Corruption of Transcription Factor Networks in

Acute Myeloid Leukaemia

Sandeep Potluri

A Thesis Submitted to the University of Birmingham for the Degree of

DOCTOR OF PHILOSOPHY

Institute of Cancer and Genomic Sciences

College of Medical and Dental Sciences

University of Birmingham

April 2021

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

Abstract

Transcription factors form highly regulated and complex networks. In Acute Myeloid Leukaemia, driver and secondary mutations result in major rewiring of these circuits, resulting in a cancer phenotype. In this thesis, we examined two prominently rewired nodes in AML, Wilms Tumour 1 (WT1) and the Activator Protein-1 (AP-1) family of Transcription factors in order to elucidate the genomic regions at which these factors bind, how perturbation affects transcription and what cellular phenotype is conferred. We studied these transcription factors in the context of two main AML subtypes, t(8;21) and FLT3-ITD mutated AML as well as in healthy CD34$^+$ stem cells.

We blocked AP-1 binding through the expression of a dominant negative FOS peptide which resulted in decreased leukaemic growth *in vitro* and *in vivo*. We found that one main mechanism for this reduced growth was a G1 cell cycle arrest, through regulation of the Cyclin D2 (*CCND2*) gene. This finding was then exploited therapeutically through use of the small molecule inhibitor Palbociclib to inhibit leukaemic growth.

We also found that overexpression of endogenous *WT1* led to increased leukaemic growth whilst knockdown of *WT1* led to decreased leukaemic growth. These effects upon leukaemic growth occurred in an isoform-specific fashion with WT1 +KTS isoforms increasing leukaemic growth in contrast to WT1 –KTS isoforms which decreased leukaemic growth. By assessing the binding sites of these isoforms and differential splicing we uncovered distinct alterations to the

epigenome and transcriptome. This also enabled us to propose models of how

these factors act within a signalling network.

Acknowledgements

Lastly, I dedicate this work to my family. My parents and sister have always been a constant course of support and my father has delighted in me following him down an academic career pathway. And finally I dedicate this work to my long-suffering but always supportive wife, Monisha, who has had to put up with me spending numerous evenings, weekends and holidays in the lab instead of with her.

Contributions

<u>Bioinformatic Analyses</u>

Salam Assi – performed all initial bioinformatics processing including trimming and alignment of reads and peak calling. Produced the heatmaps of differentially bound/ differentially open chromatin sites for the WT1 chapter. Produced all gene regulatory networks. Performed motif co-localisation analyses and transcription factor co-binding analyses.

Sophie Kellaway – produced the heatmaps of differentially bound/ differentially open chromatin sites for the AP-1 chapter.

<u>Experimental Contributions</u>

Paulynn Chin – performed several ChIP-seq experiments including all the EGR1 and RUNX1-ETO ChIP-seq experiments in Kasumi-1 cells for the WT1 chapter, as well as ChIP-seq experiments in Kasumi-1 cells for the AP-1 chapter.

Daniel Coleman – performed the initial primary FLT3-ITD and primary healthy cell transduction experiments with dnFOS. Performed the CBFβ inhibitor and FLT3 inhibitor experiments in primary FLT3-ITD AML cells, followed by RNA-seq. Also performed several primary cell DNase I-seq experiments.

Helen Blair – performed all xenotransplantation experiments.

Natalia Martinez-Soria – performed the shRNA depletion screen experiments.

Anetta Ptasinska – supervised the cloning in the construction of the dnFOS construct. Performed several ChIP-seq experiments in Kasumi-1 cells relating to the AP-1 chapter.

Rosie Imperato – performed several primary cell DNase I-seq experiments.

Anna Pickin – performed the Sp1 ChIP-seq in Kasumi-1 cells.

FACS facility: Adriana Flores-Langarica, Matthew Mackenzie – all FACS sorting.

Sequencing facility: Andrew Beggs, Celina Whalley, Charlie Poxon – running Next Generation Sequencing flow cells.

Contributions also described in the figure legends.

Table of Contents

List of Figures

18

List of Tables

25

Abbreviations

AML – Acute Myeloid Leukaemia

ATAC – Assay for Transposase-Accessible Chromatin

AP-1 – Activator Protein-1

BSA – Bovine Serum Albumin

CD – Cluster of Differentiation

ChIP – Chromatin Immunoprecipitation

C/EBPα – CCAAT/ Enhancer Binding Protein alpha

CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats

cDNA – complementary DNA

DHS – DNase I Hypersensitive Site

DNA - Deoxyribonucleic Acid

DMEM- Dulbecco's Modified Eagle Medium

DMSO - Dimethyl Sulfoxide

dnFOS – dominant negative FOS

EDTA - Ethylenediaminetetraacetic Acid

EGR1 – Early Growth Response 1

EGTA - Ethylene Glycol-bis(β-aminoethyl ether)-N,N,N′,N′-Tetraacetic Acid

ETO – Eight – Twenty One

FACS – Fluorescence-Activated Cell Sorting

FCS – Foetal Calf Serum

FLT3 – FMS-like Tyrosine Kinase 3

GFP – Green Fluorescent Protein

GRN – Gene Regulatory Network

HEK – Human Embryonic Kidney

HSC – Haematopoietic Stem Cell

IDH – Isocitrate Dehydrogenase

IL - Interleukin

IVL - Involucrin

KEGG - Kyoto Encyclopaedia of Genes and Genomes

LSC – Leukaemic Stem Cell

MAPK – Mitogen-Activated Protein Kinase

MRD - Minimal Residual Disease

NTC – Non-Targeting Control

PBS – Phosphate Buffered Saline

PBSC – Peripheral Blood Stem Cell

qPCR – quantitative Polymerase Chain Reaction

RFP – Red Fluorescent Protein

RNA - Ribonucleic Acid

RPMI - Roswell Park Memorial Institute

SDS - Sodium Dodecyl Sulphate

siRNA – small interfering RNA

shRNA – short hairpin RNA

TCGA – The Cancer Genome Atlas

TBP – TATA box Binding Protein

TF – Transcription Factor

WT1 – Wilms Tumour 1

Chapter 1: Introduction

1.1 Haematopoiesis

Haematopoiesis is a carefully regulated and intricate process by which mature blood cells are derived from Haematopoietic Stem Cells (HSCs). Stem cell differentiation permits a response to physiological stresses such as infection and bleeding by increasing numbers of whichever terminally differentiated blood cell is required (Zhao et al., 2014). As early as 1957, Conrad Waddington conceptualised cellular differentiation as a ball rolling down a bifurcating 3D landscape with the eventual cellular fate depending on which path is taken (Waddington, 1957). Lying underneath the landscape are pegs (which we now know represent transcription factors) and guy-wires (which represent the transcriptional output of these transcription factors). Transcription factors drive differentiation down this Waddington landscape through the recruitment of chromatin modifiers and remodellers (see Section 1.6).
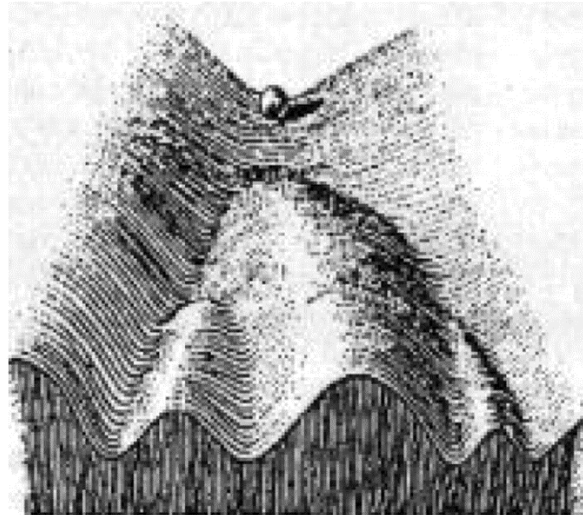
Figure 1-1: Waddington landscape. Adapted from (Waddington, 1957).

1.1.1 Haematopoietic Stem Cells

HSCs are defined by their self-renewal ability and ability to differentiate into any mature haematopoietic cell type. Experimentally, this was first demonstrated by bone marrow transplantation whereby lethally irradiated mice could be rescued by transplantation of bone marrow from another mice and this led to the production of haematopoietic colonies in the spleens of recipient mice (Till and McCulloch, 1961). HSCs can be subdivided into two populations based upon their capability to reconstitute haematopoiesis; Long-term (LT) HSCs are rare and quiescent and have reconstitutive ability for over 3 months and can differentiate to less quiescent and more abundant Short Term (ST) HSCs which have reconstitutive ability for less than one month. ST-HSCs can in turn differentiate to give rise to multipotent progenitor (MPPs) cells which have lost the ability to self-renew but can still differentiate to give any mature haematopoietic cell type (Yang et al., 2005).

31

However, even amongst HSCs, a number of studies have demonstrated considerable heterogeneity. By performing limiting dilution and single cell transplantation experiments, it was shown that different HSCs had different skews in the ratio of mature myeloid and lymphoid cells reconstituted and this suggested that lineage priming was happening even at the HSC level (Muller-Sieburg et al., 2004);(Dykstra et al., 2007). Later, by using a von Willebrand Factor (vWF) reporter mouse model, LT-HSCs which expressed high vWF were isolated and in transplantation experiments were found to have enhanced long-term platelet reconstitution as well as myeloid bias. These vWF$^+$ cells could give rise to vWF$^-$ progeny which had lymphoid bias and this suggested that vWF$^+$ cells were at the apex of the haematopoietic hierarchy (Sanjuan-Pla et al., 2013).

Such single cell transplantation experiments were refined further by cell fate mapping of individual HSCs. This involved virally barcoding HSCs, followed by the transplantation of multiple HSCs and then genome sequencing upon reconstitution. By comparing barcodes in the HSCs to mature haematopoietic cells it is possible to match up HSCs to their progeny and identify any lineage biases in HSCs (Lu et al., 2011).

However, a major limitation of transplantation studies is that they are non-physiological as chemotherapy or radiotherapy is used in recipient mice which disrupts their bone marrow niche. Steady state haematopoiesis can be investigated by inducible barcoding systems in mouse models and performing fate mapping analysis showed that most steady state haematopoiesis was established by ST-HSCs and instead LT-HSCs mainly contributed to haematopoiesis in foetal life (Busch et al., 2015).

1.1.2 Haematopoietic Differentiation

Haematopoietic stem cells differentiate into progenitors with increasingly restricted lineage potential.

It was previously thought in the classical models of haematopoiesis that the first committed stages of haematopoietic differentiation was the formation of myeloid progenitors and lymphoid progenitors. However, clonally mapping several populations in cord blood and adult marrow revealed that there were lymphoid primed multipotent progenitors (LMPPs) which could still give rise to myeloid cells but not to megakaryocyte or erythroid cells (Doulatov et al., 2010). These LMPPs could differentiate to common lymphoid progenitors (CLPs) (Kondo et al., 1997) or granulocyte-monocyte progenitors (GMPs).

Alternatively, MPPs could differentiate to form CMPs (common myeloid progenitors) (Akashi et al., 2000) rather than LMPPs. CMP can differentiate to form GMPs or megakaryocyte-erythroid precursors (MEPs).

The figure below shows a summary of haematopoiesis with immunophenotypically-defined stages defined in traditional discrete models of haematopoiesis.
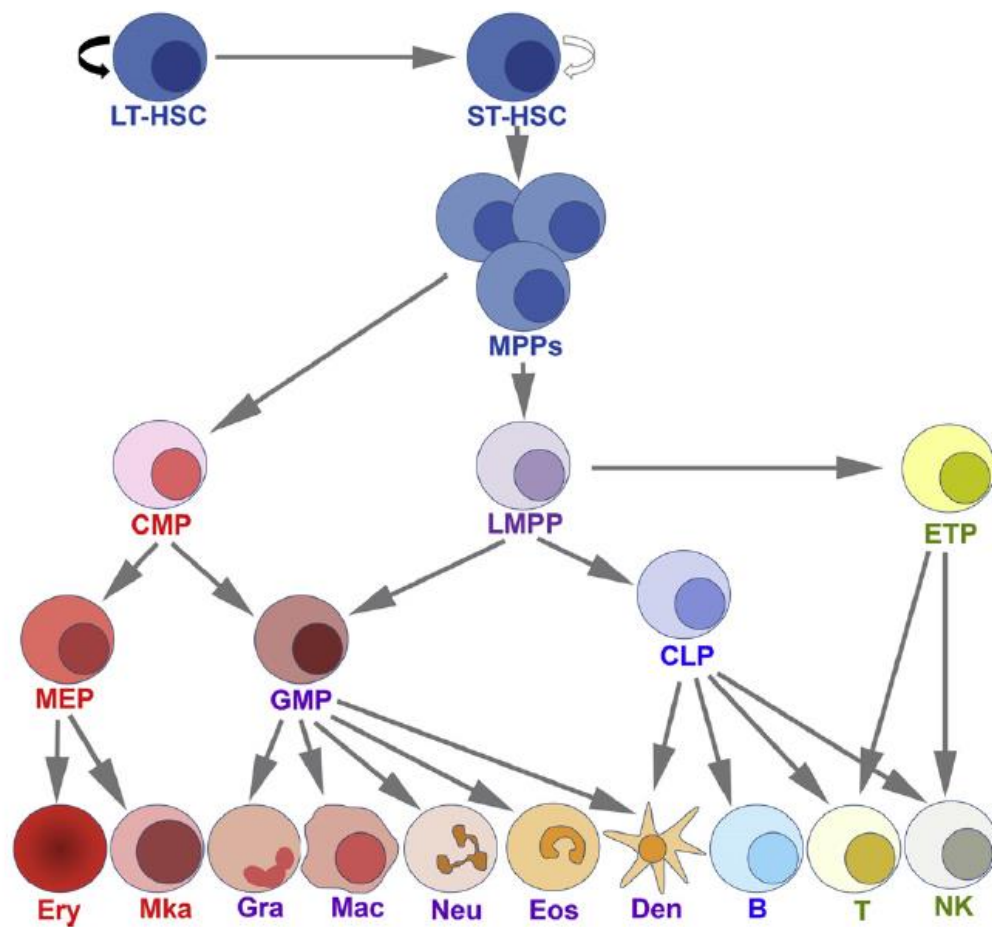
Figure 1-2: Haematopoietic hierarchy.

LT-HSC = Long Term Haematopoietic Stem Cell, ST-HSC = Short Term Haematopoietic Stem Cell, MPP = Multipotent Progenitor, CMP = Common Myeloid Progenitors, LMPP = Lymphoid-primed Multipotential Progenitors, MEP = Macrophage-Erythroid Progenitors, GMP = Granulocyte-Monocyte Progenitors, CLP = Common Lymphoid Progenitor, ETP = Early T-lineage Progenitor, Mka = Megakaryocyte Progenitor, Gra = Granulocyte Progenitors, Mac = Macrophage Progenitor, Neu = Neutrophil, Eos = Eosinophil, Den = Dendritic cells, B = Progenitor B Cell, T = Progenitor T Cell, NK = Progenitor NK Cell. From (Panigrahi and Pati, 2012).

More recently, looking at haematopoiesis at a single cell level with scRNA-seq and scATAC-seq (Buenrostro et al., 2018), it has become clear that cells go through a continuum of differentiation and it is possible to predict an individual cell's fate from their epigenetic landscape. Rather than immunophenotypically-defined stages shown above, bioinformatic clustering of cell identities based upon chromatin accessibility and the transcriptome revealed several subclasses such

as 4 stages of Common Myeloid Progenitors (CMP) and 3 stages of Granulocyte-Monocyte Progenitors (GMP). But more attractive than assigning more and more sub-stages is a continuous model of differentiation; this concept moves away from discrete models whereby at each stage there is stepwise lineage commitment but rather that cells are continuously and gradually acquiring lineage commitment over time (Busch et al., 2015).

Intriguingly, around 70% of all gene expression, methylation (Ji et al., 2010) or chromatin accessibility changes (Corces et al., 2016) between HSCs and early progenitors is independent of lineage commitment and most of these changes are related to the cell going from a quiescent (Wilson et al., 2008), autophagy-dependent, low mitochondrial and low protein translation activity cell, which confers relative resistance to stress and damage, to more differentiated progenitors which are more rapidly cell cycling and utilising oxidative metabolism.

1.2 Acute Myeloid Leukaemia

Acute Myeloid Leukaemia (AML) is a haematological cancer with an incidence of about 3000 cases per year in the UK, which occurs as a consequence of at least 2 mutations. Descriptions of leukaemia came as early as the 19th century with observations from Alfred Velpeau and others that patients had milky-looking serum. John Bennett, Rudolf Virchow and others described post-mortem findings in leukaemia such as splenomegaly and leukaemic infiltration of the liver. Furthermore, Alfred Donné and others described the first morphological descriptions of leukaemia through microscopic examination (Kampen, 2012).

Morphological descriptions today of AML are based upon the appearance of myeloid blasts, for example in the French-American-British (FAB) classification (Bennett et al., 1976). The classification is based upon extent and under which lineage leukaemic blasts have partially differentiated and this broadly corresponds with the haematopoietic progenitor cell type that is the cell of origin.

A seminal genomic sequencing study showed that AML actually demonstrates rather little mutational heterogeneity compared with many other cancers with only 23 different recurrently mutated genes. An individual patient's AML has an average of 5 recurrent mutations and a further 8 less recurrent mutations (Ley et al., 2013). However, only 2-3 of these mutations are essential for leukaemogenesis with other mutations representing random events in haematopoietic stem cells or co-operating mutations (Welch et al., 2012).

The table below summarises the various classification methods for AML mutations and mutation prevalence.

Cytogenetic and Molecular Based Class	Next Gen Sequencing Based Class	The Cancer Genome Atlas Based Class	Prevalence in AML
Class I: Activated Signalling Mutations e.g. *FLT3, KIT, RAS*	Class I: Activated Signalling Mutations e.g. *FLT3, KIT, RAS*	Class 1: Transcription Factor Fusions e.g. , t(8;21), inv(16)	18%
		Class 2: *NPM1* Mutations	27%
		Class 3: Tumour Suppressor Gene Mutations e.g. *TP53, PHF6*	16%
	Class II: Transcription and Differentiation Factor Mutations	Class 4: DNA Methylation-related Mutations e.g. *TET2, IDH1/2, DNMT3A*	44%

Class II: Transcription and Differentiation Factor Mutations e.g. C/EBPα, t(8;21), inv(16)	e.g. *C/EBPα*, t(8;21), inv(16), *RUNX1*	Class 5: Activated Signalling Mutations e.g. *FLT3, KIT, RAS*	59%
		Class 6: Chromatin Modifier Mutations e.g. *ASXL1, EZH2, MLL*	30%
	Class III: Epigenetic Modifier Mutations e.g. *TET2, DNMT3A, ASXL1*	Class 7: Myeloid Transcription Factor Mutations e.g. *C/EBPα, RUNX1*	22%
		Class 8: Cohesin Complex Mutations e.g. *STAG2, RAD21, SMC1/2*	13%
		Class 9: Spliceosome Complex Mutations e.g. *SRSF2, U2AF3S*	14%

Table 1-1: Classification and prevalence of recurrent mutations in AML.

1.2.1 Acquisition of Leukaemogenic Mutations

The order of mutations in AML has been investigated through several methods. One such method is looking at population studies analysing mutations preceding and during the development of overt AML (Abelson et al., 2018). Alternatively, the order may be inferred from variant allele frequency numbers in genome sequencing experiments whereby mutations that temporally occur earlier have the highest variant allele frequency (Papaemmanuil et al., 2016). Lastly, looking at mutation frequencies of individual AML colonies in colony formation assays allows us to reconstruct the acquisition of mutations and to look for the original founder clones with just driver mutations (Hirsch et al., 2016).

The order of mutations is critical to leukaemogenesis; for example in biallelic *CEBPA* AML, the *CEBPA* (CCCAT Enhancer Binding Protein Alpha) mutations

must precede a *CSF3R* (Colony Stimulating Factor 3 Receptor) mutation in order for AML to develop (Braun et al., 2019).

In many AML patients, the first mutation is often a Class III mutation such as *DNMT3A* (DNA cytosine-5-Methyltrasferase 3A), *TET2* (Tet methylcytosine dioxygenase 2) or *ASXL1* (ASXL transcriptional regulator 1). Such mutations can precede overt AML by around 8 years (Abelson et al., 2018) and increase in frequency with age, particularly from the age of 70 onwards (Jaiswal et al., 2014) and is termed Clonal Haematopoiesis of Indeterminate Potential (CHIP).

These Class III mutations lead to increased stem cell renewal. In murine transplant experiments, *Dnmt3a* null HSCs expanded 200 fold better than wildtype HSCs in secondary transplants without any change in proliferation or apoptosis which suggested enhanced self-renewal. Furthermore, upregulation of stem cell related genes such as *Runx1* was seen (Challen et al., 2011). A recent study looking at DNMT3A and TET2 knockout mice or patients with CHIP using single cell RNA-seq, single cell ATAC-seq and single cell methylation assays demonstrated that there was a profound differentiation defect with biases in erythroid: myelomonocytic differentiation. Intriguingly, this was linked to alterations in transcription factor binding depending on whether or not there were alterations in methylation of the motifs of CG rich motifs, due to the DNMT3A or TET2 mutations (Izzo et al., 2020). However, of note, Class III mutations are not sufficient to cause AML and further mutations are required.

The majority of AML carry a Class II mutation conferring a block in haematopoietic differentiation and enhanced self-renewal. The molecular mechanisms of Class

II mutations such as *RUNX1* and *RUNX1-ETO* will be discussed extensively later (see Sections 1.6.4 and 1.6.5). Interestingly, these Class II mutations can be congenitally acquired including those affecting *RUNX1*, *GATA2*, *CEBPA* and *ETV6* (Charrot et al., 2020). Furthermore, studies looking at newborn Guthrie Cards have identified AML-associated translocations such as *RUNX1-ETO* (Wiemels et al., 2002), *MLL-AF10* (Jones et al., 2003), *PML-RARA* and *CBFB-MYH11* (McHale et al., 2003) even at birth suggesting pre-natal origin. Most of these congenital or pre-natal changes greatly increase the risk of AML but it is important to note that some, especially *GATA2* and *CEBPA* mutations have very variable or low penetrance (Al Seraihi et al., 2018) suggesting that most Class II mutations in themselves are not sufficient to cause AML.

Lastly, most AML will carry a Class I mutation that constitutively activates intracellular signalling independent of growth factors and this allows for enhanced cell proliferation. These mutations include *KIT*, which is the receptor for Stem Cell Factor (Williams et al., 1992) and *FLT3*, which is the receptor for the FLT3 ligand (Lyman et al., 1993b).

The mechanism by which mutations may be acquired is of great interest. In t(8;21) AML, the presence of the Class II *RUNX1-ETO* translocation creates an environment of genomic instability which makes cells much more likely to acquire a Class I mutation that is needed for leukaemogenesis (Forster et al., 2015). Further to this, in the context of the analogous disease model of Acute Lymphoblastic Leukaemia (ALL), Mel Greaves and colleagues have speculated that inflammation may be the cause of acquisition of class I mutations. In their study of monozygotic twins harbouring pre-natal ALL driver translocations, they

found a surprisingly poor concordance to development of ALL (Bateman et al., 2010); (Cazzaniga et al., 2011) and speculate from epidemiological studies that inflammation may be key to mutation acquisition (Gilham et al., 2005). Alternatively, in adults it is known that systemic chronic inflammation increases with age (Sanada et al., 2018) and this might be responsible for the leukaemogenesis (and other cancers) in older adults. Intriguingly, in the CANTOS trial of IL1β blockade with canakinumab, a reduced lung cancer incidence was seen (Ridker et al., 2017).

More causal evidence implicating inflammation comes from murine models of AML whereby mesenchymal stromal cell inflammation drove disease evolution of pre-leukaemia to leukaemia (Zambetti et al., 2016). In another mouse model of inflammation with chronic arthritis, it was seen that high IL1 levels led to raised *Spi1* expression, HSC quiescence and myeloid skewing, priming the cells for myeloid leukaemia (Chavez et al., 2019). The IL1 also led to the expansion of *Cebpa* deficient MPPs (Higa et al., 2020) which also primed cells for myeloid leukaemogenesis.

More recently, with the advent of routine Next Generation or targeted panel sequencing in clinical practice and gene perturbation studies have led to increased understanding of other AML co-operating mutations such as in the Cohesin complex (Mazumdar et al., 2015) (see also Section 1.5.5). But other AML susceptibility mutations such as those affecting *ANKRD26*, *SAMD9* and *DDX41* are still relatively poorly understood. Even less well understood are Single Nucleotide Polymorphisms (SNPs) from Genome Wide Association studies such

as rs75797233 (near the *BICRA* gene), rs57706619 (near the *B3GNT3* gene) and rs2039647 (near the *KLF12* gene) (Walker et al., 2019).

1.2.2 Leukaemic and Pre-Leukaemic Stem Cells

Since there is an ordered acquisition of mutations in AML in haematopoietic stem cells, this process gives rise to intermediary cells which have not acquired all the mutations sufficient for leukaemogenesis. This was notably studied by genomic sequencing of individual cell-derived colonies, prior to the advent of single cell sequencing (Jan et al., 2012). Such pre-leukaemic cells were found to have only class III or class II mutations which gave them a competitive advantage over surrounding HSCs and allowed them to exist as an expanded pool of cells before later acquiring a class I mutation to give a leukaemic stem cell (LSC). LSCs differ from leukaemic blasts by maintaining serial transplantation ability (Bonnet and Dick, 1997).

Pre-leukaemic and leukaemic stem cells are thought to be quiescent and can survive chemotherapy and act as a reservoir for relapse termed 'minimal residual disease', as assessed by serial genotyping assays in patients undergoing chemotherapy (Corces-Zimmerman et al., 2014). Further evidence for this idea came from early RT-qPCR studies which could identify *RUNX1-ETO* oncogene mRNA even in patients in complete remission whose leukaemic blasts had been eradicated. The breakpoint at which the RUNX1-ETO translocation occurred stayed the same between presentation, remission and relapse suggesting that these stem cells were surviving chemotherapy and repopulating upon relapse. When immunophenotypically sorting these cells, it was found that pre-leukaemic stem cells resides in the Lineage$^-$ CD34$^+$ CD38$^-$ CD90$^+$ compartment whilst

leukaemic stem cells resided in the Lineage$^-$ CD34$^+$ CD38$^-$ CD90$^-$ compartment (Miyamoto et al., 2000).

It was later shown that LSCs could reside within the CD38$^+$ compartment as well. However, the CD38$^-$ compartment had a higher stem cell frequency compared with the CD38$^+$ compartment, as assessed by limiting dilution xenotransplantation assays. In addition, the CD38$^-$ LSCs were higher up in the haematopoietic hierarchy as they had LMPP potential and could differentiate to give lymphoid cells compared to the CD38$^+$ LSCs which were more GMP-like. Furthermore, the CD38$^-$ LSCs could differentiate to give CD38$^+$ LSCs but not vice-versa (Goardon et al., 2011) suggesting that CD38$^-$ LSCs are the more primitive stem cells.

More recently, single cell mutation analysis of minimal residual disease in patients being treated for AML has allowed for the identification of those patients at risk of relapse and to differentiate from those patients who only have pre-leukaemic clonal haematopoiesis (Ediriwickrema et al., 2020).

Intriguingly, others have challenged the paradigm of relapse occurring from quiescent pre-leukaemic/ leukaemic stem cell. A study of serial examinations of patients' stem cells during the course of chemotherapy have shown that stem cells are cycling and eradicated by chemotherapy (Boyd et al., 2018). Instead, they propose that relapse is derived from 'Leukaemia Regenerating Cells'; these cells in the bone marrow niche are activated by chemotherapy and bone marrow niche molecules such as dopamine.

1.2.3 Subclonal Evolution in AML

Whilst a core number of 2 or 3 mutations exist in all AML, considerable evidence points to the existence of subclones through the acquisition of further cooperating mutations. This idea was notably demonstrated through serial whole genome sequencing of patients at different time points in their chemotherapeutic treatment where it was seen that the variant allele frequencies changed over the course of treatment suggesting subclones that were expanding or regressing (Ding et al., 2012). More direct evidence has come from FACS sorting clones based upon single cell proteomic data and characterising the transcriptomes and chromatin accessibility of these subclones. Of considerable clinical interest, it was seen that different subclones responded differentially to therapy, for example a clone which has gained a FLT3-ITD mutation was sensitive to the FLT3 inhibitor Quizartinib whilst the parental clone without this mutation was not (de Boer et al., 2018).

Whilst FACS sorting provides the methodology for understanding individual clones, it is laborious and limited to larger subclones. Single cell mutation analysis allows for the understanding of which mutations truly co-occur within the same cell, for example *WT1* mutations were found to co-occur with other signalling mutations such as *KIT*, *FLT3-ITD* and *NRAS* and this could not be deduced by only analysing bulk Variant Allele Frequency data (Miles et al., 2020b).

Transcriptionally distinct clones with the same mutations have also been identified by performing single cell RNA-seq on AML (van Galen et al., 2019) and this again identified clonal selection upon use of AML therapies.

1.3 Transcription

To understand what regulates the identity of cells undergoing haematopoietic differentiation, we have to understand the mechanisms of transcription and gene regulation and how it is perturbed in cancer.

Transcription is carried out by RNA polymerase enzymes which recognise the promoter region, open up the DNA duplex, synthesise RNA and then leave the promoter, contingent upon signalling. Following promoter escape, the RNA polymerase extends the RNA chain until it reaches a termination sequence whereupon DNA and RNA are released. RNA polymerase I synthesises large ribosomal RNA, RNA polymerase II synthesises mRNA and non-coding RNA and RNA polymerase III synthesises transfer RNA and small ribosomal RNA. Each one differs in terms of regulation and associated factor but discussion will focus on RNA polymerase II and mRNA synthesis.

Firstly, RNA polymerase II needs to access the promoter region and this can only happen in open chromatin sites (Schones et al., 2008) (see Section 1.4.1). Eukaryotic genes have different types of promoter. One type is made up of CpG islands, which are regions typically over 200bp long that contain CG repeats that occupy more than 60% of the genomic sequence. CpG islands impair nucleosome formation and hence keeps chromatin open for RNA polymerase II to access (Ramirez-Carrozzi et al., 2009); this is particularly employed in the case of housekeeping genes where transcription occurs ubiquitously in most cell types. Another type of promoter contains a TATA box upstream of the transcription start

site and these tend to be found at cell type specific genes whose expression changes during differentiation including haematopoietic differentiation.

At the pre-initiation complex, RNA polymerase II is bridged to promoter elements though the assembly of initiating factors, which are required as RNA polymerase II cannot recognise promoter elements by itself. One such factor initiating factor is the TATA box binding protein (TBP), which can recruit RNA polymerase II to TATA box promoters (Bushnell et al., 2004). However, many promoter sequences lack obvious DNA recognition sequences and so transcription initiating factors must bind specifically to promoters by other mechanisms including the recognition of flanking nucleosomes (Vermeulen et al., 2007) or physical properties of DNA such as 'bendability' (Engel et al., 2017).

