1) added at intervals over an 18-hour period to maximise the digestion efficiency. A second aliquot was removed from the digests for Control 2 (the digested sample). Controls 1 and 2 were reverse crosslinked and DNA was extracted using the phenol chloroform method and the digestion efficiency of the primary restriction enzyme was then checked using an RT-qPCR assay specific to the EcoRI restriction sites.

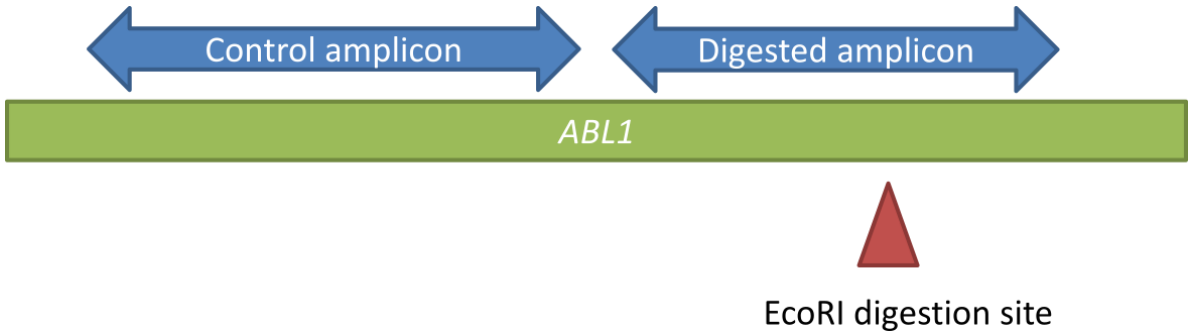


Figure 4-19: The qPCR assay accurately determines the digestion efficiency of the primary restriction enzyme¶ The real time amplification of two amplicons – one spanning an EcoRI restriction site (right) and one entirely within two EcoRI restriction site (left) – from control 1 (undigested) and control 2 (digested) were compared using the ΔΔCt method. See also 6.8.8¶

Two sets of *ABL1* primers were used; one amplicon spanned an EcoRI digestion site and one amplicon sat entirely between two EcoRI digestion sites as shown in Figure 4-19. The difference between the amplification of these controls indicates the digestion efficiency. Consistent with published observations, it was found that the cell lines showed a lower digestion efficiency (63.9-67.6%) than the primary cells (85.8 to 93.8%) as shown in Figure 4-20 (Stadhouders et al., 2013). This is likely due to nuclei clumping which inhibits access of the primary digestion enzyme to the chromatin.

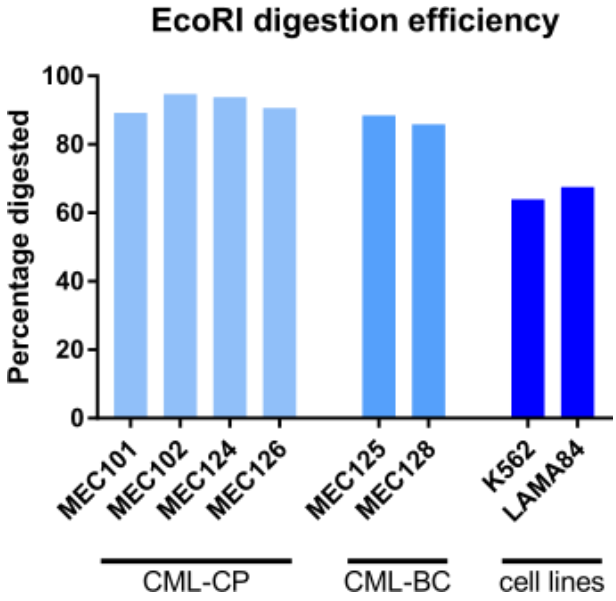


Figure 4-20: Digestion efficiency was lower in CML cell lines than primary CML cells¶ as reported in the literarture. Bar chart showing the percentage of EcoRI sites digested for primary CML samples in CP (light blue), BC (mid-blue) and cell lines (dark blue).

Meanwhile, the restriction enzyme in the remaining digests was inactivated and then a proximal ligation was carried out overnight. The libraries were again extracted using the phenol-chloroform method and quantified. At this point, the 3C libraries could be used to interrogate a suspected chromatin-chromatin interaction if both partners are known.

An aliquot of the extracted DNA was removed following primary digestion, proximal ligation, and secondary digestion and run on an agarose gel to allow for a visual check of the success of each of these steps (see Figure 4-21). The gel showed that the majority of the libraries were within the target range of 4-12Kb (for EcoRI, a six-cutter) following the first digestion, in agreement with the results of the RT-qPCR assay. This step is crucial, since inefficient digestion will result in over-representation of the uncut neighbouring fragment in the final sequencing. The proximal ligation returned the smear to a size close to genomic DNA following ligation and the second digestion resulted in the majority of the libraries within the target range of 0.1-2Kb (for DpnII, a four-cutter).

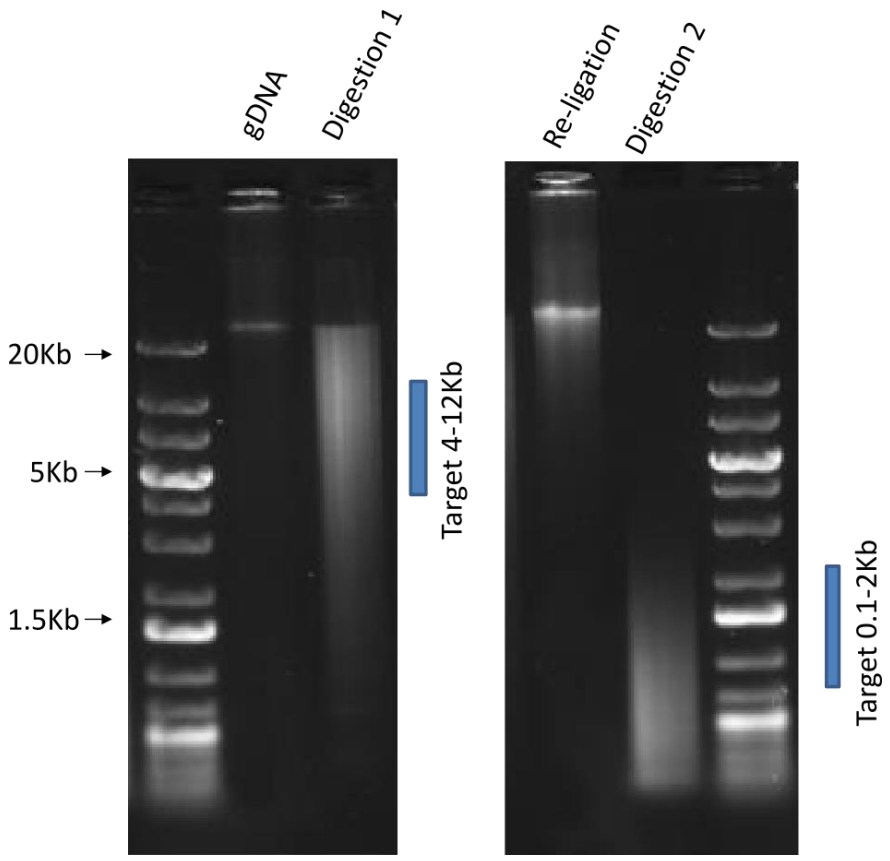


Figure 4-21: Agarose gel of a 4C library showed successful primary digestion ¶ (yielding a smear of DNA with the majority of fragments in the target range of 4-12Kb), ligation and secondary digestion (yielding a smear of DNA with the majority of fragments in the target range of 0.1-2Kb).

A 25µg aliquot of the 3C library was used to make a 4C (3C-seq) library – it was digested overnight with 50U of the secondary restriction enzyme, DpnII, at 37˚C and DNA was again extracted using the phenol-chloroform method. The libraries were circularised by dilution to a large volume and ligation at 16˚C overnight. A final phenol-chloroform DNA extraction yielded the template for 4C library construction by inverse PCR.

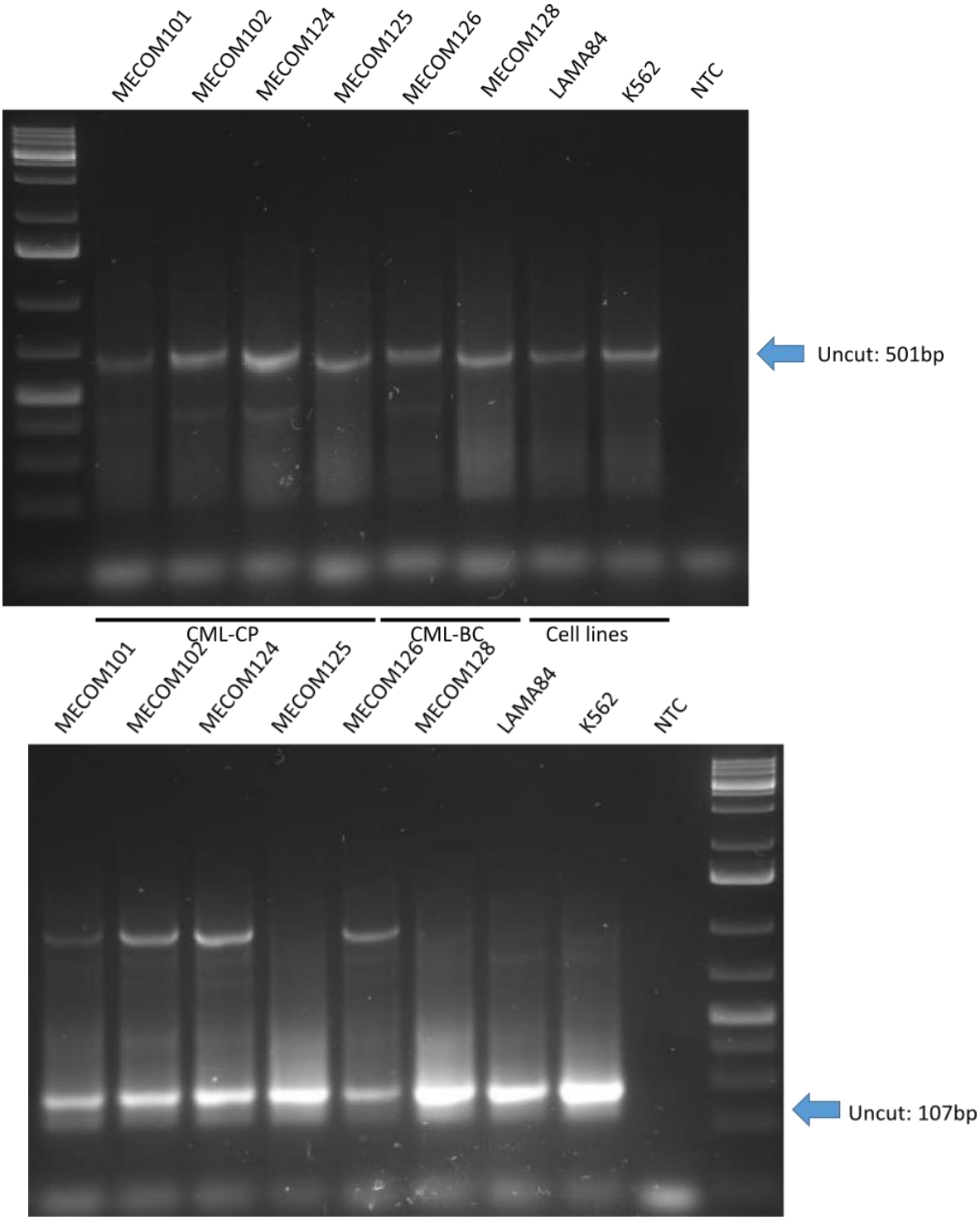


Figure 4-22: Inverse PCR of the 3C libraries yielded 4C libraries with additional bands and smears indicating potential interactions.¶Bait A (top) has a prominent band of 501bp and bait B (bottom) has a prominent band of 107bp; these are consistent with the neighbouring uncut fragment. The additional bands and smears indicate potential interactions with the promoter region. Note: both uncut bands are higher molecular weight than predicted because the library incorporates the 5' overhang necessary for Illumina sequencing.

As shown in Figure 4-22, primers for both Bait A and B yielded bands and smears in addition to the predicted neighbouring fragment from uncut restriction digest sites. These additional bands and smears indicate potential interactions which would be revealed by sequencing. The bait primers were therefore re-designed to incorporate a 5' overhang with the Illumina adapter sequence (for clustering of the library in the flow cell), the Illumina sequencing primer (from which the unknown sequence and the index sequence is read) and an index sequence (to allow for de-multiplexing, but which also introduces diversity necessary for clustering in the Illumina flow cell) as shown in Figure 4-23.

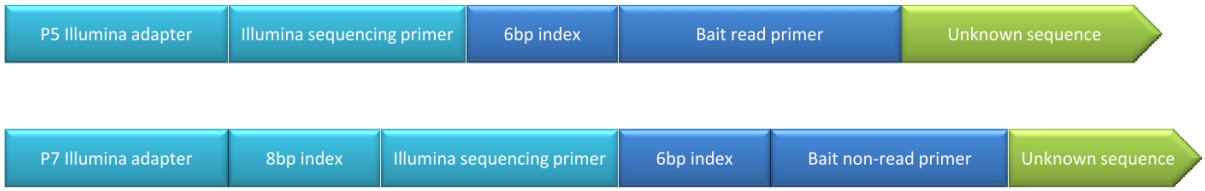


Figure 4-23: Primer design for 4C library construction. ¶The 6bp index immediately adjacent to the Illumina sequencing primer introduces sequence diversity (necessary for clustering) which the bait read and nonoread primers lack. The P7 reads are not mapped.¶

Analysis was undertaken using the 4CSeq R package as described in (Klein et al., 2015). Reads from each patient sample were de-multiplexed from each other using the unique 6 bp index sequence. The bait primer sequence, which becomes incorporated in to the read, was trimmed away. The remaining reads were aligned to human genome (build GRCh38/hg38) and the reads aligning at each *in silico* restriction fragment were counted. Variance stabilising transformation was used to normalise the reads with DESeq2 R package (Love et al., 2014). It is assumed that the count of background reads will decline symmetrically from the bait until it reaches a constant. Reads therefore underwent trend fitting based on this signal decay using the 4CSeq R Package.

To identify significant interactions, the three *EVI1*^{high} samples were compared to the same number of *EVI1*^{low} patient samples (the three with the most similar libraries). Fragments with a median of less than 40 read counts across all samples were removed to minimise the background noise. Fragments with read counts higher than expected at a given distance from the bait were given a z-score based on the residuals of the fit. Those fragments with z-scores larger above 3 in all three replicates and with an adjusted p-value of less than 0.01 in at least one replicate were called as significant.

As shown in Figure 4-24, analysis of normalised counts from Bait A revealed multiple interactions (black dot) with the bait (location of the bait is indicated with a dark red arrow). Two clusters of significant interactions (each interaction is shown as a red dot and the two

clusters are highlighted with dashed boxes) were identified at approximately 500kb and 600kb upstream of the bait in the *EVI1*^{high} patient samples. Only one interaction (shown as a green dot) of this proximal cluster (highlighted with a green dashed box) was detected in the *EVI1*^{low} samples. Interactions that did not reach the threshold of significance but that were differential between the *EVI1*^{high} and *EVI1*^{low} patient samples are shown with a blue diamond.

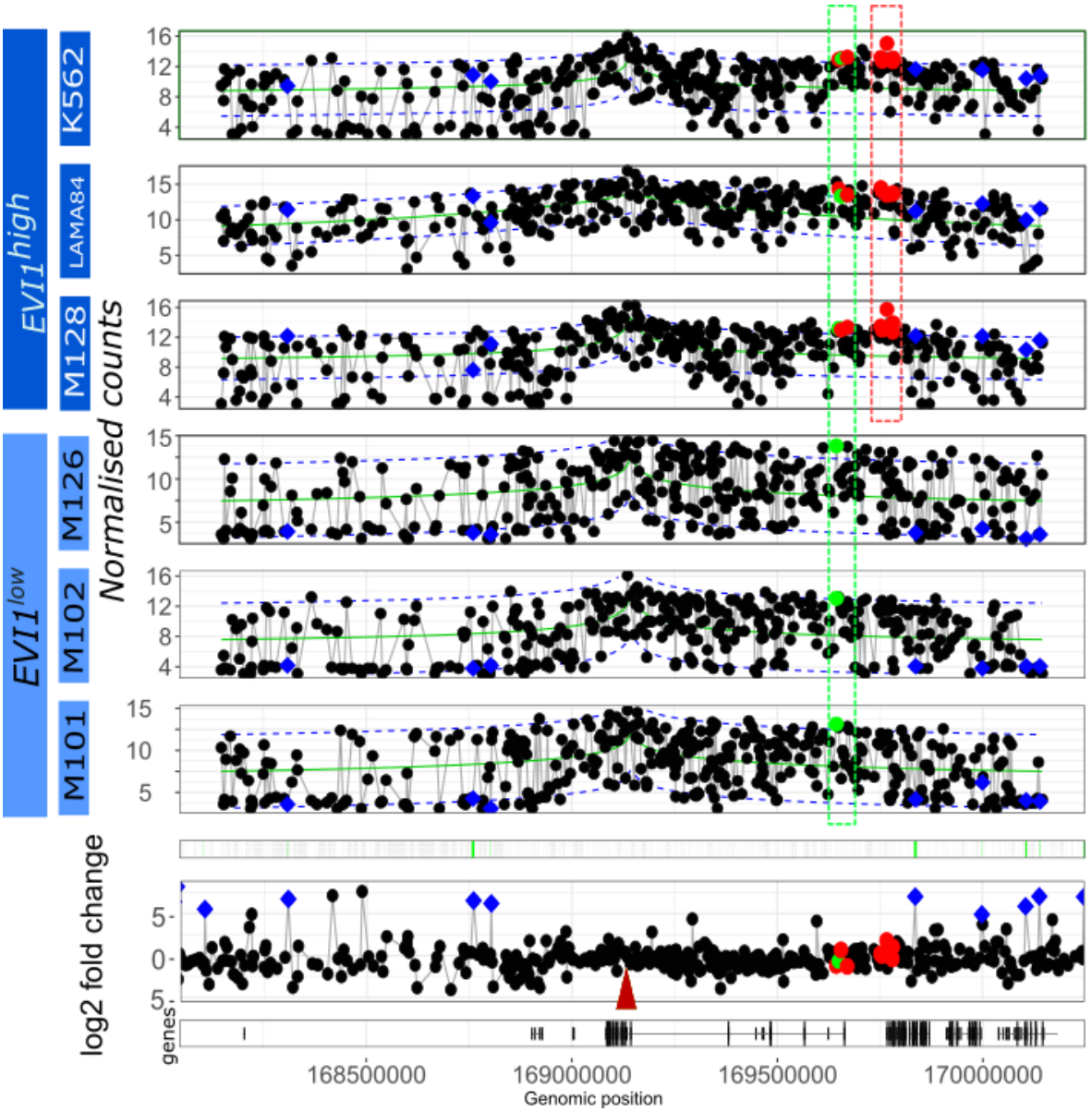


Figure 4-24: Analysis of normalised counts from Bait A revealed significant interactions in the *EVI1*^{high} samples which were not present in the *EVI1*^{low} samples.¶Custom FourCSeq plot showing the variance stabilised (normalised) counts (black dots) from 4C libraries of three¶*EVI1*^{high} samples (top) and three *EVI1*^{low} patient samples (middle). The fit of the distance dependence is shown with a green line and the dashed blue lines indicate +/-3SD. Significant interactions (those with a z-score above 3 in all replicates and a p-adjusted <0.01 in at least one replicate) are shown as a red dots (*EVI1*^{high} only) or green dots (common to both *EVI1*^{high} and *EVI1*^{low}) and highlighted with a dashed box. Differential changes in the contact profile that are not called as significant (those present in *EVI1*^{high} but not *EVI1*^{low} or vice versa) are indicated with a blue diamond. The green bars below the profile indicate where the *EVI1*^{high} has the higher signal for the differential change. The calculated log2 fold change of the differential testing per fragment and the gene maps of the region are shown at the bottom. The location of the bait is indicated with a dark red arrow. Bioinformatic analysis and figure by Dr Mark Robinson.

A second approach was taken to identify differential interactions between the *EVI1*^{high} *and* *EVI1*^{low} patient samples. Reads were processed as before, up to the variance stabilisation (normalisation) step. Differential region calling was then undertaken to identify fragments which were significantly over- or under-represented in the *EVI1*^{high} *and EVI1*^{low} patient samples. These interactions are shown in the WashU circlet plot in Figure 4-25 and span most of the long arm of chromosome 3, reaching 53Mb centromerically and 24Mb telomerically (Zhou et al., 2013a).

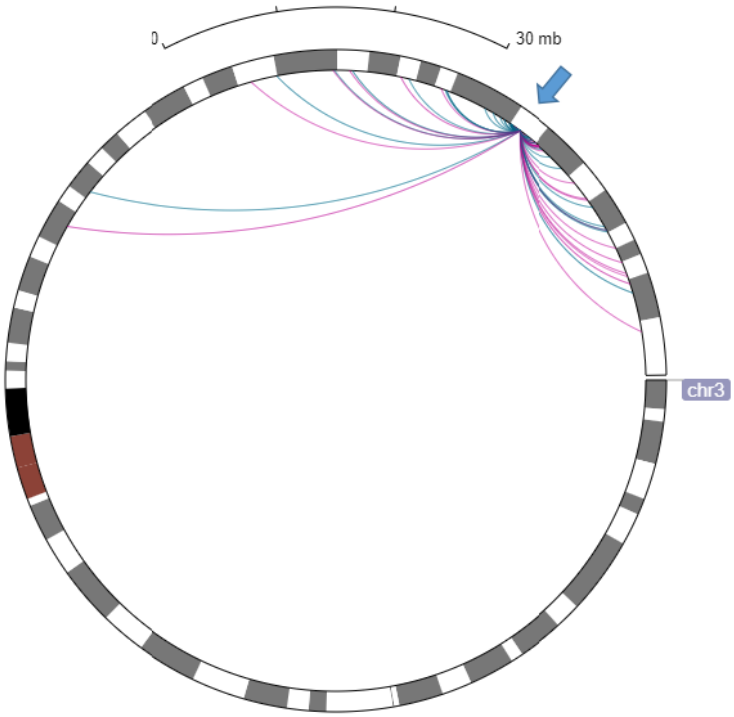


Figure 4-25: Long-range differential interactions were found but were restricted to the long arm of chromosome 3.¶Custom WashU circlet plot showing positive (blue) and negative (pink) differential interactions.

The positive (enriched in *EVI1*^{high} samples compared to *EVI1*^{low} samples) and negative (enriched in *EVI1*^{low} samples compared to *EVI1*^{high} samples) interactions are displayed alongside 4c tracks showing three merged results from *EVI1*^{high} samples and three from *EVI1*^{low} samples in Figure 4-26. A cluster of positive interactions surrounds the *GOLIM4* region downstream and a cluster of negative interactions surrounds the *PLD1* region upstream – these two regions are shown in more detail in Figure 4-27 and Figure 4-28 respectively.

The restricted size of patient samples meant that not all samples were available for all experiments – a summary is shown in Table 4-2.

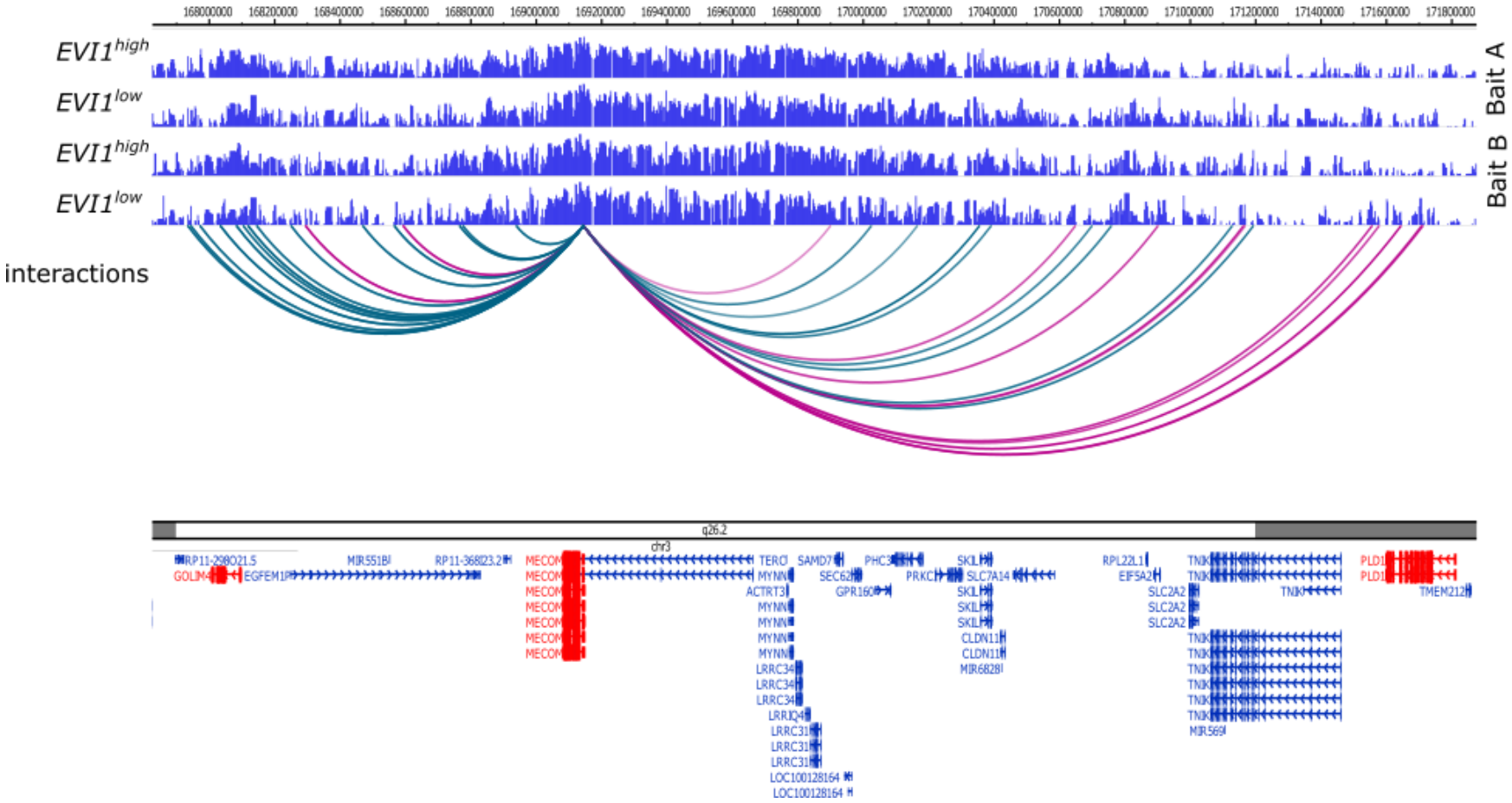


Figure 4-26: Analysis of the 4C reads revealed multiple significant differential interactions both up and downstream of *MECOM*. Custom WashU track showing H3K27Ac ChIP-seq peaks and the differential regions detected – positive interactions (those present in *EVI1*^{high} and not in *EVI1*^{low}) are indicated with blue arcs, negative interactions (those present in *EVI1*^{low} and not in *EVI1*^{high}) are shown in purple. The *EVI1* region of *MECOM* is indicated in red. Two genes in areas of interest are also highlighted in red. A cluster of positive interactions in and around *GOLIM4* (downstream) are shown in more detail in Figure 4-27, and a cluster of negative interactions in and around *PLD1* (upstream) are shown in more details in Figure 4-28.

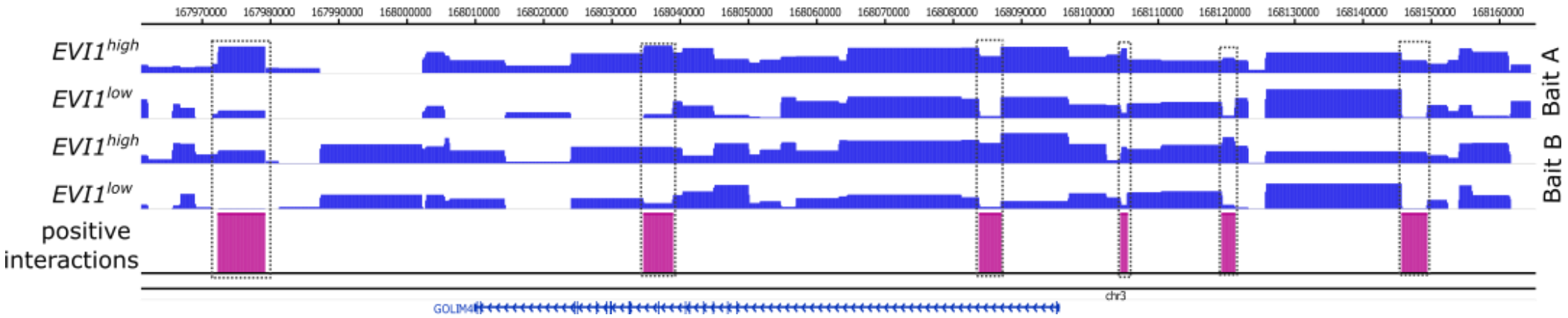


Figure 4-27: 4C revealed multiple interactions in the region of *GOLIM4* approximately 1.1Mb downstream of the baits. ¶Custom WashU track showing merged 4C reads from baits A and B in *EVI1*^{high} samples and *EVI1*^{low} samples. The positive interactions – those which were detected in the *EVI1*^{high} significantly more often than in *EVI1*^{low} - are shown with purple bars (width indicates restriction fragment size only and is not an indication of interaction significance). ¶

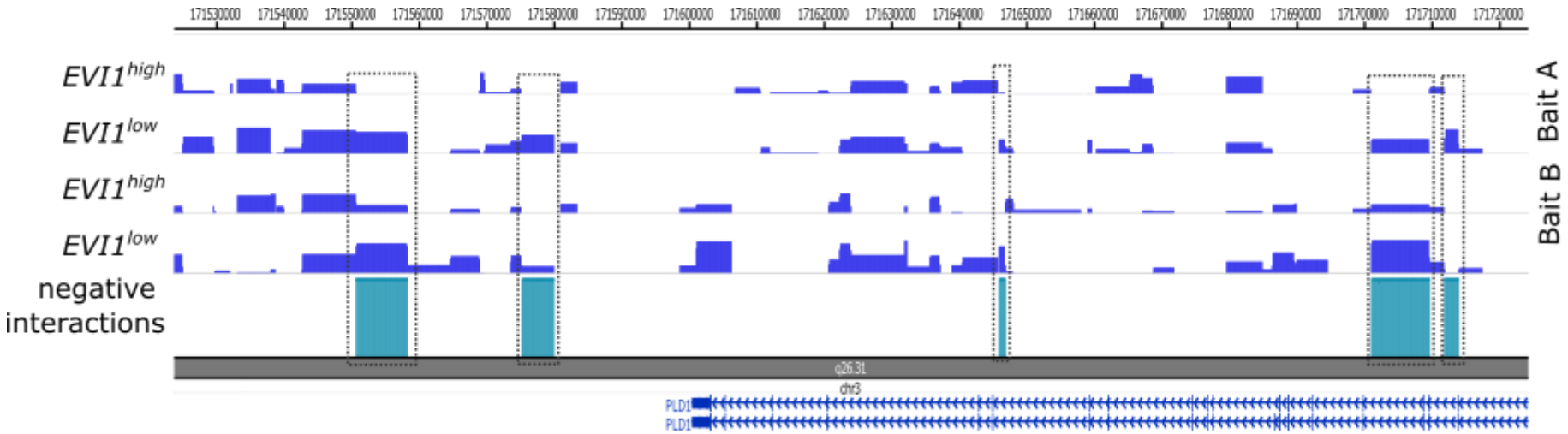


Figure 4-28: 4C revealed multiple interactions in the region of *PDL1* approximately 2.4Mb upstream of the baits. ¶Custom WashU track showing merged 4C reads from baits A and B in *EVI1*^{high} samples and *EVI1*^{low} samples. The negative interactions – those which were detected in the *EVI1*^{low} significantly more often than in *EVI1*^{high} - are shown with turquoise bars (width indicates restriction fragment size only and is not an indication of interaction significance). ¶

Table 4-2: Summary of patient samples used in ChIP-seq and 4C experiments. ¶The limited size of the patient samples meant that not all samples could be used for both H3K27Ac ChIP-seq and 4C experiments.

Sample	Disease	*EVI1* expression	4C	ChIP-seq
MEC101	CML-CP	Low	✓	✓
MEC102	CML-CP	Low	✓	✓
MEC105	CML-CP	Low		✓
MEC110	CML-CP	Low		✓
MEC121	CML-CP	Low		✓
MEC124	CML-CP	Negative	✓	✓
MEC125	CML-CP	Low	✓	
MEC126	CML-BC	Negative	✓	
MEC128	CML-BC	High	✓	✓

5 Discussion

Aim 1: Does oncogenic stress promote genetic instability associated with leukaemia progression?

It is now well recognised that oncogenic lesions in leukaemia affect lineage-specific genes (Mullighan et al., 2007; Mullighan et al., 2009b; Zhang et al., 2011). It has not been understood, however, whether these lesions arise because lineage specific genes are actively targeted by DNA damage or whether they arise through a process akin to natural selection following random DNA damage. Early indications that random abnormalities are not simply selected for came from mouse experiments where chromosomal translocations were induced in different cell lineages; successful transformation to leukaemia would occur in one lineage but not the other, such as *MLL-MLLT1* [t(11;19)] in a myeloid but not lymphoid lineage (Cano et al., 2008).