Following recognition of the promoter region, DNA needs to be opened up by separating the two strands of DNA, through the action of the ATP- dependent DNA unwinding enzyme XBP (Xeroderma Pigmentosum B), thus creating the 'transcription bubble' (Grunberg et al., 2012). Once DNA is opened up, the Mediator complex recruits CDK7 (cyclin dependent kinase 7) in order to phosphorylate the tail of RNA polymerase II (Kornberg, 2005) and allow it to start transcriptional elongation by the addition of a new nucleotides to the RNA chain and stabilises these nucleotides with a phosphodiester bond.

Once transcription has initiated, the elongation complex forms containing the RNA-DNA hybrid duplex, RNA polymerase and other elongation factors which regulate elongation. The transcription factor IIS re-aligns the elongation complex and restarts elongation whenever it is paused when facing with physical barriers

such as nucleosomes (Cheung and Cramer, 2011). RNA polymerase II often pauses around 50bp downstream of the transcription start site and is stabilised in the paused position by the factors NELF (negative elongation factor) and DSIF (DRB sensitivity inducing factor) (Yamaguchi et al., 2013). CDK9 is needed to phosphorylate and inactivate Protein Phosphatase I, which normally dephosphorylates the paused position stabilising elongation factors DSIF and NELF. Thus de-repression of a regulator of transcriptional elongation pausing, allows for continued transcriptional elongation (Marshall and Price, 1995).

Elongation stops at the end of the gene body, where it pauses at poly(A) sequences. Protein complexes CPSF (cleavage and polyadenylation specificity factor) and CSTF (cleavage stimulation factor) recognize the poly-A signal (Mandel et al., 2006) and recruits Poly(A) polymerase to add around 200 adenines to the tail.
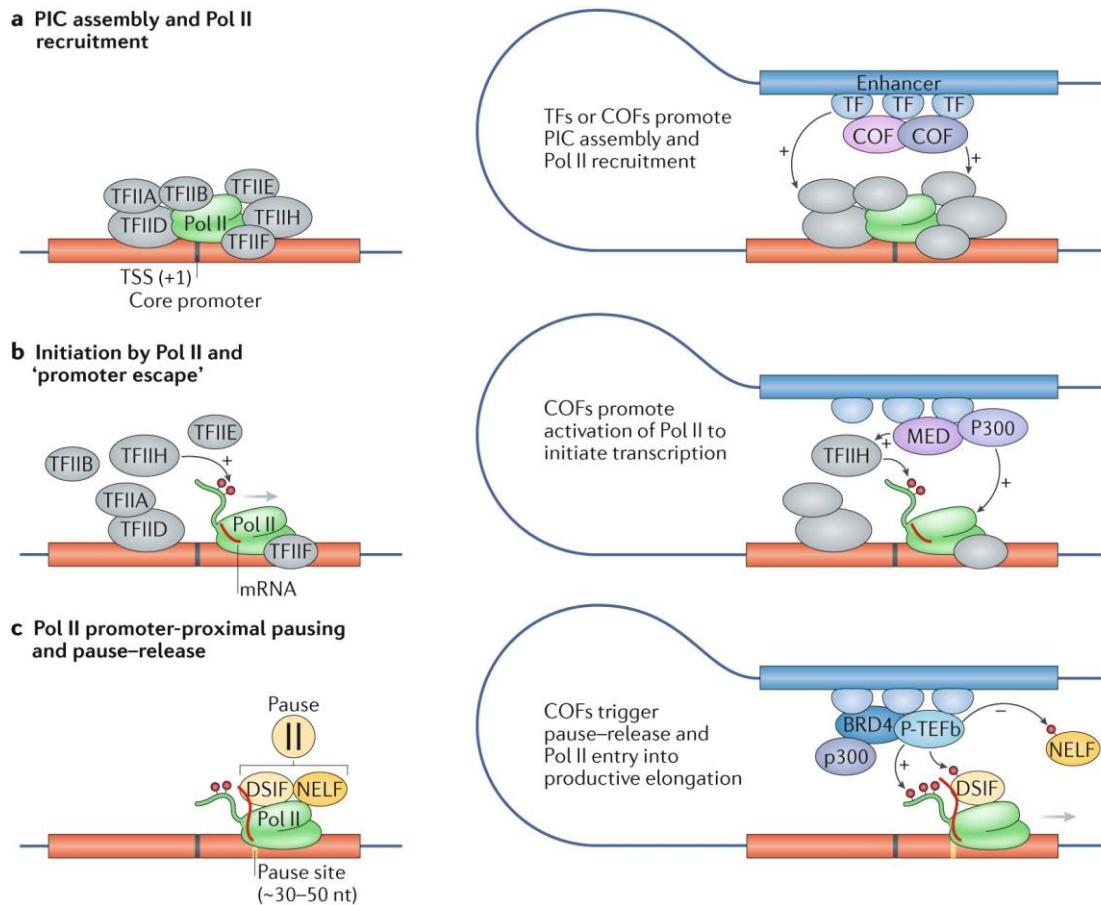
Figure 1-3: Summary of Transcription adapted from (Haberle and Stark, 2018). COF = cofactors, Pol II = RNA polymerase II, MED = Mediator.

1.3.1 Regulation of Transcription

Transcription can be regulated through several mechanisms such as chromatin structure (which impairs access of RNA polymerase II and other factors to the promoter), DNA modifications, histone modifications, cis-regulatory elements and transcription factors. These will be discussed in turn between Sections 1.3.2 and 1.6)

1.3.2 DNA Modifications

DNA may be methylated at the 5th carbon of cytosine (producing 5-methylcytosine) through the action of DNA methyltransferase enzymes (DNMT) transferring methyl groups from S-adenosyl methionine. DNMT1 mainly contributes to the maintenance of methylation upon DNA replication by preferentially binding to hemi-methylated DNA and adding the methyl group to the daughter strand (Song et al., 2012). DNMT3A/B however contributes to de nova methylation and *DNMT3A* is frequently mutated in AML. The R882H mutant DNMT3A acts as a dominant negative protein by inhibiting the homotetramerisation of DNMT3A (Russler-Germain et al., 2014).

Ten-Eleven Translocation (TET) enzymes are involved in the removal of methyl groups in DNA and catalyse the sequential oxidation of 5-methylcytosine (5mC) to 5-hydromethylcytosine (5hmC) and then 5-carboxylcytosine (5caC) and 5-formylcytosine (5fC) with the utilisation of ATP (Ito et al., 2011). *TET2* is frequently mutated in AML leading to a loss of its catalytically activity and hence reduced 5hmC levels and this was seen to promote myeloid tumourigenesis in mouse models (Ko et al., 2010). α-ketoglutarate is used as a co-factor in the oxidation of 5mC in the tricarboxylic acid cycle (TCA) by the enzyme Isocitrate dehydrogenase (IDH) 1 and 2. *IDH1* and *IDH2* are also frequently mutated in AML (Figueroa et al., 2010).

Epigenetic readers are protein that have a groove or surface that can bind to methylated DNA and modified histones. Differential binding of epigenetic readers to methyl- and hydroxymethyl- cytosine can lead to changes in transcription through histone modification (see also Section 1.4.2). Methylcytosines can be

48

read by methyl-CpG binding protein (MBD) epigenetic readers with specificity determined by flanking DNA bases. MBD1 can bind to any CpG, MeCP2 needs 4 flanking A/T around the CpG, MBD2 needs an adjacent C or G. MeCP2 can recruit repressors such as Sin3A and HDACs. MBD1 interacts with the histone methyltransferase SETDB1 to promote histone H3K9 trimethylation, which is a repressive histone mark. MBD2 has been proposed as a demethylase enzyme which can remove DNA methylation (Bhattacharya et al., 1999). MBD 3 and 4 preferentially bind to 5hmc over 5mc and can recruit the NuRD complex and promote histone deacetylation.

DNA Modifications	Nomenclature	Chromatin-Reader	Attributed Function
5-methylcytosine	5mC	MBD1/2, MeCP2	Transcription
5-hydroxymethylcytosine	5hmC	MBD3/4	Transcription
5-formylcytosine	5fC	unknown	unknown
5-carboxylcytosine	5caC	unknown	unknown

Table 1-2: DNA modifications, chromatin readers and functional consequence of the DNA modification.

1.4 Chromatin

1.4.1 Chromatin Structure

It has long been recognised from X-ray diffraction studies that chromatin structure is comprised of DNA interspaced with histones around roughly every 200bp i.e. the nucleosome (Kornberg, 2005). It was initially thought that histones served purely structural purposes in order to compact around 2m of linear DNA into a nucleus that is less than 10μm in diameter. However, it rapidly became clear that

nucleosomes impede transcription *in vitro* (Knezetic and Luse, 1986) and that deletion of either entire histones or even just their basic tails led to massive changes in transcription (Han and Grunstein, 1988). Consequently, changes in chromatin accessibility, which is dependent upon the extent to which DNA is wound around the nucleosome, is a key cellular regulatory mechanism for transcription, replication and repair of DNA.

The nucleosome consists of an octameric complex made up of histone subunits; two H3-H4 heterodimers dimerise and are bound above and below by H2A-H2B heterodimers. Basic arginine-rich grooves in each histone provide a surface with electrostatic interactions with the minor groove of DNA. 147bp of DNA is wound ~1.65 times around each nucleosome with approximately another 20-80bp linking two nucleosomes (Luger et al., 1997).
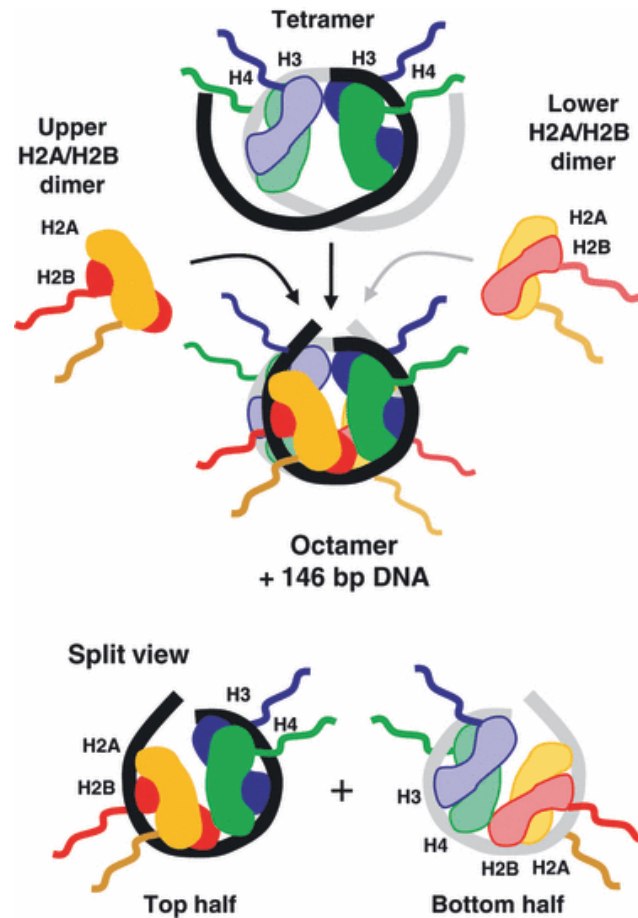
Figure 1-4: Nucleosome structure.

From (Cockerill, 2011)

Subnucleosome structures have also recently been described which contain some but not all of the subunits in the octamer. These subnucleosome structures include hexasomes, prenucleosomes, half-nucleosomes and tetrasomes but are thought to have minimal effect of transcription (Fei et al., 2015).

In vitro experiments electron microscopy experiments have shown that histone octamers spontaneously assemble into long arrays separated by linker DNA giving a 'bead on a string' 10nm structure (Olins and Olins, 2003). However, *in vivo* visualisations showed that chromatin domains actually exist as 30nm fibres

whereby nucleosomes interdigitate in a zig-zag type structure (Robinson et al., 2006). This discrepancy may have arisen from the non-physiological salt-free conditions of the *in vitro* experiments. In addition, Histone H1 which exists outside of the core octameric nucleosome can stabilise the 30nm fibre (Thoma et al., 1979).

Chromatin must de-condense from such a tightly wrapped structure in order for transcription to be permitted and it has long been recognised that regions of chromatin de-condense during mitosis called 'euchromatin', whereas there are areas that stay tightly packed called 'heterochromatin' (Heitz, 1928). In addition, distinct nuclear localisation patterns exist with the most active euchromatin localised towards the centre of the nucleus where there is the highest abundance of transcription factors and transcriptional machinery components. In contrast, heterochromatin tends to be at the periphery and can interact with the nuclear lamina which adds further stability to the closed state of the chromatin (Guelen et al., 2008).

Transcription factors are able to open up chromatin and therefore genes that are most actively transcribed tend to be in the most de-condensed regions and are more susceptible to nuclease digestion (Wood and Felsenfeld, 1982) and this forms the basis of DNase I Hypersensitive Site mapping (DNase I-seq) whereby the elements involved in transcription tend to be at hypersensitive sites . An alternative to DNase I-seq is the Assay for Transposase Accessible Chromatin (ATAC-seq) which utilises a hyperactive transposase enzyme that preferentially cuts and inserts at open chromatin site (Buenrostro et al., 2013).

1.4.2 Histone Modifications

The histone tails and globular domains are subject to a huge array of post-translational modifications including methylation, acetylation, phosphorylation and ubiquitination that affect the way chromatin condenses at any particular point. Many of these modifications affect the net charge on a nucleosome and hence its interaction with DNA. For example, histone acetylation of the lysine tail of histone H3 has been shown *in vitro* to be easier to displace DNA from (Ito et al., 2000). This is due the abrogation of the electrostatic interaction between the positively charged Lysine tail and the negatively charged phosphodiester backbone of DNA. Furthermore, all the histones can be polyADP ribosylated by polyADP ribose polymerase (PARP) which conveys a negative charge to the histone and causes electrostatic repulsion with the phosphodiester backbone, again leading to a relatively open chromatin state (Hassa et al., 2006).

Chromatin effects of other histone modifications cannot be explained as simply as changes in electrostatic interaction and instead affect the way in which other proteins interact with them such as epigenetic readers. For example, H3K4me3 which is associated with active promoters, can recruit the Chromodomain Helicase DNA Binding Protein 1 (CHD1) and the NURF complex via a Plant Homeobox Domain finger. CHD1 and the NURF complex are also ATP-dependent Chromatin Remodellers which can open up chromatin (Wysocka et al., 2006).

The table below summarises histone modifications, the epigenetic readers and the effect of these modifications on transcription.

Chromatin Modification	Nomenclature	Modifying Enzymes	Chromatin-Reader Motif	Attributed Function
Acetylation	K-ac	Histone acetyltransferases (HAT) for acetylation e.g. GNAT, MYST and CBP/p300 families, Histone deacetylases (HDACs) for deacetylation	Bromodomain, Tandem, Plant Homeodomain (PHD) fingers	transcription, repair, replication, and condensation
Methylation (lysine)	K-me1, K-me2, K-me3	Histone lysine methyltransferases for methylation e.g. DOT1, PRC1/2 Histone lysine demethylases e.g. LSD1, JARID1A, JARID1C, UTX, JMJD2	Chromodomain, Tudor domain, malignant brain tumour domain (MBT) domain, proline-tryptophan-tryptophan-proline (PWWP) domain, plant homeodomain (PHD) fingers, WD40/b propeller	transcription and repair
Methylation (arginine)	R-me1, R-me2s, R-me2a	Protein arginine methyltransferases (PRMTs) for arginine methylation	Tudor domain	transcription
Phosphorylation (serine and threonine)	S-ph, T-ph		14-3-3, BRCA1 C terminus domain (BRCT)	transcription, repair, and condensation
Phosphorylation (tyrosine)	Y-ph		SH2a	transcription and repair
Ubiquitylation	K-ub	Ubiquitin E1, E2 and E3 enzymes for ubiquitylation and de-ubiquitin isopeptidases for removal of this moiety.	ubiquitin interaction motif (UIM), inverted ubiquitin interaction motif (IUIM)	transcription and repair
Sumoylation	K-su	Ubiquitin E1, E2 and E3 enzymes for sumoylation	sumo interaction motif (SIM)	transcription and repair
ADP ribosylation	E-ar	poly-ADP ribose polymerase (PARP) for ADP ribosylation and poly-ADPribose polymerase for removal of this moiety	Macro domain, poly ADP-ribose binding zinc finger (PBZ) domain	transcription and repair
Deimination	R/Cit	Peptidyl deiminase PADI4	unknown	Transcription and decondensation
Proline isomerisation	P-cis5P-trans	Proline isomerases	unknown	transcription
Crotonylation	K-cr		unknown	transcription
Propionylation	K-pr		unknown	unknown
Butyrylation	K-bu		unknown	unknown
Formylation	K-fo		unknown	unknown
Hydroxylation	Y-oh		unknown	unknown

O-GlcNAcylation (serine and threonine)	S-GlcNAc; T-GlcNAc	O-GlcNAc Transferase for GlcNAcylation and β-N-Acetylglucosaminidase for removing this moiety	unknown	transcription

Table 1-3: Histone modifications, modifying enzymes, epigenetic readers and functional consequences of the histone modifications

Whilst Chromatin Immunoprecipitation assays have been pivotal in the understanding the role of histone modifications, many discoveries have been driven through improvements in Mass Spectrometry; increases in sensitivity and novel labelling methods have allowed for it to be a quantitative assay (Stunnenberg and Vermeulen, 2011) and allowed for mechanistic associations to be made between histone modification and whether there is active or inactive transcription.

Major chromatin modifications associated with active transcription include acetylation of the lysine tail of Histone H3 including H3K9ac, H3K14ac, H3K27ac and H3K122 by histone acetyltransferase enzymes such as the GCN5 related N-acetyltransferases (GNAT), MYST (MOZ, Ybf2, Sas2, Tip60) and CBP/p300 families. Some of the earliest evidence for this association between acetylation and transcription arose from studies of fractionated chicken embryo erythrocytes (Hebbes et al., 1988).

Several Lysine methylation modifications are also associated with active transcription including H3K4me1 which is associated with enhancers, H3K4me3

which is associated with promoters and H3K36me3 which is deposited throughout transcribed genes (Barski et al., 2007).

The best characterised chromatin modifications associated with heterochromatin and hence lack of transcription are methylation of the Lysine tail of Histone H3. In constitutive heterochromatin with permanently silenced genes such as those near centromeres and telomeres, a high level of H3K9me3 deposition is found. However, around conditionally silenced genes in facultative chromatin, high H3K27me3 deposition is found and maintained by its epigenetic reader, the Polycomb Repressive Complex (Barski et al., 2007). Such "poising" or temporary silencing of genes keeps the options open to either reactivate or permanently silence them depending on the developmental program, such as different stages of haematopoietic differentiation.

However, it must be noted that active and inactive histone modifications are not mutually exclusive, as evidences by bivalent domains. This has classically been described with the presence of both the active H3K4me3 and inactive H3K27me3 modifications found in ChIP-reChIP experiments (Bernstein et al., 2006). Here, in an embryonic stem cell differentiation model, bivalent domains were found to occur at transcription factor genes expressed only at low levels. It was proposed that such bivalent domains were keeping these transcription factors silenced so that pluripotency could be maintained but yet poised for activation when lineage-specific gene expression programs were commenced.

1.4.3 Histone Cross-Talk

Considerable histone modification 'cross-talk' occurs whereby the presence one modification affects a second modification. Such cross-talk may simply be as a result of a competition whereby a particular lysine in the H3 Lysine tail can only biochemically undergo one modification e.g. acetylation, methylation or ubiquitylation. Alternatively, the first histone modification may enhance a second modification as exemplified by the requirement of Histone H2B monoubiquitylation for robust methylation of Histone H3K4 by the COMPASS complex (Dover et al., 2002) and methylation of H3K79 by DOT1L (Wood et al., 2003). By contrast, Arginine methylation at Histone H2 prevents recruitment of the COMPASS complex (Guccione et al., 2007).

Since histone modifications are so important in regulating transcription and hence cell identity, it is also very important that they are maintained and passed onto daughter cells. This must occur despite the structural challenge of chromatin condensation during mitosis. In addition, the daughter cells must incorporate newly synthesised histones in the same places. An elegant example of how this happens is seen in the recruitment of the epigenetic reader HP1 (heterochromatin protein 1) to the constitutive heterochromatin marks H3K9me2/3 by its chromodomains. HP1 can in turn interact with SUV39 which is a histone methyltransferase that can re-methylate H3K9 in adjacent nucleosome, hence maintaining H3K9 methylation following mitosis (Bannister et al., 2001).

1.4.4 Histone Variants

Whilst the canonical nucleosome consists of DNA wound around an octamer of Histones H2A, H2B, H3 and H4, several histone variants also exist such as H3.3,

cenH3, H2A.Z, H2A.X, H2A.B and macroH2A variants of histone H2A. Whereas the canonical histone subunits are all synthesised and incorporated into DNA in the S phase of the cell cycle, histone variants are synthesised and incorporated throughout the cell cycle, perhaps conferring a greater degree of regulatory control (Talbert and Henikoff, 2017). For example, MacroH2A consists of H2A with a macro domain at the C-terminus of the protein and was correlated with transcriptional repression by interfering with SWI/SNF chromatin remodelling (see Section 1.4.5) and transcription factor recruitment (Angelov et al., 2003).

1.4.5 ATP-dependent Chromatin Remodellers

Chromatin can also be remodelled by ATP-dependent chromatin remodellers. These remodellers are classified into four families: Switch/Sucrose Non-Fermenting (SWI/SNF), Imitation Switch (IWSI), Chromodomain-Helicase DNA binding (CHD) and Inositol requiring 80 (INO80) families.

SWI/SNF remodellers, including the mammalian BAF complex, remove H2A/H2B dimers in a step-wise manner, followed by the other histone subunits in order to release naked DNA (Lorch et al., 2006). Consequently, BAF complexes are found at promoters and enhancers (Lorch et al., 2006) including super-enhancers (see Section 1.5.2).

IWSI remodellers bind to the basic patch of the N-terminal tail of Histone H4 and linker DNA and promotes DNA translocation (Clapier and Cairns, 2012). CHD remodellers form core catalytic components of histone deacetylase complexes such as the mammalian NuRD complex and can also complex with methylated

DNA binding proteins such MBD2/3 (Le Guezennec et al., 2006) (see also Section 1.3.2 on DNA modifications).

INO80 remodellers modulate gene activation/ repression through catalysing the exchange of H2AZ/H2B dimers with H2A/H2B dimers (Brahma et al., 2017) (see also Section 1.4.4 on histone variants).

1.5 Cis-Regulatory Elements

Cis-Regulatory elements drive gene expression and they consist of promoters, enhancers, silencers and insulators.

1.5.1 Promoters

Promoters are defined as the minimum DNA sequence required for the initiation of transcription via the recruitment of transcriptional machinery such as RNA polymerase and general transcription factors to form the pre-initiation complex (Sainsbury et al., 2015) (see Section 1.3). Other than TATA boxes, TATA-less promoters include GC and CCAAT boxes. GC boxes contain at least one repeat of the motif (G/T)GGGCGG(G/A)(G/A)(C/T) to which the zinc finger class of transcription factors can bind including the ubiquitously expressed Sp1 transcription factor family (Kuwahara et al., 1993); consequently these promoters tend to initiate transcription of housekeeping genes. However, other zinc fingers transcription factors including EGR (Early Growth Response) and WT1 (Wilms Tumour 1) can bind there too. CCAAT boxes bind the NFYA transcription factor and are present at the genes such at the beta globin gene and are essential for optimal transcriptional activity at these genes (Bi et al., 1997).

1.5.2 Enhancers

Enhancers are usually small (~300bp) and often distal elements (at least 1kb away) from promoters but can interact with them by DNA looping (see also Section 1.5.5). Enhancers were first identified by expressing a Simian Virus 40 (SV40) 72bp DNA sequence in cis with genes such as rabbit beta globin and seeing a huge increase gene transcripts (Banerji et al., 1981); this enhancer appeared to work bidirectionally and even over 1kb away from the promoter.

Enhancers tend to contain a high density of motifs to which transcription factors can bind which can greatly alter gene transcription depending on the type of bound transcription factors. It is therefore of no surprise that chromatin accessibility at these enhancer elements is very carefully regulated (Kim and Shiekhattar, 2015). Genes that are of important regulatory relevance such as tissue-specific genes tend to be under the control of numerous enhancers and this is well exemplified by the human beta globin gene which has 5 elements in its 'Locus Control Region' (Forrester et al., 1986) (Blom van Assendelft et al., 1989).

Enrichment for H3K27ac tends to be high at enhancers as well as a high H3K4me1:H3K4me3 ratio and consequently these histone marks have been used by the ENCODE consortium and others to imply where enhancers are (Ernst et al., 2011).
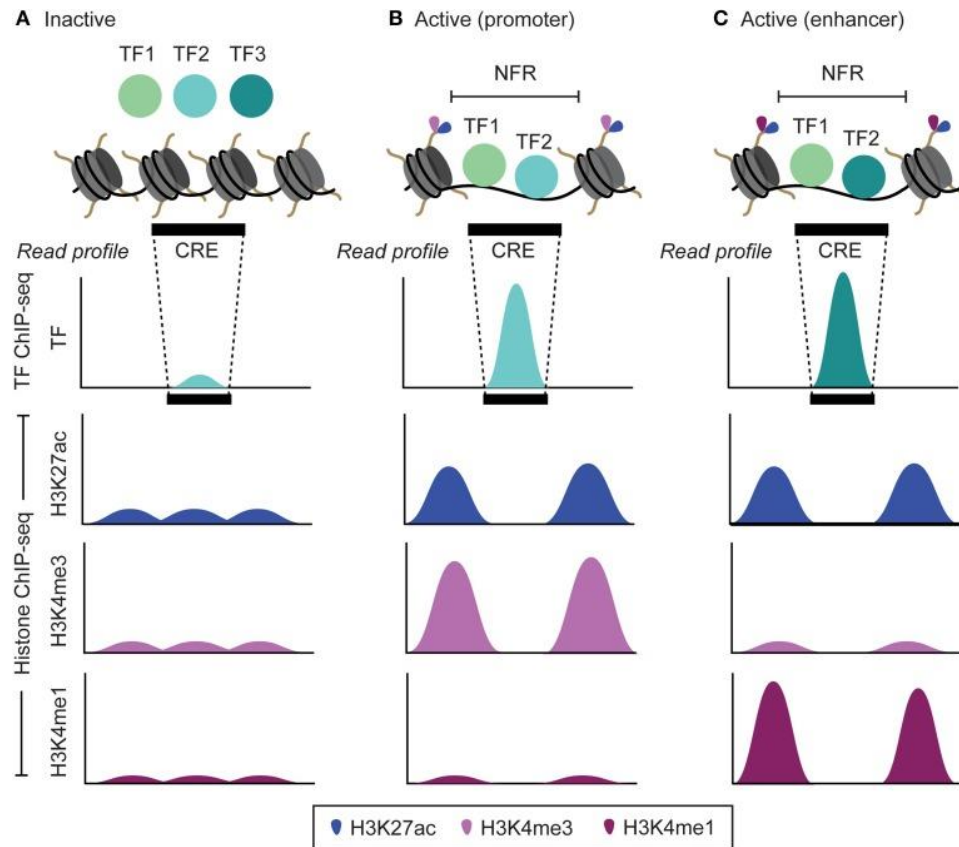
Figure 1-5: Histone landscape flanking enhancers compared to inactive regions and promoters.

NFR = Nucleosome-Free Region, CRE = Cis-Regulatory Element. From (Pundhir et al., 2015).

A more functional method of assessing enhancers is to place a candidate enhancer upstream of a minimal promoter from a reporter gene as was originally done by Banerji and colleagues (Banerji et al., 1981) but an easier reporter readout may be achieved by using a fluorescent protein reporter (Chiocchetti et al., 1997). However, the drawback of such transgenic reporter assays is that they may not represent promoter-enhancer interactions in their native chromatin state and also cannot assess interactions between multiple regulatory elements. Consequently, others have used CRISPR activation or CRISPR inactivation to

recruit activation or inhibitory cofactors to putative enhancer elements to perturb their effect upon gene expression in their native state (Simeonov et al., 2017) Each of these functional methodologies can also be scaled up to assess enhancers genome-wide through either a massively parallel reporter assay (Nguyen et al., 2016) or a CRISPR activation/inactivation library screen (Rosenbluh et al., 2017) (see also Section 1.10.1).

Super-enhancers are complex arrays of distal elements that even more strongly increase the transcription of both cell-type specific and important regulatory genes such as SOX2 and have a higher density of chromatin modifications such as H3K27 acetylation (Hnisz et al., 2013).

Whilst enhancers can recruit transcription factors and affect chromatin structure locally, the question remained for many years as to how they influenced promoter activity despite being separated by distances of at least 1kb and often much further. By using chromosome conformation capture which relied on formaldehyde-crosslinking chromatin in its native structure, it was seen that enhancers appeared to interact directly with their promoters; in higher eukaryotes this was first characterised at the beta globin Locus Control Region (Tolhuis et al., 2002). Importantly, using such chromosome conformation techniques combined with next generation sequencing such as promoter Capture HiC has allowed annotation of which enhancers or other cis-regulatory elements interact with which promoters in various healthy human blood cell types (Javierre et al., 2016), (Mifsud et al., 2015) or in AML (Assi et al., 2019).

1.5.3 Duality of Promoters and Enhancers

Despite classical definitions of promoters and enhancers, it is now appreciated that there is considerable overlap in their functions. For example, when active, both are found in nucleosome depleted regions, bound by RNA polymerase II and can divergently and bidirectionally activate transcription (Core et al., 2014). Protein coding gene promoters give rise to mRNA on their sense strands and promoter upstream transcripts (PROMPTS) on the antisense strand, which are degraded by nuclear exosome complexes (Preker et al., 2008). Enhancers give rise to enhancer RNA (eRNA) bidirectionally which are also degraded by nuclear exosomes.

But to add to this complexity, 17 of 45 identified regulatory regions with *in vivo* enhancer activity in human embryonic stem cells at the *POU5F1* locus were gene promoters (Diao et al., 2017). Furthermore, 3% of over 20,000 tested gene promoters in K562 cells had strong enhancer activity *in vitro* (Dao et al., 2017).

However, active gene promoters still produce on average 17 fold more RNA than enhancers as assessed by cap analysis of gene expression (CAGE-seq) (Andersson et al., 2014). These promoters also tend to be in CG rich areas and are enriched for H3K4me3 (Barski et al., 2007).