The data presented here collectively indicate that lineage specific genes are natural targets of oncogenic stress-related DNA damage occurring during oncogene-induced malignant transformation of B-cell precursors, and potentially throughout disease progression towards leukaemia. While genomic alterations in B-lineage leukaemia have been associated with off-target activity of the DNA-damage inducing enzymes AID and RAG1 (Teng et al., 2015), the data presented here adds oncogenic stress as a third factor to the list of potential threats to the leukaemia genome. Notably, while it was shown that oncogenic stress increases the fragility of transcribed DNA, others showed that AID (Chaudhuri et al., 2003; Meng et al., 2014; Ramiro et al., 2003) and RAG1 (Papaemmanuil et al., 2014) preferentially target transcribed loci. Lineage-specific transcriptional programs hence expose highly transcribed, lineage-specific genes to all three threats, thereby facilitating genetic alterations at these loci.

First, it was shown that healthy B-cell precursors induced to express the oncogenes *BCR-ABL1* or *MYC* experience DNA damage as evidenced by increased DNA-damage associated phosphorylation of P53 (Ser15) and H2AX (Ser139, γH2AX) compared to control cells. While oncogene expression has been shown in many cellular models related to solid tumours to promote DNA damage (Perera et al., 2016; Supek et al., 2017) its relevance for genetic instability related to leukaemia was less clear. Importantly, however, it was shown in this experiment that the DNA damage detected did not relate to the activity of *AID* or *RAG1* (2.4.1), thereby distinguishing this work from others who have investigated the mutational activity these enzymes. By excluding the effects of these sources of significant endogenous

damage, it is ensured that the damage detected experimentally derived directly from the expression of the oncogene.

Consecutive analysis by γH2AX-ChIP-seq allowed further definition and analysis of the genomic regions/loci affected by DNA damage in oncogene expressing cells. It is recognised that γH2AX histone modification can spread over extended distances around a DNA DSB (Bothmer et al., 2011), however the results of the γH2AX experiments were highly reproducible indicating the potential spread of γH2AX spreading did not confound the results.

It was interesting to note that *BCR-ABL1* and *MYC* induced DNA damage resulted in comparable γH2AX signal intensities: 99% of the peaks with a signal above background were in common between the two oncogene-induced populations with a correlation of 0.89. It was shown as early as 1986 that ectopic expression of *v-abl* eliminates interleukin-3 dependence (Mathey-Prevot et al., 1986) and later , using a temperature sensitive murine expression construct that this independence was due to constitutive activation of *myc* by *v-abl* (Cleveland et al., 1989). Later, it was shown that transformation by *BCR-ABL1* is dependent on *MYC*, since certain dominant negative forms of *MYC* prevented malignant transformation (Dang et al., 1989; Sawyers, 1993). Hence, the similarity between the range of genes damaged by the activity of the two different oncogenes may be attributable to the common *MYC* pathway. Alternatively, this may simply be caused by an overlapping mechanism – such as replicative stress triggered by the increase in proliferation – which affects both models to a similar extent.

The main conclusion from this experimental work was that the detected DNA damage was enriched at lineage specific genes. This is of importance as driver (and subsequent passenger) mutations found in the genomes of leukaemia patients frequently arise in lineage specific genes causing maturation and/or proliferative defects, which are fundamental in human leukemogenesis (Rosenbauer and Tenen, 2007). In agreement, many of the identified γH2AX regions locate to genes frequently mutated in human leukaemia (2.4.3). The predisposition to DNA damage observed in lineage specific genes thus may explain the frequent occurrence of lineage specific secondary abnormalities as a consequence of positive selection during clonal evolution as is seen in human B-ALL.

In order to confirm the ChIP-seq results, a set of γH2AX positive and γH2AX negative loci were validated by DNA FISH using dual colour break apart probes, more precisely targeting the γH2AX hotspots *PAX5* and *BLNK*, and a cold spot, *FOXP2*. Since this is a relatively small

number of loci, further genome-wide assays could be employed to corroborate this finding. This might include double strand break end capture (Lensing et al., 2016), double strand break capture sequencing (END-seq) (Canela et al., 2016) or translocation-capture sequencing (TC-seq) (Klein et al., 2011) which would have the added benefit of refining the breakpoints compared to γH2AX signals. However, in addition to the loci analysed by DNA FISH in oncogene expressing B-cell precursors, additional de novo γH2AX hotspots in lymphoid-to-myeloid converted BCR-ABL1 expressing B-cell precursors (Figure 2-47) were verified, thereby increasing the overall number of loci confirmed by DNA FISH.

While this work mostly focused on the acute response to oncogenic stress, one of the experimental models was performed at a different time point: in most experiments oncogene expression was induced in B-cell precursors and assessed at day seven post infection when cells were suffering from a period of acute oncogenic stress, but in the lymphoid-to-myeloid lineage switch experiments were conducted at day 30+ when cells were successfully transformed. Both experiments revealed transcription-associated fragility indicating the DNA damage is not restricted to the period of acute stress following oncogene activation.

Functionally, it was demonstrated that the γH2AX region formation coincided with active transcription, as determined by the H3K27Ac histone modifications at promoters of genes that overlapped with γH2AX regions, and by the association of γH2AX+ genes with RNA expression (2.4.4). Whilst transcription-associated DNA damage is not a newly recognised phenomenon – it was first noted in *E. coli* in 1971 (Herman and Dworkin, 1971)– several recent key studies have dramatically increased our understanding of the mechanism.

The orientation and proximity of transcribed genes has been the subject of intense study recently. In agreement with several of these publications, convergent and divergent genes in close proximity - which lead to exposure of single stranded DNA - were enriched in the γH2AX marked genes (Heinaniemi et al., 2016; Pannunzio and Lieber, 2016). In bacteriophage, there is evidence that two advancing elongation complexes can sidestep each other in the approaching 'transcription bubble' resulting from head to head orientation (Ma and McAllister, 2009). However, in higher eukaryotes the size, stability and/or how tightly the non-transcribed strand is held in the complex conspire to make this impossible: Hobson et al. established that two RNA PolII elongation complexes cannot pass each other in vivo, and that this was due to collision of the front edges detected in vitro (Hobson et al., 2012). It has also been proposed that the loci where transcribed regions converge are vulnerable to the reproducible 'off-target' effects of AID, and thus DNA damage (Meng et al., 2014).

In addition, we observed an overlap with γH2AX regions and established causes of transcription-associated fragility such as the presence of R-loop forming sequences and early-replicating fragile sites (2.4.5). R-loop formation – which occurs both during normal and malignant transcription – is an essential step in RNA processing that can cause damage if not resolved. Notably, when they occur in G-rich RNA, the DNA:RNA hybrid molecule is actually more stable than its DNA double helix counterpart (Lesnik and Freier, 1995) so the persistence of R-loops must be actively regulated to prevent permanent damage. RNA:DNA hybrids within R-loops can be unwound by helicases to overcome transcriptional pauses (Skourti-Stathaki et al., 2011) and conversely, topoisomerases can prevent the formation of R-loops by exposing single stranded DNA (Drolet, 2006). The suppression of either will result in accumulation of DNA damage. *FIP1L1* mutants found in chronic eosinophilic leukaemia have mRNA cleavage and polyadenylation defects; cells with defective *FIP1l1* have increased DNA-RNA hybrids and concomitant DNA damage, both of which could be reversed using RNAseH, an enzyme which degrades R-loops (Stirling et al., 2012).The preponderance of GC-rich regions at CpG islands means R-loops are omnipresent at transcription start sites and termination sequences (Ginno et al., 2012). R-loops can be accurately detected using DNA-RNA immunoprecipitation (DRIP-seq; (Chan et al., 2014)) but also predicted computationally, as described in 2.4.5. Evidence would suggest that timely formation and timely resolution are both important to minimise DNA damage.

Reports of lineage specific DNA damage are not restricted to haematopoiesis. The Cancer Genome Atlas network has recently provided a wealth of knowledge into the mutational landscape and even mutational signatures of various solid tumours. A striking example was found in whole genome sequencing of 79 lung adenocarcinomas (with matched non-tumour tissue). The top three loci found to have indel mutations were found to overlap with genes *SFTPB, SFTPA1* and *SFTPC/BMP1* (Imielinski et al., 2017). These are all genes encoding surfactant proteins which lower the surface tension to allow lungs to fill with air and participate in the innate immune response to lung infection (Haagsman and Diemel, 2001) and are thus highly expressed in a tissue-specific manner. The authors went on to extend their analysis to other solid tumours and discovered non-coding albumin mutations in gastric cancer, thyroglobulin in thyroid cancer and gastric lipase in liver cancer (Imielinski et al.; Imielinski et al., 2017)

In addition to the mechanisms discussed here, there are several other mechanisms that could potentially contribute to the DNA damage accumulation but which have not been studied. This includes the link between replication and transcription; because both processes occur on the same template, there is an inevitable collision of the transcription and the

replication machinery. This can happen as a 'head-on collision' when the lagging strand is being transcribed, or co-directional, when the leading strand is being transcribed (Kim and Jinks-Robertson, 2012). Unsurprisingly, head-on collisions are considered to be more damaging and the increase of co-directionally transcribed genes in the regions surrounding human replication origins would suggest there is a selection pressure against head-on collisions (Huvet et al., 2007). Similarly, the access of both transcription and replication machinery results in torsional stress which has to be relieved intermittently to allow fork progression. This torsional relief is achieved enzymatically by topoisomerases – type I, which breaks each strand in turn, and type II, which breaks both in concert (Wang, 2002).

The findings described above suggested that lineage-specific transcriptional programs might sensitize lineage specific genes to DNA damage in oncogene expressing cells. In order to functionally validate this hypothesis, a lymphoid-to-myeloid lineage switched cell model established by and frequently used by the Müschen laboratory (Shojaee et al., 2015) was employed. Using this model it was shown that DNA damage was increased in myeloid genes following myeloid lineage induction and concomitant transcription of myeloid genes (2.4.6). Whilst the data presented before referred primarily to B-cell precursors, the data from the lymphoid-to-myeloid switch experiment would suggest that the findings would be replicable in oncogene expressing cells of other lineages. The reciprocal experiment is technically difficult, however, due to the requirement for enforced expression of more than a single transcription factor for myeloid-to-lymphoid lineage switch (Laiosa et al., 2006).

Of note, not all highly expressed genes were subject to γH2AX-marked DNA damage. This may be because a particular gene is not G-rich and is thus not predisposed to forming an R-loop. It may be transcribed in isolation (i.e. no other transcription in close proximity), which avoids collision and/or stalling of the transcriptional machinery. Alternatively it may be due to mechanisms not discussed here, such as a poor recognition of the damage resulting in a lack of γH2AX signal or an immediate repair.

A significant question remains unanswered: how is the DNA damage detected in these experiments finally resolved, and will genomic alterations arising from this DNA damage contribute to the eventual stable transformation of the oncogene-expressing cells and hence be detectable in established monoclonal populations established in long term cultures of these cells?

Aim 2: Does stable transformation of B cell precursors depend on the adaptation of epigenetic changes?

Despite the limitations of the restricted number of biological repeats, H3K27Ac ChIP-seq did identify peaks that were highly significantly different between the day 7 conditions and the day 60+ conditions, and which were associated with a concomitant significant change in expression. As shown above, one gene featured twice: *Ctla4*. Initial studies found that soluble forms of cytotoxic T-lymphocyte antigen 4 (*Ctla4*; also known as CD152) were costimulatory with CD28 and thus resulted in T-cell activation (Linsley et al., 1992), but it is now understood that *Ctla4* is a *negative* regulator of T-cell activation (Krummel and Allison, 1995; Walunas et al., 1994). *Ctla4* is overexpressed in various cancers and leukaemias, such as chronic lymphocytic leukaemia (Motta et al., 2005), and it is constitutively expressed in about 80% of AML (Laurent et al., 2007) adding to the body of evidence which suggests that its activation may contribute to escape of tumour immune surveillance (Laurent et al., 2007). However, it is not well studied in ALL - this may be because, unlike most other haematological malignancies, flow cytometric analysis of expression of *Ctla4* in adult B-ALL showed that it was mainly restricted to the cytoplasm (Pistillo et al., 2003). *Ctla4*-blocking antibodies have been investigated as potential therapeutic strategies with some success (reviewed in (Grosso and Jure-Kunkel, 2013); engagement of the *Ctla4* receptor induced apoptosis in AML (Laurent et al., 2007). In addition, blocking of *Ctla4* was proposed as a strategy to increase graft-versus-leukaemia effect without increasing the graft-versus-host-disease; Ipilimumab (a human monoclonal antibody that block *Ctla4* binding) has been tested in a Phase I trial of various recurrent or progressive haematological malignancies which showed increased levels of T-cell activation but no corresponding increase in regulatory T-cells (Bashey et al., 2009; Zhou et al., 2011). Despite this interest, the mechanism causing constitutive activation of *Ctla4* has not been studied; the detection of novel upstream and downstream peaks in the day 60+ stably transformed cells is the first suggestion that this may be an epigenetic phenomenon. It is interesting to note that this overexpression arose *in vitro*, in the absence of T-cells. Hence, the role of *Ctla4* may not be restricted to just T-cell activation.

Not all the identified enhancers were associated with biomarkers of aggressive disease. As discussed above, stable transformation resulted in increased expression of *Galnt14* at *BCR-ABL1* induced oncogenic crisis at day 7 and at stable transformation at day 60+. The tumour necrosis factor related apoptosis inducing ligand (TRAIL), which can induce apoptosis of cancer cells, has been intensely studied recently and biomarkers which predict sensitivity to TRAIL-based therapies are highly prized. *Galnt14* is a peptidyl-O-gylcosyltransferase: these

enzymes have diverse roles including the regulation of trafficking and turnover, multimerisation and conformation of various cell surface proteins (Wagner et al., 2007). *Galnt14* expression correlates with sensitivity to TRAIL therapy; assays to measure its expression are already deployed in clinical trials of proapoptotic receptor antagonists against solid tumours such as non-small cell lung cancer and colorectal cancer (Stern et al., 2010). Up to 30% of human malignancies overexpress *Galnt14* and its overexpression can induce sensitivity to TRAIL-therapy (Wagner et al., 2007); it has not been well studied in leukaemia. Its appearance in the mouse model during oncogenic crisis at day 7 and persistence during stable transformation at day 60+ suggests its expression, perhaps in its role as regulator of cell surface markers, is required but may also indicate a vulnerability to proapoptotic therapy.

Aim 3: Are epigenetic changes responsible for the expression of *MECOM* in poor prognosis myeloid leukaemia?

The analysis of primary samples from CML patients in the search for genomic drivers of *EVI1* expression has raised several interesting findings so far. First, the single CML-BC sample tested showed a much higher expression of the *EVI1-3L* transcript than the two CML cell lines. The expression of the alternative first exons of *EVI1* is not well studied or understood, with a very limited number of publications on the subject (Aytekin et al., 2005; Lugthart et al., 2008), and some of these are restricted to *EVI1* expression in AML only. Unfortunately, most publications quantify total *EVI1* expression (with primers most commonly located in the common exon 14) and fail to make a distinction even from *MECOM*, which, according to some reports may have an antagonistic role (Maicas et al., 2017). Other reports state that this is not the case, and that expression of both *EVI1* and *MECOM* is associated with a short remission and overall poor prognosis (Haas et al., 2008). An expansion of the transcript profiling to more primary samples would therefore be useful to understand these relationships.

Second, the initial quality control steps of the H3K27Ac unexpectedly revealed what appeared to be a complete absence of H3K27Ac in all primary samples. Subsequent ChIP-seq proved that H3K27Ac was present, but at levels undetectable by immunohistochemical staining or Western blot. A lack of histone H3 and H4 acetylation has been previously reported in CD34+ fractions (Zhang et al., 2010) and results so far indicate that histone deacetylation inhibitor therapy for leukaemia (Fiskus et al., 2006; George et al., 2005; Nimmanapalli et al., 2003; Ustun et al., 2008), alone or in combination with a TKI, and for cancer more generally (Johnstone, 2002) have much clinical potential. Histone deacetylases

are frequently found to be overexpressed in cancer, leading to a global histone hypoacetylation; this is often associated with compact chromatin and aberrantly silenced genes (Ellis et al., 2009).

Finally, the H3K27Ac ChIP-seq showed that the CML-BC patient sample with *EVI1*^{high} expression had a remarkably different H3K27Ac profile throughout the *EVI1* gene body, which would indicate that the whole of this region would be open, loose chromatin that would be readily accessible to transcriptional machinery. Whilst global changes induced by the effect of the *EVI1* protein hindered the identification of candidate enhancer peaks, the hyperacetylation of the entire *EVI1* gene body in the *EVI1*^{high} patient suggested that further investigation of this area for chromatin interactions using chromosome conformation capture sequencing (4C) was worthwhile.

Despite these limitations significant interactions, and in particular, interactions which differ between the *EVI1*^{high} and *EVI1*^{low} patient samples have already been identified using two different computational methods. The proximal region coincides with the MECOM promoter. A more distal region coincides with the region in and around two upstream genes, *LRRC34* and *MYNN*. Both of these interactions will need to be confirmed in further samples and functionally verified.

Despite our growing understanding of the aetiology of *EVI1* overexpression and the poor prognosis it confers, remarkably few advances in treatment have been made. One such advance is the development of pyrrole-imidazole polyamides (PIPs) which can be designed to target any desired DNA sequence, and was used with some success to target the REL/ELK1 binding site in the *EVI1* promoter region (Syed et al., 2014). If a novel oncogenic enhancer inducing aberrant expression of *EVI1* in leukaemia were identified, it would represent a novel therapeutic target. The reversal of the universally poor prognosis that is conferred by *EVI1* expression would offer a dramatic improvement to outcome.

Limitations

At the time of publishing in *Cell Reports*, the H3K27Ac ChIP-seq analysis supporting chapter 2 was based on just one day 7 biological repeat for each oncogene. Similarly, the ChIP-seq analysis supporting chapter 3 was based on just two day 60 biological repeats for the *BCR-ABL1* transformed populations at the time of writing. It is acknowledged that a minimum of three repeats would increase the significance of these results, and as a consequence, the number of biological repeats for these experiments has subsequently been increased to n=3

or more. In addition, a single input library was used for comparison to multiple ChIP-seq libraries and this may not accurately reflect the genome accessibility across all samples.

Differential peak calling requires a minimum of two datasets forcing all D7 datasets (oncogene and empty vector day 7 conditions) to be merged for the comparison to the day 60+ stably transformed cells. However, the described unexpected similarity between the 'normal B-cell precursors' (empty vector day 7 cells) and either of the oncogene expressing, stably transformed populations at day 60+ as well as the broad difference between oncogene versus empty vector control cells in crisis stage (day 7) meant that this combination has reduced the potential to detect significant interactions specific for D60+ cells. It further limited the analysis of H3K27ac peaks specific for day 7 oncogenic crisis populations. Further ChIP-Seq repeat experiments of all conditions to achieve a minimum of n=3 have already been performed, though sequencing could not be completed before submission of this thesis.

For chapter 4, the most obvious limitation of the current 4C analysis is the limited number of samples – summarised in Table 4-2. The number of biological repeats was dependent on the receipt of suitable CML patient samples which is obviously outside the control of the laboratory. The 4C experiment investigated four CML-CP patients (three *EVI1*^{low} and one *EVI1*$^{-}$*)* but crucially only two CML-BC samples (one *EVI1*^{high} and one *EVI1*$^{-}$; all that were available when the experiment was started). Two *EVI1*^{high} cell lines were also included in both experiments but extrapolation from cell lines to primary samples is not always straightforward (Goodspeed et al., 2016). However, Imperial College Healthcare Trust is a specialist referral centre for CML patients and further suitable samples have already been received; the ethical approval for the *MECOM* BioBank has been extended and further analysis can be undertaken in the future.

In addition, *53BP1*$^{-/-}$ mice were used for the experiments, as described by others, in order to increase the sensitivity of DNA damage detection. As described, TP53-binding protein 1, *53BP1,* is an important regulator of the DNA double strand break signalling. It forms nuclear foci in response to the detection of DNA double strand breaks (Bothmer et al., 2011), where it plays a key role in checkpoint signalling, the recruitment of further double strand break response proteins, the choice of repair pathway, and also the synapsis of the broken ends in the event of NHEJ (reviewed in (Panier and Boulton, 2014)). Mice deficient in *53BP1* are slow to resolve the γH2AX mark of DNA damage (Chapman et al., 2013); the use of *53BP1* knockout in the experiments described above was therefore intended to maximise the detection of the γH2AX -marked chromatin. However, it is not clear if the *53BP1* deficient

mice simply have persistent γH2AX marks, or if thie marks spread, or if the marks may be aberrantly placed; rather than boost the detectable signal, it is possible that the use of *53BP1* has a confounding effect. It would therefore be necessary to undertake a pilot experiment using wild type *53BP1* mice to investigate this possibility by comparing γH2AX signal in *53BP1* wildtype and knockout mice.

Mice of both sexes were used for the experiments and thus it was necessary to exclude the genes on the sex chromosomes from bioinformatic analysis. However, a significant sex bias is present in human leukaemia so it is possible that the use of mixed populations has overlooked some relevant abnormalities. The use of single sex populations would be beneficial for future experiments.

Importantly, however, the genes identified in this analysis have all relied upon the presumption that an enhancer is responsible for the over expression or down regulation of the closest gene. While it is shown that gene expression of the most proximal genes is increased for the identified D60 specific enhancers, the identified enhancers might further regulate far more distant genes as enhancers can frequently exert their effect over long distances, as exemplified by the *MYC* super enhancer: the cluster of peaks was detected as highly significantly differential between the day 7 empty vector and oncogenic crisis versus the day 60+ stably transformed populations but this difference was incorrectly attributed to the most proximal gene. It is therefore important to further characterise enhancer-promoter interactions for the enhancers of interest. This could be achieved by performing for example chromatin interaction analysis by paired-end tag sequencing (ChIA-PET; (Fullwood et al., 2009)) or chromatin conformation capture sequencing (3C; (Dekker et al., 2002)) and variants.

Another interesting approach will be to combine epigenetic, genetic and phenotypic analyses to further characterise enhancers by computational and experimental methods. An elegant illustration of this point was published recently: it was revealed that a SNP within the third intron of the *PHACTR1* gene at 6p24 - which had long been significantly associated with five different vascular diseases including coronary arterial disease (Anttila et al., 2013; Deloukas et al., 2013; O'Donnell et al., 2011) - was actually within an intergenic regulator of the endothelin 1 gene (*EDN1*) located 600Kb upstream from *PHACTR1*. Deletion of the regulatory region resulted in increased expression of *EDN1* (Gupta et al., 2017).

Directions for future research

As described above, the lack of a biological repeat is a major limitation of the current

differential H3K27Ac peak analysis. This experiment has been subsequently repeated and will be sequenced shortly allowing a more informative three-way analysis of day 7 empty vector versus day 7 oncogenic crisis versus day 60+ stably transformed populations to be undertaken. This will allow identification of altered H3K27Ac peaks in acute *BCR-ABL1*-expressing oncogenic crisis, and separately, those which are altered to result in stable *BCR-ABL1*-expressing transformation. The definitive list of the most promising putative enhancers identified in the mouse model will then be confirmed to be relevant in Ph+ B-ALL human cell lines. ChIP-seq of various Ph+ve and Ph-ve B-cell lines has already been prepared for this next step.

ChIP-seq provides detailed information about binding sites but no information about specific interactions between the binding sites and other regulatory sequences (Kuo and Allis, 1999) and can have high background noise (related to, amongst other things, the efficiency of the antibody and the degree of cross-linking). The chromosome confirmation capture (3C; and variant) approaches can detect long range interactions but can also have high background noise making the analysis more complicated.

Chromatin interaction analysis using paired end tag (ChIA-PET) sequencing was developed by the Ruan group (Zhang et al., 2012) and was used initially to uncover the interaction network of the oestrogen receptor alpha transcription factor (Fullwood et al., 2009). ChIA-PET overcomes several of the above disadvantages with an additional step; the ChIA-PET reaction is split in half and each is blunt end-ligated to a different half linker (AT and CG) before being recombined for the remaining steps. Subsequently, the ATCG complexes – which obviously represent the artefactual interactions formed by ligation of DNA from two different complexes - can be discarded. This leaves ATAT and CGCG reads which are either self-ligated (i.e. ligated back to their original conformation in the correct orientation) or interligated. Interligated PETs – which may derive from the same chromosome or a different chromosome – can be mapped back and will reveal interactions over a very long range, over 10Mb and to different chromosomes (Zhang et al., 2012). The same group also released the ChIA-PET tool for visualising the interactions (Li et al., 2010), and most recently, a long-read variant. This has the significant advantage of 250bp reads which allow for more accurate mapping; this increases the chances of reading through a phased SNP and could thus allow detection of a haplotype specific interaction (Li et al., 2017). In the event that expression is caused a cis-method – a base pair mutation inducing an aberrant transcription factor binding site or focal amplification of a pre-existing enhancer, for example – acquiring haplotype specific information would be extremely useful.

Once confirmed, the enhancer-promoter interactions must be functionally validated. In 2012, the field of genetic editing was revolutionised by Charpentier and Doudna's discovery of clustered regularly interspaced short palindromic repeats, or CRISPR, a family of bacterial DNA sequences necessary for adaptive immunity from viral infection (Jinek et al., 2012). In 2013, a variant CRISPR technology was developed which used a catalytically inactive cas9 protein, which thus sterically blocked the sequence targeted by the guide RNA to prevent transcription (Larson et al., 2013). The catalytically inactive ('dead Cas9') had two point mutations introduced which ablated its nucleolytic activity *in vitro* (Jinek et al., 2012). The group used this technique – which they called CRISPR interference or CRISPRi - to knock down various genes in *E. coli* with up to 1000x fold reduction in expression, and which was both inducible and reversible (Larson et al., 2013). CRISPRi is already used by colleagues in Feldhahn group for other projects and will therefore be used to silence the putative enhancers to confirm the reduction in expression.

As discussed, the small number of primary CML samples tested with ChIP-seq and 4C is the main limitation of the search for genomic drivers of *EVI1* expression; a priority for this section is therefore the expansion to further primary CML samples.

Prior to the initiation of this project and the systematic bio banking of cross-linked cell pellets from fresh CML patient samples, multiple CML-CP and CML-BC patient samples were cryopreserved in liquid nitrogen. It may be possible to significantly expand the number of samples tested by ChIP-seq and/or 4C by thawing these samples and generating cross-linked pellets as before. However, both the freezing and the thawing of cryopreserved cells are acutely stressful and a proportion of cells will not survive the procedure (Berz et al., 2007). It will be necessary to perform a pilot experiment using crosslinked pellets from a fresh sample and a corresponding cryopreserved sample. Correlation of the ChIP-seq marks will be necessary to exclude the possibility that cryopreservation and/or thawing of cells induces histone modifications which confound the *EVI1* analysis. Dead cell depletion using magnetic activated cell sorting following thawing and before crosslinking could also be piloted if a confounding effect was observed.

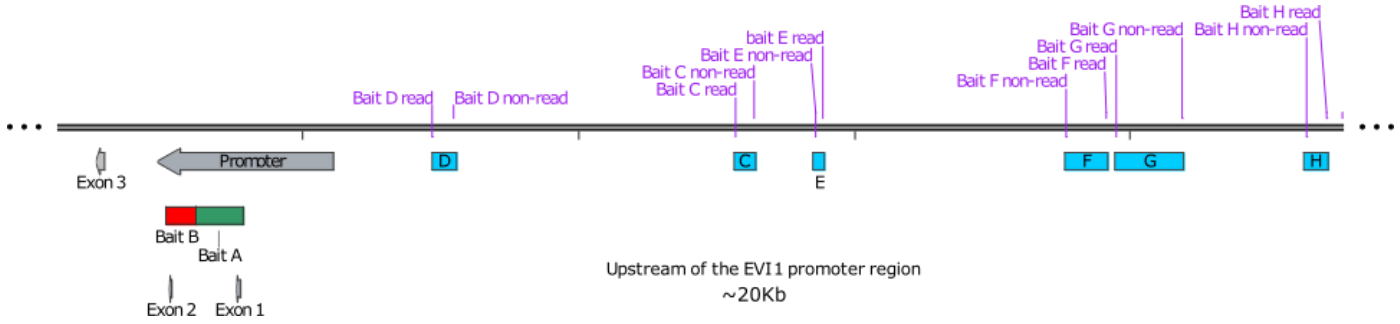


Figure 5-1: Map showing the design of six further baits¶up to 20kb upstream of the promoter region in which the existing baits A and B sit

In addition to expanding the number of patient samples analysed, the number of baits can also be increased. Six further baits in the region up to 20Kb upstream from the promoter have been designed, as shown in Figure 5-1. The primer pairs for the six new baits have been tested on the 3C library generated on CML-CP patient MEC101 (shown in Figure 5-2) and all have shown potential interactions worthy of investigation.

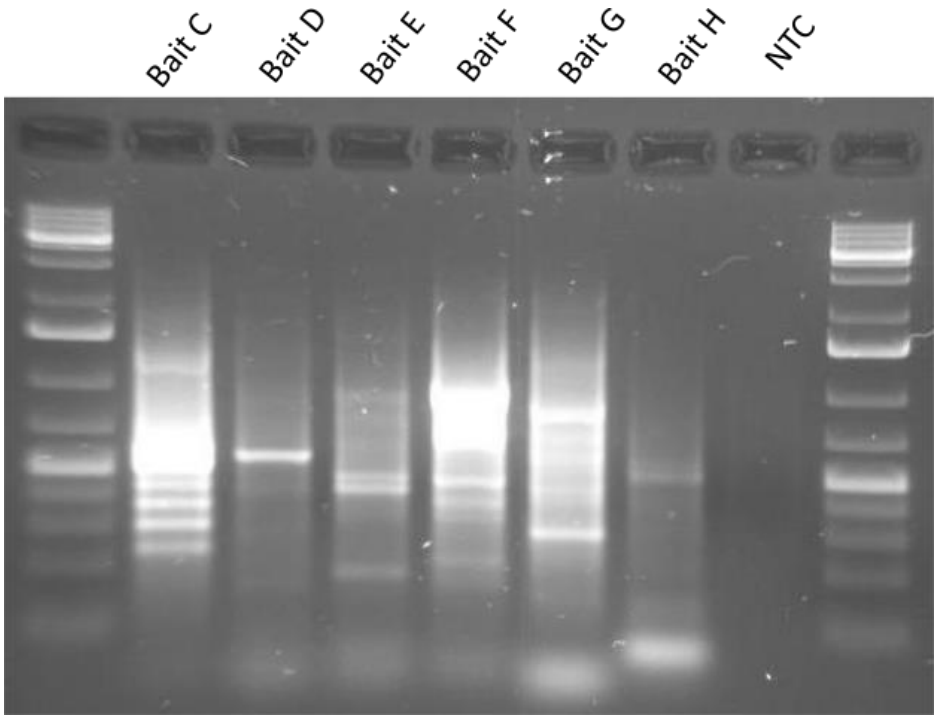


Figure 5-2: Six further baits designed to amplify the existing 3C library¶. The gel shows a test amplification of a CML-CP patient sample (MEC101) and shows additional bands and smearing indicative of potential interactions with the *EVI1* promoter region.¶These baits are all in the ~20Kb upstream of the promoter.

Likewise, alternative combinations of primary and secondary restriction enzyme can be used to make the 4C libraries – the use of a six cutter followed by a four cutter is recommended as

a starting point since this combination yields the best reproducibility. Use of a primary four cutter results in many more restriction fragments over the same genomic distance, which can lead to an interaction being spread over several restriction fragments, and thus make interpretation more difficult because of background noise (Stadhouders et al., 2013). However, the use of a four cutter primary restriction enzyme is recommended when a potential interaction is being refined.

As discussed for the potential enhancers identified in chapter 1, if candidate *EVI1* enhancers can be confidently identified CRISPRi could be used to functionally validate the interaction in cell line and/or cryopreserved primary cells.

Finally, it would be relevant to expand the search to AML patient samples that are already being biobanked; enhancers are frequently lineage specific (as exemplified by the many different lineage specific *MYC* enhancers discussed in section **Error! Reference source not found.**) so it would be very interesting to see if interactions occurring in chronic myeloid conditions also occurred in acute myeloid conditions.

The investigations above have focused on *cis* interactions but it is of course likely that at least some overexpression arises from *trans* interactions. Returning to the *TAL1* overexpression in T-ALL, in 42% of cases with *TAL1* overexpression the expression arises biallelically (Ferrando et al., 2004). This would suggest that the interaction is *trans* not *cis* which would require an entirely different approach such as chromatin proteomics (ChroP) which uses mass spectrometry to establish what proteins are bound to a region of interest (Soldi and Bonaldi, 2014).

6 Materials and Methods

6.1 Primary tissue collection and processing

6.1.1 Sample collection

Human primary samples

Leukophoresis, peripheral blood and/or bone marrow that was surplus to requirements following routine diagnostic tests or monitoring was used for all experiments requiring primary human cells.

Human cell lines

All myeloid cell lines (K562, KCL22 and LAMA84) were a kind gift from Dr Jamshid Sorouri Khorashad. All lymphoid cell lines (Kasumi-2, SMS-SB, BV173, SEM, TOM1 and SD1) were obtained from our collaborator Markus Müschen and initially purchased from the German Collection of Microorganisms and Cell Cultures GmbH; identity of cell lines was confirmed by karyotyping and/or FISH when revived, and multiple back up vials were cryopreserved once identity was confirmed. Cell lines were maintained for the minimum time required for extraction of sufficient cells for RNA, gDNA, FCS and cross-linked cell pellets and when cultured for longer than 3 months (when needed by other members of the Feldhahn Group for experiments outside the remit of this thesis) were subject to six monthly karyotyping/FISH testing (examples shown in Figure 6-1 and Figure 6-2) and six monthly Mycoplasma testing (see Figure 6-3) using the LookOut Mycoplasma PCR Detection kit (Sigma Aldrich) according to manufacturer's instructions.

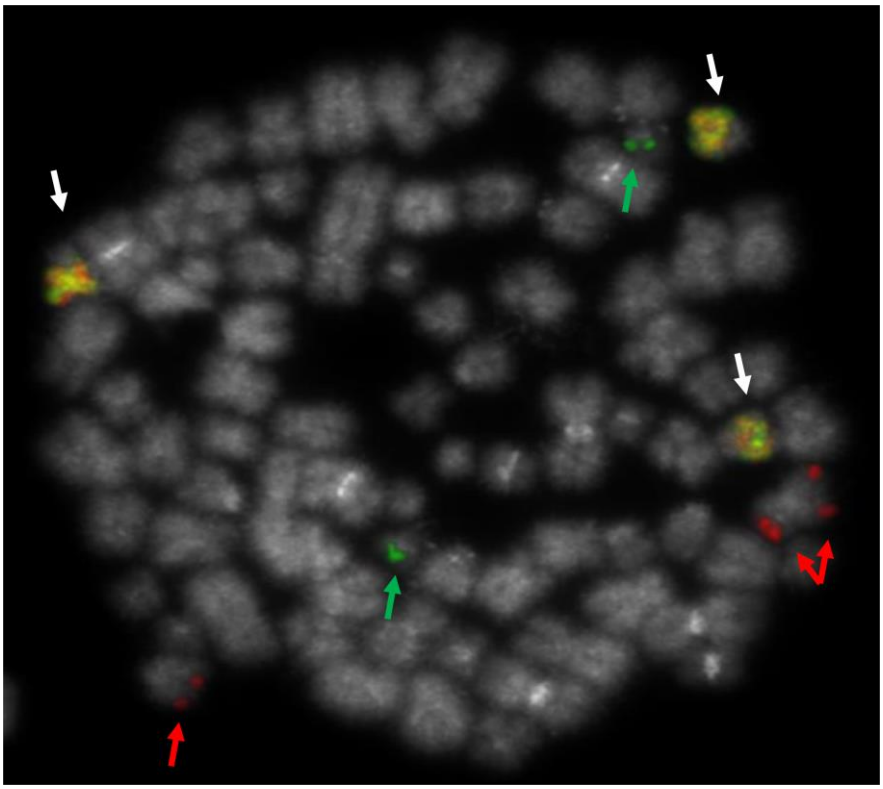


Figure 6-1: Cell line identity was confirmed by FISH where possible. ¶ Metaphase FISH analysis showing BCR-ABL1 fusion signal amplification (indicated with a white arrow) which is characteristic of the K562 cell line (Gribble et al., 2000). Red and green arrows indicate the non-translocated copies of ABL1 and BCR, respectively.

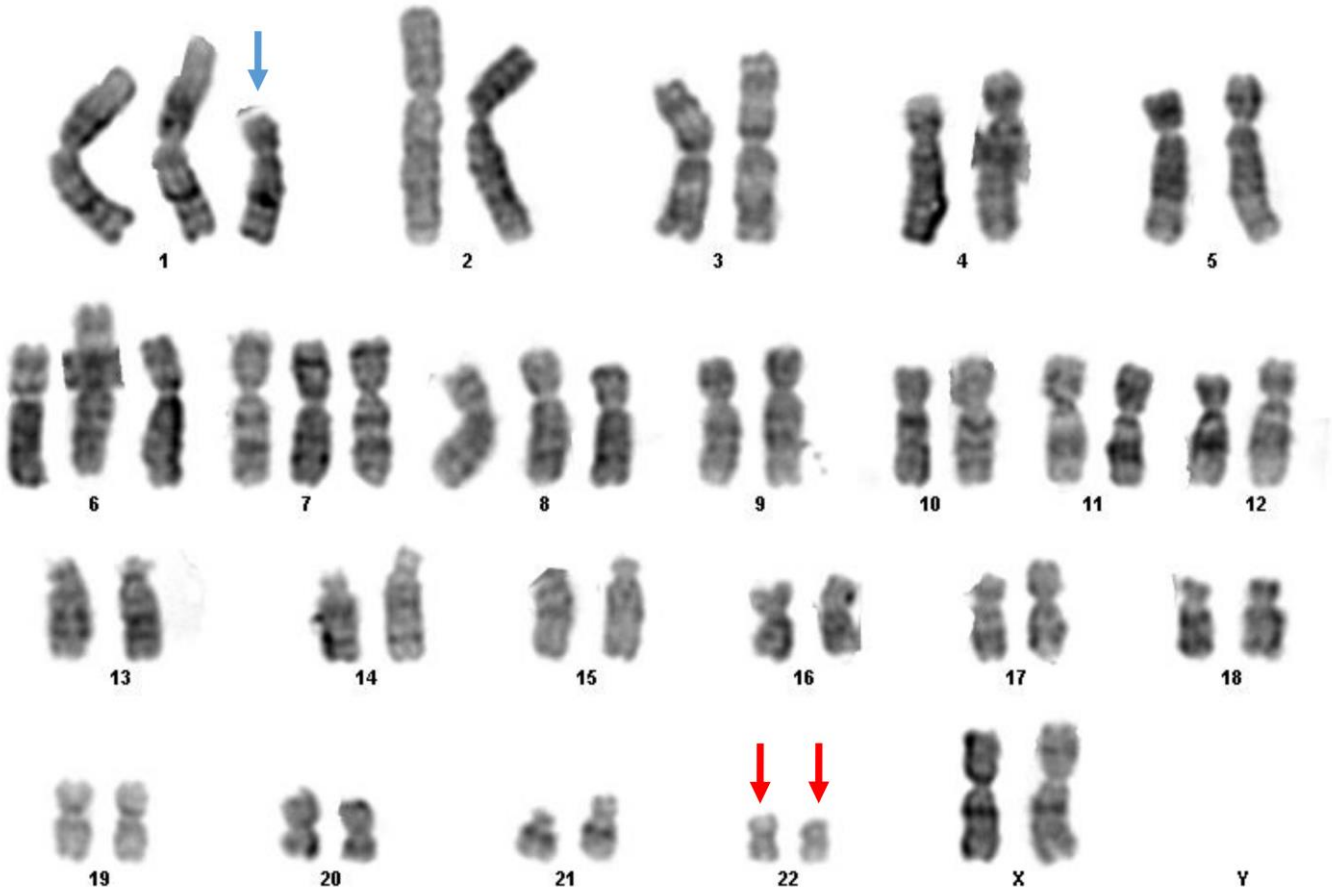


Figure 6-2: Cell line identity was confirmed by G-banded chromosome analysis as required. ¶Karyotype of the KCL-22 cell line showing two copies of the Ph chromosome (indicated by red arrows) and an additional copy of 1q (indicated with a blue arrow) as reported by (Kubonishi and Miyoshi, 1983).

Karyotype of the KCL-22 cell line showing two copies of the Ph chromosome (indicated by red arrows) and an additional copy of 1q (indicated with a blue arrow) as reported by (Kubonishi and Miyoshi, 1983).

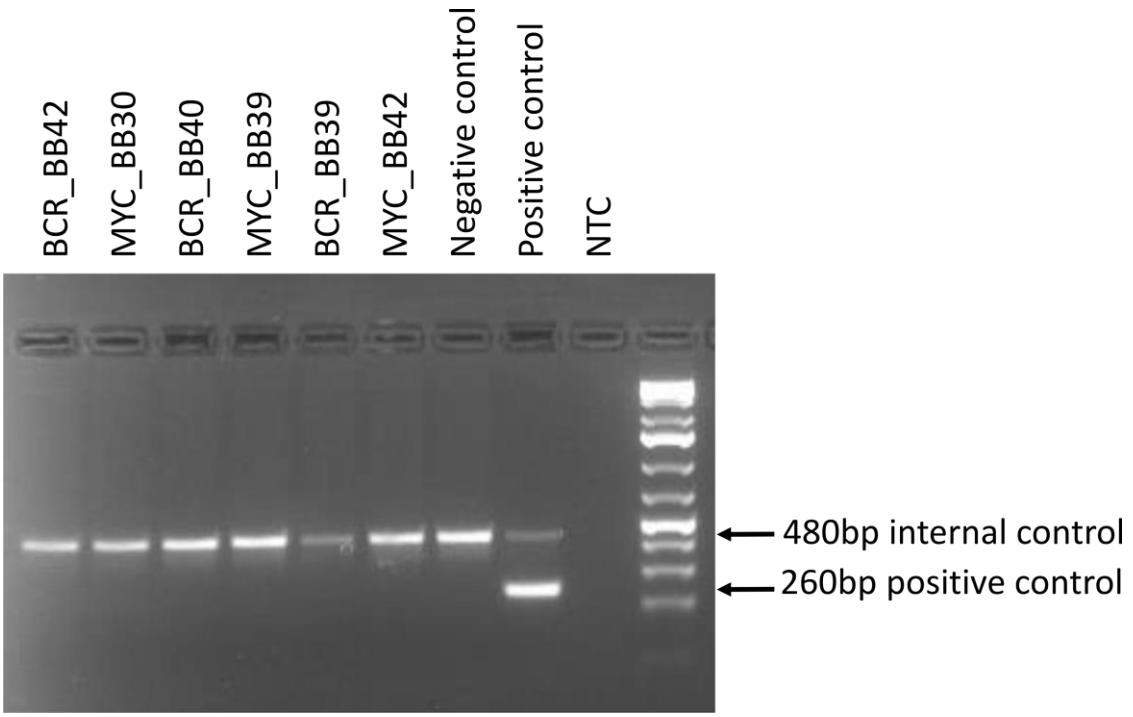


Figure 6-3: All cell lines used in this thesis were Mycoplasma negative throughout. ¶Example 2% agarose gel showing amplicons from the supernatant of murine cell lines tested using the LookOut Mycoplasma PCR Detection kit (Sigma Aldrich); all cell lines showed amplification of the internal control DNA (~480bp, confirming the PCR reaction was not inhibited) but no amplification of Mycoplasma 16S rRNA coding region (~260bp, only present in the positive control reaction). The day 60+ cell lines shown were established from the day 7 oncogenic crisis populations in Chapter 2 and used for RNA-seq and ChIP-seq in Chapter 1.

Murine primary samples

For the majority of experiments, *53BP1-/-* mice (Ward et al., 2003) were used. For the functional experiments related to forced CEBPA expression in BCR-ABL1-transformed B-cell precursors wild type mice were used. *53BP1-/-* mice were provided by Dr. Simon Boulton (The Francis Crick Institute, London, UK) and are those available from The Jackson Laboratory. All experiments were performed in agreement with Animals (Scientific Procedures) Act 1986 guidelines and regulations and protocols approved by Home Office (UK).

6.1.2 Ethics and consent

Human samples used in this research project were deposited into, stored in and subsequently retrieved from the Imperial College Healthcare Tissue Bank (ICHTB). ICHTB is supported by the National Institute for Health Research Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. ICHTB wass

approved by NRES to release human material for research in 2012 (12/WA/0196) and the approval was renewed in 2017 (17/WA/0161) – see Appendix 2 for the REC approval letters. The samples for this project were collected prospectively from surplus material following routine diagnostic investigations. A sub collection dedicated to this project was initiated (reference number MEC_AR_16_030) and access to this sub collection for this project was granted in October 2016 and renewed in November 2017 (R16065). All patients gave specific informed consent for the use of surplus tissue. All samples were linked-anonymised on receipt and given a study number.

6.1.3 Sample processing

Whole blood, bone marrow or leukophoresis was diluted 1:1 in PBS (Sigma Aldrich) and layered over Lymphoprep (Stemcell Technologies) for separation by density gradient using centrifugation according to the manufacturer's protocol. The mononuclear cells were removed and washed twice in PBS. Cells were then processed for ChIP-seq, cryopreservation and RNA extraction immediately; cell pellets were frozen at -80 for gDNA and/or protein lysate at a later date if required.

6.1.4 Cryopreservation of cells

Where necessary, cells were cryopreserved in freezing medium (90% fetal bovine serum (FBS)) supplemented with 10% dimethyl sulfoxide (DMSO; Sigma-Aldrich, Dorset, UK) at a maximum concentration of 10^7 cells/ml of freezing medium. Immediately after suspension in freezing medium, cells were gradually frozen to -80°C using an isopropanol Nalgene Cell Freezing Container (Sigma-Aldrich, Dorset, UK) for up to three days before transfer to liquid nitrogen for long term storage. When required, cryopreserved cells were incubated for 1 minute in a 37°C water bath then resuspended and washed once in warmed medium.

6.2 Cell culture

All myeloid cells were cultured in RPMI 1640 supplemented with 1% penicillin/streptomycin, 1% L-glutamine and 10% FBS (all from Sigma Aldrich). Cells were incubated at 37°C in 5% CO_2 and maintained at approximately between 0.5-3 x10^6/ml.

All lymphoid cells were cultured in Iscove's Modified Dulbecco's Medium supplemented with 1% penicillin/streptomycin, 1% L-glutamine and 20% FBS (all from Sigma Aldrich). Primary murine cells were also supplemented with 10ng/ml of murine Il7 (Peprotech). Cells were incubated at 37°C in 5% CO_2 and maintained between 0.5-3 x10^6/ml.

All BOSC cells were cultured in Dulbecco's Modified Eagle Medium supplemented with 1% penicillin/streptomycin, 1% L-glutamine and 10% FBS (all from Sigma Aldrich).

6.3 Flow cytometry

All flow cytometry was undertaken on the FACSCalibur or the LSRFortessa (BD Biosciences). Survival was indicated by forward scatter/side scatter and expression of green fluorescence protein (GFP) or DsRed. The Annexin V-PE apoptosis detection kit and 7-AAd (both BD Biosciences) was used for detailed analysis of viability and apoptosis induction. For cell surface staining of cell suspensions see Table 6-1 for the antibodies used.

Antibodies	Clone	Conjugate	Reference	Source
CD45R (B220)	RA3-6B2	PerCPCy5.5	#45-0452-82	eBioscience
CD19	eBio1D3	APC	#17-0193-82	eBioscience
IgM	II/41	APC	# 17-5790-82	eBioscience
CD43	eBioR2/60	PE	#12-0431-81	eBioscience
CD24	30-F1	FITC	#11-0241-82	eBioscience
CD11b	M1/70	PE-Cyanine7	#25-0112-81	eBioscience

Table 6-1: Antibodies used for flow cytometry

6.4 Western blot

For ChIP experiments, the whole 10µl aliquot (representing 1% of the pulled down protein) was used for Western blot. Cell pellets were lysed in 1X NuPAGE LDS sample buffer (ThermoFisher Scientific) and stored at -80˚C until required.

6.4.1 Protein quantification

Cell pellets were re-suspended in cold cell lysis buffer (see section 6.12.1) and mixed by pipetting. Following incubation on ice for 20 minutes, the samples were centrifuged at 4˚C for 15 minutes at 12,000g.

A dilution series of 0, 1, 2, 4, 6, 8, or 10µg of bovine serum albumin (Sigma Aldrich) with 5µl of cell lysis buffer or 5µl of the test cell lysate was added to 795µl distilled water and 200µl 5X Bio-Rad Protein Assay Dye Reagent Concentrate and transferred to a cuvette. After 5 minutes incubation at RT whilst protected from light the optical density of each sample was measured using a spectrometer (Eppendorf). Protein concentration was calculated from the dilution series, and 10-20µg protein was used per lane with 1X NuPAGE LDS sample buffer (ThermoFisher Scientific).

Samples were heated at 70˚C for 10 minutes then cooled on ice.

6.4.2 Protein separation

Proteins were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) using precast Tris-Glycine eXtended (TGX) stain-free gels were and the Mini-PROTEAN blotting module (all from Bio-Rad). The gels were unwrapped and the plastic

comb was removed. The gel and the wells were rinsed with 1X running buffer (see section 6.12.3). The gels were inserted to the gel module and the internal and external chambers were filled with 1X running buffer.

Protein samples were loaded and the gels were run at 200V until the loading buffer front reached the bottom of the gel (approximately 40 minutes).

6.4.3 Protein transfer

PVDF membranes were cut to size and activated for 15 seconds in methanol, two minutes in distilled water and equilibrated for 5 minutes in 1X transfer buffer (see section 6.12.4). The gel cases were opened, the gel foot was trimmed away and the gel was transferred to 1X transfer buffer for 5 minutes. Sponges and Whatman papers cut to size were also pre-soaked. The blotting sandwich (top to bottom) was set up: sponge, Whatman paper, PVDF membrane, protein gel, Whatman paper, sponge. Air bubbles were removed by gently rolling the sandwich with a clean pipette.

The sandwich was transferred to the blotting module, surrounded by ice, and blotted for 1.5 hours at 120V constant.

6.4.4 Antibody staining

Protein membranes were cut, if necessary, and slices were placed into plastic sheets. The membranes were incubated with the appropriate volume (10ml for a full membrane) of the primary antibody (listed in Table 6-2diluted 1:1000 in blocking solution (see section 6.12.7) and the plastic sheets were heat sealed. Membranes were incubated at 4˚C overnight on a rotary shaker.

Antibody	Supplier	Reference
Anti-phospho-histone H2A.X (serine 139) mouse	Millipore	05-636
Anti-phospho-P53 (Ser15) 16G8 mouse	Cell Signalling Technology	#9286S
Anti- Lamin B1 (D4Q4Z) Rabbit	Cell Signalling Technology	#12586
AID (l7E7) Mouse	Cell Signalling Technology	#4975S
Anti-H3 (acetyl K27) rabbit	Abcam	Ab4729
Anti-histone H3 rabbit	Abcam	Ab1791

Table 6-2: Antibodies used for ChIP experiments

The membranes were removed from the sealed plastic and washed for a minimum of 4 x 5 minutes in fresh PBS-T. The membranes were then re-sealed with a 1:20,000 dilution of the Horse radish peroxidase (HRP)-conjugated anti-mouse IgG (#115-035-062; from Jackson Immuno Research) or anti-rabbit IgG (HAF008; from RD&D Systems) within fresh plastic sheets and incubated at RT on a rotary shaker for 1 hour. The membrane were again removed from the sealed plastic and washed for a minimum of 4x 5 minutes in PBS-T.

6.4.5 Visualisation

HRP activity was detected chemiluminescently using the Pierce ECL Western Blotting Substrate (ThermoFisher Scientific): equal volumes of component A and B were combined and sealed with the membrane inside fresh plastic sheets. Membranes were incubated for 1 minute then detected using the ChemiDoc Imaging System (Bio-Rad).

6.5 Gene expression analysis

6.5.1 RNA extraction

RNA was extracted from $5x10^6$ cells per column using the RNeasy mini kit (Qiagen) as per manufacturer's protocol. 'RLT buffer' was supplemented with 1% β-mercaptoethanol (Sigma Aldrich) and QIAshredder (Qiagen) columns were used for cell homogenisation. The additional RNAse-free DNase (Qiagen) step to remove residual DNA contamination was included for RNA extraction destined for sequencing only. Spectrophotometric quantification of RNA yield in a 1µl aliquot was undertaken using the Nanodrop (Thermo Scientific) according to the manufacturer's protocol and RNA was stored at -80°C until required.

6.5.2 Complimentary DNA (cDNA) synthesis

Complimentary DNA was synthesised using the RevertAid First Strand cDNA synthesis kit (Thermo Scientific) according to manufacturer's protocol and stored at -20°C or -80°C until required.

6.5.3 Genomic DNA extraction

Extraction of genomic DNA was undertaken using the GeneJET genomic DNA purification kit (Thermo Scientific) according to manufacturer's protocol. Spectrophotometric quantification of DNA yield in a 1µl aliquot was undertaken using the Nanodrop (Thermo Scientific) according to the manufacturer's protocol and DNA was stored at -20°C or -80°C until required.

6.5.4 Primer design

All primers were designed using Primer 3 (Koressaar and Remm, 2007; Untergasser et al., 2012) or, where available, taken from Primer Bank (Spandidos et al., 2008; Spandidos et al.,

2010). All primers were from Sigma Aldrich and sequences are shown in Table 6-3. A stock solution in TE buffer (100μM) was stored at -20˚C; working solution in TE buffer (20μM) was made as required.

6.5.5 Endpoint PCR

Primer name	Description	Sequence 5' to 3'	Related section
NF-p84	JH4	AGAATGGCCTCTCCAGGTCT	2.4.2
NF-p85	JH4	TGCAATGTTCAGAAAACTCCATA	
NF-p86	Eμ	ACCTGGGAATGTATGGTTGTGGCT	
NF-p87	Eμ	TTACCATTTGCGGTGCCTGGTTTC	
NF-p88	CD74	TGACCAACGCGACCTCATCTCTAA	
NF-p89	CD74	AAGGGCTCTCTGCTGGTATTCACA	

Table 6-3: Primers used for endpoint PCR

6.5.6 Real-Time quantitative Polymerase Chain Reaction

All RT-qPCR was undertaken in triplicate (chapter ☐) or duplicate (chapter 4) using SYBR Jumpstart mastermix (Sigma Aldrich) according to the manufacturer's protocols. All assays were quantified using the StepOne Plus thermocycler (Applied Biosystems) at the concentrations shown in Table 6-4 in standard mode (ramp rate 60%) with the following steps: initial hotstart (2 minutes at 94°C), followed by 40 cycles of 15 seconds denaturation at 94°C and 1 minute annealing/extension at 60°C.

	Stock	Final	Volume
SYBR jumpstart mastermix	2x	1x	12.5
Forward primer	20uM	400nM	0.5
Reverse primer	20uM	400nM	0.5
ROX	100X	1x	0.25
Nuclease-free water			9.25
cDNA template			2
		TOTAL	25

Table 6-4: Reagent concentrations for RT-qPCR

All primers (see Table 6-5 and Table 6-6**Error! Reference source not found.**) were validated to ensure that they primed efficiently before use.

Primer name	Description	Sequence 5' to 3'	Reference
NF-712	Murine actin beta F	GGCTGTATTCCCCTCCATCG	N/A
NF-713	Murine actin beta R	CCAGTTGGTAACAATGCCATGT	
NF-491	Murine MYC F	GCCGATCAGCTGGAGATGA	
NF-492	Murine MYC R	GTCGTCAGGATCGCAGATGAAG	
NF-514	Human BCR-ABL1 F	TCCGCTGACCATCAAYAAGGA	
NF-515	Human BCR-ABL1 R	CACTCAGACCCTGAGGCTCAA	
NF-407	Murine RAG1 F	AGGCCTGTGGAGCAAGGTAG	
NF-408	Murine RAG1 R	TTTCATCGGGTGCAGAACTGA	
NF-411	Murine AID F	CCCTTGTACGAAGTCGATGAC	
NF-412	Murine AID R	ATCACGTGTGACATTCCAGGAG	

Table 6-5: Primers used for RT-qPCR in chapters 2 and 3

Primer name	Description	Sequence 5' to 3'	Reference
NF-645	EVI1 1A F	TATTGCTGAGTTGAGGCCATAG	(Lugthart et al., 2008)
NF-646	EVI1 1B F	TGCGGTCTGGACACGTCTC	
NF-647	EVI1 1A/B R	CTTCCAACATCTGGTTGACTGG	
NF-648	EVI1 1C F	ACCCTTTGGCTAGATTATCTTAGACGA	
NF-649	EVI1 1C R	CCAGCGAATCTAATGTACTTGAGC	
NF-650	EVI1 1D F	CTTCTTGACTAAAGCCCTTGGA	
NF-651	EVI1 1D R	GTACTTGAGCCAGCTTCCAACA	
NF-652	EVI1	GGTATCTTAGTGTATATCTTGCCCTTTGT	
NF-653	EVI 3L R	GCGCAATGTCTGCAACTACTCT	
NF-654	MEC F	GAAAGACCCCAGTTATGGATGG	
NF-655	GUSB F	GAAAATATGTGGTTGGAGAGCTCATT	
NF-656a	GUSB R	CCGAGTGAAGATCCCCTTTTTA	
NF-656b	EVI all F	AGTGCCCTGGAGATGAGTTG	
NF-657	EVI all R	TTTGAGGCTATCTGTGAAGTGC	

Table 6-6: Primers used for RT-qPCR in chapter Error! Reference source not found.

A dilution series comprising five 1:10 dilutions of a suitable control cDNA was used to create a standard curve; primers were rejected if the R^2 value was not above 0.99, if the PCR efficiency was not between 90 and 105%, and if the slope was not 3.3±10% as shown in Figure 6-4.

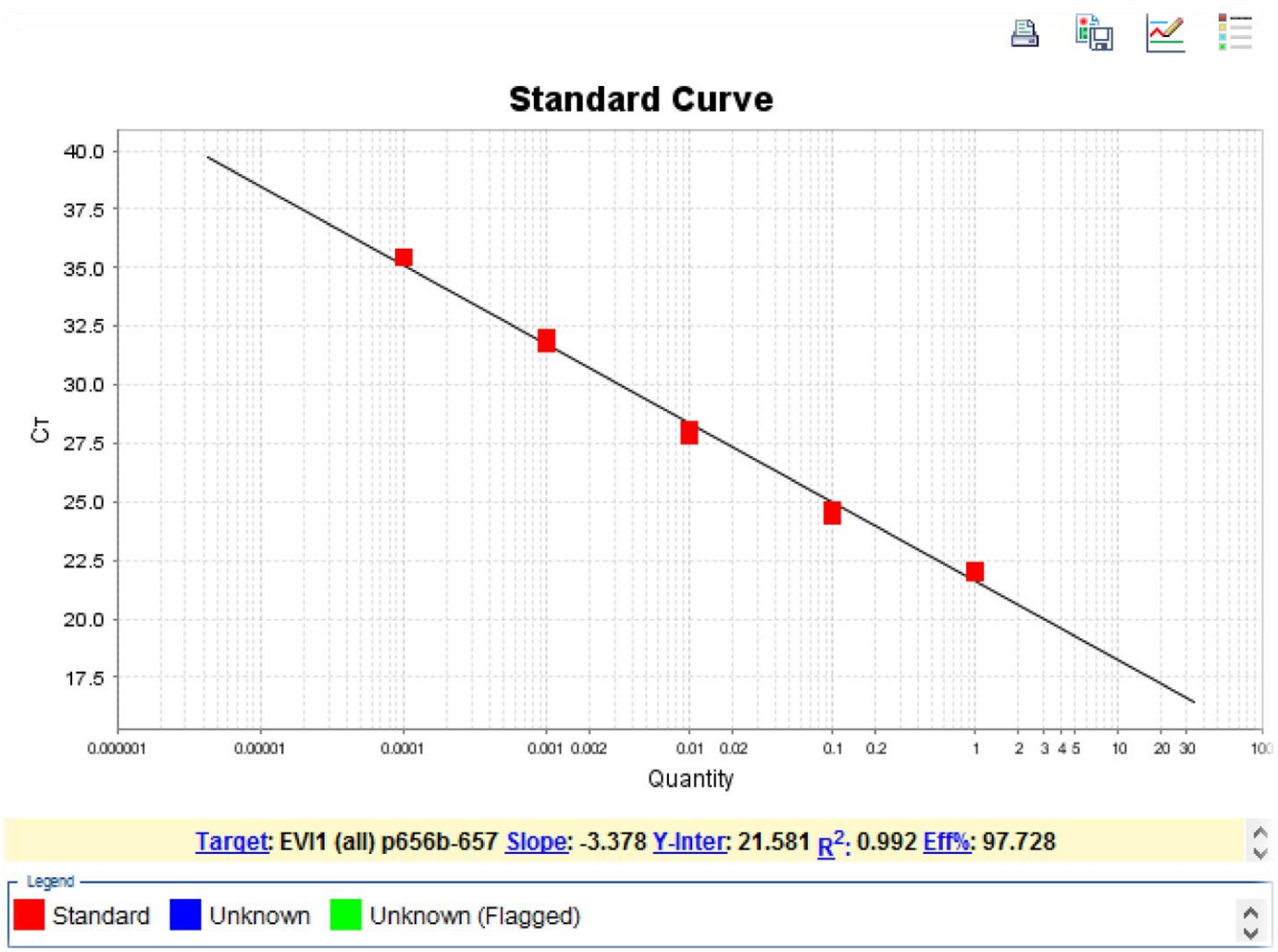


Figure 6-4: RT-qPCR primer pairs amplified efficiently.¶A standard curve showing linear amplification of five 1:10 dilutions of LAMA84 cDNA by primer pair p656b-657 (all *EVI1*). The reaction has an R^2 value over 0.99, has an efficiency of between 90 and 105%, and a slope between 3.1 and 3.6.

In addition, a melt curve analysis was undertaken of the PCR product (shown in Figure 6-5), to ensure that there was a single PCR product with no off-target binding of the primers.

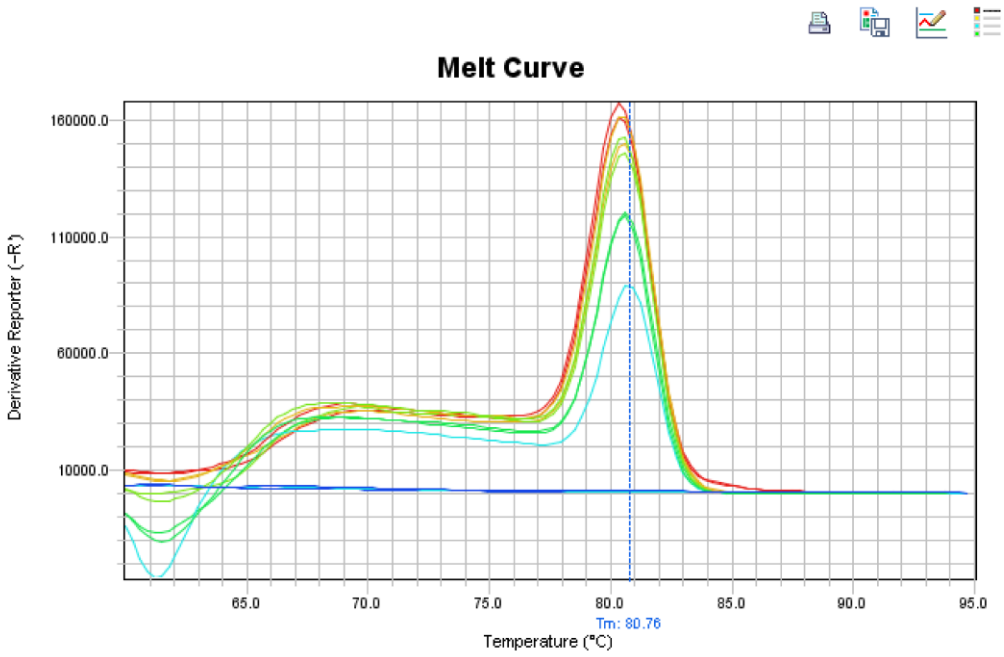


Figure 6-5: RT-qPCR primer pairs yielded a product with a uniform melt temperature, indicative of a single amplicon. ¶Melt curve analysis of the amplicon from five 1:10 dilutions of LAMA84 cDNA using primer pair p656b-657 (*EVI1* all). All amplicons melted at ~80.7˚C. ¶

6.5.7 Messenger RNA (mRNA) purification

A 1µl aliquot of total RNA was analysed using an RNA 6000 pico chip and the total eukaryote RNA assay on the Bioanalyzer (Agilent Technologies) according to the manufacturer's protocol to check for degradation prior to mRNA purification. Messenger RNA was isolated from 5µg of total RNA using the NEBNext Oligo $d(T)_{25}$ Magnetic Beads (New England Biolabs) according to the manufacturer's protocols. Beads were separated using a 0.2ml magnetic rack (Diagenode). A 1µl aliquot of the eluted mRNA was analysed using an RNA 6000 pico chip and the mRNA pico assay on the Bioanalyzer (Agilent Technologies) according to the manufacturer's protocol to confirm purity.

6.6 Retroviral transduction

6.6.1 Production of retroviral supernatant

BOSC 23 cells (a highly transfectable derivative of the human embryonic kidney 293 cell line) were cultured to 60-70% confluency in 10cm plates with 10ml media. Plasmids were thawed at RT, diluted to 1µg/µl and heated to 55˚C for 5 minutes. For each plate, 2.5µg of the retrovirus packaging vector pCL-Eco (Naviaux et al., 1996) and the appropriate empty or oncogene-expressing MIGR1 vector (MIGR1 expressing murine MYC is shown in was added to 500µl Opti-MEM (ThermoFisher Scientific) and vortexed with 15µl XtremeGENE 9

(Roche) pre-warmed to RT. The tube was incubated at RT for 15 minutes. The plasmid map of pCL-Eco is shown in Figure 6-6 and of murine MYC-expressing MYC in Figure 6-7.