1.5.4 Silencers

It has long been recognised that repressive elements, later named 'silencers' can inhibit transcription (Ptashne, 1967) and their discovery in mammals came with study of the rat insulin 1 gene locus (Laimins et al., 1986). Later, their importance in lineage specification in haematopoiesis was established by the study of how

CD4 expression was repressed in committed CD8[+] T cells (Sawada et al., 1994). Furthermore, during the GATA2 to GATA1 switch in haematopoietic differentiation, GATA1 binds to a silencer of *Kit* and there is the establishment of a silencer- *Kit* promoter chromatin loop and a loss of enhancer- *Kit* promoter loops (Jing et al., 2008).

Silencers are thought to interfere with transcription initiation by the recruitment of repressors and co-repressors such as DNMT and HDAC (Kolovos et al., 2012) or by interfering with the assembly of the General Transcription Factors that are needed to initiate transcription (Ogbourne and Antalis, 1998).

Genome-wide identification of silencers has been performed using Massively Parallel Reporter Assays on a library of candidate DNase I Hypersensitive sites that lack characteristic promoter and enhancer histone marks (Doni Jayavelu et al., 2020).

1.5.5 Insulators and DNA looping

Considering that enhancers increase transcription by directly interacting with promoters, it is essential that enhancers do not accidently interact with other promoters. Insulators are elements, which are in over 75% of cases bound by the CTCF transcription factor (Dixon et al., 2012) as well as Cohesin protein complexes (Wendt KS et al., Nature, 2008). CTCF recruits Cohesins which bridges DNA loops between CTCF sites in order to create distinct 'insulated neighbourhoods' that can constrain the spread of heterochromatin as it cascades along DNA (Dixon et al., 2012). However, DNA can be extruded through the Cohesin ring to extrude loops of ever increasing sizes that only stop growing

when a bound CTCF blocks the Cohesin ring (Sanborn et al., 2015). These loops of DNA can then interact with adjacent loops to form independent Topologically Associating Domains (TAD) within which interactions are more likely to occur.
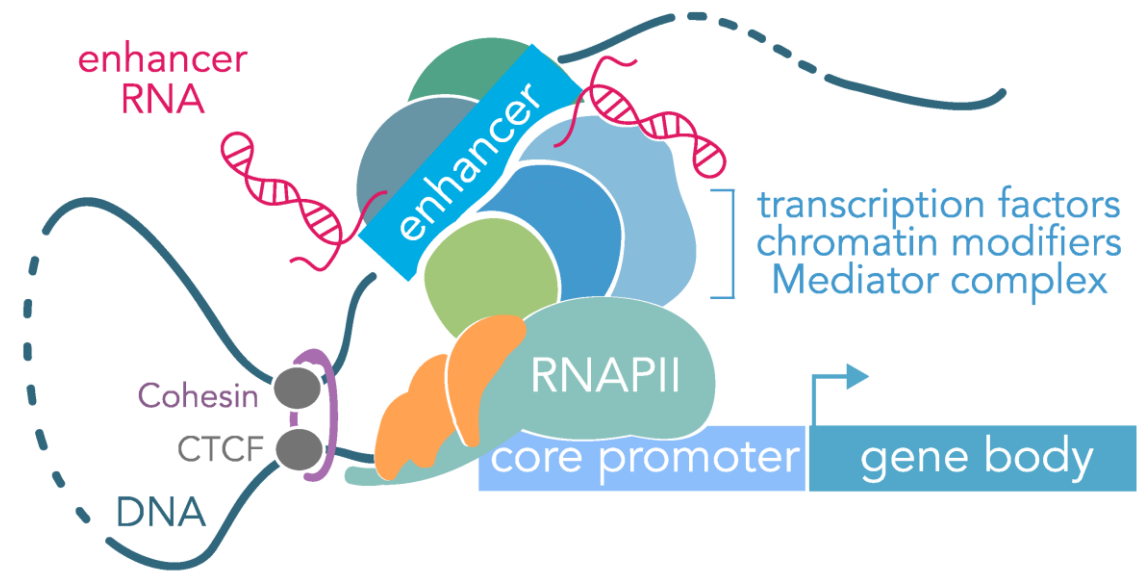
Figure 1-6: Chromatin looping.

Adapted from (Carullo and Day, 2019).

Consequently when insulators are disrupted either through germline or acquired mutations, they can have significant phenotypic consequences or cause cancer (Hnisz et al., 2016). Similarly Cohesin complex mutations such as STAG2, RAD21, SMC1 and SMC2 are mutated in as much as 13% of all AML.

However, interactions within TADs cannot be purely be explained by fleeting, chance encounters between adjacent extruded DNA loops. Transcription factor complexes between two adjacent loops certainly stabilise interactions and this notably includes the Mediator complex and the Cohesin loading factor, Nipbl

(Kagey et al., 2010). Furthermore, the zinc finger transcription factor Ying Yang 1 (YY1) is also important in stabilising interactions. YY1 acts in an analogous manner to CTCF in that it binds to a related consensus DNA sequence (GCCAT), particularly in hypomethylated DNA, and interacts with Cohesins. However, whilst CTCF is enriched at insulators, YY1 appears to present at sites of interaction between promoters and enhancers. Consequently depletion of YY1 disrupts enhancer-promoter contacts, as assessed by HiC, and alters gene expression (Weintraub et al., 2017).

However, looping does not occur purely through transcription factor complexes, recently the role of phase-separated condensates has been described (Sabari et al., 2018). Phase separation is a process by which molecules can separate out into a dense phase and a dilute phase by thermodynamic driving forces that try to exclude water. Many co-activators such as MED1 contain 'intrinsically disordered regions' which allows for their condensate formation. With high resolution fluorescence imaging, the presence of nuclear puncta with co-localisation of co-activators such as MED1 and enhancers can be visualised. Addition of 1,6-hexanediol to disrupt the phase condensates led to an 80% reduction in MED1 levels at the KLF4 superenhancer and hence a decrease in transcription. Alterations in cellular salt concentrations to disrupt phase separation also had similar effects.

1.6 Transcription Factors

Transcription factors (TFs) are DNA-binding proteins that alter the rate of transcription and haematopoiesis is exquisitely carefully regulated by them. Transcription factors can recruit cooperating factors or components of the transcription apparatus directly (Ong and Corces, 2011) or can activate transcription of the enhancers themselves producing enhancer RNAs which can promote transcription of active genes via complexes such as Mediator (Lai et al., 2013).

1.6.1 Lineage Specifying Transcription Factors

Whilst nearly half of all transcription factors encoded in the human genome are expressed in any one cell (Vaquerizas et al., 2009), a small number of master transcription factors can dramatically alter cell identity and are known as lineage specifying TFs. Such lineage specifying TFs are characterised by several features including: i) they tend to be expressed at higher levels than other TFs in a particular lineage ii) often co-occupy enhancers with other lineage specifying TFs iii) occupy most enhancers including super-enhancers of active cell-specific genes iv) autoregulate their transcriptional program through feedback loops (Lee and Young, 2013).

Within haematopoiesis, knockout experiments in mice have demonstrated the crucial roles of lineage specifying transcription factors. For example, C/EBPα is essential for differentiation to mature granulocytes (Zhang et al., 1997a) and GATA-1 for differentiation to erythroid and megakaryocyte cells (Pevny et al., 1991).

Remarkably, even after terminal differentiation, transcription factors can drastically alter haematopoietic cell fate and this is perhaps best exemplified by experiments that ectopically expressed *CEBPA* in precursor B cells or even mature B cells which was able to convert them into myeloid cells (Xie et al., 2004). The higher the *CEBPA* expression, the greater was the reprogramming efficiency. It however should be noted that the corollary of converting mature myeloid cells to other cell lineages has never been seen and that forced expression of such aberrant transcription factors normally leads to apoptosis (Iwasaki et al., 2005). Nevertheless this demonstrates the power of transcription factors.

The most convincing demonstration of the potency of transcription factors in changing cell fate were seminal experiments by Yamanaka demonstrating that their overexpression even in terminally differentiated cells could completely reverse cell fate; transgenic expression of the pluripotency factors *Oct4*, *Klf4*, *Sox2* and *Myc* in fibroblasts was found to able to reprogram them back to pluripotent stem cells (Takahashi and Yamanaka, 2006).

Transcription factor mediated reprogramming has not just been seen in cell lines. An elegant experiment where *Gata1* was overexpressed in lineage committed progenitors (Macrophage-Erythroid Progenitors), sorted straight out of mouse bone marrow, demonstrated convincingly that lineage switching can also be seen in primary cells (Heyworth et al., 2002).

Transcription factor action is remarkably tissue specific. For example, as well as its role in terminally differentiated erythroid and megakaryocyte cells, GATA

family members are also crucial for HSC development and self-renewal as evidenced by *Gata1*-null mouse experiments (Tsai et al., 1994). Such tissue-specific action is likely to be due to the presence of cooperating transcription factors such as RUNX1 which mediate cis-regulatory changes in the epigenome. Indeed, whilst either *Runx1*$^{+/-}$ or *Gata2*$^{+/-}$ heterozygotes are viable in mouse models, compound heterozygotic Runx1$^{+/-}$ and Gata2$^{+/-}$ embryos died due to lack of HSC formation (Wilson et al., 2010).

However, transcription factors frequently have antagonistic, instead of cooperating transcriptional functions. Indeed, GATA1 is known to promote erythroid-megakaryocyte differentiation whilst PU.1 is known to promote monocytic differentiation from common myeloid progenitors (Nerlov and Graf, 1998b) and consequently only one of these transcriptional programs can be active at a given time. High levels of GATA1 inhibit PU.1 by displacing its cofactor JUN (Zhang et al., 1999). Conversely, *PU.1* expression in erythroid precursors is associated with a change in active complexes such as Creb Binding Protein (CBP) into repressive complexes with the Retinoblastoma protein at GATA1 target genes such as alpha and beta globin (Stopka et al., 2005).

Another important example of cross-antagonism of lineage-specifying transcription factors is found in macrophage-erythroid progenitors. KLF1 represses FLI-1 mediated transcription and FLI-1 is capable of repressing KLF1 target genes (Starck et al., 2003). Consequently perturbations of *KLF1* or *FLI1* expression through shRNA were found to alter lineage commitment whereby knockdown of *KLF1* led to increased megakaryocytic differentiation and

69

knockdown of *FLI1* led to increased erythroid differentiation (Bouilloux et al., 2008).

Therefore, lineage-specifying transcription factors not only activate their own transcriptional program but they also shut down the actions of antagonistic transcription factors. The figure below summarises key lineage-specifying transcription factors and when cross-antagonism occurs.
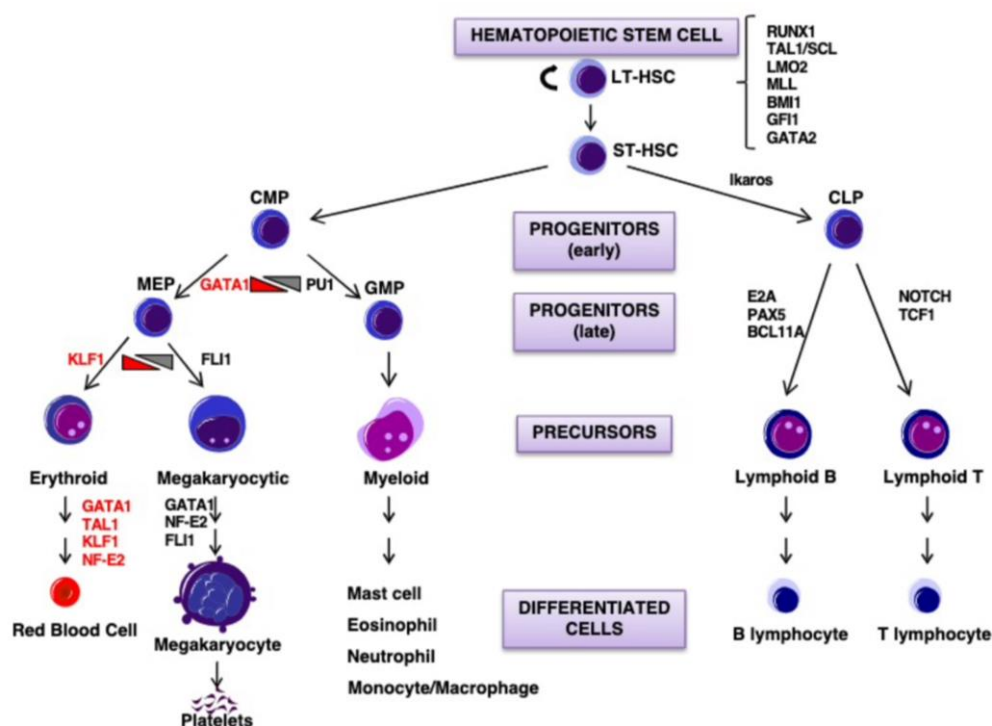
Figure 1-7: Key lineage-specifying transcription factors in haematopoiesis and cross-antagonism at branch points.

Slide from Marjorie Brand, University of Ottawa.

Most transcription factors are generally reliant on the presence of open chromatin in order to be able to access and bind to their DNA binding sites, or at least need to cooperate with another transcription factor which can recruit chromatin

remodellers. However, pioneer factors have an intrinsic ability to scan partial DNA sequence motifs that are exposed on the surface of the nucleosome and thus access sites that are not accessible to most transcription factors. They were first discovered in experiments looking at transcription factor binding to a liver-specific enhancer of the *alb1* gene in mouse embryonic development. In the extrahepatic endoderm, an area of closed chromatin as determined by DNase I hypersensitivity, FOXA (pioneer factor) binding could unexpectedly be seen (Gualdi et al., 1996). Furthermore, in *in vitro* experiments, FOXA2 could bind to mononucleosomes and dinucleosomes and help GATA4 bind to adjacent sites (Cirillo and Zaret, 1999).

Pioneer factors can open up condensed chromatin de novo by one of two mechanisms; firstly, an active process of locally unwinding chromatin such that local segments of DNA are exposed for other transcription factors to bind. Alternatively, it can also occur by a passive process whereby pioneer factors just sit on condensed chromatin; however when cooperating transcription factors are nearby they can rapidly recruit and assemble other proteins to enable chromatin modifications (Zaret and Carroll, 2011).

1.6.2 Gene Regulatory Networks

Putting together knowledge of transcription factor cooperation, where transcription factors bind and how transcription factors alter the gene expression of other transcription factors via interaction with their cis-regulatory elements, it is possible to derive Gene Regulatory Networks (GRN). Notably this was done in a large study looking at gene expression at 38 states of haematopoiesis to identify distinct but tightly connected transcription factor networks; functional validation

71

was performed through ChIP-seq and RNA interference mediated perturbation (Novershtern et al., 2011).

GRNs consist of nodes (TFs or TF families) that are highly interconnected which are important for the maintenance of a specific cell type. Other nodes appear to be more peripheral but are connected to key non-TF genes that actually specify a cell type. Many highly interconnected TFs can bind to and regulate themselves as well as other TFs which can in turn regulate the first. This feedback process allows carefully control of TF expression levels and binding patterns, which are a hallmark of self-renewing hematopoietic stem and progenitor cells.

However, within malignancies including AML, this careful balance of TF expression and binding that is maintained by the GRN is disturbed by mutations. Such mutations in turn leads to further alterations in the GRN which may lead to aberrant haematopoietic differentiation. Even different mutations in a single TF encoding gene such as *RUNX1* can lead to completely different cellular identities with distinct chromatin landscapes (Matheny et al., 2007). Malignant cells alter their GRNs in order to find a stable state that is compatible with growth and this may be through the aberrant activation of genes encoding lineage-inappropriate TFs (Assi et al., 2019, Ray et al., 2013, Somerville et al., 2015). Furthermore, mutation of one allele encoding a TF can lead to shifts in the GRN so that it now is dependent on the function of the wild-type allele (Ben-Ami et al., 2013).

However, an understanding of the GRNs can identify therapeutic vulnerabilities. Our Lab has deduced the GRNs for each mutational subtype of AML (Assi et al., 2019). By identifying highly interconnected AML-specific nodes which are not

present in healthy CD34$^+$ cells, leukaemia specific vulnerabilities may be seen. These vulnerable nodes were confirmed functionally through knockdown studies including of NFIX, FOXC1 and POU4F1 (Assi et al., 2019), IKZF2 (Park et al., 2019) and MYB (Volpe et al., 2013).

Two further highly interconnected nodes that appear to be important in all mutational subtypes of AML examined are the Activator Protein-1 (AP-1) and Wilms Tumour 1 (WT1) nodes (see section 1.6.6 and section 1.8) and which make up much of the focus of this thesis.

1.6.3 Stoichiometry between Co-activators and Co-repressors

The relative abundance of various transcription factors and co-activators and repressors is often inferred from RNA expression from RNA-seq experiments. However, we know that protein levels are not always well correlated to RNA expression for reasons such as post-translational modifications affecting protein stability. Some studies have explored relative abundance of these proteins by using single cell mass cytometry (Palii et al., 2019), which allows for comparison of relative abundance of a particular protein between two different cells or samples. However, the relative abundance between two different proteins cannot be deduced from this due to differing antibody affinities to proteins in mass cytometry.

Instead, quantitative mass spectrometry with normalisation to spiked-in synthetic isotopically labelled peptides can be employed for absolute quantification of any protein. This was performed in a seminal study (Gillespie et al., 2019) in an erythroid differentiation model from human CD34$^+$ cord blood cells. When

examining the relative levels of two antagonistic transcription factors, GATA1 and PU.1 which promote mutually exclusive erythroid and myeloid lineage commitment (as previously discussed in Section 1.6.1), there was a gradual and correlated change in the ratios of GATA1:PU.1 protein and RNA expression during differentiation and so RNA can be used as a surrogate marker for protein expression in this case. However, in the case of another well-described transcription factor switch during differentiation of GATA2 to GATA1 (Katsumura and Bresnick, 2017), RNA expression correlated very poorly with protein expression. Looking at RNA alluded to a switch in abundance of *GATA2* to *GATA1* during differentiation suggesting that lineage commitment might just be a change in competitive binding of each factor. However, proteomic analysis instead suggested that GATA1 is always more highly expressed than GATA2 at any stage of differentiation; this alludes more to a model whereby differentiation is determined by GATA2 binding to its specific (i.e. non-GATA1) binding sites and absolute GATA2 levels rather than any competition with GATA1.

Furthermore, absolute protein abundance studies can be used to explore the stoichiometry between various transcription factors, co-activators, co-repressors and enhancers. It was found that co-repressors (such as HDAC1/2/3, CHD4, DNMT1) were around 100 times more abundant than co-activators (such as CBP, EP300, UTX, DOT1L, MLL3) whereas RNA expression suggested equal levels. This could be explained by the fact that co-activators are a lot less stable that co-repressors as demonstrated in a cycloheximide chase experiment (Gillespie et al., 2019). Intriguingly, co-activators were found at such low abundance, that there were fewer co-activator molecules per cell than number of active

enhancers. This gives rise to a model whereby transcription factors compete for extremely limited numbers of co-activators in order to recruit them to specific enhancers and activate a specific transcription program, within a repressive nuclear environment with a large number of co-repressor molecules. This result indicates that gene silencing is the default state unless it is counteracted by activators.

1.6.4 RUNX1

RUNX1 is a master regulator of haematopoiesis and has been shown to be of paramount importance both in the development of HSCs (Okuda et al., 1996) and later differentiation to mature blood cells. Conditional knockout in murine models of *Runx1* led to a differentiation block of T cells in precursor states, reduced B cell numbers and vastly reduced platelet counts and this is similarly seen in patients with congenital *RUNX1* mutations with a 'Familial Platelet Disorder' (Song et al., 1999).

RUNX1 contains a RUNT domain with which it can bind to DNA via the consensus sequence TGTGGT (Speck and Baltimore, 1987). RUNX1 heterodimerises with another protein CBFβ which increases its affinity for DNA binding approximately 10 fold and structural insights into this have come from its crystal structure (Tahirov et al., 2001) whereby CBFβ restricts movement of mobile parts of the RUNT domain.

RUNX1 uses various mechanisms of increasing transcription at its binding sites. RUNX motifs are often adjacent to ETS motifs and the recruitment of ETS family members such as PU.1 further increases affinity for DNA binding (Gu et al.,

2014). Other proteins are also very important in RUNX1 function such as the adaptor proteins p300 and CBP which in turn recruit further transcription factors such as NFk-B, cMYB and AP-1 (Kamei et al., 1996). However, RUNX1 also has the ability to repress transcription in other binding sites via the recruitment of nSin3A/B which in turn recruit Histone Deacetylases (Lutterbach et al., 2000). Consequently, RUNX1 has the ability to finely activate and repress in order to very specifically reprogram the epigenome.

1.6.5 RUNX1-ETO

The t(8;21) translocation occurs in 7% of AML and is a Class II mutation conferring a differentiation block. The protein contains most of the RUNT domain fused to ETO resulting in the loss of the transactivation domain of RUNX1 (Erickson et al., 1992). Genome-wide studies have shown that this translocation resulted in the loss of Histone 3 Lysine 9 acetylation at RUNX binding sites whilst knocking down RUNX1-ETO with siRNA restored these histone marks (Ptasinska et al., 2014). This is due to the ability of RUNX1-ETO to recruit a repressive complex including Histone Deacetylases via N-CoR (Gelmetti et al., 1998).

In addition, RUNX1-ETO can directly inhibit transcription factors through protein-protein interactions (Westendorf et al., 1998) and this is best exemplified by its interaction with C/EBPα which is a key myeloid lineage-specifying transcription factor (Zhang et al., 1997a). Other direct interactions include the ETS family of protein such as ERG and FLI1 (Martens et al., 2012) as well as E proteins such as HEB (Zhang et al., 2004b) and E2A. Finally RUNX1-ETO complexes with LMO2, LDB1 and LYL1 (Wadman et al., 1997) and shRNA knockdown of any of this complex increased the survival of AML xenotransplanted mice (Sun et al.,

2013). In the β globin locus and others, the LDB1 complex augments promoter-enhancer interactions by DNA looping (Krivega et al., 2014).

The structure of the RUNX1-ETO protein is also crucial to its function. RUNX1-ETO can oligomerise through its NHR2 (Nervy Homolog Region 2) domain via a crucial α-helix. Expression of a polypeptide (NC128) which inhibits this interaction (Wichmann et al., 2007) or mutations in this region (Liu et al., 2006) caused decreased proliferation and increased cell death in t(8;21) cell line models.

ChIP-Seq experiments showed that RUNX1-ETO binds to as many as a third of all RUNX1 binding sites (Ptasinska et al., 2014). Therefore, RUNX1-ETO has a considerable impact genome-wide in reprogramming the epigenome and it is thought that it acts in a dominant negative manner. Phenotypically, the effects are tissue-specific; within the context of haematopoiesis, a profound block in haematopoietic differentiation occurred with very few neutrophils being produced (Burel et al., 2001). However, when RUNX1-ETO was knocked into the *Runx1* locus in mouse embryos, they died about 13.5 days after gestation with central nervous system haemorrhages and lack of foetal liver haematopoiesis (Okuda et al., 1998). To investigate its effects at a later stage in haematopoiesis, conditional knock-in mice via Cre-mediated recombination have shown that HSCs containing RUNX1-ETO have increased self-renewal ability as shown in both serial colony formation replating assays (Higuchi et al., 2002). Mimicking this finding *in vitro* by transducing RUNX1-ETO into cord blood, again increased HSC self-renewal is seen (Mulloy et al., 2003).

Generally, RUNX1-ETO cannot cause AML by itself and is thought to need cooperating mutations. In t(8;21) AML, the most common cooperating mutation is in KIT which is present in 19% of t(8;21) AML (Pollard et al., 2010) and leads to an aggressive AML in mice xenotransplanted with t(8;21) cells. Mechanistically, KIT mutations e.g. N822K cause increased cell proliferation, reduced apoptosis and an increase in DNA repair (Wichmann et al., 2014).

Other cooperating mutations include the Cohesin complex, activating mutations such as NRAS, KRAS, FLT3, JAK2 and CBL as well as in epigenetic modifiers such as ASXL1 and IDH1/2 (Duployez et al., 2016). Transduction of the G12D mutated NRAS into RUNX1-ETO expressing cells increased colony formation, serial replating ability, allowed for cytokine-independent growth and upregulated BCL2 to reduce apoptosis (Chou et al., 2011).

Whilst alternate forms of RUNX1-ETO such as AE9a with alternate splicing in exon 9 and AML1-ETOtr which has a C-terminal mutation causing a frameshift have been shown to be able to cause AML in mice without any co-operating mutation, this was found to be due to non-physiological levels of overexpression (Link et al., 2016). The AE9a transcript is seen in approximately 70% of human t(8;21) AML (Yan et al., 2006) whereas the AML1-ETOtr version is only found in a mutated murine model but both lack the C-terminal NH3 and NH4 domains which recruit N-CoR and other transcriptional repressors. These transcriptional repressors inhibit the Cell Cycle by downregulating CDK4 and CCND3 and so the de-repression caused by the alternate forms of RUNX1-ETO can allow for full leukaemogenesis (Yan et al., 2004).

To elucidate the effects of RUNX1-ETO without cooperating mechanisms, which characterises the pre-leukaemic state, two groups have transduced RUNX1-ETO into healthy cord blood cells. The first is the transduction of a truncated RUNX1-ETO (AE9a) as described above (Link et al., 2016). However, a superior model is transduction of full-length RUNX1-ETO into CD34$^+$ cord blood derived stem cells which promoted HSC self-renewal by the upregulation of the MPL/Thrombopoietin/Bcl-xL pathway (Chou et al., 2012). In addition, RUNX1-ETO in this model specifically upregulated the Forkhead transcription factor FOXO1 which promotes self-renewal as demonstrated by colony formation and serial replating assays. RUNX1-ETO and FOXO1 activate a significantly overlapping transcriptional programme and also share more than 50% of DNA binding sites and consequently a FOXO1 specific inhibitor AS1842856 was able to specifically abrogate growth of these pre-leukaemic cells (Lin et al., 2017). More recently, a model of RUNX1-ETO pre-leukaemia with hTERT expression was created to allow for prolongation of growth *in vitro* nearly indefinitely by overcoming telomere shortening mediated senescence. However, it does not cause transformation to AML as evidenced by xenotransplantation into NOD SCID mice (Lin et al., 2016).

The majority of work in this project has been done with the Kasumi-1 cell line which has been derived from the peripheral blood of a 7 year old boy with t(8;21) AML (Asou et al., 1991) and despite having a few other karyotypic abnormalities, well recapitulates the biology of t(8;21) AML (Ptasinska et al., 2014). Kasumi-1 cells also carry a KIT mutation.

1.7 Activator Protein-1 (AP-1)

1.7.1 AP-1 Structure and Function

Activator Protein-1 (AP-1) is a heterodimeric transcription factor complex containing the following families of proteins: FOS (FOS, FOSB, FOSL1, FOSL2), JUN (JUN, JUNB, JUND), ATF (ATF1, ATF2, ATF3) and JDP (JDP1, JDP2). AP-1 has long been identified as being critical to both basal transcription and also TPA (phorbol 12-O-tetradecanoate-13-acetate) induced transcription and hence it was named as being an activator (Angel et al., 1988).

AP-1 was originally studied through its yeast counterpart, GCN4 and was found to bind to the palindromic DNA consensus sequence TGANTCA (Hope and Struhl, 1985). It was soon after found to bind as a dimer to DNA (Hope and Struhl, 1987) and that it could bind to the first identified enhancer SV40 (Lee et al., 1987) (see also Section 1.5.2). Study of retroviral oncogenes later identified JUN and showed that it had homology in structure and function to GCN4 (Struhl, 1987). Study of viral oncogenes also identified FOS (Curran and Teich, 1982) and it was found that this could dimerise with JUN (Rauscher Iii et al., 1988). This dimerisation was through a leucine zipper (Landschulz et al., 1988) which consisted of two parallel α helices that formed a coiled coil that were stabilised though salt bridges from regularly interspersed Leucine residues (O'Shea et al., 1989).

The structure of the AP-1 family of proteins points to a highly complex biological role. It was found that JUN could form JUN-JUN homodimers but that FOS could not form FOS-FOS homodimers; however JUN-FOS heterodimers are more stable than JUN-JUN homodimers (Hai and Curran, 1991). JUN was also found

to be able to heterodimerise with the Activating Transcription Factor (ATF) family but that these dimers bound preferentially to the cAMP responsive element consensus sequence TGACGTCA. Furthermore, FOS could interact with the FOS Related Antigen (FRA) family (Franza et al., 1988). Also, a basic region next to the Leucine zipper was found to be responsible for DNA binding (Mitchell and Tjian, 1989).

A further major finding was that AP-1 was responsive to extracellular signalling. RAS, which activates the Mitogen Activated Protein Kinase (MAPK) pathway activates JUN N-terminal Kinase (JNK) which can phosphorylate JUN at positions 63 and 73 (Pulverer et al., 1993). In a mouse model of skin tumourigenesis where *Jun* was mutated to lose its JNK-mediated phosphorylation sites, tumourigenesis was impaired, even with activated Ras signalling (Behrens et al., 2000).

Within the context of haematopoietic development, AP-1 family members are absolutely crucial as *Jun* knockout mice died at the time of expected onset of haematopoiesis (Eferl et al., 1999). Similarly *Fos* knockout mice were also found to have disrupted haematopoiesis (Wang et al., 1992). Indeed AP-1 motifs were enriched in open chromatin in Haemogenic Endothelium cells which give rise to HSCs (Lichtinger et al., 2012) which may explain many of the defects in haematopoiesis. Furthermore, when reprogramming mouse embryonic fibroblasts into blood cells, *Fos* was identified as being an essential factor (Pereira et al., 2013). Further functional insights were elucidated when a dominant negative FOS which abrogated AP-1 transcriptional regulation in a mouse embryonic stem cell model, was found to be responsible in cooperation

with TEAD4 for a shift in FLK+ cells towards haematopoietic progenitors over smooth muscle (Obier et al., 2016).

1.7.2 AP-1 in Cancer

Within a cancer context, AP-1 family members often regulate cell proliferation, but importantly proliferation can be either increased or decreased depending on the AP-1 subunits employed. Experiments in mice have shown that overexpression of *Fos* led to the development of osteosarcomas (Wang et al., 1991), that *Jun* overexpression caused liver and skin cancers (Young et al., 1999), that *Fra1* overexpression led to lung tumours and that *Fra2* overexpression led to epithelial tumours (Jochum et al., 2000).

However, AP-1 factors may also act as tumour suppressors. *JunB* knockdown in the epidermis of mice promoted papilloma formation (Zenz et al., 2005). In particular, JUNB appears to alter expression of cell cycle genes such as INK4A and Cyclin D1 (Passegue and Wagner, 2000). *JunD* is also thought to be a tumour suppressor gene and in a murine study reduced fibroblast growth (Pfarr et al., 1994). But other reports suggest that JUND may act as an oncogene (Agarwal et al., 1999).

By contrast, knockdown of individual FOS family members appears to have little effect on cancer cells and it appears that FOS members can compensate for one another (Brusselbach et al., 1995).

Within the context of AML, the role of AP-1 has been established in leukaemic blasts. It has been shown that *CEBPA* is downregulated in AML and that its overexpression can permit myeloid differentiation (Loke et al., 2018). It was

previously shown that C/EBPα downregulated *JUN* expression and interfered with JUN binding to DNA by interacting with the leucine zipper domains and this is what was needed to drive differentiation (Rangatia et al., 2002). Consequently, JUN appears to maintain the leukaemic phenotype.

Also of considerable interest is how important AP-1 appears to be within the context of pre-leukaemia and LSCs. A study using a *SPI1* knockdown HSC model found that whilst *SPI1* knockdown increased serial colony formation ability, *JunB* expression in this context abrogated this increase hence reducing LSC-like behaviour (Steidl et al., 2006a).

Data from other myeloid malignancies also support the important roles of AP-1. Whilst *JunB* knockdown mice die in utero from vascular defects, partial rescue of *JunB* through transgenic expression was enough to rescue their development (Schorpp-Kistner et al., 1999). However, these rescued mice developed myeloproliferative disorders, some of which transformed through a blast crisis to AML, suggesting that *JunB* deficiency is sufficient to lead to myeloid proliferation (Passegué et al., 2001). It was later shown that the *JunB* deficiency was especially important within the long-term HSC compartment and that *JunB* overexpression led to loss of stem cell behaviour (Passegué et al., 2004). Within the context of Chronic Myeloid Leukaemia, it was shown that inhibition of FOS along with DUSP1 can eliminate the leukaemic stem cells (Kesarwani et al., 2017) illustrating the role of FOS in maintaining the stem cell phenotype.

With AP-1 being so important in cell survival, it is not surprising to see that it is controlled by multiple signalling pathways. AP-1 is activated by the RAS/MAPK

pathway and this allows for gene regulation following external stimuli activating Tyrosine Kinase Receptors. RAS signalling is generally activating and it leads to the phosphorylation of and activation of JUN and FOSL1 allowing for transcriptional activation (Mechta et al., 1997). AP-1 family members can also self-regulate and the JUN promoter for instance has an AP-1 binding site and this can allow for homeostatic regulation of AP-1 expression (Devary et al., 1992).

In this project, we used a dominant negative FOS (dnFOS) which exploits the common structure of AP-1 family members; it appends an acidic amphipathic N-terminal segment which prevents DNA binding by its basic region and also has a leucine zipper with which it can dimerise with normal AP-1 family members. These AP-1:dnFOS complexes can also compete out the binding of other AP-1 complexes in equimolar amounts (Olive et al., 1997).
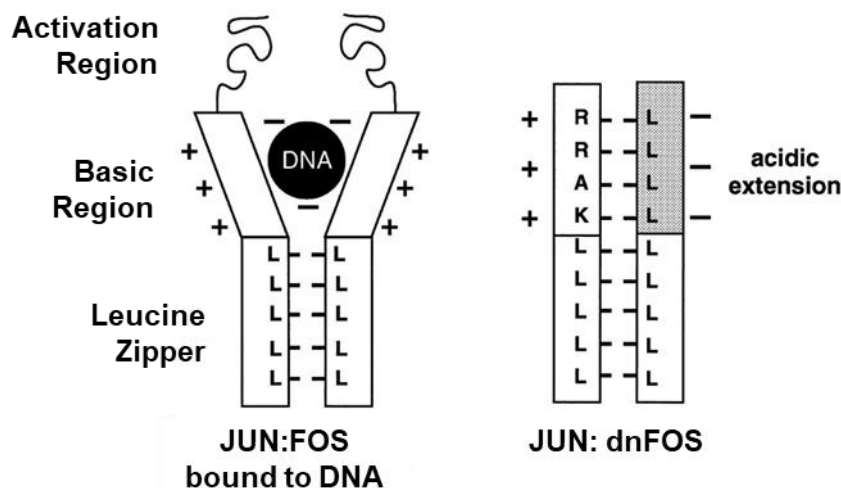
Figure 1-8: Schematic showing how AP-1 heterodimers normally bind to DNA and how DNA binding is abrogated by the dominant negative FOS (dnFOS). Adapted from (Olive et al., 1997).

1.8 Wilms Tumour 1 (WT1)

1.8.1 WT1 Structure and Function

WT1 was first identified as a tumour suppressor gene which when mutated caused the paediatric renal cancer, Wilms' Tumour (Call et al., 1990). However, *WT1* has also been implicated as an oncogene and is mutated in many AML patient cells conferring a poor prognosis (Inoue et al., 1994). It is similarly aberrantly expressed in a variety of other cancers such as breast, brain and colon cancer where it is normally not expressed at all. Functional validation of its role as an oncogene in this context was shown by RNA interference which caused cell cycle arrest and apoptosis (Tuna et al., 2005). This dual tumour-suppressor and oncogenic ability of this gene highlights the extraordinary complexity of its biology.

The gene is approximately 45kb long and this can lead to the transcription of at least 12 isoforms depending on which of 3 transcription start sites and which splice forms in exons 5 and 9 are employed. Indeed, the whole of exon 5 (51bp/17 amino acids '17AA') may be spliced in or out as well as part of exon 9 (9bp/Lysine-Threonine-Serine 'KTS') (Haber et al., 1991). Exons 6-10 code for zinc finger DNA-binding domains and consequently the KTS splice site affects DNA binding (Ullmark et al., 2017). The zinc fingers are formed by a core zinc ion that interacts with cysteines from a β sheet and histadines from an α helix to give a finger structure that can interdigitate within DNA (Pavletich and Pabo, 1991). Below is a schematic of the structure of WT1 drawn based upon the results of site-directed mutagenesis work (Wang et al., 1993).
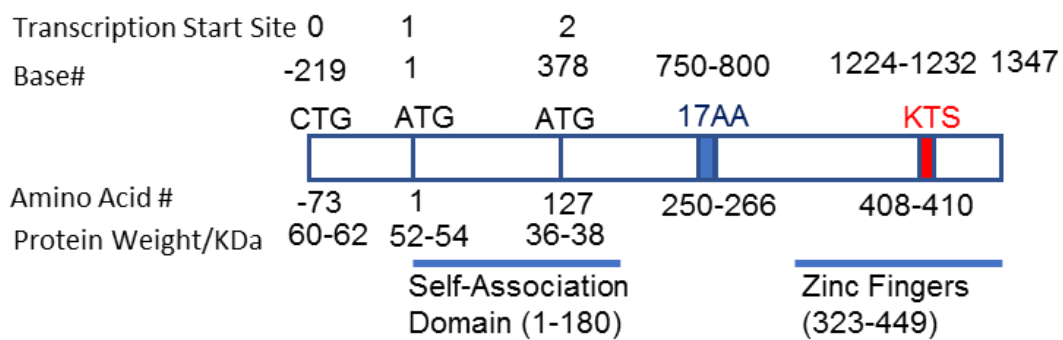
Figure 1-9: Schematic of the structure of WT1 along with alternative transcriptional start sites and alternative splice sites.

Some of the first indicators that WT1 alterations had an effect on human disease came by study of children born with Denys-Drash Syndrome; they had mutations in the Zinc finger domain that greatly reduced DNA binding and this gave a clinical phenotype of congenital nephrotic syndrome, psuedohermaphrotidism and Wilms tumour (Little et al., 1993). Equally, children born with Frasier Syndrome who have imbalances between the ratios of their *WT1* +KTS and *WT1* –KTS WT1 isoforms are severely affected (Pelletier et al., 1991). Developmentally, mice lacking either *WT1* +KTS or *WT1* –KTS isoforms die neonatally due to incomplete kidney development (Hammes et al., 2001) and together with the evidence from human disease, this shows that the ratios of *WT1* isoforms is crucial. By contrast, the presence or lack of the 17AA isoform developmentally has no phenotypic effect (Natoli et al., 2002).

WT1 is a transcription factor that has been described to bind to the consensus site GCGGGGGCG (Rauscher et al., 1990) which is identical to that of EGR1 (Early Growth Response 1). EGR1 has 3 zinc fingers which are similar to fingers

2-4 of WT1 and crystallographic studies show that only fingers 2-4 insert deeply into the DNA major grove to make base-specific contacts (Stoll et al., 2007), which may explain the identical consensus sequence. Surprisingly however, it was seen that deletion of the zinc finger 1 led to a large loss in the DNA binding strength of WT1 so it appears to stabilise interactions (Nakagama et al., 1995).

However, the WT1 +KTS isoform binds slightly differently as shown in an Electrophoretic Mobility Shift Assay (Bickmore et al., 1992). An explanation for this result came from nuclear magnetic resonance studies which showed that the KTS insertion increased the movement of a linker between zinc fingers 3 and 4 and so reduced the stability of the 4th zinc finger in the major groove of DNA (Laity et al., 2000). Consequently, when zinc finger 4 was deleted in WT1 -KTS isoforms, a large decrease in binding affinity was seen whilst when zinc finger 4 was deleted in WT1 +KTS, no difference to binding strength was apparent (Duarte et al., 1998).

Very few chromatin immunoprecipitation assays have been performed with WT1, mostly in renal glomeruli; it was shown that WT1 binds to an EGR1-like motif and can bind to both promoters and distal sites (Kann et al., 2015). ChIP-seq has also been performed on a CML blast crisis cells line, K562, but unfortunately the data quality is rather poor (Ullmark et al., 2017).

Intriguingly, WT1 binding also appears to be affected *in vitro* by signalling. Treatment of fibroblasts transgenically expressing *WT1* with forskolin to activate cAMP-dependent Protein Kinase A caused phosphorylation of Ser365 and Ser393 sites of WT1 which abrogated DNA binding (Sakamoto et al., 1997).

WT1 shows numerous protein-protein interactions which alter its activity as a transcription activator or repressor. Mutagenesis studies have identified a sub-region at residues 84-124 which acts as a repressive domain (McKay et al., 1999) by the recruitment of BASP-1 (Brain Acid Soluble Protein 1) which specifically inhibits WT1-mediated transcription and not basal transcription (Carpenter et al., 2004). Its zinc fingers can also recruit various proteins including TP53 which may explain its role in the alteration of apoptosis. WT1 can also alter the epigenetic landscape via the recruitment of TET2 which demethylates CpG (Rampal et al., 2014). Interestingly, WT1 and TET2 mutations are mutually exclusive in AML suggesting a common pathway being employed. WT1 might also bind to DNA indirectly through protein-protein interactions such as with p53 (Maheswaran et al., 1993b).

WT1 is also thought to affect post-transcriptional modification of RNA and some of the first evidence for this came from the observation that WT1 +KTS isoforms localised and interacted with splice factors (Larsson et al., 1995) and incorporated into functional spliceosomes (Davies et al., 1998). It was then shown that WT1 isoforms (WT1 +KTS better than WT1 -KTS) but not EGR1 could bind to *Igf2* RNA as the interaction occurred through zinc finger one that is not present in EGR1 (Caricasole et al., 1996). Further to this finding, WT1 may influence RNA modification through RNA binding proteins such as WTAP (Wilms Tumour Associated Protein) (Little et al., 2000) which performs m6A RNA methylation and have recently emerged as an important therapeutic targets in AML, permitting AML cells to overcome their differentiation block (Barbieri et al., 2017).

1.8.2 WT1 in Normal Haematopoiesis

Within the context of normal haematopoiesis, *WT1* transcripts levels are highest within the haematopoietic stem and progenitor cell (CD34$^+$ CD38$^-$) compartment and are undetectable in fully differentiated blood cells (Baird and Simmons, 1997). Similarly, study of mouse models showed that *Wt1* was only expressed at an early time point (day 12.5 post-conception) in the liver where stem cells are enriched (Hosen et al., 2007). Further overexpression in this compartment caused a relative expansion of this population both due to cell quiescence and inhibition of differentiation (Ellisen et al., 2001).

Study of WT1 knockout have been more challenging; *Wt1* null mice died at embryonic day 13.5 due to renal and cardiac problems (Kreidberg et al., 1993). Even a deletion of *Wt1* in 6 week old mice led to rapid death with renal failure and bone and fat loss (Chau et al., 2011). Serial murine transplantation of Wt1$^-$ foetal liver cells showed that these cells have poor engraftment and were quickly outcompeted by WT1 wildtype cells and that they also had poor colony formation ability with skews in haematopoiesis away from erythroid and megakaryocyte lineages (Alberta et al., 2003). However, other have found conflicting results and shown that even *Wt1* null foetal liver cells could fully and normally reconstitute haematopoiesis (King-Underwood et al., 2005).

1.8.3 WT1 in AML

WT1 is overexpressed in 90% of AML and mutated in about 10% of AML (King-Underwood et al., 1996). *WT1* overexpression leads to a worse relapse-free survival and response to chemotherapy in AML (Barragan et al., 2004) and similarly in *WT1* mutated AML patients, their relapse-free survival is significantly

impaired (Hou et al., 2010). However, mechanistically there is little known about how WT1 drives such a more aggressive AML; microarray data for *WT1* mutated AML patients suggests that cell proliferation pathways are perturbed but there is little more known to date (Becker et al., 2010).

In order to understand the effects of *WT1* overexpression within an AML context, a murine model of RUNX1-ETO was used and clear cooperation was seen between WT1 and RUNX1-ETO with mice overexpressing both transcripts dying rapidly of AML (Nishida et al., 2006). A WT1 GFP reporter murine study in RUNX1-ETO expressing AML showed that WT1 expression was most localised to the stem cell (LSK) compartment (Hosen et al., 2007). Following on from this result, data from the ALFA0702 trial showed that when AML patients have *WT1*-positive Minimal Residual Disease, they are more likely to relapse, even when adjusting for age, sex, presenting white cell count, cytogenetics and NPM1/FLT3 mutation status (Lambert et al., 2014). This results again implies that surviving pre-leukaemic and leukaemic stem cells following chemotherapy are responsible for the MRD from which relapse occurs and that *WT1* expression is an indicator of such cells.

WT1 mutations occur most commonly in exon 7 but can also occur in exon 8 or 9 (Krauth et al., 2015). *WT1* mutations also confer a poor prognosis (Hou et al., 2010). Again, the reason for this behaviour is unclear.

Gene regulation of *WT1* in AML is also poorly understood. In cell lines, a 3' enhancer that binds GATA1 and GATA 2 has been described (Furuhata et al., 2009). Others have described an intron 3 enhancer which binds to GATA1 and

Myb (Zhang et al., 1997b). These reports in fact directly conflict about which enhancers are important and so this is an area to be investigated further.

1.9 RNA Splicing

Alternative splicing is a major post-transcriptional mechanism of generating more than one unique mRNA from a single gene and is found in over 90% of all human genes (Pan et al., 2008). Alternative splicing leads to the generation of multiple protein isoforms in at least 37% of human genes (Kim et al., 2014). Alternative splicing can produce mRNA differing in their untranslated regions or coding regions, thus affecting mRNA translation, stability or localisation. Coding sequences may differ through exon skipping, a choice between mutually exclusive exons, intron retention or the use of alternative splice sites.

The extent to which alternative splicing occurs is determined by numerous factors including splice site strength (contain consensus sequences recognised by spliceosome machinery), the cis-regulatory sequences in pre-mRNA that favour or impair exon recognition and the protein levels of trans-acting proteins including those that are part of spliceosome machinery and RNA binding proteins.
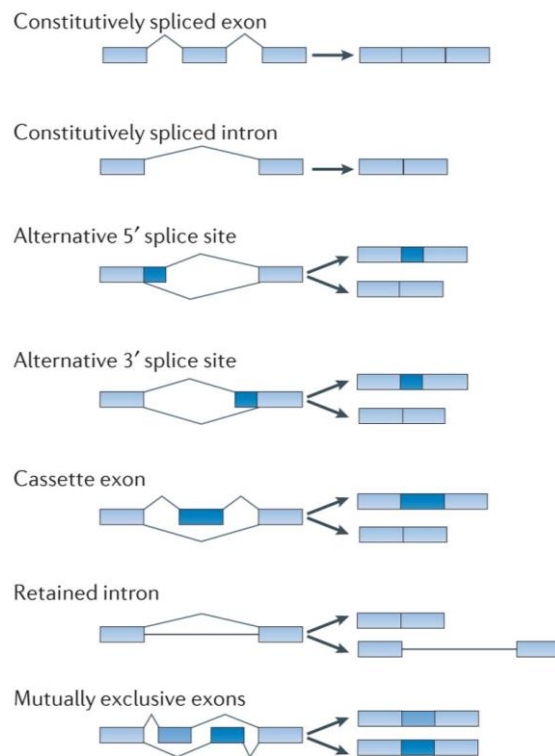
Figure 1-10: Alternative splicing mechanisms of producing different transcripts. Adapted from (Dvinge et al., 2016).

Mutations in RNA spliceosome machinery such as SRSF2, SF3B1, U2AF1 and ZRSR2 are common in cancers including AML (Yoshimi et al., 2019). These mutations all affect the 3' splice site recognition but it is unclear why these factors are recurrently mutated in cancer compared with the 30 or so other equally important 3' splice factors. On the whole (with some exceptions), Serine Rich factors such as SRSF2 promote splicing and 'heterogeneous nuclear ribonucleoproteins' and their co-factors such as U2AF1 inhibit splicing (Singh and Valcárcel, 2005). Consequently, aberrations leading to excessive or insufficient splicing may lead to cancer.

Intriguingly, such splicing factor mutations may co-operate with mutations in DNA methylation enzymes such as IDH2. For example, mutant SRSF2 binds to *INTS3*, a member of the Integrator complex and *IDH2* mutations leads to increased DNA methylation of *INTS3.* Aberrant *INTS3* splicing and methylation leads to stalling of RNA polymerase II. So in SRSF2 and IDH2 double mutant cells, these factors co-operate to promote myeloid malignancy (Yoshimi et al., 2019).

Similarly, RNA binding proteins are essential in the post-transcriptional modification of RNA (Glisovic et al., 2008) and are mutated in multiple cancers including AML (Yoshida et al., 2011). A CRISPR-Cas9 library screen in an MLL-AF9 cell line model of AML revealed one such RNA binding protein, RBM39, which when genetically or pharmacologically inhibited led to intron retention in HOXA9 target genes, many of which are known to be essential in myeloid leukaemogenesis (Wang et al., 2019). Most RNA binding proteins have RNA binding domains including K-homology domains, DEAD-box domains and RNA-recognition motifs (Lunde et al., 2007). As previously discussed, *WT1* is another such RNA binding protein.

1.10 Perturbing Gene Expression through siRNA and CRISPR-Cas9 Gene Editing

1.10.1 CRISPR-Cas9

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) was originally described as part of bacterial DNA repair systems (Makarova et al., 2002) and that spacer sequences in between CRISPR repeat elements appeared

to correlate with phage genomes, suggesting that this might be some sort of immune mechanism against bacteriophages (Mojica et al., 2005). Further study of bacteria identified the CRISPR associated protein, Cas9 (Bolotin et al., 2005).

It was then discovered that CRISPR array was transcribed and broken down into smaller crRNA (complementary region RNA) that could recognise various DNA sequences (Brouns et al., 2008) such as phage DNA but that the crRNA had to be next to a conserved Protospacer Adjacent Motif (PAM) sequence in order for Cas9 to bind and interfere (Mojica et al., 2009) and that Cas9 can make blunt double-stranded breaks 3bp upstream of the PAM sequence (Garneau et al., 2010). Lastly, tracrRNA (Transactivating complementary region RNA) was identified as a sequence of DNA needed to duplex with crRNA in order to mature it (Deltcheva et al., 2011) thus giving the 3 essential components for CRISPR-Cas9 gene editing – Cas9, tracrRNA and crRNA. A detailed biochemical understanding of CRISPR-Cas9 gene editing was elucidated showing that the Cas9-tracrRNA-crRNA complex cleaved 3bp upstream of the PAM sequence where the Cas9 is performing the cleavage, that the crRNA sequence (particularly on the 3' side) were vital to the recognition of DNA sequences and that tracrRNA and crRNA could be fused to give a single guide RNA (sgRNA) (Jinek et al., 2012).
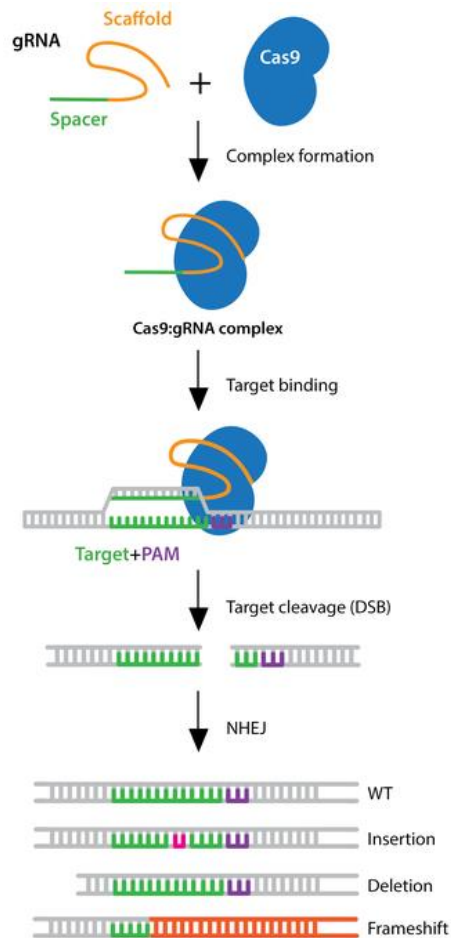
Figure 1-11: CRISPR-Cas9 gene deletion mechanism.

Image from Addgene.

CRISPR-Cas9 was then refined to allow for it to work in human cells by choosing a Streptococcus pyogenes-derived Cas9, adding a Nuclear Localisation Signal, finding the most stable 89 nucleotide tracrRNA isoform and finding that Cas9 could induce homologous recombination in human cell (Cong et al., 2013).

CRISPR-Cas9 was also repurposed to investigate gene regulation by using a catalytically dead Cas9. This was initially done to achieve CRISPR inactivation

(CRISPRi) with the catalytically dead Cas9 complex inhibiting transcriptional elongation (Qi et al., 2013). Later the catalytically dead Cas9 was fused to various chromatin modifiers to locally affect chromatin structure at any region for example Kruppel-associated box (KRAB) domains to repress gene expression (CRISPRi) and activators (CRISPRa) such as VP64 (Maeder et al., 2013) and histone acetyltransferases (Hilton et al., 2015).

Several improvements on CRISPR-Cas9 gene editing are also being investigated such as using a different Cas protein such as Cas12 which does not need tracrRNA (Zetsche et al., 2015), increasing specificity of crRNA by truncating them to 17 nucleotides (Fu et al., 2014) and base editors which can precisely change one base to another by fusing Cas9 with DNA deaminases (Komor et al., 2016).

1.10.2 RNA Interference

RNA interference is a regulatory mechanism to control gene expression through the use of small interfering RNA (siRNA). It was discovered through experiments in Caenorhabditis elegans into which double stranded RNA was injected and remarkably it was found that genes with complementary sequences were silenced (Fire et al., 1991).

In mammalian cells, longer double stranded RNA precursors are cleaved by the RNaseIII endonuclease, Dicer (Zhang et al., 2004a) to give siRNA. Alternatively microRNAs (miRNA) can be processed by Dicer to produce siRNA. Dicer-associated RNA binding proteins such as the Argonaute family transfer the processed siRNA to the RNA-induced silencing complex (RISC) which comprise

of other Argonaute family members including Ago-2 (Lee et al., 2006) and cleaves RNA into single stranded RNA; the strand which is more thermodynamically stable is retained in the RISC complex and used to scan for complementary mRNA and cut it, leading to its degradation (Khvorova et al., 2003).
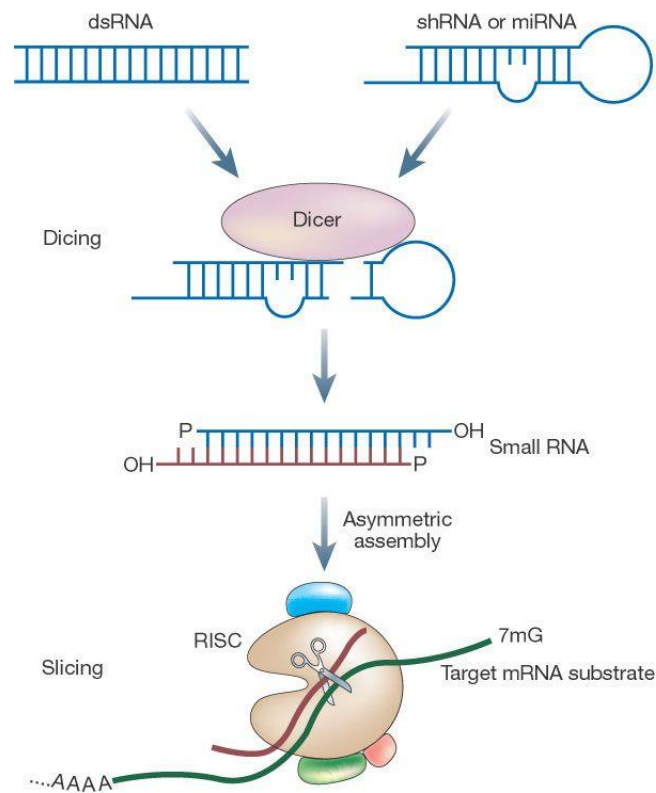
Figure 1-12: RNA interference schematic.

Figure from (Hannon and Rossi, 2004).

For gene knockdown studies, siRNA can be synthesised artificially as 21 nucleotide sequences but delivery into cells can be challenging. Alternatively, they can be produced in the cell from the processing of short hairpin RNA (shRNA) that resemble miRNA by Dicer.

1.11 Aims of the Project

There is a wealth of descriptive data describing t(8;21) and other AML subtypes based upon next generation sequencing, which have been pivotal in our understanding of the RUNX1-ETO cistrome and the transcription factors involved (Ptasinska et al., 2014). More recently, genome-wide CRISPR-Cas9 or shRNA depletion screens have provided implied conclusions about the importance of several genes in leukaemic maintenance (Martinez-Soria et al., 2018); (Tzelepis et al., 2016). However, in depth mechanistic understanding of many individual transcription factors is still lacking and consequently we cannot fully understand how TF/ Gene Regulatory Networks are altered in AML. We therefore conducted an in depth analysis of the role of two important transcription factors, AP-1 and WT1 in leukaemia maintenance.

The role of AP-1 family in t(8;21) AML

We addressed the following questions:

1. How does AP-1 regulate leukaemic growth?

2. How does it drive an AML-specific gene expression program?

3. What is the global impact of blocking AP-1 binding in t(8;21) AML?

4. How does blocking AP-1 in t(8;21) AML differ from other AMLs such as FLT3-ITD AML?

5. How does blocking AP-1 binding affect healthy stem cells?

This data will tell us whether perturbing AP-1 within leukaemia Gene Regulatory Networks allows it to be a potential therapeutic target. Since drugs selectively targeting AP-1 are not currently available, this approach may also identify candidate genes that regulate AP-1 or are regulated by AP-1 which are more druggable.

The role of WT1 and its isoforms in AML

It is clear that WT1 plays an important role in different types of AML as its upregulation or mutation leads to a worse prognosis in patients with AML. However, the precise role and mechanism of action of WT1, its isoforms and mutations is unclear, particularly within the context of leukaemia. To this end, we addressed the following objectives:

1.   Is WT1 required for leukaemic growth?

2.   How is WT1 regulated and how does it respond to outside signals?

3.   How does the overexpression of WT1 impact on leukaemic growth?

4.   What is the role of the different WT1 isoforms?

5.   Where does it bind in the genome?

6.   How does WT1 perturbation change the AML-specific gene regulatory network?

7.   How does WT1 interface with other transcription factors?

8.   Does WT1 have other roles besides transcription regulation?

This data will for the first time explain mechanistically how WT1, its isoforms and mutations facilitate leukaemic maintenance. This may also have therapeutic consequences and may elucidate candidate proteins within WT1-regulated Gene Regulatory Networks that may be targeted.

Chapter 2: Materials and Methods

2.1 Experimental Materials

2.1.1 Cell Lines and Cell Line Culture

Kasumi-1, MV4-11 cells and Human Embryonic Kidney (HEK) 293T cells were purchased from DMSZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen). Kasumi-1 and MV4-11 cells were cultured in RPMI-1640 media (Sigma-Aldrich cat# R8758) supplemented with 10% foetal calf serum (FCS) (Gibco, cat# 10500064), 2mM L-Glutamine (Sigma-Aldrich, cat# G7513), 100 U/ml Penicillin and 100 µg/ml Streptomycin (Sigma-Aldrich, cat# P4333). Human Embryonic Kidney (HEK) 293T cells were cultured in Dulbecco Modified Eagle Media (DMEM) media (Sigma-Aldrich, cat# D5546) supplemented with 10% FCS, 2mM L-Glutamine, 100 U/ml Penicillin and 100 µg/ml Streptomycin. All cells were grown in a humidified incubator (Sanyo, cat# MCO-20AIC) with 5% CO_2 at 37°C. Cells were periodically checked for Mycoplasma infection by a luminescence-based assay (Lonza, cat# LT07-318) and their identity checked by Short Tandem Repeat analysis (Eurofins Genomics).

2.1.2 Patient Samples

AML and healthy patient blood or bone marrow were obtained from either surplus diagnostic samples at the West Midlands Regional Genetics Laboratory or taken freshly from University Hospitals Birmingham with specific consent from the patients. Venepuncture was performed using a Vacutainer system (BD

Biosciences, cat# 367344) into 4ml EDTA blood tubes (BD Biosciences, cat# 367844). Bone marrow aspirates were performed at the right iliac crest using a 15 gauge aspirate needle (Mermaid Medical, cat# RAN-15-M) into 4ml EDTA blood tubes. Ethical approval was obtained from the National Health Service Research Ethics Committee (REC References 09/H0405/12 and 07/Q1206/25).