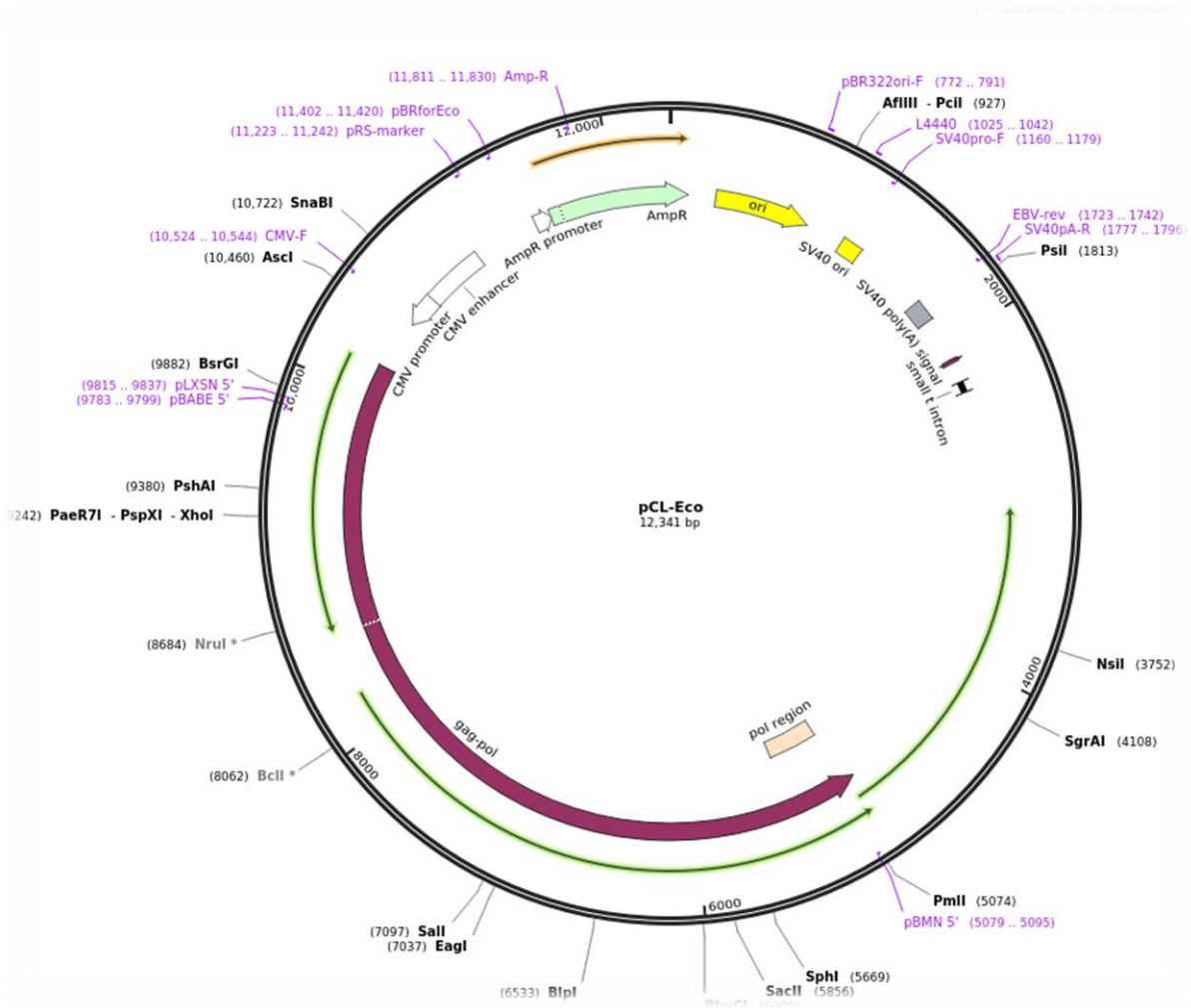


Figure 6-6: Sequence map showing the pCL-Eco plasmid

Figure 6-7: Sequence map showing the MIGR1 plasmid¶expressing murine *MYC*, cloned by Dr Niklas Feldhahn using XhoI. The same plasmid, expressing human *BCR-ABL1*, was a gift from Dr Danillo Perotti.

Meanwhile, 3ml of media was removed from each plate and reserved. The XtremeGENE/plasmid mix was added dropwise to each plate and swirled gently for 30 seconds. After 24 hours, the reserved media was replaced. Viral supernatant was harvested after 48, 72 and 96 hours, sterile filtered and stored at -80˚C until required.

6.6.2 Retroviral transduction

Cells from red-cell depleted murine bone marrow were cultured in six well plates, $10x10^6$ cells per well with 4ml lymphoid media (see section 6.2). Two rounds of retroviral infection were undertaken after 48 and 72 hours.

Sufficient retroviral supernatant was thawed to allow 1-2ml for each well of a six well plate. Polybrene (Merck Millipore; to a final concentration of 2.5µg/ml) and Hepes pH7.2 (ThermoFisher Scientific; to a final concentration of 20mM) was added. The mix was vortexed and equilibrated to RT.

Surplus media (approximately 3.5ml) was removed from the plate with care to ensure minimal disturbance to the cells and reserved. The retroviral supernatant mix was added to each plate, and the plates were centrifuged for 90 minutes at 30˚C at 800g with no brake and then incubated at 37˚C. After four hours, the retroviral supernatant was removed and the reserved media was replaced.

6.6.3 Confirmation of transduced gene expression

Transduced expression of murine *MYC* and human *BCR-ABL1* was confirmed using RT-qPCR (as described in section 6.5 using the primers listed in Table 6-5) and shown in Figure 6-8.

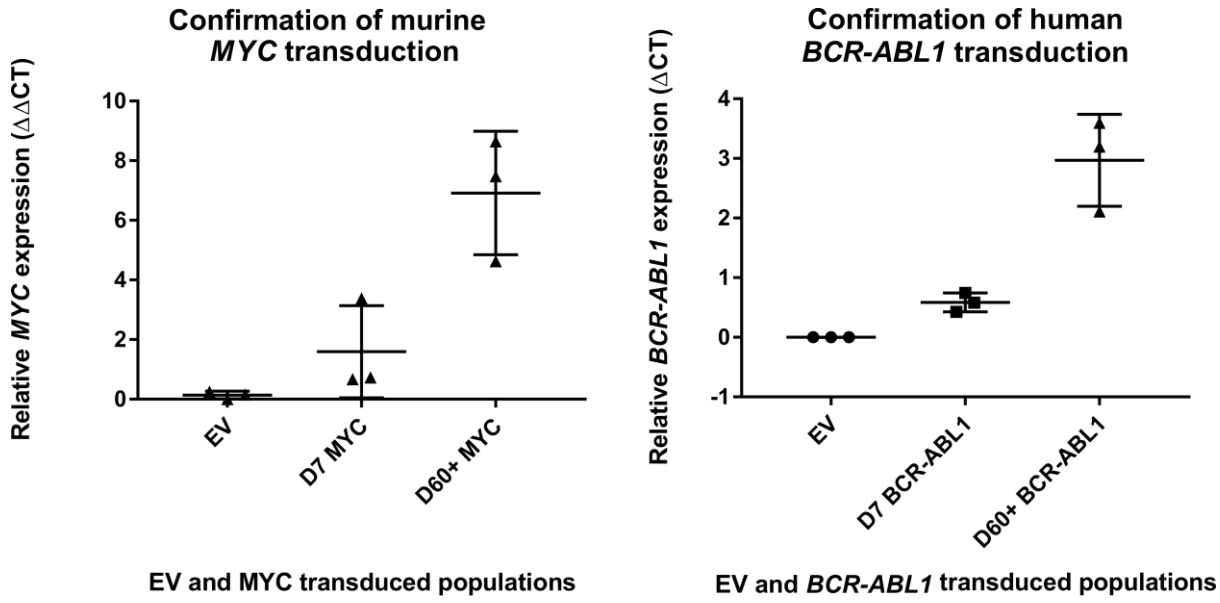


Figure 6-8: Transduction of murine *MYC* and human *BCR-ABL1* was confirmed by RT-qPCR¶ Scatter plot showing the expression of murine *MYC* (left) relative to the mean of the EV and human *BCR-ABL1*.

6.7 Cytogenetic studies

6.7.1 Fixed cell suspension (FCS)

Cells were incubated with 0.1mg/ml KaryoMAX colcemid solution (Thermo Scientific) for four hours to arrest cells at metaphase prior to harvest. Cells were then centrifuged (300g for 5 minutes). The supernatant was removed and replaced with 0.75M potassium chloride (Sigma Aldrich) hypotonic solution for 15 minutes at 37°C to swell the cell membrane. Cells were again centrifuged and the supernatant removed. Carnoy's fixative (methanol 3:1 glacial acetic acid) was added drop wise to the cell pellet with vigorous agitation. Fixation was

repeated up to three times until the suspension was clear and the cell pellet clean. FCS was stored at -20°C until required.

6.7.2 Fluorescence in situ hybridisation (FISH)

FCS was dropped on to a slide and air dried. The chromosome preparation was then dehydrated by passing the slide through an ethanol series (70%, 90% and 100%; 1 minute each). 0.5µl probe (Agilent Technologies or Kreatech Diagnostics) was diluted in 4.5µl pre-warmed Ambion ULTRAHyb (Thermo Scientific) and applied to the slide under a 22mm^2 coverslip, sealed with Parafilm (Bemis). Chromosome preparation and probe were co-denatured at 72°C for 2 minutes and then hybridised overnight at 37°C on a Hybrite (Abbott Molecular).

The next morning, coverslips were soaked off in 2xSSC (see section6.12.17) and chromosome preparations were washed in 0.4xSSC (see section 6.12.16) at 72°C for 2 minutes. Chromosome preparations were then rinsed in 2xSSC/0.3% Igepal (see section 6.12.18) and again dehydrated through an ethanol series (70%, 90% and 100%; 1 minute each). Chromosome preparations were air dried, mounted with Vectashield antifade mounting medium with 4',6-diamidino-2-phenylindole (DAPI; Vector Labs) and sealed with clear nail polish.

Interphase and metaphase images were captured by the GSL-120 (Leica Biosystems) and analysed using the CytoVision software, incorporating a Zeiss Imager M2 microscope. Probes were visualised using Spectrum Green and Spectrum Red single and dual pass filters. Slides were protected from the light and stored at 4°C until required.

6.7.3 G-banded chromosome analysis

FCS was dropped on to a slide and air dried. Slides were aged overnight at 60°C then exposed to trypsin, stained with Giemsa and mounted. Metaphase images were captured by the GSL-120 (Leica Biosystems) incorporating a Zeiss Imager M2 microscope. Karyograms were produced using the CytoVision software.

6.8 Chromatin interaction studies

The protocol for cross-linking, sonication and ChIP of murine lymphoid cells using γH2AX and H3K27Ac was optimised by Dr Bryant Boulianne. The protocols performed well for the same processed using human myeloid cells aside from sonication, which needed to be optimised.

6.8.1 Cell purification.

Cells for γH2AX ChIP-seq were depleted of the dead and/or apoptotic cellular fraction using

the Dead Cell Removal kit and column-based magnetic-activated cell sorting (Miltenyi Biotech) according to the manufacturer's protocols.

Cells with myeloid switched phenotype were enriched for CD11B expression using the CD11B microbeads and column-based magnetic-activated cell sorting (Miltenyi Biotech) according to the manufacturer's protocols.

Cells for the *EVI1* expression experiment were sorted for CD34 expression using the CD34 microbead kit and column-based magnetic-activated cell sorting (Miltenyi Biotech) according to the manufacturer's protocols.

6.8.2 Cross-linking

$15x10^6$ cells were incubated in 12ml warm complete media; 750µl of 16% formaldehyde (Alfa Aesar) was added while swirling and the sample was incubated at 37˚C or precisely 10 minutes. The cross-linking reaction was stopped with 600µl of 2.5M glycine (6.12.8) and samples were placed immediately on ice. The cells were centrifuged at 400g for 6 minutes at 4°C, washed in 10ml of cold PBS (Thermo Scientific), centrifuged, and washed again in 1.5ml cold PBS. The cells were centrifuged at 800g for 5 minutes at 4°C; all supernatant was removed and the clean cell pellet was snap-frozen on dry ice and stored at -80°C until required.

6.8.3 Sonication

The frozen cell pellet was resuspended in 130µl of cold Tris/SDS sonication buffer (6.12.10) supplemented with protease inhibitors (and phosphatase inhibitors for γH2AX ChIP only) and incubated for 20 minutes on ice. 120µl was transferred to a microTUBE AFA Fiber pre-slit snap cap tube (Covaris) and sonicated using the Covaris S220 set to 4°C. Sonication times (lymphoid 5 minutes) and settings (Peak power 105; Duty Factor 5; Cycles Burst 200) were yielded fragments of 300-400bp.

Sonication of myeloid cell pellets using the same time and settings yielded DNA which was insufficiently sonicated. Four identical cell pellets from a single CML patient were therefore used to optimised the myeloid sonication times. As shown in Figure 6-9, a ten minute sonication yielded DNA fragments of the optimal length.

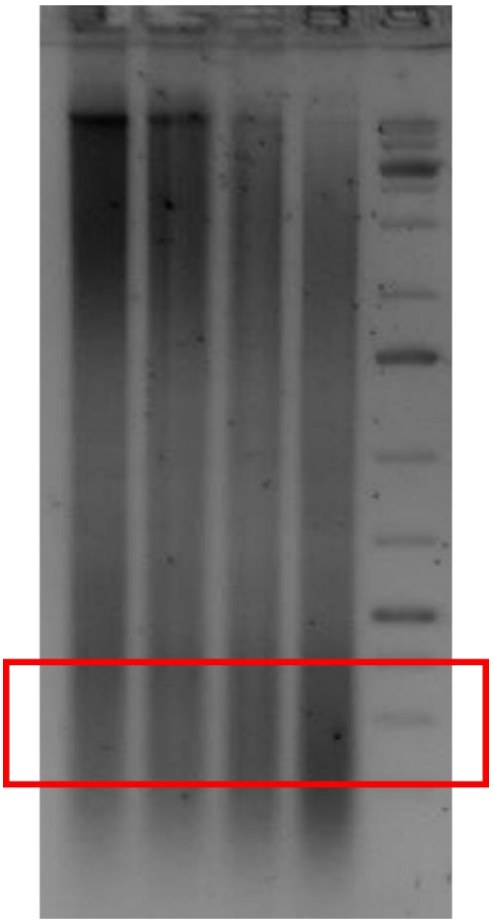


Figure 6-9: Optimisation of sonication times for myeloid cell pellets¶ 1.5% agarose gel run for 3 hours at 50V showing DNA extracted from four identical cell pellets from a single CML patient which was sonicated for (shown left to right) 4, 6, 8 and 10 minutes. The 10 minute sonication yielded fragments of the optimal 300-400bp length (indicated by a red box).

Sonicated chromatin was diluted with 880µl of cold concentrated RIPA buffer (6.12.9). A 10µl aliquot was reserved to check the efficiency of the sonication step. The samples were centrifuged at 17,000g for 10 minutes at 4°C. The supernatant was transferred to a new tube and the pellet discarded. Again, a 10µl aliquot of the sample was reserved as an input control for the reverse cross-linking step.

6.8.4 Immunoprecipitation

Protein A Dynabeads (Thermo Scientific) were used for immunoprecipitation; 40µl of Dynabeads were added to a 1.5ml Eppendorf and washed with 600µl of PBS (Thermo Scientific). The beads were removed using a DynaMag magnetic rack (Thermo Scientific) and resuspended with the sonicated cell lysate. Samples were rotated for 30 minutes at 4°C to pre-clear the chromatin.

A second aliquot of washed beads was resuspended in 100µl of PBS containing 5µg of antibody (see Table 6-7) and incubated at RT for 20 minutes.

Table 6-7: Antibodies used for ChIP experiments

Supplier	Item	Description

		code		
Anti-gamma (phospho S139) γH2AX	Abcam	Ab2893	Rabbit, polyclonal, IgG	Protein A
H3K27Ac	Abcam	Ab4729	Rabbit, polyclonal, IgG	Protein A
Isotype control	Santa Cruz Biotechnologies	Sc2027	Rabbit, polyclonal, IgG	Protein A

The histone H3 antibody-bead complex was washed with 200µl of PBS twice and resuspended with the supernatant from the pre-cleared chromatin. The tube was rotated overnight at 4°C.
The following morning, the DynaMag magnetic rack (Life Technologies) was used to perform 8 washes of the beads, all undertaken at 4°C for ten minutes:

- 2x 1ml of RIPA buffer (6.12.11)
- 2x 1ml of RIPA buffer with 0.3M NaCl (6.12.12)
- 2x 1ml of LiCl buffer (6.12.13)

The beads were resuspended with 100µl of TE with 0.2% Triton X-100. A 10µl aliquot was removed for subsequent Western blot analysis of the protein precipitate. A further 900µl TE with 0.2% Triton X-100 was added to the remaining bead suspension, and a final wash in 1ml TE was undertaken in a new tube. The beads were finally resuspended in 100µl of TE.

6.8.5 Reverse cross-linking and DNA clean up

Cross-linked protein and DNA were separated by adding 3µl of 10% SDS and 5µl of 20mg/ml proteinase K (New England Biolabs) to the beads from the immunoprecipitation step. The samples were incubated at 65°C for 4 hours, with gentle vortexing every 30 minutes to resuspend the beads.

MaXtract high density gel tubes (Qiagen) were spun at 14,000g for 1 minute to settle the gel. The beads were separated from the supernatant using the DynaMag and the supernatant was transferred to the MaXtract tubes. The beads were washed with a further 100µl of TE with 0.5M NaCl and that was also transferred to the MaXtract tube. An equal volume of phenol/chloroform as added and the tubes were shaken vigorously to mix the phases. The tubes were centrifuged at 17,000g for 5 minutes at RT and 190µl of the aqueous phase was transferred to a new 1.5ml Eppendorf, with 2µl of 20mg/ml glycogen, 20ul of 3M sodium acetate pH5.2 and 500µl of ethanol. The tubes were frozen at -80˚C for 10 minutes then centrifuged for 5 minutes at 4˚C at 17,000g. The supernatant was removed and the pellet

washed with 500µl of 70% ethanol. The DNA was again centrifuged, the supernatant was again removed and finally the DNA allowed to air dry. The dried pellet was resuspended in 30µl TE buffer. DNA yield was assessed via fluorometric quantitation using the Qubit High Sensitivity kit (Life Technologies) according to the manufacturer's protocol and stored at -80˚C until required.

6.8.6 ChIP-qPCR

All ChIP-qPCR was undertaken in triplicate using SYBR Jumpstart mastermix (Sigma Aldrich) according to the manufacturer's protocols. All assays were quantified using the StepOne Plus thermocycler (Applied Biosystems) at the concentrations shown in Table 6-8 in standard mode (ramp rate 60%) with the following steps: initial hotstart (2 minutes at 94°C), followed by 40 cycles of 15 seconds denaturation at 94°C and 1 minute annealing/extension at 60°C.

	Stock	Final	Volume
SYBR jumpstart mastermix	2x	1x	12.5
Forward primer	20uM	400nM	0.5
Reverse primer	20uM	400nM	0.5
ROX	100X	1x	0.25
Nuclease-free water			9.25
ChIP-DNA			2
		TOTAL	25

Table 6-8: Reagent concentrations for ChIP-qPCR

All primers (see Table Table 6-9) were validated to ensure that they primed efficiently before use. A dilution series comprising five 1:10 dilutions of a suitable control cDNA was used to create a standard curve; primers were rejected if the R^2 value was not above 0.99, if the PCR efficiency was not between 90 and 105%, and if the slope was not 3.3±10%. In addition, a melt curve analysis was undertaken of the PCR product, to ensure that there was a single PCR product with no off-target binding of the primers.

Table 6-9: Primers used for ChIP-qPCR¶(Sigma Aldrich)

Primer name	Description	Sequence 5' to 3'	Related section
NF-P198	DsRed F	CGAGTTCATGCGCTTCAAGG	2.4.6
NF-P199	DsRed R	GTCACCTTCAGCTTCACGGT	2.4.6

6.8.7 Chromosome conformation capture (3C)

Cross-linked pellets were resuspended in 1ml cold 3C lysis buffer and homogenised using 90

strokes of a pellet pestle (Kimble Kontes) in a 1.5ml Eppendorf. The lysate was centrifuged in a fresh Eppendorf and resuspended in 650ul of water, and split into four reactions:

Table 6-10: Composition of the digestion and control reactions

Reagent	Digestion reactions (x3)	Control 1 (undigested chromatin)
SDS (20%)	10 µl	2.5 µl
Restriction buffer (x10)	80 µl	20 µl
MilliQ Water	414 µl	111 µl
Chromatin / cells	200 µl	50 µl

The reactions were shaken horizontally at 1400rpm for 1 hour at 37°C in an Eppendorf Thermomixer C with the Thermotop heated lid. Cells were resuspended by gentle pipetting and 66µl of 20% Triton X100 was added to each digest, and 16µl to control 1 (undigested). The reactions were shaken for one hour, as before. Three additions of 500U of the primary restriction enzyme (High concentration EcoR1; New England Biolabs) were then added; the first immediately, the second at the end of the day and a third the next morning with continuous shaking as before.

A 100µl aliquot was removed from each digest to make Control 2 (digested). Controls 1 and 2 were reverse-cross-linked by addition of 3µl Proteinase K (New England Biolabs) and incubation at 65°C overnight.

The restriction enzyme in the digests was heat inactivated for 20 minutes at 65°C. Digests were then cooled on ice, and 500µl of molecular grade water, 133µl of 10X ligation buffer (Thermoscientific) and 240U of high concentration T4 DNA ligase (Thermoscientific) was added. Digests were then shaken overnight at 1400rpm at 16°C. The following night, the digests were also reverse crosslinked overnight at 65°C with 5µl proteinase K.

Meanwhile, the controls were extracted with phenol chloroform as per Table 6-11. Controls 1 and 2 were incubated at 37°C with 4µl and 6µl, respectively, of RNAse (New England Biolabs) for 30 minutes. Controls were transferred to a MaXtract tube (Qiagen) and an equal volume of phenol/chloroform was added. Tubes were shaken vigorously and centrifuged for 5 minutes at 17,000g at RT. The supernatant was transferred to a fresh 1.5ml Eppendorf and the precipitation reagents were added. The reactions were frozen at -80°C for 10 minutes then centrifuged at 17,000g for 20 minutes at 4°C. Finally, the pellet was washed in 1ml of 70% ethanol (Sigma Aldrich), air dried at RT and resuspended in 30µl water.

Table 6-11: Phenol-chloroform extraction of the digests and controls

	RNAse	Phenol/ chloroform	Ethanol precipitation	Ethanol wash	Volume for H20 reconstitution
Ligation reaction (4ml)	30ul	4ml	• Ethanol 100% 35ml • Water 7ml • 3M sodium acetate 1ml	10ml 70%	300ul
Control 1 (200ul)	4ul	200ul	• Ethanol 100% 600ul • 3M sodium acetate 24ul • Glycogen 1ul	1ml 70%	30ul
Control 2 (300ul)	6ul	300ul	• Ethanol 100% 900ul • 3M sodium acetate 36ul • Glycogen 1ul	1ml 70%	30ul

The reverse cross-linked digests were pooled in a 15ml Falcon and incubated with 30µl of RNAse for 30 minutes at 37°C. An equal volume of phenol-chloroform was added and the reaction centrifuged at 3400g for 10 minutes at RT. The supernatant was removed and transferred to a 50ml Falcon and diluted with 7ml water (to reduce the amount of DTT carried over). The DNA was precipitated with reagents as per Table 6-11 then frozen at -80°C for at least two hours. The frozen tubes were centrifuged at 3400g for 45 minutes at 4°C. The pellet was washed in 10ml 70% ethanol and transferred to a 15ml Falcon and air dried before finally being resuspended in 300µl water.

6.8.8 Assessment of 3C digestion efficiency

Digestion efficiency was quantified by qPCR and visually assessed. Three 2µl aliquots of each of the control reactions were amplified by SYBR qPCR using two sets of primer pairs (p718-p719 and p720-p721) – one amplicon spanned an EcoR1 restriction site, and one fitted within two sites. The ΔΔCT method was used to calculate the digestion efficiency as previously described (Gondor et al., 2008).

A 3ul aliquot was run on a 1.5% agarose gel stained with SYBR safe DNA gel stain (Life Technologies) and visualised using a gel doc (Analytik Jena MultiDoc-It). The target size for EcoR1 (six-cutter) was 4-12Kb.

6.9 Next Generation Sequencing libraries

6.9.1 mRNA library preparation

Messenger RNA libraries were prepared using the NEB Next mRNA library set for Illumina (New England Biolabs) according to the manufacturer's protocols. A 1µl aliquot of the fragmented mRNA was analysed using an RNA 6000 pico chip and the mRNA assay on the Bioanalyzer (Agilent Technologies) according to the manufacturer's protocol. A 1µl aliquot of the finished library was also analysed using a high sensitivity DNA chip and the high sensitivity DNA assay on the Bioanalyzer (Agilent Technologies) according to the manufacturer's protocol.

6.9.2 ChIP-seq library preparation

Libraries were prepared from 10ng ChIP-DNA using the NEB Next Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) as per the manufacturer's instructions. A 1µl aliquot of the finished library was analysed using a high sensitivity DNA chip and the high sensitivity DNA assay on the Bioanalyzer (Agilent Technologies) according to the manufacturer's protocol.

6.9.3 4C library preparation

The 3C libraries were quantified using fluorometric quantitation using the Qubit (Life Technologies) and then 25µg was digested overnight with 50U of DpnII in 1x Buffer 3.1 (both New England Biolabs) overnight at 37°C in an Eppendorf Thermomixer C. The following morning, the sample was transferred to MaXtract high density gel tubes (Qiagen) which had been centrifuged at 14,000g for 1 minute to settle the gel. An equal volume of phenol/chloroform (Sigma Aldrich) was added and the tubes were shaken vigorously to mix the phases. The tubes were centrifuged at 15,800g for 15 minutes at RT and the aqueous phase was transferred to a new 1.5ml Eppendorf, with 1µl of 20mg/ml glycogen (Thermo Scientific), 36ul of 3M sodium acetate pH5.2 (Sigma Aldrich) and 900µl of ethanol (Sigma Aldrich). The tubes were frozen at -80˚C for 10 minutes then centrifuged for 5 minutes at 4˚C at 17,000g. The supernatant was removed and the pellet washed with 1ml of 70% ethanol. The DNA was again centrifuged, the supernatant was again removed and finally the DNA allowed to air dry. The dried pellet was resuspended in 100µl molecular grade water (Sigma Aldrich). The digested DNA was analysed by running a 3µl aliquot on a 0.5% agarose gel to check that the fragments were below 1kb with the majority between 300 and 500bp.

The remaining digested DNA was circularised overnight. The sample was transferred to a 50ml Falcon with 200U T4 DNA ligase (ThermoScientific) and 1x ligation buffer made up to 14ml with molecular grade water – the increased volume increases the likelihood that each circle will comprise fewer fragments. The ligation reaction was incubated at RT overnight.

The circularised DNA was extracted with 14ml phenol/chloroform and shaken vigorously. It was then centrifuged at 3,200g at RT for 10 minutes. The aqueous phase was split into two new 50ml Falcons and the volume doubled with molecular grade water – this reduces the coprecipitation of DTT (carried over from the ligation buffer). Each tube of diluted DNA was precipitated with 2.6ml sodium acetate pH5.2 and 26ml 100% ethanol. The tubes were frozen at -80°C for at least 3 hours, until solid, then centrifuged at 3,200g at 4°C for 45 minutes. The supernatant was removed and the pellet was washed with 15ml of 70% ethanol and centrifuged as before for 15 minutes. The supernatant was again removed and pellet was air-dried. Each pellet was resuspended in 75µl of TE buffer and incubated for 30 minutes at 37°C. Split samples were recombined in an Eppendorf, and the yield was quantified using fluorometric quantitation using the Qubit Broad Range kit according to the manufacturer's protocol.

A total of 42ng per library was amplified using the reagents detailed in Table 6-12.

Table 6-12: Reagents for PCR amplification of 3C libraries

Reagent	Volume	Final Concentration
10× buffer I (Expand Long Template system)	2.5uL	1x
10mM dNTPs	0.5uL	0.2
20uM 4C-seq forward primer	2.5ul	2uM
20uM 4C-seq reverse primer	2.5ul	2uM
DNA polymerase mix	0.35uL	
42ng 3C library	16.65ul	
dH2O		
	25ul	

The reactions were first performed with standard primers and then, if successful, with PCR primers which included a 5' overhang comprising the P5 or P7 Illumina adapter sequence, an Illumina sequencing primer, a barcode and/or an index as shown in Table 6-13. These sequences were then incorporated in the PCR product to make functional libraries when amplified by the bait specific primers at the 3' end.

Primer name	Description	Sequence 5' to 3'
NF-P693	Bait A read	AGAGCGGCTCCAAAGTCG
NF-P742	Bait A non-read	TGCCAGTCAAGGAAACAAACA
NF-P759	Bait B read	TGTTGCTGGGTTTTGTTGGT
NF-P758	Bait B non-read	ACTTGAATCTGCAATTGGGTGA
NF-P847	Bait C non-read	CCGCACTAACAAAAACTCCAA
NF-P848	Bait C read	AGAGGCAGCAGTGGTAGGAA
NF-P849	Bait D non-read	AGGGTCTTCTTGCTATTTA
NF-P850	Bait D read	TGATGATGCAATGAATTC
NF-P851	Bait E non-read	TGGAAGGACTTTTAATGGAGGA
NF-P852	Bait E read	AAGGATGCCTGAATTAAACCTT
NF-P853	Bait F non-read	TCCCAGTAGTGGCAATGATTC
NF-P854	Bait F read	TAGCTGAGCCACCCCTGA
NF-P855	Bait G non-read	GGCAAGTGTATTGGGGTAGG
NF-P856	Bait G read	CTGAGGACATCAGGCAACTG
NF-P857	Bait H non-read	TGCCTCAAATTTCGTATGGT
NF-P858	Bait H read	CCCAAACAGCTGGGTCATAC
NF-P827	P5 Bait A read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTATCACGAGAGCGGCTCCAAAGTCG
NF-P828	P5 Bait A read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTCGATGTAGAGCGGCTCCAAAGTCG
NF-P829	P5 Bait A read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTTTAGGCAGAGCGGCTCCAAAGTCG
NF-P830	P5 Bait A read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTTGACCAAGAGCGGCTCCAAAGTCG
NF-P831	P5 Bait A read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTACAGTGAGAGCGGCTCCAAAGTCG
NF-P832	P5 Bait A read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTGCCAATAGAGCGGCTCCAAAGTCG
NF-P833	P5 Bait A read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTCAGATCAGAGCGGCTCCAAAGTCG
NF-P834	P5 Bait A read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTACTTGAAGAGCGGCTCCAAAGTCG
NF-P845	P7 Bait A non-read with index & Illumina sequences	CAAGCAGAAGACGGCATACGAGATATTACTCGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTATTACTCGTGCCAGTCAAGGAAACAAACA
NF-P836	P5 Bait B read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTGATCAGACTTGAATCTGCAATTGGGTGA
NF-P837	P5 Bait B read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTTAGCTTACTTGAATCTGCAATTGGGTGA
NF-P838	P5 Bait B read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTGGCTACACTTGAATCTGCAATTGGGTGA
NF-P839	P5 Bait B read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTCTTGTAACTTGAATCTGCAATTGGGTGA
NF-P840	P5 Bait B read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTAGTCAAACTTGAATCTGCAATTGGGTGA
NF-P841	P5 Bait B read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTAGTTCCACTTGAATCTGCAATTGGGTGA
NF-P842	P5 Bait B read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTATGGCAACTTGAATCTGCAATTGGGTGA
NF-P843	P5 Bait B read with index & Illumina sequences	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTCCGTCCACTTGAATCTGCAATTGGGTGA
NF-P846	P7 Bait B non-read with index & Illumina sequences	CAAGCAGAAGACGGCATACGAGATTCCGGAGAGTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCTTCCGGAGATGTTGCTGGGTTTTGTTGGT

Table 6-13: Primers used for *EVI1* 4C. ¶All relate to section 4.4.1 and 4.4.3¶

The number of cycles (detailed in Table 6-14) was minimised so as to ensure the reaction was stopped when still linear; 10µl PCR product was assessed on a 2% agarose gel to ensure the reaction was not saturated. A 1µl aliquot of a 1:10 dilution of the finished library was analysed using a high sensitivity DNA chip and the high sensitivity DNA assay on the Bioanalyzer (Agilent Technologies) according to the manufacturer's protocol in order to establish the mean fragment size.

Table 6-14: PCR cycling conditions for amplification of the 3C library

	Cycle step	Temperature	Time
1x	Initial denaturation	94°C	2 minutes
30x	Denaturation	94°C	15 seconds
	Annealing	55°C	1 minute
	Extension	68°C	2 minutes
1x	Final extension	68°C	7 minutes
	Hold	4°C	∞

6.9.4 4C library quantification

The 4C libraries were quantified by RT-qPCR using the NEBNext Library Quant Kit for Illumina (New England Biolabs) according to the manufacturer's protocol. All reactions were performed on the StepOnePlus real time system (Applied Biosystems).

6.10Next Generation Sequencing

All libraries were submitted for sequencing by the Medical Research Council London Institute of Medical Sciences at Imperial College London. Libraries were pooled at 5nM. All sequencing was undertaken using the HiSeq2500 (Illumina). All ChIP-seq experiments were subject to single end 75bp reads; all other sequencing was subject to paired end 100bp reads.

Of note, the construction of the 4C library dictates that one bait-specific primer was informative (the P5 'read') and the other was not (the P7 'non-read'). Approximately 30bp of the 4C library sequence read was 'lost' to the bait specific primer so 76bp reads were insufficient. However, a routine 100bp paired end read was cheaper than a custom 100bp single end read so the non-read P7 sequence was disregarded.

6.11 Bioinformatic analysis

All Bioinformatics analysis was undertaken by Dr Mark Robinson

6.11.1 ChIP-seq processing and custom track generation

Default Illumina quality filters and image analysis were used to obtain single end reads of 50bp. All reads were aligned to Build 38 (GRCm38/mm10) mouse genome assembly or Build 38

(GRCh38/hg38) human genome assembly, as appropriate, using default settings of Burrows-Wheeler Alignment with Maximal Exact Matches. Reads which failed to align exactly were removed with SAMtools (Li and Durbin, 2009); PCR duplicates from library amplification were also removed with Picard tools (http://broadinstitute.github.io/picard). Custom UCSC tracks for direct visualisation were generated using deepTools (Ramirez et al., 2014); coverage within 200bp windows was normalised to library and window size. Data is shown as reads per kilobase per million reads (RPKM).