Anonymised Patient Identifier	Age	Sex	White Cell Count/10^9/L	Presentation or Relapse	Cytogenetics	Mutations
t(8;21)-1	73	Male	50	Relapse	46 XY, t(8;21)(9q22;q22)	KIT, TET2
t(8;21)-2	72	Male	29	Presentation	46 XY, t(8;21)(9q22;q22)	TET2
t(8;21)-3	48	Male	36	Presentation	46 XY, t(8;21)(9q22;q22)	KIT, NOTCH
t(8;21)-4	53	Male	6	Presentation	46 XY, t(8;21)(9q22;q22)	FLT3-TKD
t(8;21)-5	45	Male	2	Presentation	46 XY, t(8;21)(9q22;q22)	Nil
t(8;21)-6	47	Female	50	Relapse	46 XX, t(8;21)(9q22;q22)	WT1
t(8;21)-7	29	Female	2	Presentation	46 XX, t(8;21)(9q22;q22)	IDH1
FLT3-ITD	54	Male	17	Relapse	46 XY	TET2, FLT3-ITD
Healthy CD34$^+$	74	Male	6	Healthy control	46 XY	Nil

Table 2-1: Demographic and genomic data of primary patient samples used in this study

2.1.3 Purification of CD34$^+$ cells from patient samples

Bone marrow was strained through a 40µm cell strainer (Falcon, cat# 352240). Blood or marrow was then diluted to 4-6x10^9/L with Phosphate Buffered Saline (PBS) (Sigma-Aldrich, cat# D8537). It was then layered on to Lymphoprep density gradient medium (StemCell Technologies, cat# 07851) and centrifuged for 20 minutes at 2300rpm, acceleration setting 4, brake setting 0, in a centrifuge

(Eppendorf, cat# 5810). Upon centrifugation, mononuclear cells had separated into several layers consisting of red blood cells, lymphoprep, mononuclear cells, granulocytes and plasma. The mononuclear cell layer was isolated using a Pasteur pipette (StarLab, cat# E1414-1311), diluted 1:2 with PBS and centrifuged for 10 minutes at 600xg and then the pellet resuspended in 300µl MACS buffer (PBS, 0.5% Bovine Serum Albumin (BSA) [Sigma-Aldrich, cat# 05470], 2mM EDTA [Sigma-Aldrich, cat# E7889]).

In order to purify CD34$^+$ cells, the mononuclear cells were incubated with 100µl FcR blocked CD34$^+$ microbeads (Miltenyl-Biotech, cat# 130-046-702) for 30 minutes at 4°C. Cells were then washed in MACS buffer before being resuspended in 500µl MACS buffer and placed in a LS column (Miltenyi-Biotech, cat# 130-042-401) on a magnetic field (Miltenyi-Biotech, cat# 130-042-303). The LS column was then washed 3 times with MACS buffer before being removed from the magnet and flushed out with 5ml MACS buffer. The eluted fraction was passed through the LS column again to increase purity. CD34$^+$ cell purity was confirmed with flow cytometry (as described in Section 2.5.3) and cells frozen in freezing medium (90% FCS, 10% Dimethyl sulphoxide [DMSO] [Sigma, cat# C6164]), TRIzol (Invitrogen, cat# 15596018) or freshly used in a downstream application such as DNase-I digestion.

2.1.4 Primary Cell Culture

Sorted CD34$^+$ cells were cultured in 90% StemSpan SFEM II (StemCell Technologies, cat# 09605), 10% CD34+ expansion supplement (StemCell Technologies, cat# 02691) and 175nM UM171 or 1µM UM729 (StemCell

Technologies, cat# 72332). Cells were grown in a humidified incubator with 5% CO_2 at 37°C.

2.2 Gene Expression Methods

2.2.1 RNA extraction

For cells lysed in 1ml of TRIzol, 200µl of chloroform was mixed in and incubated at room temperature for 3 minutes before being centrifuged at 12000xg for 15 minutes at 4°C in a benchtop centrifuge (Eppendorf, cat# 5427R). The aqueous phase was transferred to a fresh tube and 0.5ml of 100% isopropanol (Sigma-Aldrich, cat# 59304) was mixed in and incubated at room temperature for 10 minutes before being centrifuged at 12000xg for 20 minutes at 4°C. The supernatant was discarded and the RNA pellet washed in 70% ethanol (VWR, cat# 437435L) and dissolved in 17µl nuclease-free water (HyClone, cat# SH30538.01). Contaminating DNA in the RNA was then digested by adding 1µl RQ1 DNase (Promega, cat# M6101) and 2µl RQ1 buffer and incubating at 37°C on a heat block (Stuart, cat# SBH200D/3). In order to clean up the RNA, the RNeasy MinElute Kit (Qiagen, cat# 74204) was used. Briefly, 80µl water, 250µl 100% ethanol and 350µl RLT buffer were added to the DNase digested RNA. Then the mixture was added to an RNeasy spin column and centrifuged at 12000xg. Following buffer RPE and 80% ethanol washes, RNA was eluted from the column into 14µl nuclease-free water.

Alternatively, cell line RNA was extracted using the Nucleobond spin kit (Machery-Nagel, cat# 740955). Briefly, cells were lysed in 350µl RA1 buffer and

3.5µl β-mercaptoethanol (Sigma-Aldrich, cat# M6250) and loaded onto a shredder column. Following centrifugation at 12000xg for 1 minute, the flow-through was mixed with 350µl 70% ethanol and loaded onto a Nucleobond spin column and centrifuged at 12000xg for 1 minute. Following a MDB buffer wash, an on column DNase digestion was performed by adding 95µl DNase solution to the column and incubating for 15 minutes at room temperature. Following a wash with RAW2 buffer and two washes with RA3 buffer, purified RNA was eluted into 50µl of nuclease-free water.

Or in the case of very low input RNA extractions of <10,000 cells such as from rarer populations of primary patient cells, the RNeasy Plus micro kit was used (Qiagen, cat# 74034). Briefly, cells were lysed in 350µl RLT+ buffer and 3.5µl β-mercaptoethanol and added to a gDNA eliminator column. The column was centrifuged at 12000xg for 30 seconds and gDNA could be recovered by a PE buffer wash of the eliminator column followed by elution into nuclease-free water. The flow-through from the gDNA eliminator column was then mixed with 70% ethanol and loaded onto an RNeasy MinElute spin column and centrifuged at 12000xg for 1 minute. The column was then washed with RW1 buffer, RPE buffer and 80% ethanol before purified RNA was eluted into 14 µl of nuclease-free water.

Purified RNA from all methods was assessed using a Nanodrop spectrophotometer (Thermo Scientific, cat# ND-2000C), by the ratio of the absorbance at 260nM and 280nM wavelengths.

2.2.2 cDNA synthesis

In experiments when 500ng of purified RNA was available, RNA was mixed with 500ng oligo(dT)$_{15}$ primer (Promega, cat# C1101) and water to take the volume up to 15µl; this mixture was incubated at 70°C for 5 minutes in a PCR thermocycler (Biometra, cat# T3000). Then, 200 units of Moloney Murine Leukaemia Virus (M-MLV) reverse transcriptase enzyme and 5µl of M-MLV buffer (Promega, cat# M170A) as well as 25 units of recombinant RNA ribonuclease inhibitor (Promega, cat# N2511) and 50µmoles of deoxyribose nucleoside triphosphate (dNTP) (Invitrogen, cat# 10297018) were added to the mixture and incubated at 42°C for 60 minutes.

In experiments where between 10pg to 500ng of purified RNA was available, RNA was mixed with 10µmoles dNTP, 50µmoles of oligo(dT)$_{20}$ (Invitrogen, cat# 18418020) and water to take the volume up to 13µl; this mixture was incubated at 65°C for 5 minutes followed by 1 minute in ice. Then, 200 units of Superscript IV, 100 µmoles of dithiothreitol (DTT) and 4µl of SSIV buffer (Invitrogen, cat# 18090010), as well as 40 units of RNA ribonuclease inhibitor (Invitrogen, cat# 10777-019) were added to the mixture and incubated at 55°C for 10 minutes.

2.2.3 Quantitative Polymerase Chain Reaction (qPCR)

2.5µl of cDNA diluted 1:50 with water was added to 7.5µl 1x SYBR Green PCR master mix (Applied Biosystems, cat# 4309155) and 1pmol each of forward primer and reverse primer (custom synthesised by Sigma-Aldrich) in a 96 well plate (Applied Biosystems, cat# 4316813). The plate was thermal cycled (95°C for 5 minutes followed by 40 cycles of [95°C for 15 seconds and 60°C for 1 minute]) and fluorescence signal read by a StepOne Real Time PCR System

(Applied Biosystems, cat# 4376357). Experiments were performed with a minimum of 3 technical duplicates for each biological replicate and expression was quantified by using a standard curve derived from a mixture of cDNA from cell lines that express the target gene highly and with serial 1:5 dilutions and normalised to a housekeeping gene such as glyceraldehyde 3-phosphate dehydrogenase (*GAPDH).*

Primer Name	Sequence
WT1 Forward	GCTATTCGCAATCAGGGTTACAG
WT1 Reverse	TGGGATCCTCATGCTTGAATG
dnFOS Forward	TAGCATGACTGGTGGACAGC
dnFOS Reverse	AGCTCTTCGGCCTCTTTTTC
GAPDH Forward	CCCACTCCTCCACCTTTGAC
GAPDH Reverse	ACCCTGTTGCTGTAGCCAAAT
CCND2 Forward	ACCTTCCGCAGTGCTCCTA
CCND2 Reverse	CCCAGCCAAGAAACGGTCC
RUNX1-ETO Forward	TCAAAATCACAGTGGATGGGC
RUNX1-ETO Reverse	CAGCCTAGATTGCGTCTTCACA

Table 2-2: Oligonucleotides and sequences used to assess gene expression in this study.

2.2.4 RNA Library Preparation for Illumina Sequencing

Purified RNA was quantified and quality assessed using a Bioanalyzer 2100 (Agilent, cat# G2939BA) with the eukaryotic total RNA high sensitivity kit (Agilent, cat# 5067-1513). In brief, RNA is diluted down to around 2ng/µl and denatured at 70°C for 2 minutes and 1µl of the denatured RNA added to 5µl RNA Pico marker in the RNA bioanalyzer chip. Conditioning solution and gel-dye mix was also added to the chip and the chip analysed in the Bioanalyzer 2100. Peak size was assessed by comparison to the RNA ladder and this allowed for identification of

18S and 28S ribosomal RNA peaks which in turn allowed for the calculation of the RNA Integrity Number. RNA quantification was assessed by comparison to the RNA marker.

2.2.5 Illumina TruSeq Library Preparation

If at least 100ng of purified RNA was available, the TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero kit (Illumina, cat# 20020596) was used. Briefly, ribosomal RNA was removed by adding rRNA binding buffer and rRNA removal mixture with incubation at 68°C for 5 minutes. Incubation with 1.75x rRNA removal beads allowed for binding to rRNA and the remaining RNA was cleaned up using 2x RNAClean XP beads (Agencourt, cat# A63987) and following two 70% ethanol washes, eluted into 11µl elution buffer. Then 8.5µl Elute, Prime and Fragment solution was added to the solution and incubated at 94°C for a length of time determined by the RNA integrity number.

Next, first strand synthesis was achieved by adding 8µl of a first strand synthesis-actin D-superscript II mix and incubated at 25°C for 10 minutes, 42°C for 15 minutes and 70°C for 15 minutes. Then second strand synthesis was achieved by adding 5µl of resuspension buffer, 20µl second strand master mix and incubated at 16°C for 1 hour. cDNA was then cleaned up with 1.25x Solid Phase Reversible Immobilisation (SPRI) beads (Beckman Coulter, cat# A63882) and elution into 17.5µl resuspension buffer. The 3' end of cDNA was adenylated by adding 12.5µl A-tailing mix and incubating at 37°C for 30 minutes followed by 70°C for 5 minutes. Adaptor ligation was achieved by adding 2.5µl resuspension buffer, 2.5µl ligation mix and 2.5µl 1:4 diluted adaptor (Illumina, cat# 20020492) and incubating at 30°C for 10 minutes before ligation being terminated using 5µl

ligation stop mix. Then cDNA was cleaned up using a 0.85x SPRI beads step followed by a 1x SPRI bead step and elution into 20µl resuspension buffer. PCR amplification of the DNA was achieved by adding 5µl of PCR primer cocktail, 25µl of PCR master mix and incubating thermal cycling 98°C for 30 seconds, 15 cycles of [98°C for 10 seconds, 60°C for 30 seconds, 72°C for 30 seconds] followed by 72°C for 5 minutes. Lastly, a final clean up with 1x SPRI beads and elution into 30µl nuclease-free water was performed for the final RNA-seq library.

2.2.6 NEBNext Library Preparation

If between 10-100ng of purified RNA was available, the NEBNext rRNA depletion kit (New England Biolabs, cat# E6310L) and NEBNext Ultra II RNA library kit was used (New England Biolabs, cat# E7770S). RNA was mixed with 1µl rRNA depletion solution, 2µl probe hybridization buffer and water to take the volume up to 15µl and incubated at 95°C for 2 minutes followed by a temperature ramp down at 0.1°C/second to 22°C for a further 5 minute hold. Then an RNaseH digestion was performed by adding 2µl RNaseH and 2µl RNaseH reaction buffer with incubation at 37°C for 30 minutes. Then a DNase I digestion was performed by adding 22.5 µl water, 5 µl DNase I reaction buffer and 2.5 µl DNase I with incubation at 37°C for 30 minutes. RNA was then purified using 2.2x RNA purification beads and elution into 5µl water.

RNA priming and fragmentation was achieved by adding 1µl random primers and 4µl first strand reaction buffer with an incubation at 94°C for 15 minutes. First strand synthesis was achieved by adding 2µl first strand enzyme mix and 8µl water followed by incubation at 25°C for 10 minutes, 42°C for 15 minutes and 70°C for 15 minutes. Second strand synthesis was achieved by adding 8µl

second strand reaction buffer, 4µl second strand enzyme mix and 48µl water followed by incubation at 16°C for 1 hour. Then cDNA was cleaned up using 1.8x SPRI beads and elution into 0.1x TE buffer. End preparation was achieved by adding 7µl end prep reaction buffer and 3µl end prep enzyme mix with incubation at 25°C for 10 minutes, 42°C for 15 minutes and 70°C for 15 minutes. Then adaptor ligation was achieved by adding 2.5µl 1:25 diluted adaptor, 1µl ligation enhancer and 30µl ligation master mix with incubation at 20°C for 15 minutes. Then 3µl USER enzyme was added with a 37°C for 15 minutes followed by a 0.9x SPRI purification and a 15µl 0.1xTE elution. PCR amplification was performed by adding 25µl Ultra II master mix, 50pmol i5 primer and 50pmol i7 primer (New England Biotech, cat# E7335S) and thermal cycling at 98°C for 30 seconds and 12-15 cycles of [98°C for seconds, 65°C for 75 seconds], 65°C for 5 minutes. Then the library was purified by two 0.9x SPRI purifications and elution into 20µl nuclease-free water.

2.2.7 Quantification of DNA libraries and pooling

DNA libraries and average library size were quantified using a High Sensitivity DNA assay (Agilent, cat# 5067-4626) on a Bioanalyzer 2100. Briefly, 1µl of the DNA library was added to 5µl high sensitivity DNA marker in the DNA bioanalyzer chip. Gel-dye mix and DNA ladder was also added to the chip and the chip analysed in the Bioanalyzer 2100. Peak size was assessed by comparison to the DNA ladder and library quantification by comparison to the DNA marker.

Quantification was also performed using the KAPA library quantification (Roche, cat# KK4824). 4µl of 1:1000, 1:2000 and 1:4000 diluted library was mixed with 6µl of KAPA master mix and qPCR performed as described before. Sample

concentration was determined by comparison to the standard curve and library molar concentrations calculated using the average library size determined by the Bioanalyzer. 12 samples were then mixed together in equal molar concentrations.

2.2.8 Illumina Sequencing of RNA-seq libraries

Pooled libraries were quantified using the high sensitivity DNA kit (Agilent, cat# 5067-5585) on a 4200 TapeStation system (Agilent, cat# G2991AA). Then Illumina sequencing was performed by Genomics Birmingham on a NextSeq 550 (Illumina, cat# SY-415-1002) using the NextSeq 550 High Output Kit v2.5 (Illumina, 20024906). 1.3ml of 1.8pM library pool was loaded onto the flow cell and then the flow cell, buffer cartridge and reagent cartridge were loaded into the NextSeq. The NextSeq was then run in paired-end mode for 75 cycles for each of read 1 and read 2.

2.2.9 Bioinformatic Analysis of RNA-seq

Initial bioinformatic analyses were performed by or together with Dr Salam Assi in the Bonifer lab.

Raw RNA-seq reads in FASTQ file format were processed with Trimmomatic v0.32 (Bolger et al., 2014) in order to trim sequencing adaptors and low quality bases. The processed reads were then aligned to the human genome (version hg38) using Hisat2 v2.1.0 (Kim et al., 2015) with default settings to give BAM files. Gene expression was measured as fragments per kilobase of transcript per million mapped reads (FPKM) values using with Stringtie v.1.3.3 (Pertea et al., 2015) with default settings and using the RefSeq database (O'Leary et al., 2015) as the reference transcriptome. Only genes that were expressed with an FPKM

> 1 in at least one of the samples were retained for further analysis. The raw FPKM values were quantile normalised using the Limma package v3.26.9 (Ritchie et al., 2015) in R v3.5.1. The normalised data was then log2-transformed, with a pseudocount of 1 being added to each of the FPKM values prior to transformation.

Differential gene expression analysis was carried out using Limma. A gene was considered to be differentially expressed if it had a greater than 2-fold change between experimental conditions, and a Benjamini-Hochberg adjusted p-value <0.05. Kyoto Encyclopaedia for Genes and Genomes (KEGG) pathway enrichment analysis was done using the ClueGO plugin v2.5.0 (Bindea et al., 2009) for Cytoscape v.3.61 (Shannon et al., 2003). This analysis was done using a right-sided hypergeometric test, with Benjamini-Hochberg p-value correction for multiple testing. A pathway was deemed to be significantly enriched if the adjusted p-value was <0.05.

Hierarchical clustering of RNA-Seq samples and replicates was done by first calculating the Pearson correlation value for each pair of samples. The resulting correlation matrix was then hierarchically clustered using complete linkage clustering of the Euclidean distances, and finally plotted as a heatmap in R.

2.3 DNA or Chromatin Methods

2.3.1 DNase-I Hypersensitive Site Mapping

2.25×10^6 cells were centrifuged and then re-suspended in 150µl of DNaseI resuspension buffer (60mM KCl [Sigma-Aldrich, cat# P9541], 10 mM Tris pH 7.4

[Sigma-Aldrich, cat# T3253], 15 mM NaCl [Sigma-Aldrich, cat# S3014], 5 mM MgCl$_2$ [Sigma-Aldrich, cat# M8266] and 300 mM sucrose [Sigma-Aldrich, cat# S8501]). Between 10-100 units of DNase-I (Worthington, cat# LS006344) was diluted into DNase-I dilution buffer (60mM KCl, 0.4% NP40 [Sigma-Aldrich, cat# NP40], 15 mM NaCl, 5 mM MgCl$_2$, 10 mM Tris pH 7.4 and 2mM CaCl$_2$ [Sigma-Aldrich, cat# C3306]). This mixture was then incubated at 22°C for exactly 3 minutes and then 150ul of the diluted DNaseI solutions was added to each reaction and incubated for a further 3 minutes precisely before terminating the digestion using 300ul of cell lysis buffer (300 mM Sodium Acetate [Sigma-Aldrich, cat# S2889], 10 mM EDTA pH 7.4 [Sigma-Aldrich, cat# ED], 1% SDS [Sigma-Aldrich, Cat# L6026] and 1 mg/ml proteinase K [Sigma-Aldrich, Cat# P2308]). An overnight incubation at 65°C allowed for Proteinase K to digest all protein and a 30 minute incubation at 37°C with 6 µg RNaseA [Sigma-Aldrich, Cat# R6513] was used to digest all RNA. A 0.7% agarose gel was used to check the digestions by gel electrophoresis (see section 2.3.2). Those samples with optimal digestion were further validated by qPCR with primers for a highly DNase-I hypersensitive region (TBP promoter), a region of low sensitivity (Beta Actin) and a region of minimal sensitivity (a segment of Chromosome 18).

Primer Name	Sequence
TBP Forward	CTGGCGGAAGTGACATTATCAA
TBP Reverse	CCCGACCTCACTGAACCC
Beta Actin Forward	GCAATGATCTGAGGAGGGAAGGG
Beta Actin Reverse	AGCTGTCACATCCAGGGTCCTCA
Chr18 Forward	AGGTCCCAGGACATATCCATT
Chr18 Reverse	GTTCAAATTGTGTTTTGTGGTTA

Table 2-3: Oligonucleotide and sequences used for assessing DNase I Hypersenstivity in this study.

Then a size selection of DNA fragments between 50 and 250bp was carried out by running 12µg of DNA on a 1.5% gel and cut out DNA being extracted from the gel (see section 2.3.3). The size-selected samples were again analysed by qPCR using the same regions and optimal samples were then chosen to progress to library preparation.

2.3.2 Gel Electrophoresis

Agarose (Sigma-Aldrich, cat# A4718), at a concentration suitable for the DNA fragment size to be separated, was mixed in 0.5x Tris-Borate-EDTA (TBE) (Invitrogen, cat# 15581044) and microwaved until the agarose dissolved. Then Ethidium bromide (Invitrogen, cat# 15581044) was added to the solution to a concentration of 500 µg/ml and the solution set in a gel tray (Scientific Laboratory Supplies, cat# mschoice15). The gel was then placed in a gel tank with 0.5x TBE and 500 µg/ml Ethidium bromide.

DNA was mixed 1:6 with gel loading dye (New England BioLabs, cat# B7024S) and loaded into the gel, along with 0.5µg of DNA ladder (New England BioLabs, cat# N3231S) and a 70V current applied for 2 hours to allow DNA separation. The gel was then viewed on a Gel Doc XR+ (Bio-Rad, cat# BGDXR).

2.3.3 Extracting DNA from Agarose Gel

The DNA on the agarose gel was briefly visualised on a UV Transilluminator (Spectronics Corporation, cat# TE365S) and the desired size of fragments of DNA were cut out of the agarose gel using a scalpel (Appleton Woods, cat# GC686). DNA was extracted using the Qiagen MinElute Kit (Qiagen, cat# 28604).

Briefly, the gel was then incubated with 3 volumes of QG buffer for 20 minutes and then mixed with 1 volume of isopropanol. This mixture was then loaded on the MinElute column and centrifuged at 16000xg for 1 minute followed by washes with QG buffer and PE buffer before elution into 20µl EB buffer.

2.3.4 DNA Library Preparation

Library Prep was performed by the KAPA Hyper Prep kit (Roche, cat#KK8500). DNA was mixed with 7µl of End Repair and A-tailing buffer, 3µl of End Repair and A-tailing enzyme and water to take the volume up to 60µl and incubated at 20°C for 30 minutes and 65°C for 30 minutes. Adaptor ligation was achieved by adding 5µl of 1:6 diluted adaptors, 5µl water, 30µl ligation buffer and 10µl DNA ligase and incubating at 20°C for 15 minutes. Then the library was cleaned up with a 0.8x SPRI selection and eluted into 20µl water. Then 5µl of primer mix and 25µl of KAPA HiFi Hotstart master mix and thermal cycled at 98°C for 45 seconds followed by 12-15 cycles of [98°C for 15 seconds, 60°C for 30 seconds, 72°C for 30 seconds] followed by 72°C for 1 minute. The amplified library was run on a 0.7% agarose gel and the 200-300bp fragments cut out and gel-extracted followed by a 1x SPRI selection and elution into 20µl nuclease-free water.

2.3.5 Illumina Sequencing of DNase I Libraries

Following quantification and average fragment size detection by the Bioanalyzer and KAPA library quantification, 12 libraries were pooled with equal molar concentrations. Pooled libraries were quantified on a TapeStation system and then Illumina sequencing was performed by Genomics Birmingham on a NextSeq 550 using the NextSeq 550 High Output Kit v2.5 (Illumina, 20024906). 1.3ml of 1.8pM library pool was loaded onto the flow cell and then the flow cell, buffer

cartridge and reagent cartridge were loaded into the NextSeq. The NextSeq was then run in single-end mode for 75 cycles.

2.3.6 Bioinformatic Analysis of DNase I-seq

DNase I-seq analysis was performed by Dr Salam Assi in the Bonifer lab.

Raw DNase I-seq reads in FASTQ format were processed with Trimmomatic v0.32 (Bolger et al., 2014) in order to trim sequencing adaptors and low quality bases. The processed reads were then aligned to the human genome version hg38 genome using Bowtie2 (Langmead and Salzberg, 2012), in the BAM file format. Next duplicate reads and reads not uniquely aligned to chromosomal position were moved from aligned data using Picard v2.10.5 (http://broadinstitute.github.io/picard/). These filtered reads were used to generate density profiles using the genomeCoverageBed function from bedtools (Quinlan, 2014). These tag densities were displayed using the UCSC Genome Browser (Kent et al., 2002).

DNase I Hypersensitive Sites (DHSs) were called with MACS2 (Zhang et al., 2008b) using the callpeak function (nomodel, call-summits and q= 0.005 parameters). The resulting peaks were then filtered against the hg38 blacklist and simple repeat tracks from the UCSC table browser (Karolchik et al., 2004) to remove any potential artefacts. Peaks were annotated to the nearest gene or using promoter Capture HiC data, where available, (Assi et al., 2019, Ptasinska et al., 2019). They were then further annotated as either a promoter of distal element using the annotatePeaks.pl function in the HOMER software package v4.9.1. A peak was annotated as being within a gene promoter if it was within

1.5kb of a transcription start site (TSS) and as a distal element otherwise. DNase I peak unions were constructed by merging peaks that had summit positions within 400bp of each other. In these cases, peaks were combined to a single peak with a new summit position defined as the mid-point between the summit positions of the original peaks. These average peak positions were used in all further downstream analysis.

To identify regions of differential chromatin accessibility, a peak union was first created for each pair of samples being considered. The read density for these peaks was then retrieved directly from the bedGraph files produced by MACS2 using the annotatePeaks.pl function in Homer with the parameter -size 200. These tag counts were normalised as counts per million (CPM) in R, and further log2-transformed with a pseudocount of 1 added to each value prior to transformation. A peak was considered to be differentially accessible if the fold-difference of the normalised tag count was greater than 2 between experiments. Motif enrichment analysis was then carried out in these sets of peaks using the findMotifsGenome.pl function in Homer.

To create read density plots, peaks were first ordered according to fold-difference. The read density in a 2kb window centred on the peak summits was then calculated using from the bedGraph files produced by Homer using the annotatePeaks.pl file in Homer, using the options -size 2000 -hist 10 -ghist. These were then plotted as heatmaps using java TreeView v1.1 (Saldanha, 2004).

De novo motif analysis was performed on peaks using HOMER (Heinz et al., 2010). Motif lengths of 6, 8, 10, and 12 bp were identified within 200 bp from the

peak summit. The top enriched motifs with a significant p value score were recorded. The annotatePeaks function in HOMER was used to find occurrences of motifs in peaks and known motif position weight matrices (PWM) were used from the HOMER database or from ChIP-seq experiments.

Hierarchical clustering with Euclidean distance was used for clustering of DNase I peaks. The read counts for all union peaks were normalised with regards to total reads depth counts and then Pearson's correlation coefficients were calculated between samples using log2 of the normalised read counts. A correlation matrix was generated and Pearson correlation coefficients are displayed after hierarchical clustering as a heatmap.

2.3.7 gDNA extraction

gDNA was extracted either as per DNase I digested DNA (see section 2.3.1) or by the Nucleobond tissue kit (Machery-Nagel, cat# 740952). For the latter protocol, cells were lysed in 200μl T1 buffer and mixed with 25μl proteinase K solution and 200μl B3 buffer followed by incubation at 70°C for 15 minutes. Then 210μl 100% ethanol was added to the mixture before loading it on the spin column with centrifugation at 11000xg for 1 minute. The spin column was then washed with buffer BW, buffer B5 and then eluted into 50μl buffer BE.

2.3.8 Targeted Myeloid Sequencing Panel

A custom myeloid sequencing panel library preparation and sequencing was performed by the West Midlands Regional Genomics Laboratory. Except for custom oligonucleotides, reagents were obtained from the TrueSight Myeloid Sequencing Panel kit (Illumina, cat# FC-130-1010) In order to ligate the custom

oligos, 50ng of gDNA, 5μl TrueSight Oligos (TSO), 35μl of Oligo Hybridisation Solution (OHS2) and water up to a total volume of 55μl were mixed and then incubated at 95°C for 1 minute followed by 40°C for 80 minutes. Then the sample was loaded onto a filter plate unit and washed with buffer SW1 and buffer UB1 to wash off unbound oligo. Then 45μl of Extension Ligation Mix (ELM4) was added to the plate and incubated for 37°C for 45 minutes to extend and ligate bound oligos. Then 4μl of i5 index and 4μl of i7 index were added to each sample as well as 25μl of 50mM sodium hydroxide followed by an incubation at room temperature for 5 minutes. Then 22μl of a PCR master mix with DNA polymerase I was added to the samples and thermal cycling was performed with 95°C for 3 minutes, 27 cycles of [95°C for 30 seconds, 66°C for 30 seconds, 72°C for 60 seconds] followed by 72°C for 5 minutes. The libraries were then cleaned up using 0.9x SPRI beads and resuspended in 20μl of elution buffer. Then library concentrations were normalised using library normalisation beads and 8 samples were pooled together and sequenced using the MiSeq reagent kit v3 (Illumina, cat# MS-102-3001) on a MiSeq (Illumina, cat# SY-410-1003) in paired ended mode with 150 cycles for each of Read 1 and Read 2 to achieve around 5000x coverage.

Bioinformatics analysis was performed using the TruSeq Amplicon app in BaseSpace (Illumina). After demultiplexing and FASTQ file generation, the software used a custom banded SmithWaterman aligner to align the reads against the human hg38 reference genome to create BAM files. The somatic variant caller then performed variant analysis for the specified regions. The outputs are VCF files. Further filtering of called variants was performed to remove

variants with less than a 5% allelic bias, non-exonic variant, splicing variants and known Single Nucleotide Polymorphisms. Mutations were called by annotation of filtered variants to databases such as the Catalogue for Somatic Mutations in Cancer (COSMIC) (Tate et al., 2019).

2.3.9 Assay for Transposase Accessible Chromatin (ATAC)-seq

ATAC resuspension buffer (RSB) was pre-made with 10mM Tris-HCl pH7.5, 10mM NaCl and 3mM $MgCl_2$. Then between 5000 and 500,000 cells were lysed in RSB buffer with 0.1% NP-40, 0.1% Tween-20 (Sigma-Aldrich, cat# P9416) and 0.01% digitonin (Promega, cat# G9441) for 3 minutes on ice. Then the cells were washed with 1ml of ATAC wash buffer consisting of RSB with 0.1% Tween-20 and centrifuged at 500xg for 10 minutes at 4°C. Then the nuclear pellet was resuspended in ATAC transposition buffer consisting of 25µl TD buffer and a concentration of Tn5 transposase enzyme (Illumina, cat# 20034197) related to the number of input cells, 16.5µl PBS, 5µl water, 0.1% tween-20 and 0.01% digitonin and then incubated on a thermomixer (Eppendorf, cat# 444-1023) at 37°C for 30 minutes. The transposed DNA was then cleaned up using a MinElute PCR purification kit (Qiagen, cat# 28004). Briefly, 5 volumes of PB buffer were mixed with the transposed DNA and added to a MinElute column and centrifuged at 16000xg. The column was then washed with PE buffer and then DNA eluted in 20µl water.