6.11.2 ChIP-seq region calling

Enrichment was performed as described previously (Barlow et al., 2013) with some minor modifications to allow optimized calling of the enriched regions. Following pre-processing described in section 6.11.1, statistical enrichment of the uniquely aligned reads which mapped within non-overlapping windows (100bp in length) were compared to a Poissonian random distribution using the SICER program (Zang et al., 2009) designed to identify enriched domains from histone modification ChIP-seq data. The following parameters were empirically derived used: e-value 10,000 (which controls the genome-wide error rate against the random background), gap size 300 (ineligible windows between eligible ones). The statistical enrichment of 'bins' passing the above criterion were then compared to the negative control (input DNA or isotype control DNA) based on tag density fold increase, as determined from a false discovery rate of $1x10^{-5}$. To delineate enriched regions, windows that were less than 10Kb apart were then merged before application of a final empirical filtering step to remove regions with low coverage (defined as less than 1.5 RPKM), low width (less than 5Kb) or a fold change of less than 1.1. Settings were optimised for γH2AX and then applied to H3K27Ac.

6.11.3 Differential peak calling

Peaks were first identified in each sample compared to their corresponding input control using MACS2 with default settings. Peaks were filtered for a –log10 q-value of greater than 30 and combined into a common peak set by taking the union of all conditions per comparison (i.e. per oncogene). Differential binding analysis was first conducted using DiffBind R package with default settings and significantly different peaks ($p<0.05$) were called. In order to identify larger regions of differential binding typical of super enhancers, initial differentiation of peaks with a common direction of change were stitched within 1kb and reanalysed, before the re-extraction of highly significantly differentially bound regions ($p<1x10^{-4}$). To identify fragments with significantly higher read counts than expected for the distance from the bait, z-scores were calculated from the residuals of the fit.

6.11.4 Establishing the γH2AX enriched regions

γH2AX regions identified by processing of *BCR-ABL1* and *MYC* reads as described in section 6.11.2 were merged using the Granges package (Lawrence et al., 2013) using R (R Core Team 2015; https://www.R-project.org) to yield merged lists.

6.11.5 Distribution of the enriched regions

Reads per million (RPM) values for the γH2AX-enriched regions from the merged list were calculated with BEDTools (Quinlan and Hall, 2010) designed for testing correlations between different genomic features. The IdeoViz package in R and the *BCR-ABL1* RPM values were used to generate the ideogram in Figure 2-14. The total chromosome coverage, shown in Figure 2-15, was again calculated using Granges (Lawrence et al., 2013) and plotted against murine exon coverage estimates (Sakharkar et al., 2005).

6.11.6 Correlation of different ChIP-seq libraries

BEDTools (Quinlan and Hall, 2010) was used to calculate RPM values for the γH2Ax-enriched regions from the merged list. The signal from the isotype control was subtracted and the log2 signal intensity was plotted in the scatter plots shown in Figure 2-11. The red line indicates linear regression fit and Pearson's rho (r) value is shown.

6.11.7 Gene ontology analysis

Genes which overlapped the γH2AX-enriched regions from the *BCR-ABL1* and *MYC* samples were identified using the ChIPpeakAnno package in R (Zhu et al., 2010). Enriched ($p<0.01$) gene ontology categories of biological processes were then identified and associated with the four empirically identified categories of chromosome organisation (GO: 0051276), nucleic acid metabolic process (GO: 0090304), gene expression (GO: 0010467) and immune system process (GO: 0002376). The pie charts in Figure 2-18 show each enriched data set. Categories in common between the *BCR-ABL1* and *MYC* datasets were identified and the -1log10 values were plotted in the bar chart (redundant terms were removed).

6.11.8 Meta-gene analysis of the enriched regions

As shown in Figure 2-23, meta-gene profiles of the γH2AX and H3K27Ac signals were plotted for all the Ensembl protein-coding genes plus an additional 10Kb upstream and downstream (n=22,533) using the ngs.plot package (Shen et al., 2014) designed to visualise and integrate different genomic databases. Profiles were normalised to gene length by spline fitting to library sized (by RPM) and plotted as log2 fold change of the isotype control ChIP-seq library. The ChIPSeeker R package ((Yu et al., 2015) for annotation , visualisation and comparison of ChIP-seq peaks was used to determine the intersection of the enriched regions overlapping genic features.

6.11.9 The stringent lists for H2AX and H3K27Ac ChIP-seq

The stringent lists for both H2AX and H3K27Ac were merged for *BCR-ABL1* and *MYC;* RPM within the regions were calculated from the above plus isotype control samples using BEDTools (before background subtraction). For visualisation purposes, negative values were arbitrarily set at 0.1RPM and excluded during calculation of Pearson's correlation coefficients. Log10 scatterplots were generated with the overlapping and unique regions highlighted. The linear regression fit is indicated

in red.

6.11.10 Comparison of ERFS and γH2AX regions

The ChIP-seq data set published by (Barlow et al., 2013) that studied ERFS in hydroxyurea-treated mature B-cells was used in the comparison of γH2AX regions depicted in Figure 2-40 (Table S1; accession SRR648771). The ERFS regions were converted to GRCm38/mm10 for this purpose using UCSC mm9tomm10 chain file). FASTQ reads were extracted from SRA files using the SRA-toolkit and aligned to the to mm10 genome. Permutation models were masked for sex chromosomes and unmapable regions for 100,000 enumerations; observed overlaps between regions were compared to the average permuted overlaps.

6.11.11Comparison of γH2AX regions to the RLFS

The potential RLFS were identified using the QmRLFS-finder (Jenjaroenpun et al., 2015) throughout mm10. Permutation testing was then undertaken using the R package regioneR (Gel et al., 2016) with 10,000 permutations per test. Re-sampling of genomic regions within the masked mm10 genome was used to determine RLFS enrichment within the γH2AX regions. RLFS enrichment within highly expressed genes was undertaken as before, but using the top 1000 highly expressed genes (ranked by mean expression across all conditions in RNA-seq).

6.11.12 Comparison of genomic defects in leukaemia to γH2AX

For comparison of the gene sets (assuming they could be related to annotated genes, and for which a murine equivalent exists*) published datasets were used, as shown in Table 6-15.

Table 6-15: Publications used for comparison of gene sets*

Reference	Respective table	Alterations for comparison
(Mullighan et al., 2007)	Table S10	N=39
(Mullighan et al., 2009b)	Table S4	N=42
(Roberts et al., 2012)	Table S3 and S7	N=42

* iAmp21, Znf528, Ccdc26, Fam22f, C20orf94 and FLJ11273 were excluded as no murine equivalent exists, lesions such as 'all 10p' were excluded as they did not map to a single murine region

6.11.13 Comparison of replication timing to γH2AX

γH2AX regions were categorised as early or late replicating by comparison with the ENCODE repli-chip data (Mouse et al., 2012). Permutation models were performed as above.

A zero-crossing algorithm was applied to the wavelet-smoothed Repli-chip data from CH12 (murine B-cell lymphoma) cell line to define early-replicating regions; only regions called as early in both biological replicates was used to annotate the γH2AX regions.

6.11.14 RNA-seq analysis

STAR v2.5.0 (Dobin et al., 2013) was used to align reads to the mouse genome (GRCm38/mm10); indexes were supplemented with UCSC known gene reference transcript assemblies and alignment used the following settings:

- outFilterType BySJout
- outFilterMultimanNmax 20
- alignSJoverhangMin 8
- alignSJDBoverhangMin 1
- outFilterMismatchNmax 999
- outFilterMismatchNoverLmax 0.04
- alignIntronMin 20
- alignIntronMax1000000
- Differential expression analysis was performed using R (RCore Team, 2016 https://www.R-project.org) with DESeq2 v1.12.4 (Love et al., 2014).

6.11.15 Comparison of γH2AX and RNA-seq data

The median FPKM value of expressed genes for each condition was used to categorise all known UCSC genes in high or low expression. The proportion of genes overlapping the γH2AX regions categorised as high or low expression was then determined. To make a comparison between γH2AX signals in high and low categories and in BCR-ABL1 and MYC categories, all known UCSC genes were extended by 10Kb and the γH2AX signals in high and low categories and in BCR-ABL1 and MYC categories, all known UCSC genes were extended by 10Kb and the γH2AX regions was then determined using the R package bamsignals v1.4.3. Regions were normalised to library size and background signal from the isotype control ChIP-seq libraries was subtracted.

6.11.16 4C analysis

Analysis was undertaken using the 4CSeq R package as described in (Klein et al., 2015). Reads were demultiplexed and the bait primer sequence was trimmed. The remaining reads were aligned to GRCh38/hg38 and reads at each *in silico* restriction fragment were counted. Variance stabilising transformation was used to avoid skewing of reads far from or near to the bait using the DESeq2 R package (Love et al., 2014).

Then, reads underwent trend fitting based on the decay of signals with genomic distance from the bait to a constant background using the 4CSeq R Package. Fragments with a median of less than 40 read counts across all samples were removed. To identify potential interactions, fragments with read

counts higher than expected at a given distance from the bait were assigned a z-score based on the residuals of the fit. Those fragments with z-scores larger than 3 in all replicates and with an adjusted p-value of less than 0.01 in at least one replicate were called as significant. Or, reads underwent differential region calling (p=0.05) conducted with DESeq2.

6.12 Reagents and solutions

6.12.1 Cell lysis buffer

Add the following to 1ml RIPA buffer (see section 6.12.11)

- 2.5µl 0.2M PMSF
- 1µl 1M DTT
- 20µl 50X protease inhibitor (Roche)
- 50µl 20X phosphatase inhibitor (Roche)
- Plus 10µl 100X deacetylase inhibitor (Santa Cruz Biotechnology) for the Western blot in Figure 4-11 only

6.12.2 10X Running buffer

Add the following to a total of 1 litre distilled water

- 24.2g Trizma, Tris base
- 144.1g Glycine
- 10g SDS

6.12.3 1X Running buffer

- Make fresh: dilute 100ml of 10X stock in 900ml distilled water

6.12.4 1X Transfer buffer

Add the following to a total of 1 litre distilled water. Keep cold.

- 3g Trizma, Tris base
- 14.4g Glycine
- 200ml methanol

6.12.5 10X Phosphate Buffered Saline (PBS) Solution

Add the following to 800ml distilled water and adjust to pH 7.4 (all from Sigma Aldrich):

- 80g NaCl
- 2g KCl
- 14.4g sodium phosphate dibasic
- 2.4g potassium phosphate dibasic

6.12.6 PBS-T

Add the following to 900ml distilled water:

- 100ml 10X PBS
- 1ml Tween20

6.12.7 Blocking solution

- Add 5g bovine serum albumin (Sigma Aldrich) to 100ml PBS-T

6.12.8 2.5M glycine solution

Dissolve 9.39g glycine in a total volume of 50ml PBS (both from Sigma Aldrich)

6.12.9 Concentrated RIPA buffer

Add the following to 37.3ml molecular grade water (all from Sigma Aldrich):

- 435µl 1M Tris-HCl pH7.6
- 100µl 0.5M EDTA
- 175µl 10% SDS
- 5ml 10% TritonX-100
- 500µl 10% Na-Dec

6.12.10 Tris/SDS sonication buffer

Add the following to 6.1ml molecular grade water (all from Sigma Aldrich)

- 325µl 10% SDS
- 65µl1M Tris-HCl pH7.6

6.12.11 RIPA buffer

Add the following to 43.4ml molecular grade water (all from Sigma Aldrich):

- 500µl 1M Tris-HCL pH7.6
- 100µl 0.5M EDTA
- 500µl 10% SDS
- 5ml 10% TritonX-100
- 500µl 10% Na-Dec

6.12.12 RIPA buffer with 0.3M NaCl

Add the following to 43.4ml molecular grade water (all from Sigma Aldrich):

- 500µl 1M Tris-HCl pH7.6
- 100µl 0.5M EDTA
- 500µl 10% SDS
- 5ml 10% TritonX-100
- 500µl 10% Na-Dec

- 0.88g NaCl

6.12.13 LiCl buffer

Add the following to 43.4ml molecular grade water (all from Sigma Aldrich):

- 1.56ml 8M LiCl
- 2.5ml 10% Nonidet P40
- 2.5ml 10% Na-Dec

6.12.14 3C lysis buffer

Prepare fresh each time. Add the following to 49.25ml molecular grade water and rotate at 4°C until dissolved. All from Sigma Aldrich unless otherwise stated:

- 500ul 1M Tris pH8
- 125µl 4M NaCl
- 100µl 100% NP-40
- 1 tablet complete protein inhibitor (Roche)

6.12.15 20x SSC stock solution

Add the following to 500ml distilled water

- Dissolve 132g SSC (Abbott Molecular)

6.12.16 0.4x SSC working solution

- Make up 10ml of 20X SSC stock solution to 500ml with distilled water.
- Adjust pH if necessary with hydrochloric acid or sodium hydroxide to pH7.0 ±0.2.

6.12.17 2x SSC working solution

- Make up 50ml of 20X SSC stock solution to 500ml with distilled water.
- Adjust pH if necessary with hydrochloric acid or sodium hydroxide to pH7.0 ±0.2.

6.12.18 2x SSC/Igepal working solution

- Make up 50ml of 20X SSC stock solution to 500ml with distilled water.
- Add 500ul Igepal (Sigma Aldrich) – pipette slowly as viscous
- Adjust pH if necessary with hydrochloric acid or sodium hydroxide to pH7.0 ±0.2.

6.13 Statistics

All statistical tests (outside of the bioinformatic methods detailed in section 6.11) were undertaken using GraphPad Prism software (GraphPad Prism version 7.0 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com).

The DNA FISH (in **Error! Reference source not found.** and Figure 2-47) comparisons were analysed using a paired Student's t-test +/- SEM. The comparison of H2AX regions to the published B-ALL gene defect datasets (in Table 1-1) was by two-by-two contingency table analysis, as described by (Swaminathan et al., 2015).

Scatterplots were analysed using Pearson's correlation shown in the formula below.

$$r = \frac{\sum (x - m_x)(y - m_y)}{\sqrt{\sum (x - m_x)^2 \sum (y - m_y)^2}}$$

Where x and y are two vectors of length n, and m_x and m_y are the means of x and y. The p-value of the correlation is determined by using the correlation coefficient table for the degrees of freedom (df=n-2) where n is the number of observations in the x and y variables.

7 References

Adam, R.C., Yang, H., Rockowitz, S., Larsen, S.B., Nikolova, M., Oristian, D.S., Polak, L., Kadaja, M., Asare, A., and Zheng, D. (2015). Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. Nature *521*, 366-370.

Agazie, Y.M., Burkholder, G.D., and Lee, J.S. (1996). Triplex DNA in the nucleus: Direct binding of triplex-specific antibodies and their effect on transcription, replication and cell growth. Biochem J *316*, 461-466.

Aguilera, A., and Gomez-Gonzalez, B. (2008). Genome instability: a mechanistic view of its causes and consequences. Nat Rev Genet *9*, 204-217.

Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., and Stratton, M.R. (2015). Clock-like mutational processes in human somatic cells. Nature Genet *47*, 1402-1407.

Alexandrov, L.B., and Stratton, M.R. (2014). Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. Current opinion in genetics & development *24*, 52-60.

Alter, B.P. (1996). Fanconi's anemia and malignancies. Am J Hematol *53*, 99-110.

Alvarnas, J.C., Brown, P.A., Aoun, P., Ballen, K.K., Barta, S.K., Borate, U., Boyer, M.W., Burke, P.W., Cassaday, R., Castro, J.E., Coccia, P.F., Coutre, S.E., Damon, L.E., DeAngelo, D.J., Douer, D., Frankfurt, O., Greer, J.P., Johnson, R.A., Kantarjian, H.M., Klisovic, R.B., Kupfer, G., Litzow, M., Liu, A., Rao, A.V., Shah, B., Uy, G.L., Wang, E.S., Zelenetz, A.D., Gregory, K., and Smith, C. (2015). Acute Lymphoblastic Leukemia, Version 2.2015. Journal of the National Comprehensive Cancer Network : JNCCN *13*, 1240-1279.

Anttila, V., Winsvold, B.S., Gormley, P., Kurth, T., Bettella, F., McMahon, G., Kallela, M., Malik, R., de Vries, B., Terwindt, G., Medland, S.E., Todt, U., McArdle, W.L., Quaye, L., Koiranen, M., Ikram, M.A., Lehtimaki, T., Stam, A.H., Ligthart, L., Wedenoja, J., Dunham, I., Neale, B.M., Palta, P., Hamalainen, E., Schurks, M., Rose, L.M., Buring, J.E., Ridker, P.M., Steinberg, S., Stefansson, H., Jakobsson, F., Lawlor, D.A., Evans, D.M., Ring, S.M., Farkkila, M., Artto, V., Kaunisto, M.A., Freilinger, T., Schoenen, J., Frants, R.R., Pelzer, N., Weller, C.M., Zielman, R., Heath, A.C., Madden, P.A.F., Montgomery, G.W., Martin, N.G., Borck, G., Gobel, H., Heinze, A., Heinze-Kuhn, K., Williams, F.M.K., Hartikainen, A.-L., Pouta, A., van den Ende, J., Uitterlinden, A.G., Hofman, A., Amin, N., Hottenga, J.-J., Vink, J.M., Heikkila, K., Alexander, M., Muller-Myhsok, B., Schreiber, S., Meitinger, T., Wichmann, H.E., Aromaa, A., Eriksson, J.G., Traynor, B.J., Trabzuni, D., North American Brain Expression, C., Consortium, U.K.B.E., Rossin, E., Lage, K., Jacobs, S.B.R., Gibbs, J.R., Birney, E., Kaprio, J., Penninx, B.W., Boomsma, D.I., van Duijn, C., Raitakari, O., Jarvelin, M.-R., Zwart, J.-A., Cherkas, L., Strachan, D.P., Kubisch, C., Ferrari, M.D., van den Maagdenberg, A.M.J.M., Dichgans, M., Wessman, M., Smith, G.D., Stefansson, K., Daly, M.J., and the International Headache Genetics, C. (2013). Genome-wide meta-analysis identifies new susceptibility loci for migraine. Nat Genet *45*, 912-917.

Aplan, P.D., Lombardi, D.P., Ginsberg, A.M., Cossman, J., Bertness, V.L., and Kirsch, I.R. (1990). DISRUPTION OF THE HUMAN SCL LOCUS BY ILLEGITIMATE V-(D)-J RECOMBINASE ACTIVITY. Science *250*, 1426-1429.

Appelbaum, F.R., Gundacker, H., Head, D.R., Slovak, M.L., Willman, C.L., Godwin, J.E., Anderson, J.E., and Petersdorf, S.H. (2006). Age and acute myeloid leukemia. Blood *107*, 3481-3485.

Arber, D.A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M.J., Le Beau, M.M., Bloomfield, C.D., Cazzola, M., and Vardiman, J.W. (2016). The 2016 revision to the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia. Blood, blood-2016-2003-643544.

Arico, M., Valsecchi, M.G., Camitta, B., Schrappe, M., Chessells, J., Baruchel, A., Gaynon, P., Silverman, L., Janka-Schaub, G., Kamps, W., Pui, C.H., and Masera, G. (2000). Outcome of treatment in children with Philadelphia chromosome-positive acute lymphoblastic leukemia. N Engl J Med *342*, 998-1006.

Aytekin, M., Vinatzer, U., Musteanu, M., Raynaud, S., and Wieser, R. (2005). Regulation of the expression of the oncogene EVI1 through the use of alternative mRNA 5 '-ends. Gene *356*, 160-168.

Balgobind, B.V., Lugthart, S., Hollink, I.H., Arentsen-Peters, S.T.J.C.M., van Wering, E.R., de Graaf, S.S.N., Reinhardt, D., Creutzig, U., Kaspers, G.J.L., de Bont, E.S.J.M., Stary, J., Trka, J., Zimmermann, M., Beverloo, H.B., Pieters, R., Delwel, R., Zwaan, C.M., and van den Heuvel-Eibrink, M.M. (2010). EVI1 overexpression in distinct subtypes of pediatric acute myeloid leukemia. Leukemia *24*, 942-949.

Banerji, J., Olson, L., and Schaffner, W. (1983). A LYMPHOCYTE-SPECIFIC CELLULAR ENHANCER IS LOCATED DOWNSTREAM OF THE JOINING REGION IN IMMUNOGLOBULIN HEAVY-CHAIN GENES. Cell *33*, 729-740.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). EXPRESSION OF A BETA-GLOBIN GENE IS ENHANCED BY REMOTE SV40 DNA-SEQUENCES. Cell *27*, 299-308.

Barjesteh van Waalwijk van Doorn-Khosrovani, S., Erpelinck, C., van Putten, W.L., Valk, P.J., van der Poel-van de Luytgaarde, S., Hack, R., Slater, R., Smit, E.M., Beverloo, H.B., Verhoef, G., Verdonck, L.F., Ossenkoppele, G.J., Sonneveld, P., de Greef, G.E., Lowenberg, B., and Delwel, R. (2003). High EVI1 expression predicts poor survival in acute myeloid leukemia: a study of 319 de novo AML patients. Blood *101*, 837-845.

Barlow, J.H., Faryabi, R.B., Callen, E., Wong, N., Malhowski, A., Chen, H.T., Gutierrez-Cruz, G., Sun, H.W., McKinnon, P., Wright, G., Casellas, R., Robbiani, D.F., Staudt, L., Fernandez-Capetillo, O., and Nussenzweig, A. (2013). Identification of early replicating fragile sites that contribute to genome instability. Cell *152*, 620-632.

Bartkova, J., Horejsi, Z., Koed, K., Kramer, A., Tort, F., Zieger, K., Guldberg, P., Sehested, M., Nesland, J.M., Lukas, C., Orntoft, T., Lukas, J., and Bartek, J. (2005). DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. Nature *434*, 864-870.

Bashey, A., Medina, B., Corringham, S., Pasek, M., Carrier, E., Vrooman, L., Lowy, I., Solomon, S.R., Morris, L.E., and Holland, H.K. (2009). CTLA4 blockade with ipilimumab to treat relapse of malignancy after allogeneic hematopoietic cell transplantation. Blood *113*, 1581-1588.

Bassan, R., and Hoelzer, D. (2011). Modern therapy of acute lymphoblastic leukemia. J Clin Oncol *29*, 532-543.

Bassan, R., Rossi, G., Pogliani, E.M., Di Bona, E., Angelucci, E., Cavattoni, I., Lambertenghi-Deliliers, G., Mannelli, F., Levis, A., Ciceri, F., Mattei, D., Borlenghi, E., Terruzzi, E., Borghero, C., Romani, C., Spinelli, O., Tosi, M., Oldani, E., Intermesoli, T., and Rambaldi, A. (2010). Chemotherapy-Phased Imatinib Pulses Improve Long-Term Outcome of Adult Patients With Philadelphia Chromosome-Positive Acute Lymphoblastic Leukemia: Northern Italy Leukemia Group Protocol 09/00. J Clin Oncol *28*, 3644-3652.

Baxevanis, A.D., and Landsman, D. (1996). Histone sequence database: A compilation of highly-conserved nucleoprotein sequences. Nucleic Acids Res *24*, 245-247.

Behrendt, H., Charrin, C., Gibbons, B., Harrison, C.J., Hawkins, J.M., Heerema, N.A., Horschler-Bötel, B., Huret, J.L., Laï, J.L., and Lampert, F. (1995). Dicentric (9;12) in acute lymphocytic leukemia and other hematological malignancies: report from a dic(9;12) study group. Leukemia *9*, 102-106.

Belton, J.M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. Methods (San Diego, Calif) *58*, 268-276.

Bennett, J., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D., Gralnick, H., and Sultan, C. (1991). Proposal for the recognition of minimally differentiated acute myeloid leukaemia (AML-MO). Br J Haematol *78*, 325-329.

Bennett, J.H. (1845). Case of Hypertrophy of the Spleen and Liver: In which Death Took Place from Suppuration of the Blood (Stark and Comp.).

Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R., and Sultan, C. (1976). Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. Br J Haematol *33*, 451-458.

Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A.G., Gralnick, H.R., and Sultan, C. (1982). PROPOSALS FOR THE CLASSIFICATION OF THE MYELODYSPLASTIC SYNDROMES. Br J Haematol *51*, 189-199.

Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A.G., Gralnick, H.R., and Sultan, C. (1985). CRITERIA FOR THE DIAGNOSIS OF ACUTE-LEUKEMIA OF MEGAKARYOCYTE LINEAGE (M7) - A REPORT OF THE FRENCH-AMERICAN-BRITISH COOPERATIVE GROUP. Annals of Internal Medicine *103*, 460-462.

Berger, M., Krebs, P., Crozat, K., Li, X., Croker, B.A., Siggs, O.M., Popkin, D., Du, X., Lawson, B.R., Theofilopoulos, A.N., Xia, Y., Khovananth, K., Moresco, E.M., Satoh, T., Takeuchi, O., Akira, S., and Beutler, B. (2010). An Slfn2 mutation causes lymphoid and myeloid immunodeficiency due to loss of immune cell quiescence. Nature immunology *11*, 335-343.

Berz, D., McCormack, E.M., Winer, E.S., Colvin, G.A., and Quesenberry, P.J. (2007). Cryopreservation of Hematopoietic Stem Cells. Am J Hematol *82*, 463-472.

Bindels, E.M.J., Havermans, M., Lugthart, S., Erpelinck, C., Wocjtowicz, E., Krivtsov, A.V., Rombouts, E., Armstrong, S.A., Taskesen, E., Haanstra, J.R., Beverloo, H.B., Döhner, H., Hudson, W.A., Kersey, J.H., Delwel, R., and Kumar, A.R. (2012). EVI1 is critical for the pathogenesis of a subset of MLL-AF9–rearranged AMLs. Blood *119*, 5838-5849.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & development *16*, 6-21.

Boseley, P., Moss, T., Mächler, M., Portmann, R., and Birnstiel, M. (1979). Sequence organization of the spacer DNA in a ribosomal gene unit of Xenopus laevis. Cell *17*, 19-31.

Bothmer, A., Robbiani, D.F., Di Virgilio, M., Bunting, S.F., Klein, I.A., Feldhahn, N., Barlow, J., Chen, H.T., Bosque, D., Callen, E., Nussenzweig, A., and Nussenzweig, M.C. (2011). Regulation of DNA end joining, resection, and immunoglobulin class switch recombination by 53BP1. Mol Cell *42*, 319-329.

Boulianne, B., Robinson, M.E., May, P.C., Castellano, L., Blighe, K., Thomas, J., Reid, A., Muschen, M., Apperley, J.F., Stebbing, J., and Feldhahn, N. (2017). Lineage-Specific Genes Are Prominent DNA Damage Hotspots during Leukemic Transformation of B Cell Precursors. Cell Rep *18*, 1687-1698.

Bouman, A., Knegt, L., Groschel, S., Erpelinck, C., Sanders, M., Delwel, R., Kuijpers, T., and Cobben, J.M. (2016). Congenital thrombocytopenia in a neonate with an interstitial microdeletion of 3q26.2q26.31. Am J Med Genet A *170a*, 504-509.

Boveri, T. (1914). Zur frage de Entstehung Maligner Tumoren.

Bower, H., Bjorkholm, M., Dickman, P.W., Hoglund, M., Lambert, P.C., and Andersson, T.M. (2016). Life Expectancy of Patients With Chronic Myeloid Leukemia Approaches the Life Expectancy of the General Population. J Clin Oncol *34*, 2851-2857.

Bradbury, E.M., Inglis, R.J., and Matthews, H.R. (1974). CONTROL OF CELL-DIVISION BY VERY LYSINE RICH HISTONE (F1) PHOSPHORYLATION. Nature *247*, 257-261.

Bransteitter, R., Pham, P., Scharff, M.D., and Goodman, M.F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proceedings of the National Academy of Sciences of the United States of America *100*, 4102-4107.

Breit, T.M., Mol, E.J., Wolverstettero, I.L.M., Ludwig, W.D., Vanwering, E.R., and Vandongen, J.J.M. (1993). SITE-SPECIFIC DELETIONS INVOLVING THE TAL-1 AND SIL GENES ARE RESTRICTED TO CELLS OF THE T-CELL RECEPTOR ALPHA/BETA LINEAGE - T-CELL RECEPTOR-DELTA GENE DELETION MECHANISM AFFECTS MULTIPLE GENES. Journal of Experimental Medicine *177*, 965-977.

Brooks, D.J., Woodward, S., Thompson, F.H., DosSantos, B., Russell, M., Yang, J.M., Guan, X.Y., Trent, J., Alberts, D.S., and Taetle, R. (1996). Expression of the zinc finger gene EVI-1 in ovarian and other cancers. British Journal of Cancer *74*, 1518-1525.

Brown, L., Cheng, J.T., Chen, Q., Siciliano, M.J., Crist, W., Buchanan, G., and Baer, R. (1990). SITE-SPECIFIC RECOMBINATION OF THE TAL-1 GENE IS A COMMON OCCURRENCE IN HUMAN T-CELL LEUKEMIA. Embo Journal *9*, 3343-3351.

Brownson, R.C., Novotny, T.E., and Perry, M.C. (1993). Cigarette smoking and adult leukemia: a meta-analysis. Archives of internal medicine *153*, 469-475.

Buonamici, S., Li, D.L., Chi, Y.Q., Zhao, R., Wang, X.R., Brace, L., Ni, H.Y., Saunthararajah, Y., and Nucifora, G. (2004). EVI1 induces myelodysplastic syndrome in mice. J Clin Invest *114*, 713-719.

Burnett, A.K., Hills, R.K., Milligan, D., Kjeldsen, L., Kell, J., Russell, N.H., Yin, J.A., Hunter, A., Goldstone, A.H., and Wheatley, K. (2010). Identification of patients with acute myeloblastic leukemia who benefit from the addition of gemtuzumab ozogamicin: results of the MRC AML15 trial. J Clin Oncol *29*, 369-377.

Cammenga, J., Mulloy, J.C., Berguido, F.J., MacGrogan, D., Viale, A., and Nimer, S.D. (2003). Induction of C/EBPalpha activity alters gene expression and differentiation of human CD34+ cells. Blood *101*, 2206-2214.

Canela, A., Maman, Y., Jung, S., Wong, N., Callen, E., Day, A., Kieffer-Kwon, K.-R., Pekowska, A., Zhang, H., Rao, S.S.P., Huang, S.-C., McKinnon, P.J., Aplan, P.D., Pommier, Y., Aiden, E.L., Casellas, R., and Nussenzweig, A. (2017). Genome Organization Drives Chromosome Fragility. Cell *170*, 507-+.

Canela, A., Sridharan, S., Sciascia, N., Tubbs, A., Meltzer, P., Sleckman, B.P., and Nussenzweig, A. (2016). DNA Breaks and End Resection Measured Genome-wide by End Sequencing. Mol Cell *63*, 898-911.

Cano, F., Drynan, L.F., Pannell, R., and Rabbitts, T.H. (2008). Leukaemia lineage specification caused by cell-specific Mll-Enl translocations. Oncogene *27*, 1945-1950.

Carroll, A.J., Raimondi, S.C., Williams, D.L., Behm, F.G., Borowitz, M., Castleberry, R.P., Harris, M.B., Patterson, R.B., Pullen, D.J., and Crist, W.M. (1987). TDIC(9-12) - A NONRANDOM CHROMOSOME ABNORMALITY IN CHILDHOOD B-CELL PRECURSOR ACUTE LYMPHOBLASTIC-LEUKEMIA - A PEDIATRIC ONCOLOGY GROUP-STUDY. Blood *70*, 1962-1965.

Cartwright, R.A., Gurney, K.A., and Moorman, A.V. (2002). Sex ratios and the risks of haematological malignancies. Br J Haematol *118*, 1071-1077.

Chan, L.N., Chen, Z., Braas, D., Lee, J.-W., Xiao, G., Geng, H., Cosgun, K.N., Hurtz, C., Shojaee, S., Cazzaniga, V., Schjerven, H., Ernst, T., Hochhaus, A., Kornblau, S.M., Konopleva, M., Pufall, M.A., Cazzaniga, G., Liu, G.J.,

Milne, T.A., Koeffler, H.P., Ross, T.S., Sánchez-García, I., Borkhardt, A., Yamamoto, K.R., Dickins, R.A., Graeber, T.G., and Müschen, M. (2017). Metabolic gatekeeper function of B-lymphoid transcription factors. Nature *542*, 479.

Chan, Y.A., Aristizabal, M.J., Lu, P.Y., Luo, Z., Hamza, A., and Kobor, M.S. (2014). Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. P Lo S Genet *10*, e1004288.

Chapman, J.R., Barral, P., Vannier, J.-B., Borel, V., Steger, M., Tomas-Loba, A., Sartori, A.A., Adams, I.R., Batista, F.D., and Boulton, S.J. (2013). RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. Molecular cell *49*, 858-871.

Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., and Alt, F.W. (2003). Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature *422*, 726-730.

Chen, H.T., Bhandoola, A., Difilippantonio, M.J., Zhu, J., Brown, M.J., Tai, X.G., Rogakou, E.P., Brotz, T.M., Bonner, W.M., Ried, T., and Nussenzweig, A. (2000). Response to RAG-mediated V(D)J cleavage by NBS1 and gamma-H2AX. Science *290*, 1962-1964.

Chen, Z., Shojaee, S., Buchner, M., Geng, H., Lee, J.W., Klemm, L., Titz, B., Graeber, T.G., Park, E., Tan, Y.X., Satterthwaite, A., Paietta, E., Hunger, S.P., Willman, C.L., Melnick, A., Loh, M.L., Jung, J.U., Coligan, J.E., Bolland, S., Mak, T.W., Limnander, A., Jumaa, H., Reth, M., Weiss, A., Lowell, C.A., and Muschen, M. (2015). Signalling thresholds and negative B-cell selection in acute lymphoblastic leukaemia. Nature *521*, 357-361.

Chizuka, A., Suda, M., Shibata, T., Kusumi, E., Hori, A., Hamaki, T., Kodama, Y., Horigome, K., Kishi, Y., Kobayashi, K., Matsumura, T., Yuji, K., Tanaka, Y., and Kami, M. (2006). Difference between hematological malignancy and Solid tumor research articles published in four major medical journals. Leukemia *20*, 1655.

Chou, S.H., Chin, K.H., and Wang, A.H.J. (2003). Unusual DNA duplex and hairpin motifs. Nucleic Acids Res *31*, 2461-2474.

Churchman, M.L., Low, J., Qu, C., Paietta, E.M., Kasper, L.H., Chang, Y., Payne-Turner, D., Althoff, M.J., Song, G., and Chen, S.-C. (2015). Efficacy of retinoids in IKZF1-mutated BCR-ABL1 acute lymphoblastic leukemia. Cancer cell *28*, 343-356.

Cinghu, S., Yang, P.Y., Kosak, J.P., Conway, A.E., Kumar, D., Oldfield, A.J., Adelman, K., and Jothi, R. (2017). Intragenic Enhancers Attenuate Host Gene Expression. Molecular Cell *68*, 104-+.

Clausen, B.E., Burkhardt, C., Reith, W., Renkawitz, R., and Forster, I. (1999). Conditional gene targeting in macrophages and granulocytes using LysMcre mice. Transgenic research *8*, 265-277.

Cleveland, J.L., Dean, M., Rosenberg, N., Wang, J.Y., and Rapp, U.R. (1989). Tyrosine kinase oncogenes abrogate interleukin-3 dependence of murine myeloid cells through signaling pathways involving c-myc: conditional regulation of c-myc transcription by temperature-sensitive v-abl. Molecular and Cellular Biology *9*, 5685-5695.

Collins, A.R. (1999). Oxidative DNA damage, antioxidants, and cancer. Bioessays *21*, 238-246.

Cortes, J.E., Kantarjian, H., Shah, N.P., Bixby, D., Mauro, M.J., Flinn, I., O'Hare, T., Hu, S., Narasimhan, N.I., Rivera, V.M., Clackson, T., Turner, C.D., Haluska, F.G., Druker, B.J., Deininger, M.W.N., and Talpaz, M. (2012). Ponatinib in Refractory Philadelphia Chromosome-Positive Leukemias. N Engl J Med *367*, 2075-2088.

Cortes, J.E., Kim, D.W., Pinilla-Ibarz, J., le Coutre, P., Paquette, R., Chuah, C., Nicolini, F.E., Apperley, J.F., Khoury, H.J., Talpaz, M., DiPersio, J., DeAngelo, D.J., Abruzzese, E., Rea, D., Baccarani, M., Mueller, M.C.,

Gambacorti-Passerini, C., Wong, S., Lustgarten, S., Rivera, V.M., Clackson, T., Turner, C.D., Haluska, F.G., Guilhot, F., Deininger, M.W., Hochhaus, A., Hughes, T., Goldman, J.M., Shah, N.P., Kantarjian, H., and Investigators, P. (2013). A Phase 2 Trial of Ponatinib in Philadelphia Chromosome-Positive Leukemias. N Engl J Med *369*, 1783-1796.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., Boyer, L.A., Young, R.A., and Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America *107*, 21931-21936.

Cuenco, G.M., and Ren, R.B. (2004). Both AML1 and EVI1 oncogenic components are required for the cooperation of AML1/MDS1/EVI1 with BCR/ABL in the induction of acute myelogenous leukemia in mice. Oncogene *23*, 569-579.

Cunniff, C., Bassetti, J.A., and Ellis, N.A. (2017). Bloom's syndrome: clinical spectrum, molecular pathogenesis, and cancer predisposition. Molecular syndromology *8*, 4-23.

d'Adda di Fagagna, F., Reaper, P.M., Clay-Farrace, L., Fiegler, H., Carr, P., Von Zglinicki, T., Saretzki, G., Carter, N.P., and Jackson, S.P. (2003). A DNA damage checkpoint response in telomere-initiated senescence. Nature *426*, 194-198.

D'Andrea, A.D., and Grompe, M. (2003). The Fanconi anaemia/BRCA pathway. Nature Reviews Cancer *3*, 23.

Daghistani, M., Marin, D., Khorashad, J.S., Wang, L.H., May, P.C., Paliompeis, C., Milojkovic, D., De Melo, V.A., Gerrard, G., Goldman, J.M., Apperley, J.F., Clark, R.E., Foroni, L., and Reid, A.G. (2010). EVI-1 oncogene expression predicts survival in chronic-phase CML patients resistant to imatinib treated with second-generation tyrosine kinase inhibitors. Blood *116*, 6014-6017.

Dang, C.V., McGuire, M., Buckmire, M., and Lee, W.M.F. (1989). Involvement of the 'leucine zipper' region in the oligomerization and transforming activity of human c-myc protein. Nature *337*, 664.

Deininger, M.W.N., Bose, S., Gora-Tybor, J., Yan, X.-H., Goldman, J.M., and Melo, J.V. (1998). Selective Induction of Leukemia-associated Fusion Genes by High-Dose Ionizing Radiation. Cancer Research *58*, 421-425.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.

Deklein, A., Hagemeijer, A., Bartram, C.R., Houwen, R., Hoefsloot, L., Carbonell, F., Chan, L., Barnett, M., Greaves, M., Kleihauer, E., Heisterkamp, N., Groffen, J., and Grosveld, G. (1986). BCR REARRANGEMENT AND TRANSLOCATION OF THE C-ABL ONCOGENE IN PHILADELPHIA POSITIVE ACUTE LYMPHOBLASTIC-LEUKEMIA. Blood *68*, 1369-1375.

Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastritis, E., Gilpatrick, T., Paranal, R.M., and Qi, J. (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Cell *146*, 904-917.

Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T.L., Thompson, J.R., Ingelsson, E., Saleheen, D., Erdmann, J., Goldstein, B.A., Stirrups, K., König, I.R., Cazier, J.B., Johansson, Å., Hall, A.S., Lee, J.Y., Willer, C.J., Chambers, J.C., Esko, T., Folkersen, L., Goel, A., Grundberg, E., Havulinna, A.S., Ho, W.K., Hopewell, J.C., Eriksson, N., Kleber, M.E., Kristiansson, K., Lundmark, P., Lyytikäinen, L.P., Rafelt, S., Shungin, D., Strawbridge, R.J., Thorleifsson, G., Tikkanen, E., Van Zuydam, N., Voight, B.F., Waite, L.L., Zhang, W., Ziegler, A., Absher, D., Altshuler, D., Balmforth, A.J., Barroso, I., Braund, P.S., Burgdorf, C., Claudi-Boehm, S., Cox, D., Dimitriou, M., Do, R., Doney, A.S.F., El Mokhtari, N., Eriksson, P., Fischer, K., Fontanillas, P., Franco-Cereceda, A., Gigante, B., Groop, L., Gustafsson, S., Hager, J., Hallmans, G., Han, B.G., Hunt, S.E., Kang, H.M., Illig, T., Kessler, T., Knowles,

J.W., Kolovou, G., Kuusisto, J., Langenberg, C., Langford, C., Leander, K., Lokki, M.L., Lundmark, A., McCarthy, M.I., Meisinger, C., Melander, O., Mihailov, E., Maouche, S., Morris, A.D., Müller-Nurasyid, M., Nikus, K., Peden, J.F., Rayner, N.W., Rasheed, A., Rosinger, S., Rubin, D., Rumpf, M.P., Schäfer, A., Sivananthan, M., Song, C., Stewart, A.F.R., Tan, S.T., Thorgeirsson, G., Van Der Schoot, C.E., Wagner, P.J., Wells, G.A., Wild, P.S., Yang, T.P., Amouyel, P., Arveiler, D., Basart, H., Boehnke, M., Boerwinkle, E., Brambilla, P., Cambien, F., Cupples, A.L., De Faire, U., Dehghan, A., Diemert, P., Epstein, S.E., Evans, A., Ferrario, M.M., Ferrières, J., Gauguier, D., Go, A.S., Goodall, A.H., Gudnason, V., Hazen, S.L., Holm, H., Iribarren, C., Jang, Y., Kähönen, M., Kee, F., Kim, H.S., Klopp, N., Koenig, W., Kratzer, W., Kuulasmaa, K., Laakso, M., Laaksonen, R., Lee, J.Y., Lind, L., Ouwehand, W.H., Parish, S., Park, J.E., Pedersen, N.L., Peters, A., Quertermous, T., Rader, D.J., Salomaa, V., Schadt, E., Shah, S.H., Sinisalo, J., Stark, K., Stefansson, K., Trégouët, D.A., Virtamo, J., Wallentin, L., Wareham, N., Zimmermann, M.E., Nieminen, M.S., Hengstenberg, C., Sandhu, M.S., Pastinen, T., Syvänen, A.C., Hovingh, G.K., Dedoussis, G., Franks, P.W., Lehtimäki, T., Metspalu, A., Zalloua, P.A., Siegbahn, A., Schreiber, S., Ripatti, S., Blankenberg, S.S., Perola, M., Clarke, R., Boehm, B.O., O'Donnell, C., Reilly, M.P., März, W., Collins, R., Kathiresan, S., Hamsten, A., Kooner, J.S., Thorsteinsdottir, U., Danesh, J., Palmer, C.N.A., Roberts, R., Watkins, H., Schunkert, H., and Samani, N.J. (2013). Large-scale association analysis identifies new risk loci for coronary artery disease. Nature Genet *45*, 25-33.

Dengler, M.A., Staiger, A.M., Gutekunst, M., Hofmann, U., Doszczak, M., Scheurich, P., Schwab, M., Aulitzky, W.E., and van der Kuip, H. (2011). Oncogenic stress induced by acute hyper-activation of Bcr-Abl leads to cell death upon induction of excessive aerobic glycolysis. PLoS One *6*, e25139.

Di Noia, J.M., and Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. Annual review of biochemistry *76*, 1-22.

Di Tullio, A., Manh, T.P.V., Schubert, A., Castellano, G., Månsson, R., and Graf, T. (2011). CCAAT/enhancer binding protein α (C/EBPα)-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. Proceedings of the National Academy of Sciences *108*, 17016-17021.

Dickstein, J., Senyuk, V., Premanand, K., Laricchia-Robbio, L., Xu, P., Cattaneo, F., Fazzina, R., and Nucifora, G. (2010). Methylation and silencing of miRNA-124 by EVI1 and self-renewal exhaustion of hematopoietic stem cells in murine myelodysplastic syndrome. Proceedings of the National Academy of Sciences of the United States of America *107*, 9783-9788.

Dimri, G.P., Lee, X., Basile, G., Acosta, M., Scott, G., Roskelley, C., Medrano, E.E., Linskens, M., Rubelj, I., Pereira-Smith, O., and et al. (1995). A biomarker that identifies senescent human cells in culture and in aging skin in vivo. Proc Natl Acad Sci U S A *92*, 9363-9367.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Dohner, H., Estey, E., Grimwade, D., Amadori, S., Appelbaum, F.R., Buchner, T., Dombret, H., Ebert, B.L., Fenaux, P., Larson, R.A., Levine, R.L., Lo-Coco, F., Naoe, T., Niederwieser, D., Ossenkoppele, G.J., Sanz, M., Sierra, J., Tallman, M.S., Tien, H.F., Wei, A.H., Lowenberg, B., and Bloomfield, C.D. (2017). Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood *129*, 424-447.

Drolet, M. (2006). Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. Mol Microbiol *59*, 723-730.

Druker, B.J., Talpaz, M., Resta, D.J., Peng, B., Buchdunger, E., Ford, J.M., Lydon, N.B., Kantarjian, H., Capdeville, R., Ohno-Jones, S., and Sawyers, C.L. (2001). Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med *344*, 1031-1037.

Druker, B.J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G.M., Fanning, S., Zimmermann, J., and Lydon, N.B. (1996). Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. Nature Medicine *2*, 561-566.

Ellis, L., Atadja, P.W., and Johnstone, R.W. (2009). Epigenetics in cancer: targeting chromatin modifications. Mol Cancer Ther *8*, 1409-1420.

Epstein, M.A., Achong, B.G., and Barr, Y.M. (1964). Virus particles in cultured lymphoblasts from Burkitt's lymphoma. The Lancet *283*, 702-703.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047-3048.

Fabarius, A., Leitner, A., Hochhaus, A., Müller, M.C., Hanfstein, B., Haferlach, C., Göhring, G., Schlegelberger, B., Jotterand, M., Reiter, A., Jung-Munkwitz, S., Proetel, U., Schwaab, J., Hofmann, W.-K., Schubert, J., Einsele, H., Ho, A.D., Falge, C., Kanz, L., Neubauer, A., Kneba, M., Stegelmann, F., Pfreundschuh, M., Waller, C.F., Spiekermann, K., Baerlocher, G.M., Lauseker, M., Pfirrmann, M., Hasford, J., Saussele, S., and Hehlmann, R. (2011). Impact of additional cytogenetic aberrations at diagnosis on prognosis of CML: long-term observation of 1151 patients from the randomized CML Study IV. Blood *118*, 6760-6768.

Fears, S., Mathieu, C., Zeleznik-Le, N., Huang, S., Rowley, J.D., and Nucifora, G. (1996). Intergenic splicing of MDS1 and EVI1 occurs in normal tissues as well as in myeloid leukemia and produces a new member of the PR domain family. Proc Natl Acad Sci U S A *93*, 1642-1647.

Feinberg, A.P., and Vogelstein, B. (1983a). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature *301*, 89.

Feinberg, A.P., and Vogelstein, B. (1983b). Hypomethylation of ras oncogenes in primary human cancers. Biochem Biophys Res Commun *111*, 47-54.

Feingold, K.R., Kazemi, M.R., Magra, A.L., McDonald, C.M., Chui, L.G., Shigenaga, J.K., Patzek, S.M., Chan, Z.W., Londos, C., and Grunfeld, C. (2010). ADRP/ADFP and Mal1 expression are increased in macrophages treated with TLR agonists. Atherosclerosis *209*, 81-88.

Feldhahn, N., Henke, N., Melchior, K., Duy, C., Soh, B.N., Klein, F., von Levetzow, G., Giebel, B., Li, A.H., Hofmann, W.K., Jumaa, H., and Muschen, M. (2007). Activation-induced cytidine deaminase acts as a mutator in BCR-ABL1-transformed acute lymphoblastic leukemia cells. Journal of Experimental Medicine *204*, 1157-1166.

Ferrando, A.A., Herblot, S., Palomero, T., Hansen, M., Hoang, T., Fox, E.A., and Look, A.T. (2004). Biallelic transcriptional activation of oncogenic transcription factors in T-cell acute lymphoblastic leukemia. Blood *103*, 1909-1911.

Fialkow, P.J. (1974). The origin and development of human tumors studied with cell markers. N Engl J Med *291*, 26-35.

Fialkow, P.J. (1978). Chronic myelocytic leukemia: Origin of some lymphocytes from leukemic stem cells. The Journal of clinical investigation *62*, 815-823.

Fielding, A.K., Richards, S.M., Chopra, R., Lazarus, H.M., Litzow, M.R., Buck, G., Durrant, I.J., Luger, S.M., Marks, D.I., Franklin, I.M., McMillan, A.K., Tallman, M.S., Rowe, J.M., and Goldstone, A.H. (2007). Outcome of 609 adults after relapse of acute lymphoblastic leukemia (ALL); an MRC UKALL12/ECOG 2993 study. Blood *109*, 944-950.

Fiskus, W., Pranpat, M., Bali, P., Balasis, M., Kumaraswamy, S., Boyapalle, S., Rocha, K., Wu, J., Giles, F., Manley, P.W., Atadja, P., and Bhalla, K. (2006). Combined effects of novel tyrosine kinase inhibitor AMN107 and histone deacetylase inhibitor LBH589 against Bcr-Abl-expressing human leukemia cells. Blood *108*, 645-652.

Fraga, M.F., and Esteller, M. (2002). DNA methylation: a profile of methods and applications. Biotechniques *33*, 632-649.

Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J., and Schatz, D.G. (2000). The rag proteins and V(D)J recombination: Complexes, ends, and transposition. Annual Review of Immunology *18*, 495-527.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., and Mei, P.H. (2009). An oestrogen-receptor-α-bound human chromatin interactome. Nature *462*, 58-64.

Funabiki, T., Kreider, B.L., and Ihle, J.N. (1994). THE CARBOXYL DOMAIN OF ZINC FINGERS OF THE EVI-1 MYELOID TRANSFORMING GENE BINDS A CONSENSUS SEQUENCE OF GAAGATGAG. Oncogene *9*, 1575-1581.

Galton, D., and Dacie, J. (1975). Classification of the acute leukaemias. In Unclassifiable Leukemias (Springer), pp. 17-24.

Gao, J.-S., Zhang, Y., Tang, X., Tucker, L.D., Tarwater, P.M., Quesenberry, P.J., Rigoutsos, I., and Ramratnam, B. (2011). The Evi1, microRNA-143, K-Ras axis in colon cancer. Febs Letters *585*, 693-699.

Garcia-Manero, G., Daniel, J., Smith, T.L., Kornblau, S.M., Lee, M.S., Kantarjian, H.M., and Issa, J.P.J. (2002). DNA methylation of multiple promoter-associated CpG islands in adult acute lymphocytic leukemia. Clin Cancer Res *8*, 2217-2224.

Garcia-Manero, G., Jeha, S., Daniel, J., Williamson, J., Albitar, M., Kantarjian, H.M., and Issa, J.P.J. (2003). Aberrant DNA methylation in pediatric patients with acute lymphocytic leukemia. Cancer *97*, 695-702.

Gel, B., Diez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., and Malinverni, R. (2016). regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics *32*, 289-291.

George, P., Bali, P., Annavarapu, S., Scuto, A., Fiskus, W., Guo, F., Sigua, C., Sondarva, G., Moscinski, L., Atadja, P., and Bhalla, K. (2005). Combination of the histone deacetylase inhibitor LBH589 and the hsp90 inhibitor 17-AAG is highly active against human CML-BC cells and AML cells with activating mutation of FLT-3. Blood *105*, 1768-1776.

Gerhardt, T.M., Schmahl, G.E., Flotho, C., Rath, A.V., and Niemeyer, C.M. (1997). Expression of the Evi-1 gene in haemopoietic cells of children with juvenile myelomonocytic leukaemia and normal donors. Br J Haematol *99*, 882-887.

German, J. (1993). Bloom syndrome: a mendelian prototype of somatic mutational disease. Medicine *72*, 393-406.

Giles, F.J., le Coutre, P.D., Pinilla-Ibarz, J., Larson, R.A., Gattermann, N., Ottmann, O.G., Hochhaus, A., Radich, J.P., Saglio, G., Hughes, T.P., Martinelli, G., Kim, D.W., Novick, S., Gillis, K., Fan, X., Cortes, J., Baccarani, M., and Kantarjian, H.M. (2013). Nilotinib in imatinib-resistant or imatinib-intolerant patients with chronic myeloid leukemia in chronic phase: 48-month follow-up results of a phase II study. Leukemia *27*, 107-112.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chedin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell *45*, 814-825.

Godley, L.A., and Shimamura, A. (2017). Genetic predisposition to hematologic malignancies: management and surveillance. Blood, blood-2017-2002-735290.

Goldman, J.M. (2010). Chronic Myeloid Leukemia: A Historical Perspective. Seminars in Hematology *47*, 302-311.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., and Caligiuri, M.A. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science *286*, 531-537.

Gomez-Benito, M., Conchillo, A., Garcia, M.A., Vazquez, I., Maicas, M., Vicente, C., Cristobal, I., Marcotegui, N., Garcia-Orti, L., Bandres, E., Calasanz, M.J., Alonso, M.M., and Odero, M.D. (2010). EVI1 controls proliferation in acute myeloid leukaemia through modulation of miR-1-2. British Journal of Cancer *103*, 1292-1296.

Gondor, A., Rougier, C., and Ohlsson, R. (2008). High-resolution circular chromosome conformation capture assay. Nat Protocols *3*, 303-313.

Goodspeed, A., Heiser, L.M., Gray, J.W., and Costello, J.C. (2016). Tumor-derived cell lines as molecular models of cancer pharmacogenomics. Mol Cancer Res *14*, 3-13.

Gorgoulis, V.G., Vassiliou, L.V., Karakaidos, P., Zacharatos, P., Kotsinas, A., Liloglou, T., Venere, M., Ditullio, R.A., Jr., Kastrinakis, N.G., Levy, B., Kletsas, D., Yoneta, A., Herlyn, M., Kittas, C., and Halazonetis, T.D. (2005). Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. Nature *434*, 907-913.

Goto, T., Macdonald, P., and Maniatis, T. (1989). Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. Cell *57*, 413-422.

Goyama, S., Yamamoto, G., Shimabe, M., Sato, T., Ichikawa, M., Ogawa, S., Chiba, S., and Kurokawa, M. (2008). Evi-1 is a critical regulator for hematopoietic stem cells and transformed leukemic cells. Cell Stem Cell *3*, 207-220.

Graham, S.M., Vass, J.K., Holyoake, T.L., and Graham, G.J. (2007). Transcriptional analysis of quiescent and proliferating CD34+ human hemopoietic cells from normal and chronic myeloid leukemia sources. Stem cells *25*, 3111-3120.

Greaves, M.F., and Wiemels, J. (2003). Origins of chromosome translocations in childhood leukaemia. Nature reviews Cancer *3*, 639-649.

Greenberg, P.L., Tuechler, H., Schanz, J., Sanz, G., Garcia-Manero, G., Solé, F., Bennett, J.M., Bowen, D., Fenaux, P., and Dreyfus, F. (2012). Revised international prognostic scoring system for myelodysplastic syndromes. Blood *120*, 2454-2465.

Gribble, S.M., Roberts, I., Grace, C., Andrews, K.M., Green, A.R., and Nacheva, E.P. (2000). Cytogenetics of the chronic myeloid leukemia-derived cell line K562: karyotype clarification by multicolor fluorescence in situ hybridization, comparative genomic hybridization, and locus-specific fluorescence in situ hybridization. Cancer Genet Cytogenet *118*, 1-8.

Groeschel, S., Lugthart, S., Schlenk, R.F., Valk, P.J.M., Eiwen, K., Goudswaard, C., van Putten, W.J.L., Kayser, S., Verdonck, L.F., Luebbert, M., Ossenkoppele, G.-J., Germing, U., Schmidt-Wolf, I., Schlegelberger, B., Krauter, J., Ganser, A., Doehner, H., Loewenberg, B., Doehner, K., and Delwel, R. (2010). High EVI1 Expression Predicts Outcome in Younger Adult Patients With Acute Myeloid Leukemia and Is Associated With Distinct Cytogenetic Abnormalities. J Clin Oncol *28*, 2101-2107.

Groeschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A.M., Erpelinck, C., van der Velden, V.H.J., Havermans, M., Avellino, R., van Lom, K., Rombouts, E.J., van Duin, M., Doehner, K., Beverloo, H.B., Bradner, J.E., Doehner, H., Lowenberg, B., Valk, P.J.M., Bindels, E.M.J., de Laat, W., and Delwel, R. (2014). A Single Oncogenic Enhancer Rearrangement Causes Concomitant EVI1 and GATA2 Deregulation in Leukemia. Cell *157*, 369-381.

Groffen, J., Stephenson, J.R., Heisterkamp, N., Deklein, A., Bartram, C.R., and Grosveld, G. (1984). PHILADELPHIA CHROMOSOMAL BREAKPOINTS ARE CLUSTERED WITHIN A LIMITED REGION, BCR, ON CHROMOSOME-22. Cell *36*, 93-99.

Grosso, J.F., and Jure-Kunkel, M.N. (2013). CTLA-4 blockade in tumor models: an overview of preclinical and translational research. Cancer Immunity Archive *13*, 5.

Gundem, G., Van Loo, P., Kremeyer, B., Alexandrov, L.B., Tubio, J.M., Papaemmanuil, E., Brewer, D.S., Kallio, H.M., Högnäs, G., and Annala, M. (2015). The evolutionary history of lethal metastatic prostate cancer. Nature *520*, 353.

Gupta, R.M., Hadaya, J., Trehan, A., Zekavat, S.M., Roselli, C., Klarin, D., Emdin, C.A., Hilvering, C.R.E., Bianchi, V., Mueller, C., Khera, A.V., Ryan, R.J.H., Engreitz, J.M., Issner, R., Shoresh, N., Epstein, C.B., de Laat, W., Brown, J.D., Schnabel, R.B., Bernstein, B.E., and Kathiresan, S. (2017). A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. Cell *170*, 522-533.e515.

Gurley, L.R., Danna, J.A., Barham, S.S., Deaven, L.L., and Tobey, R.A. (1978). HISTONE PHOSPHORYLATION AND CHROMATIN STRUCTURE DURING MITOSIS IN CHINESE-HAMSTER CELLS. Eur J Biochem *84*, 1-15.

Haagsman, H.P., and Diemel, R.V. (2001). Surfactant-associated proteins: functions and structural variation. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology *129*, 91-108.

Haas, K., Kundi, M., Sperr, W.R., Esterbauer, H., Ludwig, W.-D., Ratei, R., Koller, E., Gruener, H., Sauerland, C., Fonatsch, C., Valent, P., and Wieser, R. (2008). Expression and prognostic significance of different mRNA 5 ' -End variants of the oncogene EVII in 266 patients with de novo AML: EVII and MDSI/EVII overexpression both predict short remission duration. Gene Chromosomes Cancer *47*, 288-298.

Hahn, C.N., Chong, C.-E., Carmichael, C.L., Wilkins, E.J., Brautigan, P.J., Li, X.-C., Babic, M., Lin, M., Carmagnac, A., and Lee, Y.K. (2011). Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. Nature Genet *43*, 1012.

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

Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E., and Taipale, J. (2006). Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. Cell *124*, 47-59.

Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T., and Cimprich, K.A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. Cell *170*, 774-+.

Hansen, J.A., Gooley, T.A., Martin, P.J., Appelbaum, F., Chauncey, T.R., Clift, R.A., Petersdorf, E.W., Radich, J., Sanders, J.E., Storb, R.F., Sullivan, K.M., and Anasetti, C. (1998). Bone marrow transplants from unrelated donors for patients with chronic myeloid leukemia. N Engl J Med *338*, 962-968.

Hayflick, L., and Moorhead, P.S. (1961). SERIAL CULTIVATION OF HUMAN DIPLOID CELL STRAINS. Experimental Cell Research *25*, 585-&.

Heinaniemi, M., Vuorenmaa, T., Teppo, S., Kaikkonen, M.U., Bouvy-Liivrand, M., Mehtonen, J., Niskanen, H., Zachariadis, V., Laukkanen, S., Liuksiala, T., Teittinen, K., and Lohi, O. (2016). Transcription-coupled genetic instability marks acute lymphoblastic leukemia structural variation hotspots. Elife *5*, 1-26.

Heintz, N. (1991). THE REGULATION OF HISTONE GENE-EXPRESSION DURING THE CELL-CYCLE. Biochimica Et Biophysica Acta *1088*, 327-339.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., Ching, K.A., Antosiewicz-Bourget, J.E., Liu, H., Zhang, X.M., Green, R.D., Lobanenkov, V.V., Stewart, R., Thomson, J.A., Crawford, G.E., Kellis, M., and Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y.T., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C.X., Ching, K.A., Wang, W., Weng, Z.P., Green, R.D., Crawford, G.E., and Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genet *39*, 311-318.

Helgason, G.V., Young, G.A.R., and Holyoake, T.L. (2010). Targeting Chronic Myeloid Leukemia Stem Cells. Current Hematologic Malignancy Reports *5*, 81-87.

Henderson, C.A., Jr. (2003). Imatinib: the promise of a "magic bullet" for cancer fulfilled. Journal of the Medical Association of Georgia *92*, 12-14, 22.

Herman, R.K., and Dworkin, N.B. (1971). Effect of gene induction on the rate of mutagenesis by ICR-191 in Escherichia coli. J Bacteriol *106*, 543-550.

Herranz, D., Ambesi-Impiombato, A., Palomero, T., Schnell, S.A., Belver, L., Wendorff, A.A., Xu, L., Castillo-Martin, M., Llobet-Navás, D., and Cordon-Cardo, C. (2014). A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. Nature medicine *20*, 1130-1137.

Hobson, D.J., Wei, W., Steinmetz, L.M., and Svejstrup, J.Q. (2012). RNA polymerase II collision interrupts convergent transcription. Mol Cell *48*, 365-374.

Hochhaus, A., Baccarani, M., Deininger, M., Apperley, J.F., Lipton, J.H., Goldberg, S.L., Corm, S., Shah, N.P., Cervantes, F., Silver, R.T., Niederwieser, D., Stone, R.M., Dombret, H., Larson, R.A., Roy, L., Hughes, T., Mueller, M.C., Ezzeddine, R., Countouriotis, A.M., and Kantarjian, H.M. (2008). Dasatinib induces durable cytogenetic responses in patients with chronic myelogenous leukemia in chronic phase with resistance or intolerance to imatinib. Leukemia *22*, 1200-1206.

Hödl, M., and Basler, K. (2012). Transcription in the absence of histone H3. 2 and H3K4 methylation. Current Biology *22*, 2253-2257.

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. Nature Genet *45*, 242-252.

Hou, A.W., Zhao, L.Z., Zhao, F.Z., Wang, W.L., Niu, J.Y., Li, B.X., Zhou, Z.J., and Zhu, D.Y. (2016). Expression of MECOM is associated with unfavorable prognosis in glioblastoma multiforme. Oncotargets and Therapy *9*, 315-320.

Hoyt, P.R., Bartholomew, C., Davis, A.J., Yutzey, K., Gamer, L.W., Potter, S.S., Ihle, J.N., and Mucenski, M.L. (1997). The Evil proto-oncogene is required at midgestation for neural, heart, and paraxial mesenchyme development. Mechanisms of Development *65*, 55-70.

Hübner, S., Cazzaniga, G., Flohr, T., Van der Velden, V., Konrad, M., Pötschger, U., Basso, G., Schrappe, M., van Dongen, J., and Bartram, C. (2004). High incidence and unique features of antigen receptor gene rearrangements in TEL–AML1-positive leukemias. Leukemia *18*, 84.

Huppert, J.L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. Nucleic Acids Res *33*, 2908-2916.

Hussain, S.P., Hofseth, L.J., and Harris, C.C. (2003). Radical causes of cancer. Nature reviews Cancer *3*, 276-285.

Huvet, M., Nicolay, S., Touchon, M., Audit, B., d'Aubenton-Carafa, Y., Arneodo, A., and Thermes, C. (2007). Human gene organization driven by the coordination of replication and transcription. Genome Res *17*, 1278-1285.

Igwe, I.J., Yang, D.Y., Merchant, A., Merin, N., Yaghmour, G., Kelly, K., and Ramsingh, G. (2017). The presence of Philadelphia chromosome does not confer poor prognosis in adult pre-B acute lymphoblastic leukaemia in the tyrosine kinase inhibitor era - a surveillance, epidemiology, and end results database analysis. Br J Haematol *179*, 618-626.

Imielinski, M., Guo, G., and Meyerson, M. Insertions and Deletions Target Lineage-Defining Genes in Human Cancers. Cell *168*, 460-472.e414.

Imielinski, M., Guo, G., and Meyerson, M. (2017). Insertions and Deletions Target Lineage-Defining Genes in Human Cancers. Cell *168*, 460-472.e414.

International Standing Committee on Human Cytogenetic, N., Shaffer, L.G., Slovak, M.L., and Campbell, L.J. (2009). ISCN 2009 : an international system for human cytogenetic nomenclature (2009) (Basel; Unionville, CT: Karger).

Jabbour, E., Kantarjian, H., Ravandi, F., Thomas, D., Huang, X., Faderl, S., Pemmaraju, N., Daver, N., Garcia-Manero, G., Sasaki, K., Cortes, J., Garris, R., Yin, C.C., Khoury, J.D., Jorgensen, J., Estrov, Z., Bohannan, Z., Konopleva, M., Kadia, T., Jain, N., DiNardo, C., Wierda, W., Jeanis, V., and O'Brien, S. (2015). Combination of hyper-CVAD with ponatinib as first-line therapy for patients with Philadelphia chromosome-positive acute lymphoblastic leukaemia: a single-centre, phase 2 study. The lancet oncology *16*, 1547-1555.

Jabbour, E., Kantarjian, H.M., Saglio, G., Luis Steegmann, J., Shah, N.P., Boque, C., Chuah, C., Pavlovsky, C., Mayer, J., Cortes, J., Baccarani, M., Kim, D.-W., Bradley-Garelik, M.B., Mohamed, H., Wildgust, M., and Hochhaus, A. (2014). Early response with dasatinib or imatinib in chronic myeloid leukemia: 3-year follow-up from a randomized phase 3 trial (DASISION). Blood *123*, 494-500.

Jabbour, E.J., Faderl, S., and Kantarjian, H.M. (2005). Adult acute lymphoblastic leukemia. Mayo Clinic proceedings *80*, 1517-1527.

Jaffe, E., Swerdlow, S.H.C.E., Campo, E., Pileri, S., Thiele, J., Lee Harris, N., Ht, S., and Wardiman, J.W. (2008). WHO Classification of Tumours of the Haematopoietic and Lymphoid Tissues, Vol 2.

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., and Chavez, A. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med *371*, 2488-2498.

Jamieson, C.H., Barroga, C.F., and Vainchenker, W.P. (2008). Miscreant myeloproliferative disorder stem cells. Leukemia *22*, 2011-2019.

Jenjaroenpun, P., Wongsurawat, T., Yenamandra, S.P., and Kuznetsov, V.A. (2015). QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. Nucleic Acids Res *43*, W527-W534.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jin, G., Yamazaki, Y., Takuwa, M., Takahara, T., Kaneko, K., Kuwata, T., Miyata, S., and Nakamura, T. (2007). Trib1 and Evi1 cooperate with Hoxa and Meis1 in myeloid leukemogenesis. Blood *109*, 3998-4005.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Johnstone, R.W. (2002). Histone-deacetylase inhibitors: novel drugs for the treatment of cancer. Nature reviews Drug discovery *1*, 287-299.

Jones, R.M., Mortusewicz, O., Afzal, I., Lorvellec, M., Garcia, P., Helleday, T., and Petermann, E. (2013). Increased replication initiation and conflicts with transcription underlie Cyclin E-induced replication stress. Oncogene *32*, 3744-3753.