PCR amplification and library preparation was achieved by mixing the purified transposed DNA with 62.5 pmoles adaptor 1, 62.5 pmoles indexed adaptor 2 and 25µl of NEBNext high fidelity master mix and thermal cycled at 72°C for 5 minutes, 98°C for 30 seconds followed by 5 cycles of [98°C for 10 seconds, 63°C

for 30 seconds and 72ºC for 1 minute]. Then a further number of PCR cycles corresponding to ¼ of maximum amplification were used as assessed in a qPCR side reaction. The library was purified by a MinElute PCR purification and size selected with a 1x SPRI selection. The libraries were finally quantified on a Bioanalyzer, the Kapa library quantification kit by qPCR, pooled and run on an Illumina Flow Cell in as per the DNase I sequencing protocol.

2.3.10 Bioinformatic Analysis of ATAC-seq

Bioinformatic analysis of ATAC-seq was performed by Dr Salam Assi in the Bonifer Lab. Analysis was performed as per DNase I seq analysis with a few exceptions:

1) Bowtie2 was used with the parameter --very-sensitive-local
2) Peak calling with MACS2 v2.1.1 used the settings --nomodel --nolambda -B --trackline.

2.3.11 Chromatin Immunoprecipitation

Cells were washed in PBS and then crosslinked in 1% formaldehyde (Thermo Scientific, cat# 28906) for 10 minutes at room temperature before being quenched by 0.4M Glycine (VWR, cat# 10119CU). Cells were lysed for 10 minutes in 1 ml cell lysis buffer (5mM PIPES pH 8.0 [Sigma-Aldrich, cat# P6757], 85mM KCl, 0.5% NP-40, 1:200 Protease Inhibitory Cocktail [PIC] [Sigma-Aldrich, cat# P8340) before undergoing nuclear lysis for 10 minutes in 300 µl nuclear lysis buffer (50mM Tris pH 8.0, 10mM EDTA, 1% SDS, 1:100 PIC). Nuclei were then sonicated in a Bioruptor Pico (Diagenode, cat# B01060010) in Pico microtubes (Diagenode, cat# C30010016) for several cycles of 30 sec on, 30 sec off, until

chromatin was sheared to a 200-600bp fragment size. 5% of this sheared chromatin was saved for input and mixed with 6mM Tris-HCL pH 7.5, 1.5mM EDTA and 15mM NaCl as well as 3µg RNaseA and incubated for 30 minutes at 37°C. Next SDS was added to a final concentration of 1% as well as 50µg of proteinase K and input was reverse cross-linked for 4-16 hours in a 65°C water bath (Grant, cat# B8R05659) and DNA purified with 1.8x SPRI beads. 10% of the input was run on a 1.4% agarose gel to confirm size distribution.

The remaining 95% of sheared chromatin was diluted in blocking buffer (16.7mM Tris pH 8.0, 167mM NaCl, 1.2mM EDTA, 1.1% Triton X-100 [Sigma-Aldrich, cat# T8787], 0.01% SDS and 1:100 PIC). 15ul of Dynabeads G (Invitrogen, cat# 10007D) were pre-blocked for 2 hours at 4°C with blocking buffer consisting of 0.5% BSA, 30µg glycogen (Roche, cat# 10901393001) and ChIP grade antibody at a suitable concentration, before being incubated overnight 4°C with the diluted chromatin. Then the beads were washed in sonication wash buffer (50mM Tris pH 8.0, 10mM EDTA, 1% SDS), twice in Buffer A (20mM Tris, pH 8.0, 0.1% SDS, 1% Triton-X 100, 2mM EDTA, 150mM NaCl), twice in Buffer B (20mM Tris, pH 8.0, 0.1% SDS, 1% Triton-X 100, 2mM EDTA, 150mM NaCl) and if the immunoprecipitated protein is highly expressed, with Lithium buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl [Sigma-Aldrich, cat# L9650], 1 mM EDTA pH 8.0, 0.5 % NP-40 and 0.5 % sodium deoxycholate [Sigma-Aldrich, cat# 6750]). Beads were then washed once in 1xTE. Chromatin was then eluted in 100mM NaHCO$_3$ (Sigma-Aldrich, cat# S8875) and 1% SDS and reverse crosslinked for 4-16 hours at 65°C with 50µg Proteinase K. DNA was then cleaned up with 1.8x SPRI beads.

qPCR amplification of immunoprecipitated DNA was compared to input at several loci of putative protein binding sites as well as non-binding sites to assess the success of ChIP.

Primer name	Sequence
NAB2 Promoter Forward	CACCTCGGTCCCCAATTC
NAB2 Promoter Reverse	GCTTAGAGACTGGGAGAGG
PU.1 -14kb Enhancer Forward	AACAGGAAGCGCCCAGTCA
PU.1 -14kb Enhancer Reverse	TGTGCGGTGCCTGTGGTAAT
IVL Forward	GCCGTGCTTTGGAGTTCTTA
IVL Reverse	CCTCTGCTGCTGCCACTT
LAT2 Forward	AAACCCAGAACAACCCAGGC
LAT2 Reverse	ATGAGGAAGGATGTGTGTGCGG
IGFBP7 Forward	GTCAAGCACTAAAAGGACAAACCG
IGFBP7 Reverse	TGAATGCCACTGGGAG
CSF1R FIRE Forward	GCCTGACGCCAACAATGTG
CSF1R FIRE Reverse	GGCAAAGGAGGGAAGTGAGAG

Table 2-4: Oligonucleotides and sequences used to assess ChIP in this study.

Immunoprecipitated DNA was then library prepped with the KAPA Hyperprep kit as described before, except with a 200-450bp size selection of the final library and Illumina sequenced as described before.

Antibody	Amount	Manufacturer and Catalogue Number
WT1	1.5µg	Abcam, cat# ab89901
H3K27Ac	0.5µg	Abcam, cat# ab4729
ETO (C-terminus)	1µg	Santa Cruz, cat# sc9737
RUNX1 (C-terminus)	1µg	Abcam, cat# ab23980
EGR-1	10µl	Cell Signaling, cat# 4153
Sp1	1µg	Sigma-Aldrich, cat# 17-601

Table 2-5: Antibodies used for ChIP in this study

2.3.12 Bioinformatic Analysis of ChIP-seq

ChIP-analyses were performed by Dr Salam Assi in the Bonifer lab.

Reads from ChIP-Seq experiments were processed, aligned to the human genome and filtered in the same way as with DNase I-seq experiments. Peaks from transcription factor ChIP-seq experiments were identified using MACS2 with default settings. These peaks were then compared to the DNase I-seq or ATAC-Seq data, with only peaks that occurred within open chromatin regions being retained for further analysis. To identify differential binding of transcription factors such as WT1, initially union of all peaks was constructed by merging peaks that had summits within 100bp of each other. The read density in these peaks was then retrieved using the annoatePeaks.pl function in HOMER and normalised as counts per million in R. Peaks that had a fold-difference of at least 2 were considered to be differentially bound between experiments. The target genes for these peaks were identified by annotating each peak to its closest TSS using the annotatePeaks.pl function in Homer or using promoter Capture HiC data, where available (Assi et al., 2019, Ptasinska et al., 2019). Motif enrichment analysis was performed as with DNase I-seq data.

Average profiles of ChIP-Seq data were constructed by first normalising each of the alignment tracks as counts per million (CPM) using the bamCoverage function in deepTools v3.2.0 (Ramírez et al., 2014). These were then plotted using the plotProfile function in deepTools.

2.4 Gene Perturbation Methods

2.4.1 siRNA knockdown

siRNA was designed against the breakpoint of RUNX1 and ETO in order to specifically target RUNX1-ETO and was custom ordered from Sigma-Aldrich. 5×10^6 cells were centrifuged at 300xg for 5 minutes and resuspended in 350µl of whichever media is normally used to culture the cells and transferred to a 4mm electroporation cuvette (Geneflow, cat# E6-0070). 200nM of siRNA was added to the cuvette and the cuvette incubated for 5 minutes at room temperature. Then, the cuvette was placed in an electroporator (Fischer, cat# EPI 3500) and electroporation took place for 10ms at 350V. Following a further 5 minute incubation, the cell were resuspended to a concentration of 5×10^5/ml and cultures as normal. Electroporation was repeated every 2 days.

siRNA	Sequence
RUNX1-ETO sense strand	CCUCGAAAUCGUACUGAGAAG
RUNX1-ETO antisense strand	UCUCAGUACGAUUUCGAGGUU
Mismatch sense strand	CCUCGAAUUCGUUCUGAGAAG
Mismatch antisense strand	UCUCAGAACGAAUUCGAGGUU

Table 2-6: Oligonucleotides and sequences used for siRNA in this study.

2.4.2 Cloning of shRNA constructs

Custom linker A and linker B oligonucleotides were synthesised by Sigma-Aldrich containing the sense and antisense sequences as well as a loop to create a

hairpin recognised by Drosha enzymes in the nucleus and restriction enzyme overhangs for XhoI and MluI.

Linker A: TCGAG–AAGGTATATTGCTGTTGACAGTGAGCG-C–sense sequence–TAGTGAAGCCACAGATGTA-antisense sequence–A–TGCCTACTGCCTCG-A

Linker B: CGCGT–CGAGGCAGTAGGCA–antisense sequence-TACATCTGTGGCTTCACTA–sense sequence–G–CGCTCACTGTCAACAGCAATATACCTT-C

100pmoles of linker A and linker B were mixed together with 1x T4 DNA ligase buffer (New England Biolabs, cat# B0202), 5 units of T4 Polynucleotide Kinase (New England Biolabs, cat# M0201S) and water to take the volume up to 10µl; this mixture was incubated for 30 minutes at 37°C in order to phosphorylate the oligos followed by a 5 minute 95°C incubation and a gradual cooling ramp down to 25°C at 5°C/minute in a thermocycler in order to anneal the oligonucleotides. 15µg of pTRIPZ (Dharmacon, cat# RHS4744) was digested with 80 units of XhoI (New England Biolabs, cat# R0146S) and 40 units MluI (New England Biolabs, cat# R0198S) restriction enzymes in 1x buffer 3.1 (New England Biolabs, cat# B7203S) and water to take the volume up to 40µl; this mixture was incubated at 37°C for 4 hours. Then the plasmid was dephosphorylated with 30 units of Calf Intestinal Alkaline Phosphatase (New England Biolabs, cat# M0290) with a one hour incubation at 37°C. The oligonucleotides and 100ng of the digested plasmid were then ligated in a 4:1 molar ratio using 800 units of T4 DNA Ligase (New England Biolabs, cat# M0202S), 1x T4 DNA ligase buffer and water to take the volume up to 10µl; this mixture was incubated for 20 minutes at room temperature followed by an inactivation step of 65°C for 10 minutes.

Oligonucleotide	Sequence
WT1 shRNA_A sense	GGTGTTGATCTTACAAGATAT
WT1 shRNA_A antisense	ATATCTTGTAAGATCAACACC
WT1 shRNA_B sense	GCATCTGAGACCAGTGAGAAA
WT1 shRNA_B antisense	TTTCTCACTGGTCTCAGATGC
Non Targeting Control sense	TCTCGCTTGGGCGAGAGTAA
Non Targeting Control antisense	TTACTCTCGCCCAAGCGAGA

Table 2-7: Oligonucleotides and sequences used for shRNA cloning.
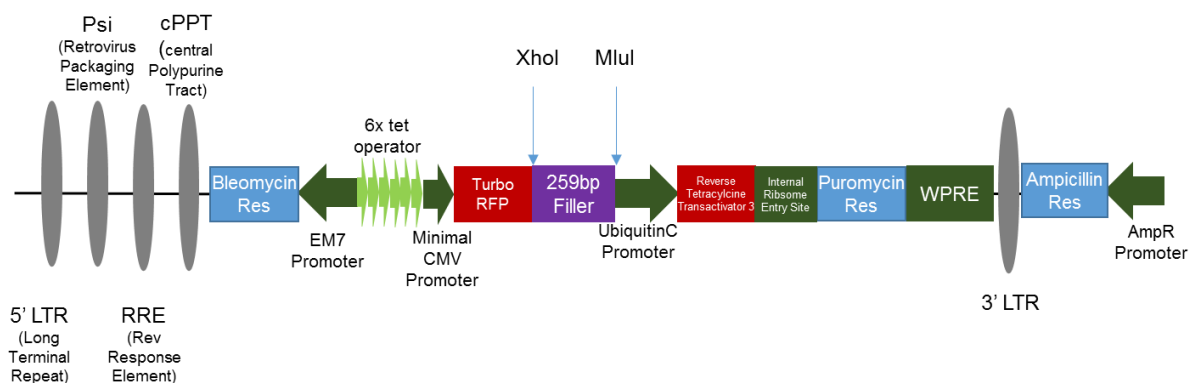
Figure 2-1: Schematic of pTRIPZ plasmid and relevant restriction enzyme sites

2.4.3 Transformation of Competent Bacteria

10ng of ligated plasmid were pipetted into 50µl of the DH5α strain of E. coli (Bioline, cat# BIO-85046) which have been made chemically competent and incubated on ice for 30 minutes. Then the bacteria were heat shocked by placing the mixture in a 42°C waterbath for 40 seconds followed by an incubation on ice for two minutes. Then the bacteria were mixed with 250µl of SOC (super optimal broth with catabolite repression) medium (Sigma-Aldrich, cat# S1797) and incubated at 37°C with 200 rpm rotation in a shaking incubator (Infor HT, cat# Multitron) for one hour. Then 125µl of the mixture was spread onto agar plates consisting of LB (Luria-Bertani) broth (Sigma-Aldrich, cat# L3522) and agar

(Sigma-Aldrich, cat# A1296) with a 100μg/ml ampicillin (Sigma-Aldrich, cat# A0166) and incubated for 16 hours at 37°C. Then individual colonies were picked and grown in 5ml LB broth containing 100μg/ml carbenicillin (Sigma-Aldrich, cat# C1389) at 37°C for 16 hours and 4.5ml used for minipreps.

2.4.4 Miniprep

Plasmid DNA was extracted using the Qiagen spin miniprep kit (Qiagen, cat# 27104). Briefly, bacteria grown in LB broth was centrifuged at 4000xg for 5 minutes and then resuspended in 250μl buffer P1. Then buffer P2 was added to lyse the bacterial cells for 4 minutes and the lysis was terminated by adding 350μl buffer N3; this mixture was centrifuged at 16000xg for 10 minutes to pellet cell wall, nucleus and other bacterial components whilst the supernatant containing the plasmid DNA was loaded onto a QIAprep spin column. The column was centrifuged at 16000xg for 1 minute and then washed with buffer PE and eluted into 50μl buffer EB.

Purified plasmids then underwent an analytical digest with combinations of one or more restriction enzymes followed by gel electrophoresis and fragment sizes compared to those expected, in order to confirm the success of cloning. Furthermore, the purified plasmid was sent for Sanger sequencing, using the pTRIPZ sequencing primer (sequence: GGAAAGAATCAAGGAGG) by Source Biosciences, in order to confirm that the plasmids had the correct shRNA inserts.

For successful clones, the remaining 0.5ml of bacterial in LB broth was grown in 250ml of LB broth with 100μg/ml carbenicillin and incubated in a shaking

incubator at 37°C, 250 rpm for 16 hours. The resultant bacterial culture then underwent a maxiprep.

2.4.5 Maxiprep

In order to extract plasmid DNA, the Nucleobond Xtra Midiprep kit was used (Machery-Nagel, cat# 740420.50). The bacterial culture was centrifuged at 4000xg for 20 minutes at 4°C and then the pellet was resuspended in 16ml buffer RES-EF. The bacterial cells were lysed by adding 16ml buffer LYS-EF for 5 minutes and the lysis terminated by adding 16 ml of buffer NEU-EF. The Nucleobond column was equilibrated with buffer EQU-EF and then the lysed bacteria suspension loaded onto the column and allowed to drain under gravity. The column filter was washed with buffer FIL-EF and then discarded. The column was then washed with buffer ENDO-EF to remove bacterial endotoxins and buffer WASH-EF. Then elution from the column was undertaken by adding 5ml of ELU-EF and collecting the eluent after it had drained by gravity from the column. Then 3.5ml isopropanol was added to the eluent and the mixture centrifuged at 20000xg for 45 minutes. The supernatant was discarded, the pellet washed with 70% ethanol and finally resuspended in buffer TE-EF.

The purified plasmid was again checked by an analytical restriction enzyme digest and Sanger sequencing, as before.

2.4.6 Cloning of Overexpression Constructs

WT1 isoform cDNA was a gift from Naohiko Seki, Chiba University, Japan (Moriya, 2008). The cDNA was amplified by PCR using primers with SalI and NotI restriction enzyme overhangs. 100ng of cDNA was mixed with 10μmoles of

dNTP, 1 unit Phusion DNA polymerase (New England Biolabs, cat# M0530S), 1x Phusion HF buffer, 25pmoles of forward primer and 25pmoles of reverse primer and water to make the volume up to 50µl; this mixture was thermal cycled at 98°C for 1 minute, then 30 cycles of [98°C for 10 seconds, 65°C for 30 seconds, 72°C for 90 seconds] and a final extension of 72°C for 10 minutes. The PCR product was cleaned up using the Qiagen PCR clean-up kit and then digested with 20 units SalI-HF (New England Biolabs, cat# R3138S), 20 units of NotI-HF, 1x Cutsmart buffer (New England Biolabs, cat# B7204S) and water to take the volume up to 50µl with an incubation of 37°C for 4 hours. The digested products were again cleaned up with the Qiagen PCR clean-up kit.

Dominant negative FOS (dnFOS) cDNA was a gift from Charles Vinson, National Cancer Institute, Bethesda, USA (Olive et al., 1997). The cDNA was amplified by PCR using primers with SalI and NotI restriction enzyme overhangs, using the same Phusion polymerase protocol as with WT1 but with thermal cycler conditions of 98°C for 30 seconds, then 30 cycles of [98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds] and a final extension of 72°C for 10 minutes. The PCR product was cleaned up, digested with SalI and NotI as with WT1 and then once again cleaned up.

Oligonucleotide Name	Sequence
dnFOS cloning primer Forward	ATTGTCGACGCCACCATGGACTACAAGGACGACG
dnFOS cloning primer Reverse	TAGGCGGCCGCGAATTAATCAGGGATC
WT1 cloning primer Forward	ATTGTCGACGCCACCATGTTTCCTAACGCGCCCTACCTGC
WT1 cloning primer Reverse	TAGGCGGCCGCGAATTACCTTTTGAATAGACTTTAATTGAGAGCAAAGTG

Table 2-8: Oligonucleotide and sequences used for cloning in expression constructs in this study.

25µg of pENTR-GFP plasmid (modified from the Invitrogen pENTR plasmid by Ben Edginton-White, Bonifer lab) was digested with 60 units of SalI and 60 units of NotI, in 1x Cutsmart buffer and water to take the volume up to 100µl, at 37°C for 4 hours. Agarose gel electrophoresis was undertaken on the product on a 0.8% agarose gel and the larger 3925 bp fragment was gel extracted. The pENTR-GFP backbone was then ligated to the dnFOS or *WT1* isoform insert in a 1:4 molar ratio using 400 units of T4 DNA ligase, 1x T4 DNA ligase buffer and water to make the volume up to 20µl and then incubated at 16°C for 16 hours. 10ng of ligated product was transformed into competent bacteria, plated onto agar plates containing 50µg/ml kanamycin (Sigma-Aldrich, cat# K0244), colonies grown in LB broth with kanamycin and miniprepped as before. Clones were screened by analytical restriction enzyme digests and Sanger sequencing.

The plasmids with the desired inserts ligated into pENTR-GFP were then recombined in the tet-on plasmid pCW57.1 (David Root, Addgene plasmid #41393) using Gateway technology. 150ng of insert plasmid, 150ng of pCW57.1, 2µl of LR Clonase II (Invitrogen, cat# 11791020) and TE buffer to take the volume up to 10µl and incubated at 25°C for 1 hour to permit recombination. Recombination was terminated by adding 1µl proteinase K and incubating at 37°C for 10 minutes. 20ng of recombined product was then transformed into competent bacteria, plated onto agar plates with ampicillin, cultured in LB broth, miniprepped and maxiprepped to give the final plasmid suitable for transfection.

The plasmid was checked with an analytical restriction enzyme digests and Sanger sequencing.
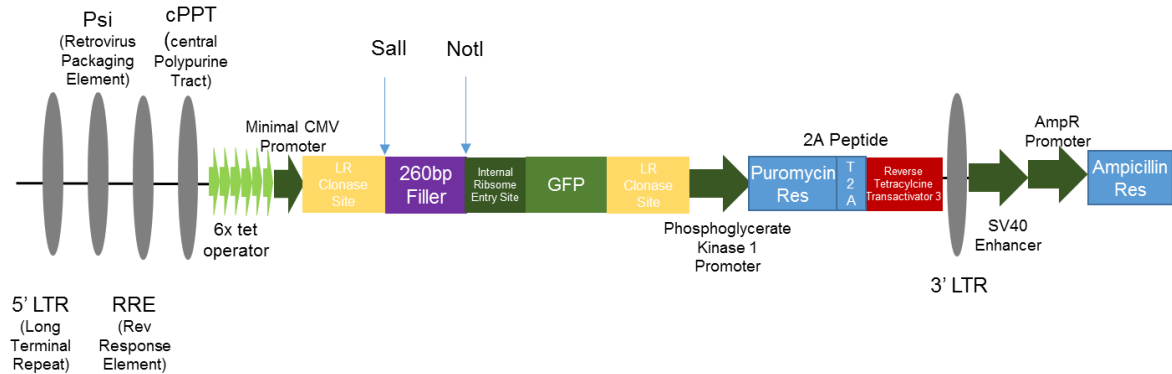
Figure 2-2: Schematic of pENTR in pCW57.1 plasmid and relevant restriction enzyme sites.

2.4.7 *WT1* Mutant Cloning

WT1 mutant plasmids were created with stop codons in exons 7, 8 or 9 by site-directed mutagenesis of pENTR-GFP plasmids with the *WT1* TSS 1, Exon 5 containing insert.

10ng of plasmid DNA was mixed with 1 unit of Q5 DNA polymerase (New England Biolabs, cat# M0491S), 1x Q5 reaction buffer, 10nmoles dNTP, 25pmoles forward primer, 25pmoles reverse primer and water up to 50µl; this mixture was thermal cycled at 98oC for 30 seconds, then 30 cycles of [98°C for 10 seconds, annealing temperature for 20 seconds, 72°C for 90 seconds] followed by a final extension of 72°C for 2 minutes. The annealing temperature was 68°C for the exon 7_A and exon 7_B primer PCRs and 60°C for the exon 8 and exon 9 primer PCRs.

Oligonucleotide	Sequence
WT1 Exon 7_A SDM Forward	TAGTAGCCCCGACTCTTGTACG
WT1 Exon 7_A SDM Reverse	TCCAGGCACACGTCGCAC
WT1 Exon 7_B SDM Forward	TAGCCCCGACTCTTGTACGG
WT1 Exon 7_B SDM Reverse	TACTCCAGGCACACGTCG
WT1 Exon 8 SDM Forward	TGACTGTGAACGAAGGTTTTC
WT1 Exon 8 SDM Reverse	CCTTGAAGTCACACTGGTATG
WT1 Exon 9 SDM Forward	TGAAACTTGTCAGCGAAAGTTC
WT1 Exon 9 SDM Reverse	TACACTGGAATGGTTTCAC

Table 2-9: Oligonucleotides and sequences used for site directed mutagenesis in this study.

Following PCR, the product was loaded on a 0.8% agarose gel and separated by electrophoresis. The 3.9kb fragment was excised from the gel and gel extracted. In order to digest the parent plasmid template and circularise the PCR fragment, half of the PCR product was mixed with 400 units of T4 DNA ligase, 10 units of T4 PNK, 1x T4 DNA ligase buffer, 20 units of DpnI (New England Biolabs, cat# R0176S) and water to take the volume up to 10µl; this mixture was incubated at room temperature for 1 hour and then transformed into chemically competent bacteria. This mixture was then plated on agar plates with 50µg/ml kanamycin and expanded as before, recombined into pCw57.1 using Gateway technology as before, expanded under 100µg/ml ampicillin selection and maxiprepped.

2.4.8 CRISPR-Cas9 Gene Editing

sgRNA were designed using the Broad Institute sgRNA Design Tool (Doench et al., 2016). The Protospacer adjacent motif (PAM) sequence was removed and well as 2 nucleotides from the opposite side to the PAM sequence to increase specificity (Fu et al., 2014). Then restriction enzyme overhangs for BsmBI were

added and Linker A and Linker B oligonucleotides synthesised by Sigma-Aldrich as below.

Name	Sense Sequence
WT1 upstream	TGGCAAACGTCAGCGAGT
WT1 downstream	AGCCGGGCGTATTTCGAT

Table 2-10: Sense sequence used in oligonucleotides for CRISPR cloning.

Linker A oligo: CACCG-sense sequence

Linker B oligo: C-antisense sequence-CAAA

The linker A and B oligos were phosphorylated and annealed in the same way as shRNA oligos.

The LentiCRISPRv2 plasmid (Feng Zhang, Addgene plasmid #52961) was digested with 30 units of FastDigest BsmBI (Thermo Scientific, cat# FD0454) in the presence of 10mM DTT and concurrently dephosphorylated with 30 units of FastAP (Fermantas) with a 30 minute incubation at 37°C. The oligonucleotides and digested plasmid were then ligated in a 3:1 molar ratio using 2000 units of Quick Ligase (NEB) for 10 minutes at room temperature. Ligated plasmid was transformed into chemically competent bacteria, plated onto agar plates with 100µg/ml ampicillin and grown in LB broth. The bacteria were miniprepped and maxiprepped as before and clones screened by analytical restriction enzyme digests and Sanger sequencing beginning at the U6 promoter of the plasmid (sequence: GAGGGCCTATTTCCCATGATT).
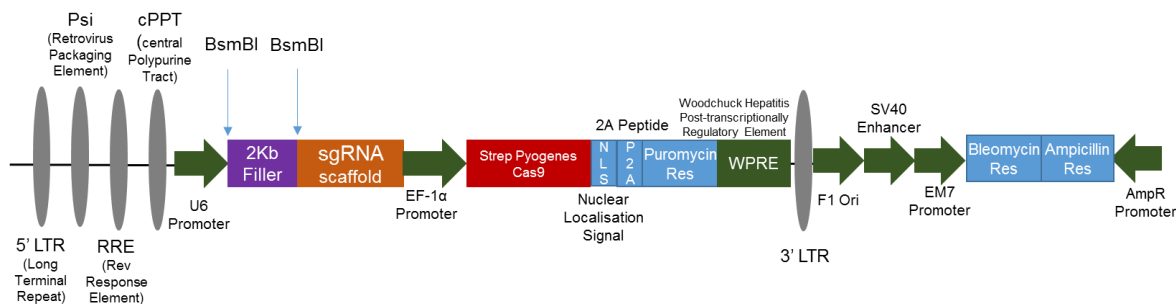
Figure 2-3: Schematic of lentiCRISPRv2 plasmid and relevant restriction enzyme sites

Kasumi-1 cells were lentivirally transduced with the plasmid (see section 2.4.10) and single cell clones were screened for CRISPR-Cas9 deletion using primers Clones were screened by performing PCR using primers designed to amplify the wildtype allele and primers designed to amplify the allele with a deleted *WT1* gene. PCR was performed as per the Phusion polymerase protocol as previously described. Thermal cycling conditions were 98°C for 30 seconds followed by 30 cycles of [98°C for 10 seconds, 62°C for 20 seconds, 72°C for 30 seconds] followed by a final extension of 72°C for 5 minutes.

Primer Name	Sequence
Wildtype Allele PCR Forward	TGCATTGGTGTGTTTGTCCT
Wildtype Allele PCR Reverse	CAGGGGGAATGTCTCAAAGA
Deletion Allele PCR Forward	GGGTAGAGGGCAATGGACTT
Deletion Allele PCR Reverse	TTCCCTCCTAAACTAGCCGC

Table 2-11: Oligonucleotides used in PCR to confirm CRISPR-Cas9 gene editing

2.4.9 CRISPR Activation

2µg of each of the LentiCRISPRv2 plasmid and p300-dCas9 (Charles Gersbach, Addgene #83889) were digested using 5 units RsrII (New England Biolabs, cat# R0501S) and AgeI (New England Biolabs, cat# R0552S) in 1x Cutsmart buffer at 37°C for 5 hours. The 6241bp band was gel extracted from p300-dCas9 and the 10495bp band from LentiCRISPRv2. These two fragments were then ligated in a 3:1 molar ratio using 20units of T4 DNA ligase in 1x T4 buffer, with water to take the volume up to 20µl and incubated at room temperature for 30 minutes. The ligated plasmid 'lentiCRISPR-p300' was transformed into chemically competent bacteria, plated onto agar plates with 100µg/ml ampicillin and grown in LB broth. The bacteria were miniprepped and maxiprepped as before and clones screened by analytical restriction enzyme digests.

Linker A and linker B oligos were annealed as with CRISPR-Cas9 gene editing and cloned into plasmid LentiCRISPR-p300 in the same way and similarly maxiprepped. The maxipreps for 3 designs for each DHS was then combined so that a tiled CRISPRa strategy could be employed.

Name	Sense Sequence
3' DHS	ACTGTTAATTATAGCGAGTG
3' DHS	CATGGCAAACGTCAGCGAGT
3' DHS	TTGATTGTTTCAACTGCACA
CTCF bound DHS	TCTACTGACAAAGCTTATCG
CTCF bound DHS	CGCTGGAGAGACATCACGGG
CTCF bound DHS	TGTCCTCACATATGCACAGG
Intron 3 DHS	TTGGCTTACAGCGACCACTG
Intron 3 DHS	TGAGGTGGTGGCACCTACAG
Intron 3 DHS	TACTGGGCCGAAGCATGCTC
Intron 8 DHS	GGACAGAGAAGGTCTAGCCT
Intron 8 DHS	AGTGAGATCCCACTAGCAGT
Intron 8 DHS	CAACCATTCCTTAACCACAG

Table 2-12: Sense sequences used in oligonucleotides for CRISPRa cloning.
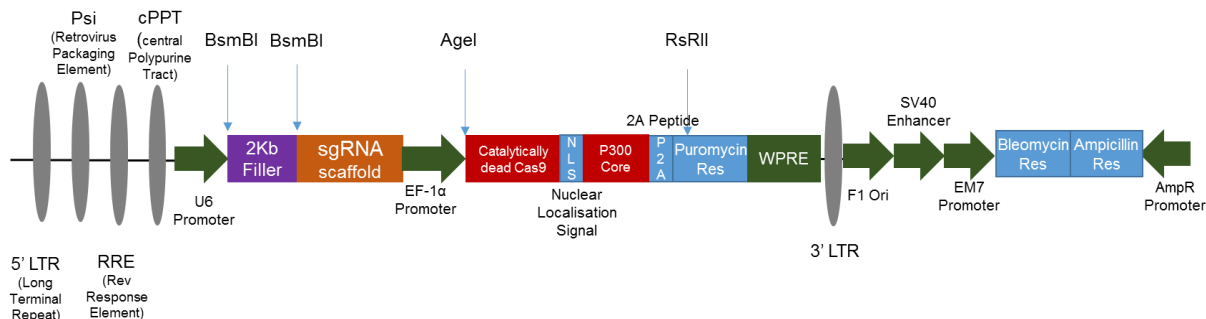
Figure 2-4: Schematic of LentiCRISPR-p300 plasmid and relevant restriction enzyme sites.