Kantarjian, H., O'brien, S., Ravandi, F., Cortes, J., Shan, J., Bennett, J.M., List, A., Fenaux, P., Sanz, G., and Issa, J.P. (2008). Proposal for a new risk model in myelodysplastic syndrome that accounts for events not considered in the original International Prognostic Scoring System. Cancer *113*, 1351-1361.

Kataoka, K., Sato, T., Yoshimi, A., Goyama, S., Tsuruta, T., Kobayashi, H., Shimabe, M., Arai, S., Nakagawa, M., Imai, Y., Kumano, K., Kumagai, K., Kubota, N., Kadowaki, T., and Kurokawa, M. (2011). Evi1 is essential for hematopoietic stem cell self-renewal, and its expression marks hematopoietic cells with long-term multilineage repopulating activity. Journal of Experimental Medicine *208*, 2402-2415.

Katayama, S., Suzuki, M., Yamaoka, A., Keleku-Lukwete, N., Katsuoka, F., Otsuki, A., Kure, S., Engel, J.D., and Yamamoto, M. (2017). GATA2 haploinsufficiency accelerates EVI1-driven leukemogenesis. Blood *130*, 908-919.

Katerndahl, C.D.S., Heltemes-Harris, L.M., Willette, M.J.L., Henzler, C.M., Frietze, S., Yang, R., Schjerven, H., Silverstein, K.A.T., Ramsey, L.B., Hubbard, G., Wells, A.D., Kuiper, R.P., Scheijen, B., van Leeuwen, F.N., Muschen, M., Kornblau, S.M., and Farrar, M.A. (2017). Antagonism of B cell enhancer networks by STAT5 drives leukemia and poor patient survival. Nature immunology.

Kenderian, S., Ruella, M., Shestova, O., Klichinsky, M., Aikawa, V., Morrissette, J., Scholler, J., Song, D., Porter, D., and Carroll, M. (2015). CD33-specific chimeric antigen receptor T cells exhibit potent preclinical activity against human acute myeloid leukemia. Leukemia *29*, 1637-1647.

Kerr, J.F., Wyllie, A.H., and Currie, A.R. (1972). Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. British journal of cancer *26*, 239.

Kim, N., and Jinks-Robertson, S. (2012). Transcription as a source of genome instability. Nat Rev Genet *13*, 204-214.

Kim, Y., Schulz, V.P., Satake, N., Gruber, T.A., Teixeira, A.M., Halene, S., Gallagher, P.G., and Krause, D.S. (2014). Whole-exome sequencing identifies a novel somatic mutation in MMP8 associated with a t(1;22)-acute megakaryoblastic leukemia. Leukemia *28*, 945-948.

Klein, F.A., Pakozdi, T., Anders, S., Ghavi-Helm, Y., Furlong, E.E., and Huber, W. (2015). FourCSeq: analysis of 4C sequencing data. Bioinformatics *31*, 3085-3091.

Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., Casellas, R., and Nussenzweig, M.C. (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell *147*, 95-106.

Klemm, L., Duy, C., Iacobucci, I., Kuchen, S., von Levetzow, G., Feldhahn, N., Henke, N., Li, Z.Y., Hoffmann, T.K., Kim, Y.M., Hofmann, W.K., Jumaa, H., Groffen, J., Heisterkamp, N., Martinelli, G., Lieber, M.R., Casellas, R., and Muschen, M. (2009). The B Cell Mutator AID Promotes B Lymphoid Blast Crisis and Drug Resistance in Chronic Myeloid Leukemia. Cancer Cell *16*, 232-245.

Knudson, A.G. (1996). Hereditary cancer: Two hits revisited. Journal of Cancer Research and Clinical Oncology *122*, 135-140.

Komori, T., Okada, A., Stewart, V., and Alt, F.W. (1993). LACK OF N-REGIONS IN ANTIGEN RECEPTOR VARIABLE REGION GENES OF TDT-DEFICIENT LYMPHOCYTES. Science *261*, 1171-1175.

Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. Bioinformatics *23*, 1289-1291.

Kotsantis, P., Silva, L.M., Irmscher, S., Jones, R.M., Folkes, L., Gromak, N., and Petermann, E. (2016). Increased global transcription activity as a mechanism of replication stress in cancer. Nat Commun *7*, 13087.

Kreider, B.L., Orkin, S.H., and Ihle, J.N. (1993). LOSS OF ERYTHROPOIETIN RESPONSIVENESS IN ERYTHROID PROGENITORS DUE TO EXPRESSION OF THE EVI-1 MYELOID-TRANSFORMING GENE. Proceedings of the National Academy of Sciences of the United States of America *90*, 6454-6458.

Krummel, M.F., and Allison, J.P. (1995). CD28 and CTLA-4 have opposing effects on the response of T cells to stimulation. Journal of Experimental Medicine *182*, 459-465.

Kuang, S.Q., Tong, W.G., Yang, H., Lin, W., Lee, M.K., Fang, Z.H., Wei, Y., Jelinek, J., Issa, J.P., and Garcia-Manero, G. (2008). Genome-wide identification of aberrantly methylated promoter associated CpG islands in acute lymphocytic leukemia. Leukemia *22*, 1529-1538.

Kubonishi, I., and Miyoshi, I. (1983). Establishment of a Ph1 chromosome-positive cell line from chronic myelogenous leukemia in blast crisis. International journal of cell cloning *1*, 105-117.

Kuo, M.-H., and Allis, C.D. (1999). In vivo cross-linking and immunoprecipitation for studying dynamic protein: DNA associations in a chromatin environment. Methods (San Diego, Calif) *19*, 425-433.

Kurahashi, S., Hayakawa, F., Miyata, Y., Yasuda, T., Minami, Y., Tsuzuki, S., Abe, A., and Naoe, T. (2011). PAX5-PML acts as a dual dominant-negative form of both PAX5 and PML. Oncogene *30*, 1822-1830.

Kurokawa, M., Mitani, K., Imai, Y., Ogawa, S., Yazaki, Y., and Hirai, H. (1998). The t(3;21) fusion product, AML1/Evi-1, interacts with Smad3 and blocks transforming growth factor-beta-mediated growth inhibition of myeloid cells. Blood *92*, 4003-4012.

Laiosa, C.V., Stadtfeld, M., and Graf, T. (2006). Determinants of lymphoid-myeloid lineage diversification. Annual Review of Immunology *24*, 705-738.

Laricchia-Robbio, L., Fazzina, R., Li, D., Rinaldi, C.R., Sinha, K.K., Chakraborty, S., and Nucifora, G. (2006). Point mutations in two EVI1 Zn fingers abolish EVI1-GATA1 interaction and allow erythroid differentiation of murine bone marrow cells. Molecular and Cellular Biology *26*, 7658-7666.

Laricchia-Robbio, L., and Nucifora, G. (2008). Significant increase of self-renewal in hematopoietic cells after forced expression of EVI1. Blood Cells Molecules and Diseases *40*, 141-147.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc *8*, 2180-2196.

Laurent, S., Palmisano, G.L., Martelli, A.M., Kato, T., Tazzari, P.L., Pierri, I., Clavio, M., Dozin, B., Balbi, G., Megna, M., Morabito, A., Lamparelli, T., Bacigalupo, A., Gobbi, M., and Pistillo, M.P. (2007). CTLA-4 expressed by chemoresistant, as well as untreated, myeloid leukaemia cells can be targeted with ligands to induce apoptosis. Br J Haematol *136*, 597-608.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS computational biology *9*, e1003118.

Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. Nature *396*, 643-649.

Lensing, S.V., Marsico, G., Hansel-Hertsch, R., Lam, E.Y., Tannahill, D., and Balasubramanian, S. (2016). DSBCapture: in situ capture and sequencing of DNA breaks. Nat Methods *13*, 855-857.

Leroyviard, K., Vinit, M.A., Lecointe, N., Mathieumahul, D., and Romeo, P.H. (1994). DISTINCT DNASE-I HYPERSENSITIVE SITES ARE ASSOCIATED WITH TAL-1 TRANSCRIPTION IN ERYTHROID AND T-CELL LINES. Blood *84*, 3819-3827.

Lesnik, E.A., and Freier, S.M. (1995). Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. Biochemistry *34*, 10807-10815.

Ley, T.J., Mardis, E.R., Ding, L., Fulton, B., McLellan, M.D., Chen, K., Dooling, D., Dunford-Shore, B.H., McGrath, S., and Hickenbotham, M. (2008). DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature *456*, 66.

Li, G., Fullwood, M.J., Xu, H., Mulawadi, F.H., Velkov, S., Vega, V., Ariyaratne, P.N., Bin Mohamed, Y., Ooi, H.S., Tennakoon, C., Wei, C.L., Ruan, Y.J., and Sung, W.K. (2010). ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biology *11*, 13.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, X., Luo, O.J., Wang, P., Zheng, M., Wang, D., Piecuch, E., Zhu, J.J., Tian, S.Z., Tang, Z., Li, G., and Ruan, Y. (2017). Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. Nat Protocols *12*, 899-915.

Liew, E., and Owen, C. (2011). Familial myelodysplastic syndromes: a review of the literature. Haematologica *96*, 1536-1542.

Lin, C.Y., Erkek, S., Tong, Y., Yin, L., Federation, A.J., Zapatka, M., Haldipur, P., Kawauchi, D., Risch, T., and Warnatz, H.-J. (2016). Active medulloblastoma enhancers reveal subgroup-specific cellular origins. Nature *530*, 57-62.

Linsley, P.S., Greene, J., Tan, P., Bradshaw, J., Ledbetter, J.A., Anasetti, C., and Damle, N.K. (1992). Coexpression and functional cooperation of CTLA-4 and CD28 on activated T lymphocytes. Journal of Experimental Medicine *176*, 1595-1604.

Lo-Coco, F., Avvisati, G., Vignetti, M., Thiede, C., Orlando, S.M., Iacobelli, S., Ferrara, F., Fazi, P., Cicconi, L., and Di Bona, E. (2013). Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. N Engl J Med *369*, 111-121.

Lombardo, L.J., Lee, F.Y., Chen, P., Norris, D., Barrish, J.C., Behnia, K., Castaneda, S., Cornelius, L.A.M., Das, J., Doweyko, A.M., Fairchild, C., Hunt, J.T., Inigo, I., Johnston, K., Kamath, A., Kan, D., Klei, H., Marathe, P., Pang, S.H., Peterson, R., Pitt, S., Schieven, G.L., Schmidt, R.J., Tokarski, J., Wen, M.L., Wityak, J., and Borzilleri, R.M. (2004). Discovery of N-(2-chloro-6-methylphenyl)-2-(6-(4-(2-hydroxyethyl)-piperazin-1-yl)-2-m ethylpyrimidin-4-ylamino)thiazole-5-carboxamide (BMS-354825), a dual Src/Abl kinase inhibitor with potent antitumor activity in preclinical assays. Journal of Medicinal Chemistry *47*, 6658-6661.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology *15*, 550.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997a). Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. Nature *389*, 251-260.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997b). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251.

Lugthart, S., van Drunen, E., van Norden, Y., van Hoven, A., Erpelinck, C.A.J., Valk, P.J.M., Beverloo, H.B., Lowenberg, B., and Delwel, R. (2008). High EVI1 levels predict adverse outcome in acute myeloid leukemia: prevalence of EVI1 overexpression. and chromosome 3q26 abnormalities underestimated. Blood *111*, 4329-4337.

Ma, N., and McAllister, W.T. (2009). In a Head-on Collision, Two RNA Polymerases Approaching One Another on the Same DNA May Pass by One Another. J Mol Biol *391*, 808-812.

Mah, L.J., El-Osta, A., and Karagiannis, T.C. (2010). gamma H2AX: a sensitive molecular marker of DNA damage and repair. Leukemia *24*, 679-686.

Mahadevaiah, S.K., Turner, J.M.A., Baudat, F., Rogakou, E.P., de Boer, P., Blanco-Rodriguez, J., Jasin, M., Keeney, S., Bonner, W.M., and Burgoyne, P.S. (2001). Recombinational DNA double-strand breaks in mice precede synapsis. Nature Genet *27*, 271-276.

Mahieux, R., and Gessain, A. (2003). HTLV-1 and associated adult T-cell leukemia/lymphoma. Reviews in clinical and experimental hematology *7*, 336-361.

Maicas, M., Vazquez, I., Alis, R., Marcotegui, N., Urquiza, L., Cortes-Lavaud, X., Cristobal, I., Garcia-Sanchez, M.A., and Odero, M.D. (2017). The MDS and EVI1 complex locus (MECOM) isoforms regulate their own transcription and have different roles in the transformation of hematopoietic stem and progenitor cells. Biochim Biophys Acta *1860*, 721-729.

Maicas, M., Vazquez, I., Cortes-Layaud, X., Garcia-Sanchez, M.A., Marcotegui, N., Urquiza, L., Cristobal, I., Calasanz, M.J., and Odero, M.D. (2012). The EVI1 human protein regulates its own transcription. Role of the different isoforms. Cancer Research *72*.

Maicas, M., Vazquez, I., Vicente, C., Garcia-Sanchez, M.A., Marcotegui, N., Urquiza, L., Calasanz, M.J., and Odero, M.D. (2013). Functional characterization of the promoter region of the human EVI1 gene in acute myeloid leukemia: RUNX1 and ELK1 directly regulate its transcription. Oncogene *32*, 2069-2078.

Malcovati, L., Della Porta, M.G., Strupp, C., Ambaglio, I., Kuendgen, A., Nachtkamp, K., Travaglino, E., Invernizzi, R., Pascutto, C., and Lazzarino, M. (2011). Impact of the degree of anemia on the outcome of patients with myelodysplastic syndrome and its integration into the WHO classification-based Prognostic Scoring System (WPSS). haematologica *96*, 1433-1440.

Manachai, N., Saito, Y., Nakahata, S., Bahirvani, A.G., Osato, M., and Morishita, K. (2017). Activation of EVI1 transcription by the LEF1/β-catenin complex with p53-alteration in myeloid blast crisis of chronic myeloid leukemia. Biochem Biophys Res Commun *482*, 994-1000.

Mansour, M.R., Abraham, B.J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A.D., Etchin, J., Lawton, L., Sallan, S.E., Silverman, L.B., Loh, M.L., Hunger, S.P., Sanda, T., Young, R.A., and Look, A.T. (2014). Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science *346*, 1373-1377.

Maru, Y., and Witte, O.N. (1991). THE BCR GENE ENCODES A NOVEL SERINE THREONINE KINASE-ACTIVITY WITHIN A SINGLE EXON. Cell *67*, 459-468.

Mathey-Prevot, B., Nabel, G., Palacios, R., and Baltimore, D. (1986). Abelson virus abrogation of interleukin-3 dependence in a lymphoid cell line. Mol Cell Biol *6*, 4133-4135.

Maude, S.L., Frey, N., Shaw, P.A., Aplenc, R., Barrett, D.M., Bunin, N.J., Chew, A., Gonzalez, V.E., Zheng, Z., and Lacey, S.F. (2014). Chimeric antigen receptor T cells for sustained remissions in leukemia. N Engl J Med *371*, 1507-1517.

McGhee, J.D., and Felsenfeld, G. (1980). Nucleosome structure. Annual review of biochemistry *49*, 1115-1156.

McHale, C.M., Zhang, L., and Smith, M.T. (2011). Current understanding of the mechanism of benzene-induced leukemia in humans: implications for risk assessment. Carcinogenesis *33*, 240-252.

McNally, R.J.Q., Roman, E., and Cartwright, R.A. (1999). Leukemias and lymphomas: time trends in the UK, 1984–93. Cancer Causes & Control *10*, 35-42.

Meng, F.-L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.-R., Meyers, Robin M., Amor, C., Wasserman, Caitlyn R., Neuberg, D., Casellas, R., Nussenzweig, Michel C., Bradner, James E., Liu, X.S., and Alt, Frederick W. (2014). Convergent Transcription at Intragenic Super-Enhancers Targets AID-Initiated Genomic Instability. Cell *159*, 1538-1548.

Miller, R.A., Maloney, D.G., Warnke, R., and Levy, R. (1982). Treatment of B-cell lymphoma with monoclonal anti-idiotype antibody. N Engl J Med *306*, 517-522.

Mirro, J., Kitchingman, G., Williams, D., Lauzon, G.J., Lin, C.C., Callihan, T., and Zipf, T.F. (1986a). CLINICAL AND LABORATORY CHARACTERISTICS OF ACUTE-LEUKEMIA WITH THE 4-11 TRANSLOCATION. Blood *67*, 689-697.

Mirro, J., Kitchingman, G.R., Williams, D.L., Murphy, S.B., Zipf, T.F., and Stass, S.A. (1986b). MIXED LINEAGE LEUKEMIA - THE IMPLICATIONS FOR HEMATOPOIETIC DIFFERENTIATION. Blood *68*, 597-598.

Mirro, J., Zipf, T.F., Pui, C.H., Kitchingman, G., Williams, D., Melvin, S., Murphy, S.B., and Stass, S. (1985). ACUTE MIXED LINEAGE LEUKEMIA - CLINICOPATHOLOGIC CORRELATIONS AND PROGNOSTIC-SIGNIFICANCE. Blood *66*, 1115-1123.

Mitani, K., Ogawa, S., Tanaka, T., Miyoshi, H., Kurokawa, M., Mano, H., Yazaki, Y., Ohki, M., and Hirai, H. (1994). GENERATION OF THE AML1-EVI-1 FUSION GENE IN THE T(3 21)(Q26 Q22) CAUSES BLASTIC CRISIS IN CHRONIC MYELOCYTIC-LEUKEMIA. Embo Journal *13*, 504-510.

Mitelman, F., Johansson, B., and Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. Nature Reviews Cancer *7*, 233-245.

Miyoshi, H., Shimizu, K., Kozu, T., Maseki, N., Kaneko, Y., and Ohki, M. (1991). T(8-21) BREAKPOINTS ON CHROMOSOME-21 IN ACUTE MYELOID-LEUKEMIA ARE CLUSTERED WITHIN A LIMITED REGION OF A SINGLE GENE, AML1. Proceedings of the National Academy of Sciences of the United States of America *88*, 10431-10434.

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. (2007). 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*, 3189-3197.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M., and Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res *9*, 6047-6068.

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. Proceedings of the National Academy of Sciences *99*, 8242-8247.

Morishita, K., Parganas, E., Bartholomew, C., Sacchi, N., Valentine, M.B., Raimondi, S.C., Lebeau, M.M., and Ihle, J.N. (1990a). THE HUMAN EVI-1 GENE IS LOCATED ON CHROMOSOME 3Q24-Q28 BUT IS NOT REARRANGED IN 3 CASES OF ACUTE NONLYMPHOCYTIC LEUKEMIAS CONTAINING T(3-5)(Q25-Q34) TRANSLOCATIONS. Oncogene Research *5*, 221-231.

Morishita, K., Parganas, E., Douglass, E.C., and Ihle, J.N. (1990b). UNIQUE EXPRESSION OF THE HUMAN EVI-1 GENE IN AN ENDOMETRIAL CARCINOMA CELL-LINE - SEQUENCE OF CDNAS AND STRUCTURE OF ALTERNATIVELY SPLICED TRANSCRIPTS. Oncogene *5*, 963-971.

Morishita, K., Parganas, E., Matsugi, T., and Ihle, J.N. (1992a). EXPRESSION OF THE EVI-1 ZINC FINGER GENE IN 32DCL3 MYELOID CELLS BLOCKS GRANULOCYTIC DIFFERENTIATION IN RESPONSE TO GRANULOCYTE COLONY-STIMULATING FACTOR. Molecular and Cellular Biology *12*, 183-189.

Morishita, K., Parganas, E., Willman, C.L., Whittaker, M.H., Drabkin, H., Oval, J., Taetle, R., Valentine, M.B., and Ihle, J.N. (1992b). ACTIVATION OF EVI1 GENE-EXPRESSION IN HUMAN ACUTE MYELOGENOUS LEUKEMIAS BY TRANSLOCATIONS SPANNING 300-400 KILOBASES ON CHROMOSOME BAND-3Q26. Proceedings of the National Academy of Sciences of the United States of America *89*, 3937-3941.

Motta, M., Rassenti, L., Shelvin, B.J., Lerner, S., Kipps, T.J., Keating, M.J., and Wierda, W.G. (2005). Increased expression of CD152 (CTLA-4) by normal T lymphocytes in untreated patients with B-cell chronic lymphocytic leukemia. Leukemia *19*, 1788-1793.

Mouse, E.C., Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., Kaul, R., Canfield, T., Giste, E., Johnson, A., Zhang, M., Balasundaram, G., Byron, R., Roach, V., Sabo, P.J., Sandstrom, R., Stehling, A.S., Thurman, R.E., Weissman, S.M., Cayting, P., Hariharan, M., Lian, J., Cheng, Y., Landt, S.G., Ma, Z., Wold, B.J., Dekker, J., Crawford, G.E., Keller, C.A., Wu, W., Morrissey, C., Kumar, S.A., Mishra, T., Jain, D., Byrska-Bishop, M., Blankenberg, D., Lajoie, B.R., Jain, G., Sanyal, A., Chen, K.B., Denas, O., Taylor, J., Blobel, G.A., Weiss, M.J., Pimkin, M., Deng, W., Marinov, G.K., Williams, B.A., Fisher-Aylor, K.I., Desalvo, G., Kiralusha, A., Trout, D., Amrhein, H., Mortazavi, A., Edsall, L., McCleary, D., Kuan, S., Shen, Y., Yue, F., Ye, Z., Davis, C.A., Zaleski, C., Jha, S., Xue, C., Dobin, A., Lin, W., Fastuca, M., Wang, H., Guigo, R., Djebali, S., Lagarde, J., Ryba, T., Sasaki, T., Malladi, V.S., Cline, M.S., Kirkup, V.M., Learned, K., Rosenbloom, K.R., Kent, W.J., Feingold, E.A., Good, P.J., Pazin, M., Lowdon, R.F., and Adams, L.B. (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol *13*, 418.

Moysich, K.B., Menezes, R.J., and Michalek, A.M. (2002). Chernobyl-related ionising radiation exposure and cancer risk: an epidemiological review. The lancet oncology *3*, 269-279.

Mrózek, K., Marcucci, G., Nicolet, D., Maharry, K.S., Becker, H., Whitman, S.P., Metzeler, K.H., Schwind, S., Wu, Y.-Z., and Kohlschmidt, J. (2012). Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia. J Clin Oncol *30*, 4515-4523.

Mucenski, M.L., Taylor, B.A., Ihle, J.N., Hartley, J.W., Morse, H.C., Jenkins, N.A., and Copeland, N.G. (1988). IDENTIFICATION OF A COMMON ECOTROPIC VIRAL INTEGRATION SITE, EVI-1, IN THE DNA OF AKXD MURINE MYELOID TUMORS. Molecular and Cellular Biology *8*, 301-308.

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*, 758-764.

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. (2008). BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature *453*, 110-+.

Mullighan, C.G., Su, X., Zhang, J., Radtke, I., Phillips, L.A., Miller, C.B., Ma, J., Liu, W., Cheng, C., and Schulman, B.A. (2009a). Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med *360*, 470-480.

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*, 470-480.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553-563.

Nanjundan, M., Nakayama, Y., Cheng, K.W., Lahad, J., Liu, J., Lu, K., Kuo, W.-L., Smith-McCune, K., Fishman, D., Gray, J.W., and Mills, G.B. (2007). Amplification of MDS1/EVI1 and EVI1, Located in the 3q26.2 Amplicon, Is Associated with Favorable Patient Prognosis in Ovarian Cancer. Cancer Research *67*, 3074-3084.

Narlikar, G.J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. Cell *154*, 490-503.

Natoli, G., and Andrau, J.-C. (2012). Noncoding transcription at enhancers: general principles and functional models. Annual review of genetics *46*, 1-19.

Naviaux, R.K., Costanzi, E., Haas, M., and Verma, I.M. (1996). The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. Journal of virology *70*, 5701-5705.

Nielsen, M., Vermont, C.L., Aten, E., Ruivenkamp, C.A., van Herrewegen, F., Santen, G.W., and Breuning, M.H. (2012). Deletion of the 3q26 region including the EVI1 and MDS1 genes in a neonate with congenital thrombocytopenia and subsequent aplastic anaemia. J Med Genet *49*, 598-600.

Nimer, S.D. (2008). Myelodysplastic syndromes. Blood *111*, 4841-4851.

Nimmanapalli, R., Fuino, L., Bali, P., Gasparetto, M., Glozak, M., Tao, J.G., Moscinski, L., Smith, C., Wu, J., Jove, R., Atadja, P., and Bhalla, K. (2003). Histone deacetylase inhibitor LAQ824 both lowers expression and promotes proteasomal degradation of Bcr-Abl and induces apoptosis of imatinib mesylate-sensitive or -refractory chronic myelogenous leukemia-blast crisis cells. Cancer Research *63*, 5126-5135.

Nishitani, H., and Lygerou, Z. (2002). Control of DNA replication licensing in a cell cycle. Genes to cells : devoted to molecular & cellular mechanisms *7*, 523-534.

Norris, D., and Stone, J. (2008). WHO classification of tumours of haematopoietic and lymphoid tissues.

Northcott, P.A., Lee, C., Zichner, T., Stütz, A.M., Erkek, S., Kawauchi, D., Shih, D.J., Hovestadt, V., Zapatka, M., and Sturm, D. (2014). Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Nature *511*, 428-434.

Nowell, P.C. (1976). The clonal evolution of tumor cell populations. Science *194*, 23-28.

Nowell, P.C., and Hungerford (1960a). A minute chromosome in human chronic granulocytic leukemia. Science *132*, 1497.

Nowell, P.C., and Hungerford, D.A. (1960b). Chromosome studies on normal and leukemic human leukocytes. J Natl Cancer Inst *25*, 85-109.

Nutt, S.L., Heavey, B., Rolink, A.G., and Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature *401*, 556-562.

O'Donnell, C.J., Kavousi, M., Smith, A.V., Kardia, S.L.R., Feitosa, M.F., Hwang, S.-J., Sun, Y.V., Province, M.A., Aspelund, T., Dehghan, A., Hoffmann, U., Bielak, L.F., Zhang, Q., Eiriksdottir, G., van Duijn, C.M., Fox, C.S., de Andrade, M., Kraja, A.T., Sigurdsson, S., Elias-Smale, S.E., Murabito, J.M., Launer, L.J., van der Lugt, A., Kathiresan, S., Krestin, G.P., Herrington, D.M., Howard, T.D., Liu, Y., Post, W., Mitchell, B.D., O'Connell, J.R., Shen, H., Shuldiner, A.R., Altshuler, D., Elosua, R., Salomaa, V., Schwartz, S.M., Siscovick, D.S., Voight, B.F., Bis, J.C., Glazer, N.L., Psaty, B.M., Boerwinkle, E., Heiss, G., Blankenberg, S., Zeller, T., Wild, P.S., Schnabel, R.B., Schillert, A., Ziegler, A., Münzel, T.F., White, C.C., Rotter, J.I., Nalls, M., Oudkerk, M., Johnson, A.D., Newman, A.B., Uitterlinden, A.G., Massaro, J.M., Cunningham, J., Harris, T.B., Hofman, A., Peyser, P.A., Borecki, I.B., Cupples, L.A., Gudnason, V., and Witteman, J.C.M. (2011). Genome-Wide Association Study for Coronary Artery Calcification With Follow-Up in Myocardial Infarction. Circulation *124*, 2855-2864.

Ogawa, S., Kurokawa, M., Tanaka, T., Tanaka, K., Hangaishi, A., Mitani, K., Kamada, N., Yazaki, Y., and Hirai, H. (1996a). Increased Evi-1 expression is frequently observed in blastic crisis of chronic myelocytic leukemia. Leukemia *10*, 788-794.

Ogawa, S., Mitani, K., Kurokawa, M., Matsuo, Y., Minowada, J., Inazawa, J., Kamada, N., Tsubota, T., Yazaki, Y., and Hirai, H. (1996b). Abnormal expression of Evi-1 gene in human leukemias. Human cell *9*, 323-332.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent Enhancers Activated by Stimulation in Differentiated Cells. Cell *152*, 157-171.

Ottmann, O.G., Druker, B.J., Sawyers, C.L., Goldman, J.M., Reiffers, J., Silver, R.T., Tura, S., Fischer, T., Deininger, M.W., Schiffer, C.A., Baccarani, M., Gratwohl, A., Hochhaus, A., Hoelzer, D., Fernandes-Reese, S., Gathmann, I., Capdeville, R., and O'Brien, S.G. (2002). A phase 2 study of imatinib in patients with relapsed or refractory Philadelphia chromosome-positive acute lymphoid leukemias. Blood *100*, 1965-1971.

Panier, S., and Boulton, S.J. (2014). Double-strand break repair: 53BP1 comes into focus. Nature reviews Molecular cell biology *15*, 7.

Pannunzio, N.R., and Lieber, M.R. (2016). Dissecting the Roles of Divergent and Convergent Transcription in Chromosome Instability. Cell Rep *14*, 1025-1031.

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*, 116-125.

Park, Y.-J., Yoon, S.-J., Suh, H.-W., Kim, D.O., Park, J.-R., Jung, H., Kim, T.-D., Yoon, S.R., Min, J.-K., and Na, H.-J. (2013). TXNIP deficiency exacerbates endotoxic shock via the induction of excessive nitric oxide synthesis. PLoS pathogens *9*, e1003646.

Patel, H., Marley, S.B., and Gordon, M.Y. (2006). Detection in primary chronic myeloid leukaemia cells of p210BCR-ABL1 in complexes with adaptor proteins CBL, CRKL, and GRB2. Genes, Chromosomes and Cancer *45*, 1121-1129.

Patel, P., Hanson, D.L., Sullivan, P.S., Novak, R.M., Moorman, A.C., Tong, T.C., Holmberg, S.D., and Brooks, J.T. (2008). Incidence of types of cancer among HIV-infected persons compared with the general population in the United States, 1992–2003. Annals of internal medicine *148*, 728-736.

Paul, S.P., Taylor, L.S., Stansbury, E.K., and McVicar, D.W. (2000). Myeloid specific human CD33 is an inhibitory receptor with differential ITIM function in recruiting the phosphatases SHP-1 and SHP-2. Blood *96*, 483-490.

Peeters, P., Wlodarska, I., Baens, M., Criel, A., Selleslag, D., Hagemeijer, A., VandenBerghe, H., and Marynen, P. (1997). Fusion of ETV6 to MDS1/EVI1 as a result of t(3;12)(q26;p13) in myeloproliferative disorders. Cancer Research *57*, 564-569.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N., and Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. Nature *514*, 389-393.

Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., and Lewis, K.D. (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature *444*, 499-502.

Perera, D., Poulos, R.C., Shah, A., Beck, D., Pimanda, J.E., and Wong, J.W.H. (2016). Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. Nature *532*, 259-263.

Perkins, A.S., Fishel, R., Jenkins, N.A., and Copeland, N.G. (1991a). EVI-1, A MURINE ZINC FINGER PROTOONCOGENE, ENCODES A SEQUENCE-SPECIFIC DNA-BINDING PROTEIN. Molecular and Cellular Biology *11*, 2665-2674.

Perkins, A.S., Mercer, J.A., Jenkins, N.A., and Copeland, N.G. (1991b). PATTERNS OF EVI-1 EXPRESSION IN EMBRYONIC AND ADULT TISSUES SUGGEST THAT EVI-1 PLAYS AN IMPORTANT REGULATORY ROLE IN MOUSE DEVELOPMENT. Development *111*, 479-487.

Petermann, E., Orta, M.L., Issaeva, N., Schultz, N., and Helleday, T. (2010). Hydroxyurea-Stalled Replication Forks Become Progressively Inactivated and Require Two Different RAD51-Mediated Pathways for Restart and Repair. Molecular Cell *37*, 492-502.

Pfeifer, G. (2006). Mutagenesis at methylated CpG sequences. In DNA methylation: basic mechanisms (Springer), pp. 259-281.

Phillips, L.L. (1956). Effect of Free Radicals on Chromosomes of Barley. Science *124*, 889-890.

Pistillo, M.P., Tazzari, P.L., Palmisano, G.L., Pierri, I., Bolognesi, A., Ferlito, F., Capanni, P., Polito, L., Ratta, M., and Pileri, S. (2003). CTLA-4 is not restricted to the lymphoid cell lineage and can function as a target molecule for apoptosis induction of leukemic cells. Blood *101*, 202-209.

Pruss, D., Hayes, J.J., and Wolffe, A.P. (1995). NUCLEOSOMAL ANATOMY - WHERE ARE THE HISTONES. Bioessays *17*, 161-170.

Pui, C.-H., and Evans, W.E. (1998). Acute lymphoblastic leukemia. N Engl J Med *339*, 605-615.

Pui, C.-H., Williams, D.L., Raimondi, S.C., Rivera, G.K., Look, A.T., Dodge, R.K., George, S.L., Behm, F.G., Crist, W.M., and Murphy, S.B. (1987). Hypodiploidy is associated with a poor prognosis in childhood acute lymphoblastic leukemia. Blood *70*, 247-253.

Qu, G.-z., Grundy, P.E., Narayan, A., and Ehrlich, M. (1999). Frequent hypomethylation in Wilms tumors of pericentromeric DNA in chromosomes 1 and 16. Cancer Genet Cytogenet *109*, 34-39.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Raghavan, S.C., Swanson, P.C., Ma, Y.M., and Lieber, M.R. (2005). Double-strand break formation by the RAG complex at the bcl-2 major breakpoint region and at other non-B DNA structures in vitro. Molecular and Cellular Biology *25*, 5904-5919.

Rahman, S., Magnussen, M., Leon, T.E., Farah, N., Li, Z., Abraham, B.J., Alapi, K.Z., Mitchell, R.J., Naughton, T., Fielding, A.K., Pizzey, A., Bustraan, S., Allen, C., Popa, T., Pike-Overzet, K., Garcia-Perez, L., Gale, R.E., Linch, D.C., Staal, F.J.T., Young, R.A., Look, A.T., and Mansour, M.R. (2017). Activation of the LMO2 oncogene through a somatically acquired neomorphic promoter in T-cell acute lymphoblastic leukemia. Blood *129*, 3221-3226.