Following lentiviral transduction of Kasumi-1 cells, cells were assessed for H3K27Ac deposition by the dCas9-p300 by ChIP-qPCR using the following validation primers. Enrichment was normalised to a known site of H3K27Ac deposition, the PU1 -14kb enhancer and to enrichment of the empty vector control.

Primer	Sequence
PU.1 -14kb Enhancer Forward	AACAGGAAGCGCCCAGTCA
PU.1 -14kb Enhancer Reverse	TGTGCGGTGCCTGTGGTAAT
3' DHS Forward	TCTGTTCTTCCCTGTGCAGT
3' DHS Reverse	GCCCCTCTTATTTTGCATCTGG
CTCF bound DHS Forward	CGCCACGATAAGCTTTGTCA
CTCF bound DHS Reverse	CAGAATGCCACTGGATGTCC
Intron 3 DHS Forward	ACACCCAGCTCTCATCATGG
Intron 3 DHS Reverse	TGAAATGCCTACGTGACCAC
Intron 8 DHS Forward	CCTCGGCCCTAACAATGTGG
Intron 8 DHS Reverse	GTGAGAAGCTGGACCTCTGT

Table 2-13: Oligonucleotides uses to confirm CRISPRa by ChIP.

2.4.10 Generation of Lentiviral Particles using HEK293T Cells

tat, rev, gag/pol and vsv-g plasmids were a gift from Richard Mulligan, Harvard medical school, USA (Mostoslavsky et al., 2005). HEK 293T cells were grown in adherent culture at a 40-60% confluency and transfected using calcium phosphate co-precipitation: five plasmids (24µg pCW57.1 with cDNA insert, 1.2µg tat, 1.2µg rev, 1.2µg gag/pol and 2.4µg vsv-g) were mixed with 250mM Calcium Chloride and 2.5mM HEPES (Sigma-Aldrich, cat# H4034). This mixture was then slowly added, whilst vortexing, to a buffer consisting of 50mM HEPES, 280mM Sodium Chloride and 1.5mM Na_2HPO_4 and incubated at room temperature for 35 minutes and then promptly added dropwise to the plates of HEK 293T cells. Viral supernatant was then harvested after 24 hours and subsequently every 12 hours for 36 hours prior to concentration. Supernatant was centrifuged at 300xg for 5 minutes and then filtered through a 0.45µm filter (Starlab, cat# E4780-1453).

If there was a large volume of viral supernatant, virus was concentrated using a Centricon Plus-70 100kDa filter (Millipore, cat# UFC710008). 60ml of viral supernatant was added to the filter at a time and the filter centrifuged at 2000xg for 30-50 minutes with flow-through discarded. To elute concentrated virus from the column, the filter was turned upside-down and centrifuged at 300xg for 1 minute.

If there was a small volume of viral supernatant, virus was concentrated by ultracentrifugation. 30ml of viral supernatant was loaded into a centrifuge tube (Beckman Coulter, cat# 358126) and along with a 25mm tube adaptor (Beckman Coulter, cat# 358156) placed in a SW 32 Ti rotor (Beckman Coulter, cat# 369694) and centrifuged for 2 hours at 25,000xg at 4°C in a Optima centrifuge (Beckman Coulter, cat# L100-XP). Supernatant was discarded and the viral pellet resuspended in cell culture medium.

2.4.11 Lentiviral Transduction

$1x10^6$ cells from a cell line or sorted CD34$^+$ primary cells were mixed with viral supernatant (half a 150mm plate's worth of virus for cell lines or three 150mm plate's worth of virus for primary cells. 8µg/ml polybrene (Sigma-Aldrich, cat# TR-1003) was added to the mixture and cells spinoculated at 1500xg at 32°C for 90 minutes for cell lines and at 1000xg at 32°C for 45 minutes for primary cells. After 12–16 hours of incubation in an incubator, viral media was exchanged for fresh media. If the viral constructs contain a puromycin resistance cassette, cells were then puromycin (Sigma-Aldrich, cat# P8833) selected for 5-7 days at the minimum concentration at which 100% of untransduced cells would be killed.

Cells were then sorted on a FACS Aria Custom for GFP positive cells, where appropriate or single cell sorted into 96 well plates in order to grow individual clones. Single cell clones were grown for 3-4 weeks in 96 well plates in RPMI-1640 with 20% foetal calf serum, 2mM L-Glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin. Clones were then screened by FACS for GFP positive cells upon the addition of doxycycline (Sigma-Aldrich, cat# D9891) and by RT-qPCR using primers against the cDNA insert.

2.5 Phenotypic Methods

2.5.1 Colony Formation Assays

In the case of cell lines, 2000-5000 cells were seeded into 1.1ml of a pre-mix of 2.6% Methocult methylcellulose (StemCell Technologies, cat# 04100), 20% foetal calf serum, 0.4mM L-Glutamine, 25mM HEPES and 40% Iscove's Modified Dulbecco Medium (Sigma-Aldrich, cat# I2911) with or without 2 µg/ml of Doxycycline, in triplicate. Colonies were assessed and counted 10 days after incubation at 37°C.

In the case of primary cells, 2000 cells were seeded into 1.1ml of Methocult Complete (StemCell Technologies, cat# 04434) with or without 2 µg/ml of Doxycycline, in triplicate. Colonies were assessed and counted 14 days after incubation at 37°C.

2.5.2 Growth Curves of Cell Lines

2.5×10^5 cells/ml were cultured in usual culture media. Cells were mixed with an equal volume of 0.4% trypan blue (Sigma-Aldrich, cat# T8154). Live cells which did not uptake trypan blue were counted and cells were split every 2 days to 2.5×10^5 cells/ml and 2 µg/ml of doxycycline was added where appropriate.

2.5.3 Flow Cytometry

All flow cytometry was performed on a CyAn ADP machine (Beckman Coulter) and analysed on Summit 4.3 (Beckman Coulter).

Cell differentiation was assessed by antibody staining for the stem cell marker CD34 and the differentiated myeloid cell marker CD11b. $1x10^5$ cells were washed in MACS buffer and stained with 2µl of CD34-PE (Miltenyi Biotech, cat# 130-113-179) and 2µl of CD11b-APC PE (Miltenyi Biotech, cat# 130-110-554) for 20 minutes at 4°C. Cells were then washed in MACS buffer and resuspended in 400µl MACS buffer before being analysed on the CyAn flow cytometer.

Apoptosis was assessed by using Annexin V kits with Annexin V-FITC (BD Biosciences, cat# 556547) or Annexin V-APC (BD Biosciences, cat# 550474) in experiments in which cells were already GFP positive. $1x10^5$ cells were washed in PBS and Binding buffer and stained with 5µl of Annexin V and 5 µl of Propidium iodide for 15 minutes at room temperature. The cells were then diluted in binding buffer to a total volume of 500µl and then run on the CyAn flow cytometer. Propidium iodide is a DNA binding stain that is permeable only when the cell membrane is damaged in processes such as apoptosis or necrosis, and is detected in the PE channel. Annexin V binds to phosphatidylserine which is flipped to the outer plasma membrane during apoptosis.

Cell cycle analysis to distinguish the G1, S and G2-M phases of the cell cycle was performed by analysing cellular DNA content as assessed by fluorescence from the DNA intercalating stain 7-aminoactinomycin D (7-AAD) (Sigma-Aldrich, cat# SML1633). $1x10^6$ cells were washed in PBS, centrifuged at 300xg for 5 minutes and then resuspended in 250µl PBS. Then 750µl of ice cold 100% ethanol was added dropwise whilst vortexing and then cells were incubated at 4°C for at least 16 hours. Cells were then centrifuged at 500xg for 5 minutes and resuspended in 500µl PBS and allowed to rehydrate over 20 minutes during an

incubation at room temperature. Then 100µg/ml RNaseA was added and incubated for 20 minutes at room temperature in order to degrade all RNA (which would otherwise also be stained by 7-AAD). Cells were then washed in PBS and resuspended in 400µl PBS. 10µg/ml 7-AAD was then added to the cells and then the cells were run on the CyAn flow cytometer at a low speed of 100-200 events per second. 7-AAD fluorescence was detected in the PE-Texas Red channel.

G0 cell cycle analysis was performed by staining 1 million cells in 1ml of cell culture medium with 10µg/ml Hoescht 33342 (Sigma-Aldrich, cat#14533) with a 45 minute incubation at 37°C followed by 0.5µg/ml of Pyronin Y (Sigma-Aldrich, cat# 213519) with a 15 minute incubation at 37°C. The cells were then run on the CyAn flow cytometer. Hoescht 33342 is a DNA intercalating stain and was detected in the e450 channel. Pyronin Y is an RNA intercalating stain and was detected in the PE channel. The cells in G0 phase were defined as those with low RNA content but diploid DNA content.

2.5.4 Preparing Wright-Giemsa stained slides

10000 cells were washed in PBS and then resuspended in 100µl PBS. Cells were then loaded onto a cytospin funnel which is clipped onto a glass microscope slide (Thermo Scientific, cat# 12392098) and filter; this apparatus was loaded onto a Cytospin 4 cytocentrifuge (Thermo Scientific, cat# A78300003). The cytospin machine was used at 800rpm for 4 minutes with the low acceleration setting. Cells were then fixed onto slides by adding 100µl of ice cold methanol (Sigma-Aldrich, cat# 34860) for 1 minute and then letting the slide air-dry. Then 0.5ml of Wright-Giemsa stain (Sigma-Aldrich, cat# WG16) was added onto the slide and incubated for 1 minute before 0.5ml of PBS was also added and mixed in and

incubated for a further 3 minutes. Then the slides were washed with distilled water.

2.5.5 Western Blot

Protein extraction was carried out by resuspending 1×10^7 cells per 1ml of RIPA (Radioimmunoprecipitation assay) buffer and rotating on a wheel at 4°C for 4 hours. RIPA buffer consists of 150mM NaCl, 10mM Tris-HCl pH 8.0, 0.1% SDS, 1% Triton X-100, 1% Sodium deoxycholate, 5mM EDTA) with 1:100 Protease Inhibitory Cocktail. The extracted protein was quantified by using the Pierce BCA Protein Assay Kit (Thermo Scientific, cat# 23225); 25µl of either BSA protein standard or sample was mixed with 200µl of reagent A and B mix in a 96 well plate. After an incubation at 37°C for 30 minutes, the plate was read in a Victor 2 plate reader (Perkin Elmer, cat# wallac1420) and protein concentrations of the samples determined by using a standard curve. 30µg of protein was mixed with Laemmli buffer (0.3125M Tris-HCl pH6.8, 10% β2 mercaptoethanol, 10% SDS, 25% glycerol and 0.05% bromophenol blue [Sigma-Aldrich, cat# B0126]) and was denatured by heated at 95°C for 10 minutes. The protein was then loaded onto a 5-20% Mini-Protean TGX gel (Bio-Rad, cat# 4561096), as well as 10µl protein ladder (Thermo Scientific, cat# 26623). The gel was placed in SDS-PAGE running buffer (0.025M Tris pH 8.3, 0.192M glycine and 0.1% SDS) in a Mini-Protean tetra vertical electrophoresis cell (Bio-Rad, cat# 1658004); this gel was run at 25V for 90 minutes to separate the protein by molecular weight. The protein was then transferred onto a nitrocellulose membrane using a TGX transfer gel pack (Bio-Rad, cat# 1704158) on a Trans-Blot Turbo transfer system (Bio-Rad, cat# 1704150), utilising the 7 min mixed Molecular Weight program. The

membrane was then blocked with 10% milk (Sigma-Aldrich, cat# M7409) in TBST (0.5% Tween, 0.075M NaCl and 0.01M Tris pH 7.5) for 1 hour at room temperature before being incubated with a primary antibody diluted in 5% milk at 4°C for at least 16 hours. Following this, the membrane was washed three times in TBST and incubated with the appropriate Horse Radish Peroxidase (HRP) conjugated secondary antibody in 5% milk for 1 hour at room temperature. The membrane was then developed by adding 1ml of reagent 1 and 2 from the Pierce Enhanced Chemiluminescence kit (Thermo Scientific, cat# 32106) and imaged on a Bio-Rad ChemiDoc XRS+.

Antibody	Dilution	Manufacturer and Catalogue Number
Mouse anti-CCND2 Primary	1:1000	Proteintech, cat# 10934-1-AP
Mouse anti-GAPDH Primary	1:10000	Abcam, cat# ab8245
Rabbit anti-WT1 Primary	1:2000	Abcam, cat# ab89901
Rabbit Anti-mouse HRP Secondary	1:7500	Abcam, cat# ab97046
Goat Anti-rabbit HRP Secondary	1:7500	Abcam, cat# ab6721

Table 2-14: Antibodies used for Western blot in this study.

2.6 Xenotransplantation

Xenotransplantation experiments were conducted by Dr Helen Blair and Prof Olaf Heidenreich at the University of Newcastle. Immunodeficient Rag2$^{-/-}$, Il2r$\gamma^{-/-}$ 129 × Balb/c (RG) mice were housed under pathogen-free conditions and all animal

work was conducted in accordance with Home Office Project License PPL60/4552. 2.5x10^5 Kasumi-1 cells were intrahepatically injected into 14 newborn RG mice and 12 days later, mice were randomised into two treatment groups, with one given 50 mg/kg doxycycline three times per week intraperitoneally in an unblinded fashion until the experimental endpoint. 5×10^5 MV4-11 cells were intrafemorally injected into RG mice and similarly randomised 12 days later into two groups. For the doxycycline group, doxycycline was added at 2 mg/ml for the first 3 days and at 0.2 mg/ml subsequently to drinking water containing 2% sucrose. Controls mice were given just water containing 2% sucrose. Animals were humanely killed upon clinical signs of illness or at defined experimental endpoints.

RESULTS

Chapter 3: A crucial role of AP-1 in AML

3.1 The AP-1 family of transcription factors is vital for leukaemogenesis

3.1.1 AP-1 family members are upregulated in AML relative to healthy CD34[+] cells

Large consortia studies in AML such as The Cancer Genome Atlas (TCGA, 2013) have profiled gene expression in AML. However, a major limitation of such studies is that RNA is extracted from whole bone marrow and so contaminating normal cells may alter the gene expression profiles of the leukaemic cells. We have collected a large number of AML patient samples and purified leukaemic blasts based upon CD34 or CD117 stem cell surface protein expression and assessed gene expression of these purified cells (Assi et al., 2019). Whilst TCGA does not reveal any major changes in the gene expression in the AP-1 family of transcription factors, within our data we show upregulation of FOS-, JUN-, JDP- and ATF- subtypes of AP-1 in different mutational subtypes of AML relative to healthy CD34[+] cells (Figure 3-1).
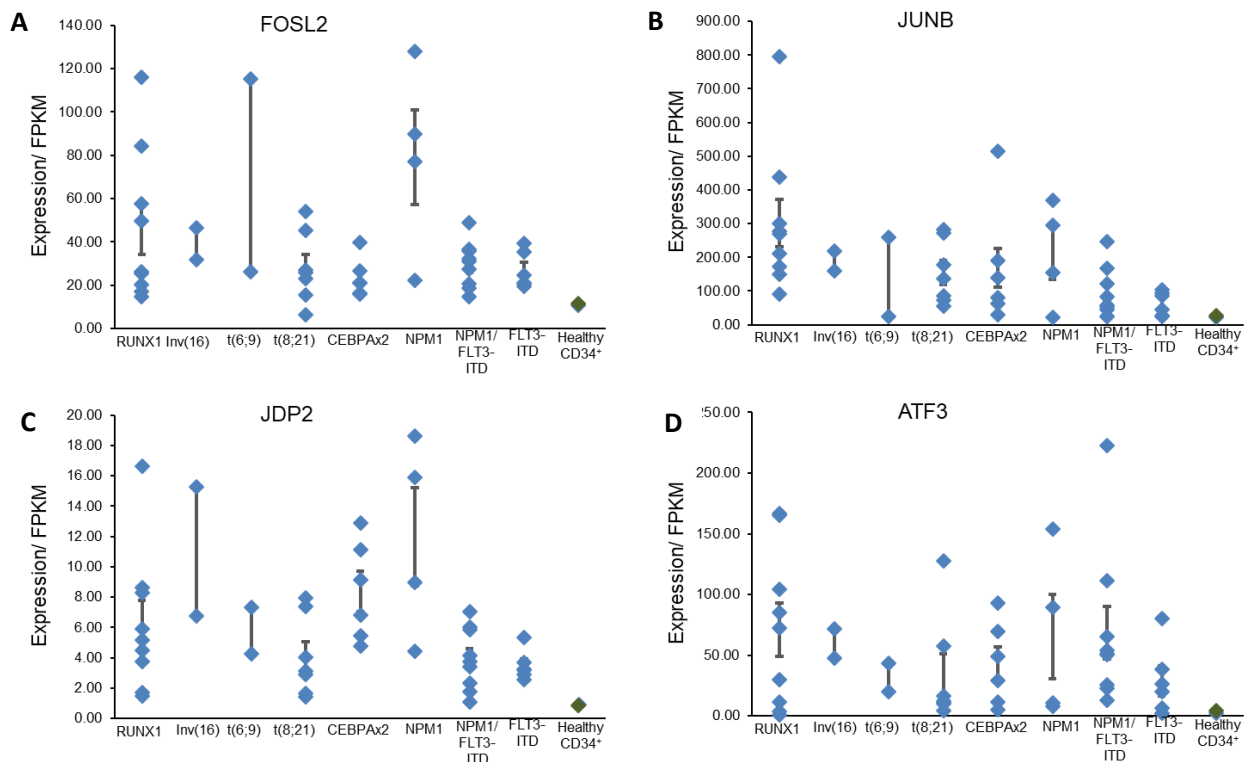
Figure 3-1: AP-1 Family members are upregulated in AML relative to healthy CD34+ cells.

Expression of (A) *FOSL2*, (B) *JUNB*, (C) *JDP2* or (D) *ATF3* mRNA in purified primary AML leukaemic blasts with different mutational subtypes, as measured by RNA-seq. Error bars show standard error of the mean. Some RNA-seq experiments were performed by Dr Rosie Imperato (Assi et al., 2019).

3.1.2 AP-1 is essential for t(8;21) leukaemogenesis

In order to establish the significance of the observed AP-1 upregulation, functional experiments are vital. As part of a collaborative project, the Heidenreich Lab at Newcastle University performed an shRNA depletion screen against a core set of genes regulated by the RUNX1-ETO oncoprotein in t(8;21) AML including *JUN* (Martinez-Soria et al., 2018). Figure 3-2 A-B shows the experimental setup for the shRNA depletion screen in serial colony formation assays and with long-

term culture. With a negative control of shRNA against a non-targeting control (NTC), no depletion of shRNA barcodes was seen upon the induction of doxycycline. In addition, with the positive control of shRNA against the *RUNX1-ETO* driver oncogene, large depletion of shRNA barcodes was seen upon induction of doxycycline (Figure 3-2C). Figure 3-2D shows the experimental strategy for the *in vivo* shRNA depletion screen with the pSLIEW plasmid for the expression of luciferase and Green Fluorescent Protein (GFP) as well as the pTRIPZ plasmid for the doxycycline-induced expression of Red Fluorescent Protein (RFP) in Kasumi-1 cells. This strategy allowed for the tracking of leukaemic tumour formation through FACS (GFP) and bioluminescence imagining (luciferase) as well as confirmation of the induction of shRNA through FACS looking at RFP (Figure 3-2E).

We analysed the original sequencing data and when we looked at the barcode representation for shRNAs against *JUN*, we found considerable heterogeneity in the importance of *JUN* in leukaemic maintenance between the experiments and between different shRNA constructs (Figure 3-2F). Intriguingly, the most significant barcode depletion was seen in the *in vivo* experiments suggesting that *JUN* may have a role in leukaemic blast survival within the bone marrow microenvironment.
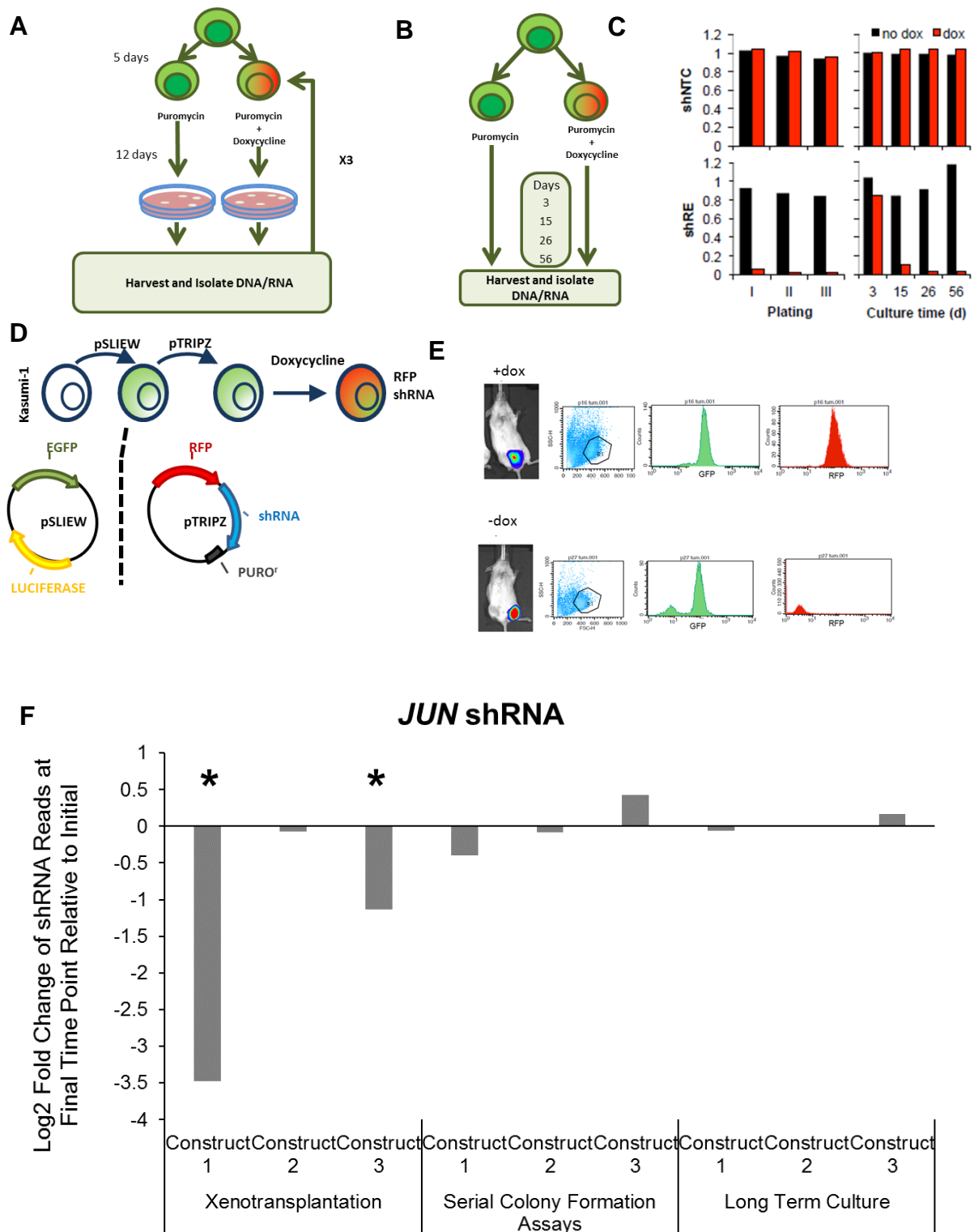
Figure 3-2: shRNA targeting *JUN* are depleted *in vivo* in a shRNA depletion screen.

(A-B) Experimental strategy of Kasumi-1 cell shRNA depletion screen in (A) serial colony formation and (B) liquid culture (C) positive control of depletion of RUNX1-

ETO (D-E) Experimental strategy of *in vivo* shRNA depletion screen using pSLIEW and pTRIPZ plasmids with validation (E) showing doxycycline-induced RFP cells. (F) Log2 fold change of shRNA reads at the final time point relative to initial time point in xenotransplantation, serial colony formation assays or long term culture. * indicates p-value < 0.05. Experiments performed by Dr Natalia Martinez-Soria, Newcastle University (Martinez-Soria et al., 2018).

When further examining *JUN* expression in t(8;21) AML, the upregulation in *JUN* appeared to be driven by the RUNX1-ETO driver oncoprotein as *RUNX1-ETO* knockdown in Kasumi-1 cells led to reduction in *JUN*. But of note, the gene expression of other AP-1 family members including *FOSL1* and *FOSL2* were RUNX1-ETO dependent as well (Figure 3-3).
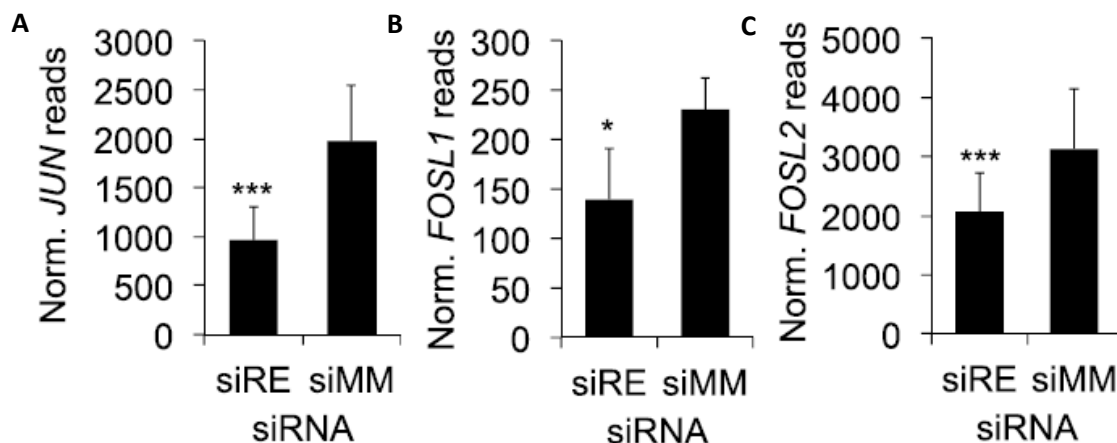
Figure 3-3: Gene expression of AP-1 family members is downregulated by *RUNX1-ETO* knockdown.

(A-C) Histograms of quantile normalised RNA-seq reads of (A) *JUN*, (B) *FOSL1* and (C) *FOSL2* mRNA expression from Kasumi-1 cells transfected with siRNA against RUNX1-ETO (siRE) or a mismatch control (siMM). n=3 and error bars show standard deviation. Significance was determined using a 2 tailed student's t-test. * indicates p-value <0.05, ** indicates p-value <0.005 and *** indicates p-value <0.0005.

These findings that the driver oncoprotein RUNX1-ETO was upregulating several AP-1 family members, coupled with the large degree of heterogeneity in the role of JUN in leukaemic maintenance suggested to us that JUN was not alone necessary and sufficient for the survival of leukaemic blasts. AP-1 members are known to be largely functionally redundant and form multiple types of heterodimers (Mechta-Grigoriou et al., 2001) and so we were required to use a method of reducing all AP-1 activity in order to fully understand the role of these factors. To this end, we expressed a doxycycline-inducible dominant negative FOS (dnFOS) (Olive et al., 1997) in Kasumi-1 cells as a way of blocking all AP-1 binding.

Doxycycline-mediated induction of dnFOS led to a decrease in cell growth (Figure 3-4A) whilst induction of an Empty Vector control did not (Figure 3-4B), suggesting that AP-1 transcription factors are important for leukaemogenesis. Induction of dnFOS led to a G1 cell cycle arrest whereas the Empty Vector control had no effect upon cell cycle phase (Figure 3-4C). However, this cell cycle arrest was entirely reversible and when doxycycline was washed off and the experiment repeated 3 days later, the cell cycle distribution reverted to that of un-induced cells (Figure 3-4D). Furthermore, the colony formation ability of Kasumi-1 cells was reduced upon induction of dnFOS whilst the Empty Vector control was unaffected (Figure 3-4E).
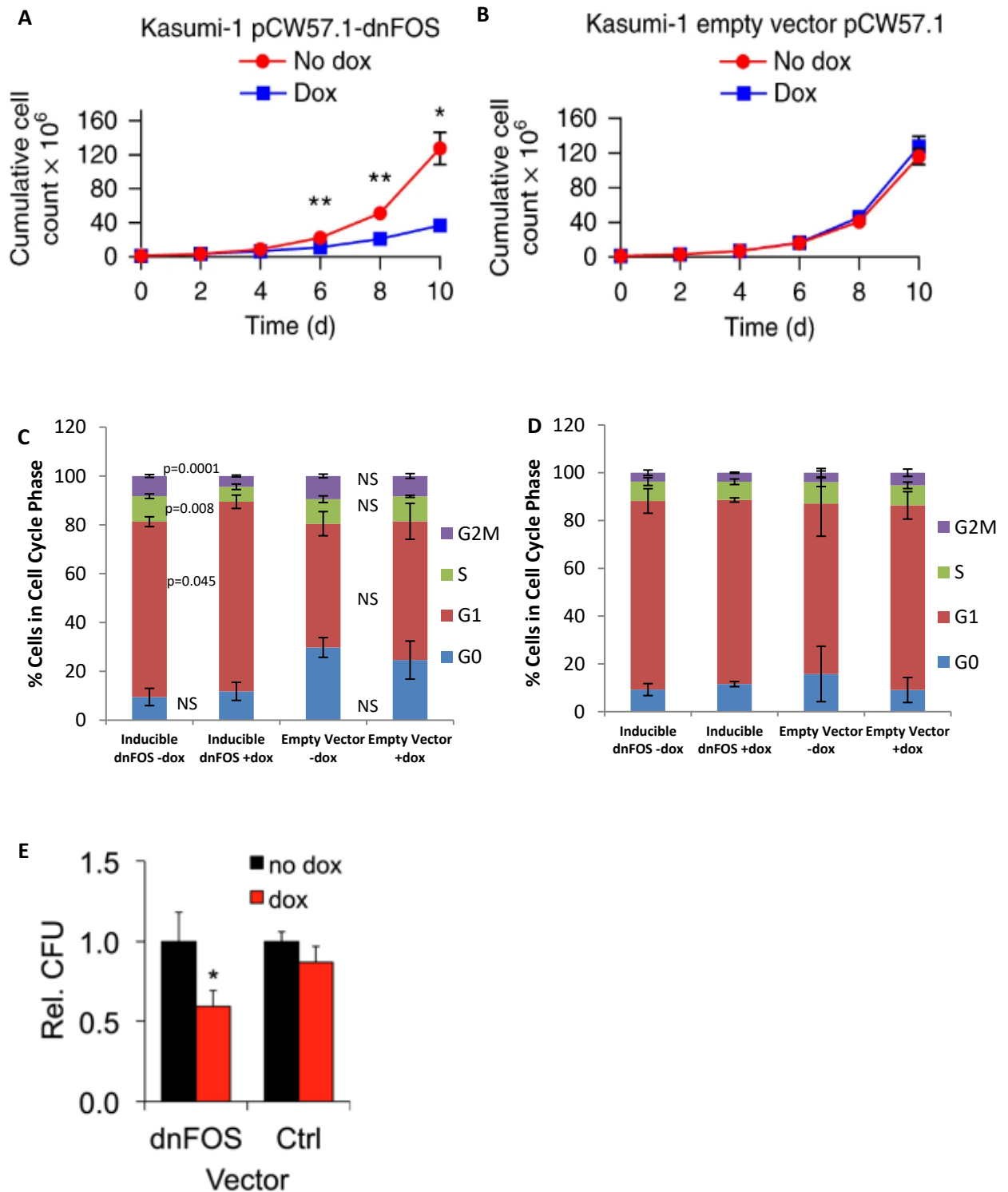
Figure 3-4: Induction of dnFOS leads to reduced growth and clonogenicity and a G1 cell cycle arrest.