Raimondi, S.C., Chang, M.N., Ravindranath, Y., Behm, F.G., Gresik, M.V., Steuber, C.P., Weinstein, H.J., and Carroll, A.J. (1999). Chromosomal abnormalities in 478 children with acute myeloid leukemia: clinical characteristics and treatment outcome in a cooperative pediatric oncology group study-POG 8821. Blood *94*, 3707-3716.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res *42*, W187-191.

Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.T., McBride, K.M., Eisenreich, T.R., Chen, J.J., Dickins, R.A., Lowe, S.W., Nussenzweig, A., and Nussenzweig, M.C. (2006). Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. Nature *440*, 105-109.

Ramiro, A.R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., Muramatsu, M., Hongo, T., Nussenzweig, A., and Nussenzweig, M.C. (2004). AID is required for c-myc/IgH chromosome translocations in vivo. Cell *118*, 431-438.

Ramiro, A.R., Stavropoulos, P., Jankovic, M., and Nussenzweig, M.C. (2003). Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. Nature immunology *4*, 452-456.

Ravandi, F., O'Brien, S., Thomas, D., Faderl, S., Jones, D., Garris, R., Dara, S., Jorgensen, J., Kebriaei, P., Champlin, R., Borthakur, G., Burger, J., Ferrajoli, A., Garcia-Manero, G., Wierda, W., Cortes, J., and Kantarjian, H. (2010). First report of phase 2 study of dasatinib with hyper-CVAD for the frontline treatment of patients with Philadelphia chromosome-positive (Ph+) acute lymphoblastic leukemia. Blood *116*, 2070-2077.

Ribera, J.M. (2013). Optimal approach to treatment of patients with Philadelphia chromosome-positive acute lymphoblastic leukemia: how to best use all the available tools. Leuk Lymphoma *54*, 21-27.

Ribera, J.M., Oriol, A., Gonzalez, M., Vidriales, B., Brunet, S., Esteve, J., Del Potro, E., Rivas, C., Moreno, M.J., Tormo, M., Martin-Reina, V., Sarra, J., Parody, R., de Oteyza, J.P., Bureo, E., and Bernal, M.T. (2010). Concurrent intensive chemotherapy and imatinib before and after stem cell transplantation in newly diagnosed Philadelphia chromosome-positive acute lymphoblastic leukemia. Final results of the CSTIBES02 trial. Haematologica *95*, 87-95.

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*, 153-166.

Robertson, K.D. (2005). DNA methylation and human disease. Nature Reviews Genetics *6*, 597.

Rodriguez, R., Miller, K.M., Forment, J.V., Bradshaw, C.R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S., and Jackson, S.P. (2012). Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. Nature chemical biology *8*, 301-310.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem *273*, 5858-5868.

Rosenbauer, F., and Tenen, D.G. (2007). Transcription factors in myeloid development: balancing differentiation with transformation. Nature reviews Immunology *7*, 105-117.

Rowley, J.D. (1973). NEW CONSISTENT CHROMOSOMAL ABNORMALITY IN CHRONIC MYELOGENOUS LEUKEMIA IDENTIFIED BY QUINACRINE FLUORESCENCE AND GIEMSA STAINING. Nature *243*, 290-293.

Rowley, J.D., and Potter, D. (1976). CHROMOSOMAL BANDING-PATTERNS IN ACUTE NONLYMPHOCYTIC LEUKEMIA. Blood *47*, 705-721.

Roy, S., Jorgensen, H.G., Roy, P., El Baky, M.A., Melo, J.V., Strathdee, G., Holyoake, T.L., and Bartholomew, C. (2012). BCR-ABL1 tyrosine kinase sustained MECOM expression in chronic myeloid leukaemia. Br J Haematol *157*, 446-456.

Russell, M., List, A., Greenberg, P., Woodward, S., Glinsmann, B., Parganas, E., Ihle, J., and Taetle, R. (1994). EXPRESSION OF EVI1 IN MYELODYSPLASTIC SYNDROMES AND OTHER HEMATOLOGIC MALIGNANCIES WITHOUT 3Q26 TRANSLOCATIONS. Blood *84*, 1243-1248.

Sakharkar, M.K., Perumal, B.S., Sakharkar, K.R., and Kangueane, P. (2005). An analysis on gene architecture in human and mouse genomes. In silico biology *5*, 347-365.

Sanda, T., Lawton, L.N., Barrasa, M.I., Fan, Z.P., Kohlhammer, H., Gutierrez, A., Ma, W., Tatarek, J., Ahn, Y., Kelliher, M.A., Jamieson, C.H.M., Staudt, L.M., Young, R.A., and Look, A.T. (2012). Core Transcriptional Regulatory Circuit Controlled by the TAL1 Complex in Human T Cell Acute Lymphoblastic Leukemia. Cancer Cell *22*, 209-221.

Sato, T., Goyama, S., Kataoka, K., Nasu, R., Tsuruta-Kishino, T., Kagoya, Y., Nukina, A., Kumagai, K., Kubota, N., Nakagawa, M., Arai, S., Yoshimi, A., Honda, H., Kadowaki, T., and Kurokawa, M. (2014). Evil defines leukemia-initiating capacity and tyrosine kinase inhibitor resistance in chronic myeloid leukemia. Oncogene *33*, 5028-5038.

Sawyers, C.L. (1993). The role of myc in transformation by BCR-ABL. Leuk Lymphoma *11 Suppl 1*, 45-46.

Sekeres, M.A. (2011). Epidemiology, natural history, and practice patterns of patients with myelodysplastic syndromes in 2010. Journal of the National Comprehensive Cancer Network : JNCCN *9*, 57-63.

Sellick, G., Spendlove, H., Catovsky, D., Pritchard-Jones, K., and Houlston, R. (2005). Further evidence that germline CEBPA mutations cause dominant inheritance of acute myeloid leukaemia. Leukemia *19*, 1276.

Seo, J., Kim, S.C., Lee, H.S., Kim, J.K., Shon, H.J., Salleh, N.L., Desai, K.V., Lee, J.H., Kang, E.S., Kim, J.S., and Choi, J.K. (2012). Genome-wide profiles of H2AX and gamma-H2AX differentiate endogenous and exogenous DNA damage hotspots in human cells. Nucleic Acids Res *40*, 5965-5974.

Shah, N.P., Kantarjian, H.M., Kim, D.-W., Rea, D., Dorlhiac-Llacer, P.E., Milone, J.H., Vela-Ojeda, J., Silver, R.T., Khoury, H.J., Charbonnier, A., Khoroshko, N., Paquette, R.L., Deininger, M., Collins, R.H., Otero, I., Hughes, T., Bleickardt, E., Strauss, L., Francis, S., and Hochhaus, A. (2008). Intermittent target inhibition with dasatinib 100 mg once daily preserves efficacy and improves tolerability in imatinib-resistant and -intolerant chronic-phase chronic myeloid leukemia. J Clin Oncol *26*, 3204-3212.

Shah, N.P., Kim, D.-W., Kantarjian, H., Rousselot, P., Dorlhiac Llacer, P.E., Enrico, A., Vela-Ojeda, J., Silver, R.T., Khoury, H.J., Mueller, M.C., Lambert, A., Matloub, Y., and Hochhaus, A. (2010). Potent, transient inhibition of BCR-ABL with dasatinib 100 mg daily achieves rapid and durable cytogenetic responses and high transformation-free survival rates in chronic phase chronic myeloid leukemia patients with resistance, suboptimal response or intolerance to imatinib. Haematol-Hematol J *95*, 232-240.

Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs. plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC genomics *15*, 284.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Shi, J., Whyte, W.A., Zepeda-Mendoza, C.J., Milazzo, J.P., Shen, C., Roe, J.S., Minder, J.L., Mercan, F., Wang, E., Eckersley-Maslin, M.A., Campbell, A.E., Kawaoka, S., Shareef, S., Zhu, Z., Kendall, J., Muhar, M., Haslinger, C., Yu, M., Roeder, R.G., Wigler, M.H., Blobel, G.A., Zuber, J., Spector, D.L., Young, R.A., and Vakoc, C.R. (2013). Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. Genes Dev *27*, 2648-2662.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet *15*, 272-286.

Shojaee, S., Caeser, R., Buchner, M., Park, E., Swaminathan, S., Hurtz, C., Geng, H., Chan, L.N., Klemm, L., Hofmann, W.K., Qiu, Y.H., Zhang, N., Coombes, K.R., Paietta, E., Molkentin, J., Koeffler, H.P., Willman, C.L., Hunger, S.P., Melnick, A., Kornblau, S.M., and Muschen, M. (2015). Erk Negative Feedback Control Enables Pre-B Cell Transformation and Represents a Therapeutic Target in Acute Lymphoblastic Leukemia. Cancer Cell *28*, 114-128.

Shteper, P., Siegfried, Z., Asimakopoulos, F., Palumbo, G., Rachmilewitz, E., Ben-Neriah, Y., and Ben-Yehuda, D. (2001). ABL1 methylation in Ph-positive ALL is exclusively associated with the P210 form of BCR-ABL. Leukemia *15*, 575.

Siegel, R., Naishadham, D., and Jemal, A. (2012). Cancer Statistics, 2012. CA-Cancer J Clin *62*, 10-29.

Siegel, R.L., Miller, K.D., and Jemal, A. (2015). Cancer statistics, 2015. CA: a cancer journal for clinicians *65*, 5-29.

Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science *350*, 978-981.

Sill, H., Olipitz, W., Zebisch, A., Schulz, E., and Wolfler, A. (2011). Therapy-related myeloid neoplasms: pathobiology and clinical characteristics. Br J Pharmacol *162*, 792-805.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell *42*, 794-805.

Snyder, R. (2012). Leukemia and benzene. International journal of environmental research and public health *9*, 2875-2893.

Soderholm, J., Kobayashi, H., Mathieu, C., Rowley, J.D., and Nucifora, G. (1997). The leukemia-associated gene MDS1/EVI1 is a new type of GATA-binding transactivator. Leukemia *11*, 352-358.

Soldi, M., and Bonaldi, T. (2014). The ChroP approach combines ChIP and mass spectrometry to dissect locus-specific proteomic landscapes of chromatin. J Vis Exp.

Sollier, J., and Cimprich, K.A. (2015). Breaking bad: R-loops and genome integrity. Trends Cell Biol *25*, 514-522.

Song, F., Smith, J.F., Kimura, M.T., Morrow, A.D., Matsuyama, T., Nagase, H., and Held, W.A. (2005). Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proceedings of the National Academy of Sciences of the United States of America *102*, 3336-3341.

Song, L., Zhang, Z., Grasfeder, L.L., Boyle, A.P., Giresi, P.G., Lee, B.-K., Sheffield, N.C., Gräf, S., Huss, M., and Keefe, D. (2011). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome research *21*, 1757-1767.

Soverini, S., De Benedittis, C., Papayannidis, C., Paolini, S., Venturi, C., Iacobucci, I., Luppi, M., Bresciani, P., Salvucci, M., Russo, D., Sica, S., Orlandi, E., Intermesoli, T., Gozzini, A., Bonifacio, M., Rigolin, G.M., Pane, F., Baccarani, M., Cavo, M., and Martinelli, G. (2014). Drug Resistance and BCR-ABL Kinase Domain Mutations in Philadelphia Chromosome- Positive Acute Lymphoblastic Leukemia From the Imatinib to the Second-Generation Tyrosine Kinase Inhibitor Era. Cancer *120*, 1002-1009.

Spandidos, A., Wang, X., Wang, H., Dragnev, S., Thurber, T., and Seed, B. (2008). A comprehensive collection of experimentally validated primers for Polymerase Chain Reaction quantitation of murine transcript abundance. BMC Genomics *9*, 633.

Spandidos, A., Wang, X., Wang, H., and Seed, B. (2010). PrimerBank: a resource of human and mouse PCR primer pairs for gene expression detection and quantification. Nucleic Acids Res *38*, D792-D799.

Speck, N.A., and Gilliland, D.G. (2002). Core-binding factors in haematopoiesis and leukaemia. Nature reviews Cancer *2*, 502-513.

Sperling, A.S., Gibson, C.J., and Ebert, B.L. (2016). The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. Nature Reviews Cancer *17*, 5.

Stadhouders, R., Kolovos, P., Brouwer, R., Zuin, J., van den Heuvel, A., Kockx, C., Palstra, R.-J., Wendt, K.S., Grosveld, F., van Ijcken, W., and Soler, E. (2013). Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. Nat Protocols *8*, 509-524.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., Tiwari, V.K., and Schubeler, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490-495.

Steinleitner, K., Rampetsreiter, P., Koeffel, R., Ramanathan, G., Mannhalter, C., Strobl, H., and Wieser, R. (2012). EVI1 and MDS1/EVI1 Expression During Primary Human Hematopoietic Progenitor Cell Differentiation into Various Myeloid Lineages. Anticancer Research *32*, 4883-4889.

Stephens, P.J., Greenman, C.D., Fu, B.Y., Yang, F.T., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., McLaren, S., Lin, M.L., McBride, D.J., Varela, I., Nik-Zainal, S., Leroy, C., Jia, M.M., Menzies, A., Butler, A.P., Teague, J.W., Quail, M.A., Burton, J., Swerdlow, H., Carter, N.P., Morsberger, L.A., Iacobuzio-Donahue, C., Follows, G.A., Green, A.R., Flanagan, A.M., Stratton, M.R., Futreal, P.A., and Campbell, P.J. (2011). Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. Cell *144*, 27-40.

Stern, H.M., Padilla, M., Wagner, K., Amler, L., and Ashkenazi, A. (2010). Development of immunohistochemistry assays to assess GALNT14 and FUT3/6 in clinical trials of dulanermin and drozitumab. Clin Cancer Res *16*, 1587-1596.

Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., and Sipahimalani, P. (2012). R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. Genes Dev *26*, 163-175.

Strehl, S., Konig, M., Dworzak, M.N., Kalwak, K., and Haas, O.A. (2003). PAX5/ETV6 fusion defines cytogenetic entity dic(9;12)(p13;p13). Leukemia *17*, 1121-1123.

Supek, F., Lehner, B., Supek, F., Lehner, B., Supek, F., Lehner, B., and Supek, F. (2017). Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes. Cell *170*, 534-547.e523.

Swaminathan, S., Klemm, L., Park, E., Papaemmanuil, E., Ford, A., Kweon, S.M., Trageser, D., Hasselfeld, B., Henke, N., Mooster, J., Geng, H., Schwarz, K., Kogan, S.C., Casellas, R., Schatz, D.G., Lieber, M.R., Greaves, M.F., and Muschen, M. (2015). Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia. Nature immunology *16*, 766-774.

Swerdlow, S.H., Campo, E., Harris, N.L., Jaffe, E.S., Pileri, S.A., Stein, H., Thiele, J., and Vardiman, J.W. (2008). WHO classification of tumours of haematopoietic and lymphoid tissues (Lyon: International Agency for Research on Cancer).

Syed, J., Pandian, G.N., Sato, S., Taniguchi, J., Chandran, A., Hashiya, K., Bando, T., and Sugiyama, H. (2014). Targeted Suppression of EVI1 Oncogene Expression by Sequence-Specific Pyrrole-Imidazole Polyamide. Chemistry & Biology *21*, 1370-1380.

Tapper, W., Jones, A.V., Kralovics, R., Harutyunyan, A.S., Zoi, K., Leung, W., Godfrey, A.L., Guglielmelli, P., Callaway, A., Ward, D., Aranaz, P., White, H.E., Waghorn, K., Lin, F., Chase, A., Baxter, E.J., Maclean, C., Nangalia, J., Chen, E., Evans, P., Short, M., Jack, A., Wallis, L., Oscier, D., Duncombe, A.S., Schuh, A., Mead, A.J., Griffiths, M., Ewing, J., Gale, R.E., Schnittger, S., Haferlach, T., Stegelmann, F., Dohner, K., Grallert, H., Strauch, K., Tanaka, T., Bandinelli, S., Giannopoulos, A., Pieri, L., Mannarelli, C., Gisslinger, H., Barosi, G., Cazzola, M., Reiter, A., Harrison, C., Campbell, P., Green, A.R., Vannucchi, A., and Cross, N.C. (2015). Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. Nat Commun *6*, 6691.

Tasian, S.K., Loh, M.L., and Hunger, S.P. (2017). Philadelphia chromosome–like acute lymphoblastic leukemia. Blood *130*, 2064-2072.

Taub, R., Kirsch, I., Morton, C., Lenoir, G., Swan, D., Tronick, S., Aaronson, S., and Leder, P. (1982). Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. Proceedings of the National Academy of Sciences *79*, 7837-7841.

Teng, G., Maman, Y., Resch, W., Kim, M., Yamane, A., Qian, J., Kieffer-Kwon, K.R., Mandal, M., Ji, Y., Meffre, E., Clark, M.R., Cowell, L.G., Casellas, R., and Schatz, D.G. (2015). RAG Represents a Widespread Threat to the Lymphocyte Genome. Cell *162*, 751-765.

Terwilliger, T., and Abdul-Hay, M. (2017). Acute lymphoblastic leukemia: a comprehensive review and 2017 update. Blood Cancer J *7*, e577.

Thatcher, T.H., and Gorovsky, M.A. (1994). PHYLOGENETIC ANALYSIS OF THE CORE HISTONES H2A, H2B, H3, AND H4. Nucleic Acids Res *22*, 174-179.

Thomas, D.A., Faderl, S., Cortes, J., O'Brien, S., Giles, F.J., Kornblau, S.M., Garcia-Manero, G., Keating, M.J., Andreeff, M., Jeha, S., Beran, M., Verstovsek, S., Pierce, S., Letvak, L., Salvado, A., Champlin, R., Talpaz, M., and Kantarjian, H. (2004). Treatment of Philadelphia chromosome-positive acute lymphocytic leukemia with hyper-CVAD and imatinib mesylate. Blood *103*, 4396-4407.

Thomas, E.D. (1999). Bone marrow transplantation: a review. Paper presented at: Seminars in hematology.

Trifa, A.P., Banescu, C., Bojan, A.S., Voina, C.M., Popa, S., Visan, S., Ciubean, A.D., Tripon, F., Dima, D., Popov, V.M., Vesa, S.C., Andreescu, M., Torok-Vistai, T., Mihaila, R.G., Berbec, N., Macarie, I., Colita, A., Iordache, M., Catana, A.C., Farcas, M.F., Tomuleasa, C., Vasile, K., Truica, C., Todinca, A., Pop-Muntean, L., Manolache, R., Bumbea, H., Vladareanu, A.M., Gaman, M., Ciufu, C.M., and Popp, R.A. (2018). MECOM, HBS1L-MYB, THRB-RARB, JAK2, and TERT polymorphisms defining the genetic predisposition to myeloproliferative neoplasms: A study on 939 patients. Am J Hematol *93*, 100-106.

Tsai, A.G., Lu, H., Raghavan, S.C., Muschen, M., Hsieh, C.-L., and Lieber, M.R. (2008). Human Chromosomal Translocations at CpG Sites and a Theoretical Basis for Their Lineage and Stage Specificity. Cell *135*, 1130-1142.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3—new capabilities and interfaces. Nucleic Acids Res *40*, e115-e115.

Ustun, C., Fiskus, W., Wang, Y., Rao, R., Lee, P., Kolhe, R., Fernandez, P., Jillella, A.P., Buser, C., and Bhalla, K.N. (2008). Co-treatment with aurora kinase inhibitor MK-0457 and pan-histone deacetylase inhibitor vorinostat: A novel targeted treatment for AML and CML. J Clin Oncol *26*.

van de Werken, H.J.G., de Vree, P.J.P., Splinter, E., Holwerda, S.J.B., Klous, P., de Wit, E., and de Laat, W. (2012). Chapter Four - 4C Technology: Protocols and Data Analysis. In Methods in Enzymology, W. Carl, and C.D. Allis, eds. (Academic Press), pp. 89-112.

Vázquez, I., Maicas, M., Marcotegui, N., Conchillo, A., Guruceaga, E., Roman-Gomez, J., Calasanz, M.J., Agirre, X., Prosper, F., and Odero, M.D. (2010). Silencing of hsa-miR-124 by EVI1 in cell lines and patients with acute myeloid leukemia. Proceedings of the National Academy of Sciences *107*, E167-E168.

Veillette, A., Thibaudeau, E., and Latour, S. (1998). High expression of inhibitory receptor SHPS-1 and its association with protein-tyrosine phosphatase SHP-1 in macrophages. J Biol Chem *273*, 22719-22728.

Visel, A., Rubin, E.M., and Pennacchio, L.A. (2009). Genomic views of distant-acting enhancers. Nature *461*, 199-205.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546-1558.

Volkert, S., Schnittger, S., Zenger, M., Kern, W., Haferlach, T., and Haferlach, C. (2014). Amplification of EVI1 on cytogenetically cryptic double minutes as new mechanism for increased expression of EVI1. Cancer genetics *207*, 103-108.

Wagner, K.W., Punnoose, E.A., Januario, T., Lawrence, D.A., Pitti, R.M., Lancaster, K., Lee, D., von Goetz, M., Yee, S.F., and Totpal, K. (2007). Death-receptor O-glycosylation controls tumor-cell sensitivity to the proapoptotic ligand Apo2L/TRAIL. Nature medicine *13*, 1070-1077.

Walker, L.C., Stevens, J., Campbell, H., Corbett, R., Spearing, R., Heaton, D., Macdonald, D.H., Morris, C.M., and Ganly, P. (2002). A novel inherited mutation of the transcription factor RUNX1 causes thrombocytopenia and may predispose to acute myeloid leukaemia. Br J Haematol *117*, 878-881.

Walunas, T.L., Lenschow, D.J., Bakker, C.Y., Linsley, P.S., Freeman, G.J., Green, J.M., Thompson, C.B., and Bluestone, J.A. (1994). CTLA-4 can function as a negative regulator of T cell activation. Immunity *1*, 405-413.

Wang, G.G., Allis, C.D., and Chi, P. (2007). Chromatin remodeling and cancer, Part I: Covalent histone modifications. Trends in molecular medicine *13*, 363-372.

Wang, H., Queisser, A., Hagedorn, S., Konantz, M., Schafer, T., Alavi, S., Vogel, W., Massenhausen, A.V., Kristiansen, G., Duensing, S., Kirfel, J., Perner, S., and Lengerke, C. (2016). EVI1-a novel oncogene in prostate carcinoma. Oncology Research and Treatment *39*, 267-267.

Wang, H., Schaefer, T., Konantz, M., Braun, M., Varga, Z., Paczulla, A.M., Reich, S., Jacob, F., Perner, S., Moch, H., Fehm, T., Kanz, L., Schulze-Osthoff, K., and Lengerke, C. (2017). Prominent oncogenic roles of EVI1 in breast carcinoma. Cancer Res.

Wang, J.C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. Nat Rev Mol Cell Biol *3*, 430-440.

Wang, Y.L., Lee, J.W., Cesarman, E., Jin, D.K., and Csernus, B. (2006). Molecular monitoring of chronic myelogenous leukemia - Identification of the most suitable internal control gene for real-time quantification of BCR-ABL transcripts. J Mol Diagn *8*, 231-239.

Ward, I.M., and Chen, J.J. (2001). Histone H2AX is phosphorylated in an ATR-dependent manner in response to replicational stress. J Biol Chem *276*, 47759-47762.

Ward, I.M., Minn, K., van Deursen, J., and Chen, J. (2003). p53 Binding protein 53BP1 is required for DNA damage responses and tumor suppression in mice. Mol Cell Biol *23*, 2556-2563.

Wassmann, B., Pfeifer, H., Goekbuget, N., Beelen, D.W., Beck, J., Stelljes, M., Bornhauser, M., Reichle, A., Perz, J., Haas, R., Ganser, A., Schmid, M., Kanz, L., Lenz, G., Kaufmann, M., Binckebanck, A., Bruck, P., Reutzel, R., Gschaidmeier, H., Schwartz, S., Hoelzer, D., and Ottmann, O.G. (2006). Alternating versus concurrent schedules of imatinib and chemotherapy as front-line therapy for Philadelphia-positive acute lymphoblastic leukemia (Ph+ ALL). Blood *108*, 1469-1477.

Wechsler, J., Greene, M., McDevitt, M.A., Anastasi, J., Karp, J.E., Le Beau, M.M., and Crispino, J.D. (2002). Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. Nature Genet *32*, 148.

West, M.H.P., and Bonner, W.M. (1980). HISTONE-2A, A HETEROMORPHOUS FAMILY OF 8 PROTEIN SPECIES. Biochemistry *19*, 3238-3245.

Westin, E.H., Wongstaal, F., Gelmann, E.P., Dellafavera, R., Papas, T.S., Lautenberger, J.A., Eva, A., Reddy, E.P., Tronick, S.R., Aaronson, S.A., and Gallo, R.C. (1982). EXPRESSION OF CELLULAR HOMOLOGS OF RETROVIRAL ONC GENES IN HUMAN HEMATOPOIETIC-CELLS. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences *79*, 2490-2494.

White, H.E., Matejtschuk, P., Rigsby, P., Gabert, J., Lin, F., Wang, Y.L., Branford, S., Muller, M.C., Beaufils, N., Beillard, E., Colomer, D., Dvorakova, D., Ehrencrona, H., Goh, H.G., El Housni, H., Jones, D., Kairisto, V., Kamel-Reid, S., Kim, D.W., Langabeer, S., Ma, E.S.K., Press, R.D., Romeo, G., Wang, L.H., Zoi, K., Hughes, T., Saglio, G., Hochhaus, A., Goldman, J.M., Metcalfe, P., and Cross, N.C.P. (2010). Establishment of the first World Health Organization International Genetic Reference Panel for quantitation of BCR-ABL mRNA. Blood *116*, E111-E117.

Widom, J. (1989). Toward a unified model of chromatin folding. Annual review of biophysics and biophysical chemistry *18*, 365-395.

Wong, R.W.J., Ngoc, P.C.T., Leong, W.Z., Yam, A.W.Y., Zhang, T., Asamitsu, K., Iida, S., Okamoto, T., Ueda, R., Gray, N.S., Ishida, T., and Sanda, T. (2017). Enhancer profiling identifies critical cancer genes and characterizes cell identity in adult T-cell leukemia. Blood *130*, 2326-2338.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H.-w., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. Nature immunology *12*, 62-69.

Yamane, A., Robbiani, D.F., Resch, W., Bothmer, A., Nakahashi, H., Oliveira, T., Rommel, P.C., Brown, E.J., Nussenzweig, A., and Nussenzweig, M.C. (2013). RPA accumulation during class switch recombination represents 5′–3′ DNA-end resection during the S–G2/M phase of the cell cycle. Cell reports *3*, 138-147.

Yamazaki, H., Suzuki, M., Otsuki, A., Shimizu, R., Bresnick, Emery H., Engel, James D., and Yamamoto, M. (2014). A Remote GATA2 Hematopoietic Enhancer Drives Leukemogenesis in inv(3)(q21;q26) by Activating EVI1 Expression. Cancer Cell *25*, 415-427.

Yokoyama, T., Kanno, Y., Yamazaki, Y., Takahara, T., Miyata, S., and Nakamura, T. (2010). Trib1 links the MEK1/ERK pathway in myeloid leukemogenesis. Blood *116*, 2768-2775.

Yoshimi, A., Goyama, S., Watanabe-Okochi, N., Yoshiki, Y., Nannya, Y., Nitta, E., Arai, S., Sato, T., Shimabe, M., Nakagawa, M., Imai, Y., Kitamura, T., and Kurokawa, M. (2011). Evi1 represses PTEN expression and activates PI3K/AKT/mTOR via interactions with polycomb proteins. Blood *117*, 3617-3628.

Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., and Wang, S. (2010). GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics *26*, 976-978.

Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382-2383.

Yuasa, H., Oike, Y., Iwama, A., Nishikata, I., Sugiyama, D., Perkins, A., Mucenski, M.L., Suda, T., and Morishita, K. (2005). Oncogenic transcription factor Evi1 regulates hematopoietic stem cell proliferation through GATA-2 expression. Embo Journal *24*, 1976-1987.

Yukawa, M., Akiyama, T., Franke, V., Mise, N., Isagawa, T., Suzuki, Y., Suzuki, M.G., Vlahovicek, K., Abe, K., and Aburatani, H. (2014). Genome-wide analysis of the chromatin composition of histone H2A and H3 variants in mouse embryonic stem cells. PloS one *9*, e92689.

Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature *518*, 556-559.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics *25*, 1952-1958.

Zhang, B., Strauss, A.C., Chu, S., Li, M., Ho, Y., Shiang, K.D., Snyder, D.S., Huettner, C.S., Shultz, L., Holyoake, T., and Bhatia, R. (2010). Effective targeting of quiescent chronic myelogenous leukemia stem cells by histone deacetylase inhibitors in combination with imatinib mesylate. Cancer Cell *17*, 427-442.

Zhang, J., Mullighan, C.G., Harvey, R.C., Wu, G., Chen, X., Edmonson, M., Buetow, K.H., Carroll, W.L., Chen, I.M., Devidas, M., Gerhard, D.S., Loh, M.L., Reaman, G.H., Relling, M.V., Camitta, B.M., Bowman, W.P., Smith, M.A., Willman, C.L., Downing, J.R., and Hunger, S.P. (2011). Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. Blood *118*, 3080-3087.

Zhang, J.Y., Poh, H.M., Peh, S.Q., Sia, Y.Y., Li, G.L., Mulawadi, F.H., Goh, Y.F., Fullwood, M.J., Sung, W.K., Ruan, X.A., and Ruan, Y.J. (2012). ChIA-PET analysis of transcriptional chromatin interactions. Methods (San Diego, Calif) *58*, 289-299.

Zhang, M., and Swanson, P.C. (2008). V(D)J recombinase binding and cleavage of cryptic recombination signal sequences identified from lymphoid malignancies. J Biol Chem *283*, 6717-6727.

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/EBPα. Immunity *21*, 853-863.

Zhang, W., Konopleva, M., Shi, Y.-x., McQueen, T., Harris, D., Ling, X., Estrov, Z., Quintás-Cardama, A., Small, D., and Cortes, J. (2008). Mutant FLT3: a direct target of sorafenib in acute myelogenous leukemia. JNCI: Journal of the National Cancer Institute *100*, 184-198.

Zhang, X., Choi, P.S., Francis, J.M., Gao, G.F., Campbell, J.D., Ramachandran, A., Mitsuishi, Y., Ha, G., Shih, J., Vazquez, F., Tsherniak, A., Taylor, A.M., Zhou, J., Wu, Z., Berger, A.C., Giannakis, M., Hahn, W.C., Cherniack, A.D., and Meyerson, M. (2018). Somatic Superenhancer Duplications and Hotspot Mutations Lead to Oncogenic Activation of the KLF5 Transcription Factor. Cancer discovery *8*, 108-125.

Zhang, X., Choi, P.S., Francis, J.M., Imielinski, M., Watanabe, H., Cherniack, A.D., and Meyerson, M. (2016). Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. Nature Genet *48*, 176-182.

Zhao, J.H., Bacolla, A., Wang, G.L., and Vasquez, K.M. (2010). Non-B DNA structure-induced genetic instability and evolution. Cell Mol Life Sci *67*, 43-62.

Zhao, Y., Xiong, Z., Lechner, E.J., Klenotic, P.A., Hamburg, B.J., Hulver, M., Khare, A., Oriss, T., Mangalmurti, N., and Chan, Y. (2014). Thrombospondin-1 triggers macrophage IL-10 production and promotes resolution of experimental lung injury. Mucosal immunology *7*, 440.

Zhou, C.Y., Johnson, S.L., Gamarra, N.I., and Narlikar, G.J. (2016). Mechanisms of ATP-Dependent Chromatin Remodeling Motors. Annual Review of Biophysics *45*, 153-181.

Zhou, J., Bashey, A., Zhong, R., Corringham, S., Messer, K., Pu, M., Ma, W., Chut, T., Soiffer, R., and Mitrovich, R.C. (2011). CTLA-4 blockade following relapse of malignancy after allogeneic stem cell transplantation is associated with T cell activation but not with increased levels of T regulatory cells. Biology of Blood and Marrow Transplantation *17*, 682-692.

Zhou, X., Lowdon, R.F., Li, D., Lawson, H.A., Madden, P.A., Costello, J.F., and Wang, T. (2013a). Exploring long-range genome interactions using the WashU Epigenome Browser. Nature methods *10*, 375-376.

Zhou, X., Sun, L., de Oliveira, F.B., Qi, X., Brown, W.J., Smolka, M.B., Sun, Y., and Hu, F. (2015). Prosaposin facilitates sortilin-independent lysosomal trafficking of progranulin. J Cell Biol *210*, 991-1002.

Zhou, Y., Kurukuti, S., Saffrey, P., Vukovic, M., Michie, A.M., Strogantsev, R., West, A.G., and Vetrie, D. (2013b). Chromatin looping defines expression of TAL1, its flanking genes, and regulation in T-ALL. Blood *122*, 4199-4209.

Zhou, Y., Kurukuti, S., Saffrey, P., Vukovic, M., Michie, A.M., Strogantsev, R., West, A.G., and Vetrie, D. (2013c). Chromatin looping defines expression of TAL1, its flanking genes, and regulation in T-ALL. Blood *122*, 4199-4209.

Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC bioinformatics *11*, 237.

Zion, M., Benyehuda, D., Avraham, A., Cohen, O., Wetzler, M., Melloul, D., and Benneriah, Y. (1994). PROGRESSIVE DE-NOVO DNA METHYLATION AT THE BCR-ABL LOCUS IN THE COURSE OF CHRONIC MYELOGENOUS LEUKEMIA. Proceedings of the National Academy of Sciences of the United States of America *91*, 10722-10726.

8 Appendix