(A-B) Time course of cumulative cell growth of Kasumi-1 cells with or without doxycycline induction of a (A) Dominant Negative FOS (dnFOS) or (B) Empty

Vector control. (C-D) Cell cycle phase of Kasumi-1 cells with or without doxycycline induction of an Empty Vector control or Dominant Negative FOS (dnFOS) (C) 3 days after induction or (D) a further 3 days post-washout of doxycycline. (E) Relative colony formation (CFU) per 2000 cells seeded with Kasumi-1 cells with or without doxycycline induction of an Empty Vector control or Dominant Negative FOS (dnFOS), normalised to no doxycycline counts. For all panels, n=3 and error bars show standard deviation. Significance was determined using a 2 tailed student's t-test. * indicates p-value <0.05, ** indicates p-value <0.005.

3.1.3 AP-1 is essential for FLT3-ITD leukaemogenesis but is not required for healthy stem cell survival

Similarly, we wanted to confirm that AP-1 was also essential in other mutational subtypes of AML. Consequently, we expressed the Empty Vector and dnFOS in the FLT3-ITD AML cell line, MV4-11. Whilst induction of an Empty Vector control (Figure 3-5A) had no effect upon cell growth, induction of dnFOS greatly reduced cell growth (Figure 3-5B) suggesting an important role of AP-1 factors in FLT3-ITD leukaemogenesis.

However, a limitation of using MV4-11 cells is the fact that they also contain the MLL-AF9 driver oncoprotein, which also has significant effect upon cell behaviour, whereas we wanted to study the contribution of FLT3-ITD alone. Consequently, we transduced viral vectors carrying dnFOS or Empty Vector control into primary FLT3-ITD AML cells and found that dnFOS greatly reduced colony formation ability (Figure 3-5C). However, when this experiment was repeated in healthy donor derived CD34$^+$ Peripheral Blood Stem Cells (PBSC), the expression of dnFOS did not affect colony formation (Figure 3-5D). This result

suggests that AP-1 factor binding is uniquely important in the survival of leukaemic cells and normal haematopoiesis is not affected.

A MV4-11 empty vector pCW57.1

B MV4-11 pCW57.1-dnFOS

C FLT3-ITD⁺ primary cell CFA

D PBSC primary cell CFA
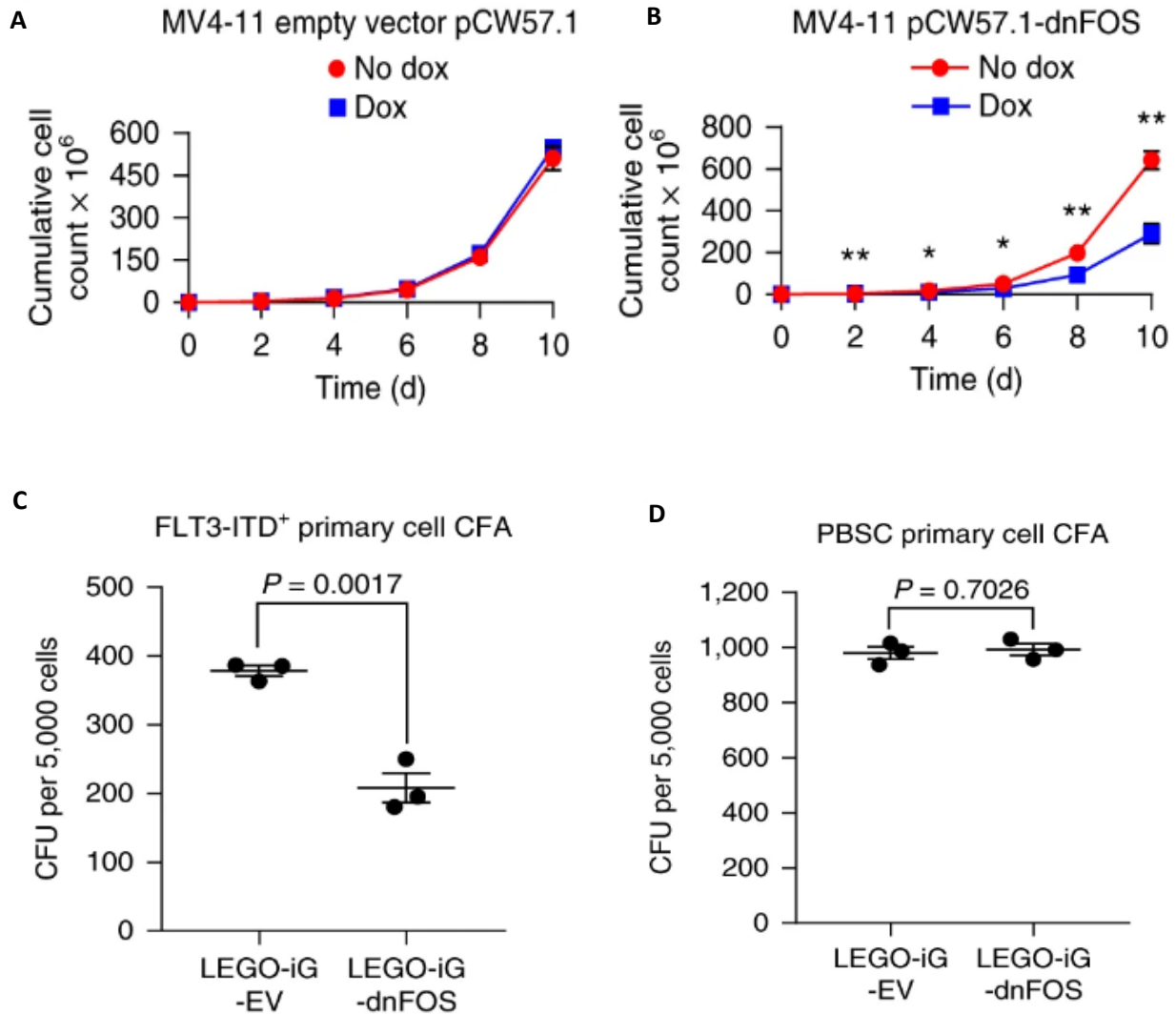
<u>Figure 3-5: Primary FLT3-ITD AML but not healthy peripheral blood stem cells reduce their growth or colony formation after dnFOS induction.</u>

(A-B) Time course of cumulative cell growth of MV4-11 cells with or without doxycycline induction of an (A) Empty Vector control or (B) Dominant Negative FOS (dnFOS). (C-D) Colony formation (CFU) per 5000 cells seeded with (C) primary patient-derived FLT3-ITD AML or (D) primary healthy donor derived Peripheral Blood Stem Cells (PBSC) transduced with an Empty Vector control or Dominant Negative FOS (dnFOS). For all panels, n=3 and error bars show standard deviation. Significance was determined using a 2 tailed student's t-test.

* indicates p-value <0.05, ** indicates p-value <0.005. Primary cell experiments (lower two panels) performed together with Dr Daniel Coleman.

3.1.4 AP-1 is essential for leukaemogenesis for two AML subtypes *in vivo*

AP-1 is activated by oncogenic signalling and its activity is increased by activating mutations in signalling molecules such as *KIT, FLT3, NRAS and KRAS* (Papaemmanuil et al., 2016). An important aspect of signalling is the interaction with the bone marrow niche (Zhou et al., 2016), which we cannot model through *in vitro* experimentation. Therefore, we performed xenotransplantation experiments whereby cells transduced with doxycycline-induced dnFOS were injected into MISTRG mice which can secrete human cytokines and most closely replicate the human bone marrow niche (Rongvaux et al., 2014).

Beginning with Kasumi-1 cells, mice with doxycycline in their feed had decreased tumour masses compared with those that did not (Figure 3-6A). The only two mice with doxycycline in their feed that developed tumours, did not express GFP in their tumours which suggests that non-transduced Kasumi-1 cells were responsible for the tumours. Unfortunately due to a respiratory infection in the mouse colony, the mice had to be culled and an analysis of tumour progression-free survival could not be conducted.

When the experiments were repeated with the FLT3-ITD mutated MV4-11 cells, a full Kaplan-Meier survival analysis could be conducted, which showed that mice with doxycycline in their feed and hence induction of dnFOS survived for longer than mice which did not (Figure 3-6B).

Both of these *in vivo* findings were more dramatic than might have been expected based upon the *in vitro* experiments (Sections 3.1.2 - 3.1.3). Consequently, AP-1 signalling probably has a major role in leukaemogenesis within the bone marrow niche.
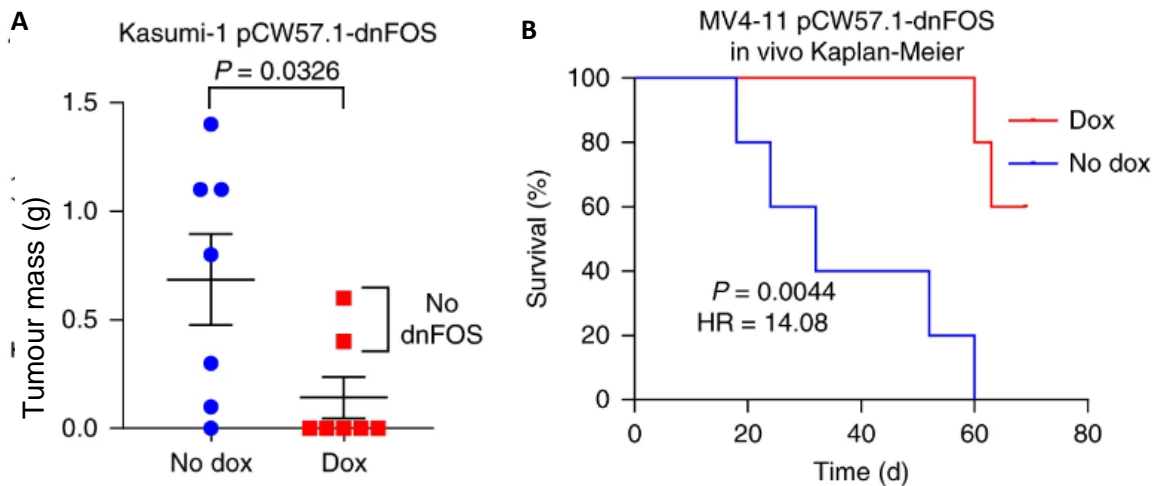
Figure 3-6: Induction of dnFOS decreases tumour engraftment *in vivo.*

(A) Tumour mass of tumours formed in mice xenotransplanted with Kasumi-1 cells transduced with doxycycline-induced dnFOS, with or without the addition of doxycycline to the mouse feed. (B) Kaplan-Meier curve showing the survival of mice xenotransplanted with MV4-11 cells transduced with doxycycline-induced dnFOS, with or without the addition of doxycycline to the mouse feed. Xenotransplantation aspects of the experiments were performed by Dr Helen Blair. Published in (Assi et al., 2019).

3.2 AP-1 regulates the Cyclin D2 *(CCND2) gene*

We previously discussed in Section 3.1.2 that induction of dnFOS led to a G1 cell cycle arrest. The G1 cell cycle transition occurs through Cyclin Dependent Kinases (CDK) and Cyclin expression, particularly through the CDK4 - CDK6 -

Cyclin D complex (Sherr et al., 2016). Therefore we explored these factors further as to how they may interact with AP-1.

3.2.1 AP-1 binding is important for the expression of Cyclin D2

Induction of dnFOS (Figure 3-7A) led to a decrease in Cyclin D2 (*CCND2)* mRNA expression (Figure 3-7B) but not in the Empty Vector control. We found the same results when looking for at CCND2 protein expression (Figure 3-7C).
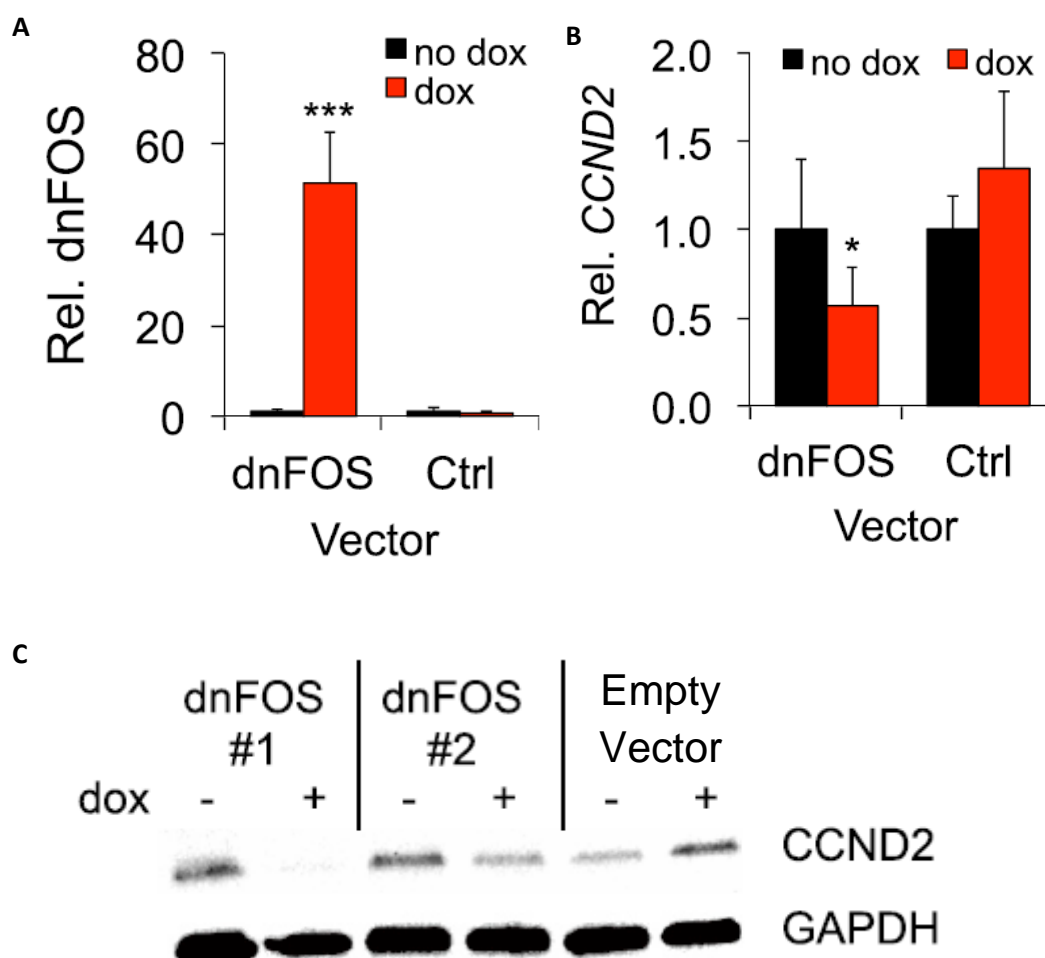
Figure 3-7: Induction of dnFOS decreases *CCND2* (Cyclin D2) mRNA and protein expression.

(A-B) RT-qPCR measuring (A) dnFOS and (B) *CCND2* in Kasumi-1 cells transduced with doxycycline-induced dnFOS, with or without the addition of

doxycycline. (C) Western blot showing protein expression of CCND2 and GAPDH in two Kasumi-1 clones transduced with doxycycline-induced dnFOS or with an Empty Vector control, with or without the addition of doxycycline. n=3 and error bars show standard deviation. * indicates p-value <0.05, ** indicates p-value <0.005, *** indicates p-value <0.0005.

Next, we wanted to confirm that AP-1 family members were directly able to regulate *CCND2* expression by showing that they could bind to *CCND2* cis-regulatory elements. We chose JUND out of the various AP-1 family members due to the technical aspect of a JUND antibody being the only ChIP-grade antibody available. JUND ChIP-seq demonstrated that JUND could bind to a -31kb and a -4kb *CCND2* cis-regulatory element as well as the *CCND2* promoter (Figure 3-8A). We discussed in Section 3.1.2 that AP-1 family member gene expression appears to be regulated by RUNX1-ETO. We therefore wanted to see whether any of the three *CCND2* JUND binding sites were dependent on RUNX1-ETO. RUNX1-ETO can bind to the -31kb cis-regulatory element and promoter of *CCND2*. Of these two sites, it appears to be at the promoter where JUND binding is abrogated upon *RUNX1-ETO* knockdown. This finding was also confirmed using ChIP-qPCR (Figure 3-8B).
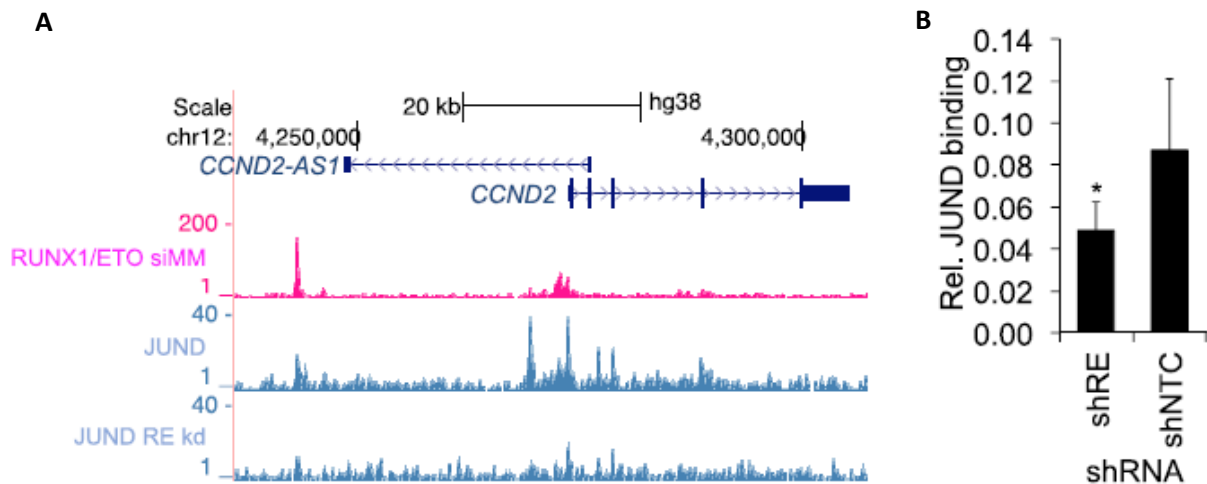
Figure 3-8: JUND binds to the *CCND2* promoter.

(A) UCSC genome browser screenshot of RUNX1-ETO and JUND ChIP-seq at the *CCND2* locus of Kasumi-1 cells transfected with siRNA targeting *RUNX1-ETO* or a mismatch control. (B) ChIP-qPCR of JUND at the *CCND2* promoter of Kasumi-1 cells transduced with shRNA against *RUNX1-ETO* or a non-targeting control. n=3 and error bars show standard deviation. Significance was determined using a 2 tailed student's t-test. * indicates p-value <0.05. ChIP experiments performed by Dr Paulynn Chin (Martinez-Soria et al., 2018).

3.2.2 Cyclin D2 is a therapeutic vulnerability in AML

The CDK4-CDK6-CyclinD2 complex is targetable using the drug Palbociclib, which is already routinely used in patients with metastatic breast cancer (Finn et al., 2015). Having established that t(8;21) AML cells require AP-1 binding to *CCND2* cis-regulatory elements to prevent a G1 cell cycle arrest, we wanted to examine if this feature could be exploited therapeutically. Use of Palbociclib led to reduction in cell numbers of t(8;21) AML cell lines, Kasumi-1 and SKNO-1 in a dose-dependent manner in nanomolar concentrations, which are achievable in humans from a pharmacokinetics point of view (Figure 3-9A). Unfortunately,

Palbociclib was also found to be toxic to healthy CD34$^+$ cells (Figure 3-9B). Pancytopenias have been reported in breast cancer patients treated with this drug too but these side effects could be managed with supportive therapies (Finn et al., 2015).

To confirm the efficacy of Palbociclib in t(8;21) AML, the drug was used on primary patient-derived t(8;21) AML cells and it again led to a dose-dependent reduction in cell numbers (Figure 3-9C). Furthermore, in *in vivo* experiments where Kasumi-1 cells were injected into MISTRG mice, three periods of treatment with Palbociclib could increase survival compared with an Empty Vector control (Figure 3-9D).
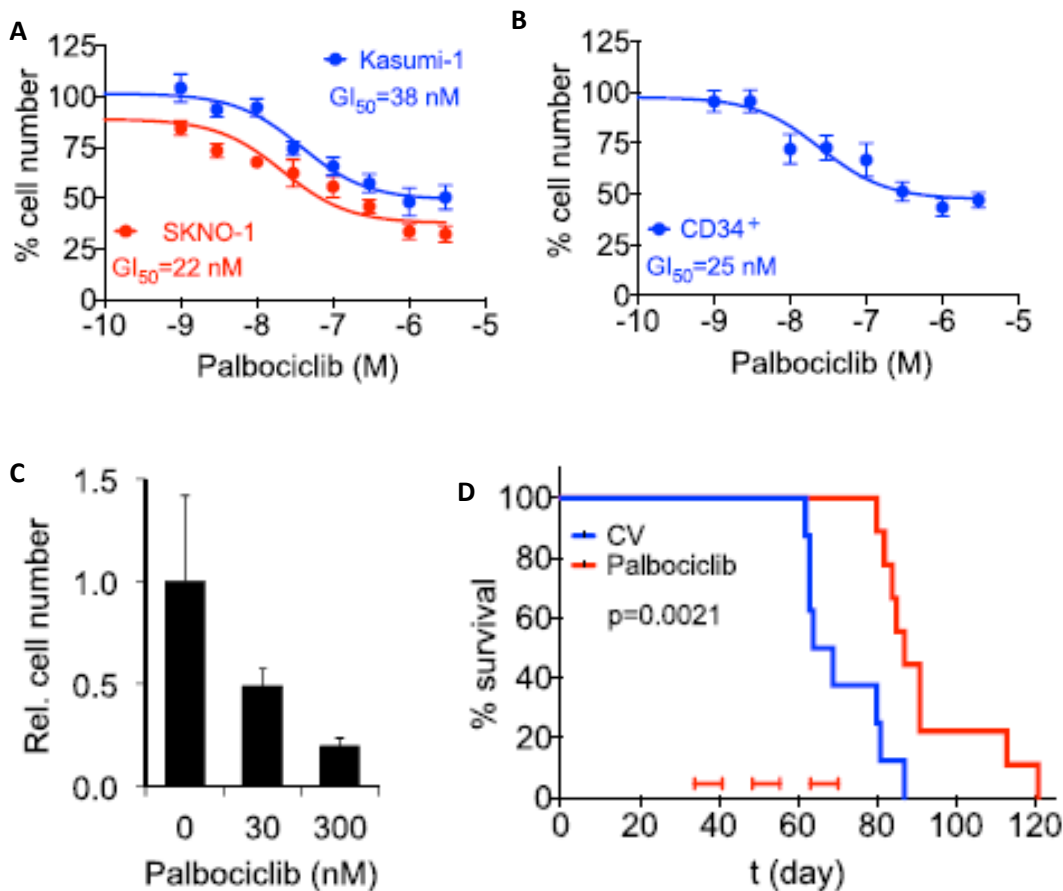
(A-B) Dose-response curves of (A) SKNO-1 and Kasumi-1 cells or (B) primary healthy CD34+ cells treated with Palbociclib (C) Relative colony formation of primary t(8;21) AML cells treated with different concentrations of Palbociclib. (D) Kaplan-Meier curve of Kasumi-1 cells xenotransplanted into mice and treated with Palbociclib or a Control Vehicle (CV). Horizontal red bars indicate the periods of drug treatment. n=3 and error bars show standard deviation. Experiments performed by Dr Lynsey McKenzie and Dr Helen Blair, Newcastle University (Martinez-Soria et al., 2018).

3.3 The genome-wide role of AP-1 family members in AML

3.3.1 Induction of dnFOS in Kasumi-1 cells leads to aberrant cell differentiation

We next investigated the gene expression changes observed after dnFOS induction using RNA-seq. When examining the expression of various differentiation markers in the RNA-seq data of Kasumi-1 cells, we observed an upregulation of genes coding for lymphoid cell surface markers such as *CD52*, *CD79A*, *CD79B*, *CD74*, *CD48* and *CD96* was seen with dnFOS induction (Figure 3-10A) suggesting aberrant cell differentiation. Kasumi-1 cells already aberrantly express the B cell marker CD19 due to PAX5 upregulation (Walter et al., 2010) and the expression of other lymphoid differentiation markers is seen after induction of dnFOS. Examining protein expression by FACS, no change in expression was seen of the stem cell marker CD34 and the mature myeloid marker CD11b after dnFOS induction, suggesting no increase in myeloid differentiation (Figure 3-10B). However, expression of the lymphoid marker CD52

increased. Consequently, if AP-1 could be targeted with drug therapy, targeting CD52 with currently available therapies such as Alemtuzumab, (Tibes et al., 2006) may represent an avenue for synthetic lethality.

We derived KEGG (Kyoto Encyclopaedia of Genes and Genomes) pathways of upregulated (Figure 3-10C) and downregulated (Figure 3-10D) genes after dnFOS induction to examine cellular pathways changing after AP-1 activity was blocked with dnFOS. Firstly, numerous lymphocyte related KEGG pathways including B cell signalling, T cell differentiation and immunodeficiency related genes were upregulated, concurring with the upregulated lymphoid differentiation markers data (Figure 3-10A).

The RAS signalling pathway was also upregulated after dnFOS induction. The RAS/Mitogen-Activated Protein Kinase (MAPK) pathway is known to activate AP-1 (Zhang and Liu, 2002). This results points to a negative feedback loop whereby with reduced AP-1 binding, RAS/MAPK are de-repressed and upregulated.

Furthermore, within the 'transcriptional misregulation in cancer' upregulated KEGG pathway, *MYCN* was upregulated nearly 7-fold. MYCN has previously been shown to contribute to leukaemic proliferation in mouse models (Kawagoe et al., 2007) through increased JUN signalling, suggesting an additional feedback loop maintaining the activity of pathways downstream of AP-1.

Downregulated KEGG pathways after dnFOS induction included those related to inflammation with genes within disease pathways such as Chagas disease, Leishmaniasis, Pertussis, Tuberculosis and Legionellosis emerging. AML cells are known to flourish in an inflammatory microenvironment which both drives

clonal evolution and promotes leukaemic self-renewal (Hemmati et al., 2017). AP-1 as a downstream effector of MAPK signalling is a mediator of inflammatory cytokines and so inflammatory pathways are downregulated when dnFOS is induced.

Furthermore, downregulated KEGG pathways included several pathways demonstrating altered metabolism involving steroid biosynthesis, butanoate, pyruvate, tryptophan, arginine and proline metabolism. In order for the AML cells to replicate and grow rapidly they are in need of amino acid and micronutrients and our data appears the AP-1 pathway upregulates these pathways.

C

dnFOS upregulated KEGG pathways

D

dnFOS downregulated KEGG pathways
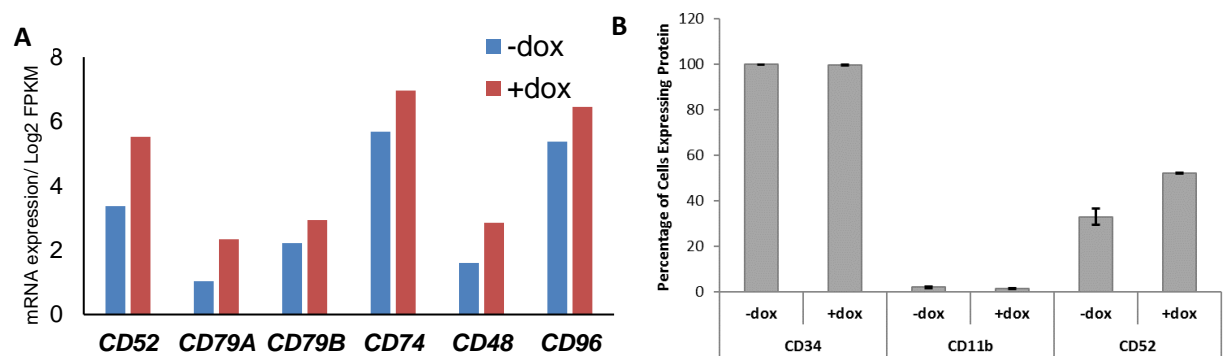
Figure 3-10: Gene and protein expression changes with dnFOS induction.

(A) mRNA expression level of the indicated genes as determined by RNA-seq of Kasumi-1 cells with or without doxycycline-mediated induction of dnFOS. (B) Protein expression of level of CD34, CD11b or CD52 of Kasumi-1 cells with or without doxycycline-mediated induction of dnFOS for 8 days, as determined by FACS. (C-D) Enriched KEGG pathway analysis derived from (C) upregulated or (D) downregulated genes from RNA-seq of Kasumi-1 cell induced for dnFOS expression. n=3 and error bars show standard deviation.

To obtain further evidence of aberrant cell differentiation and to home in on the molecular mechanisms driving it, we performed DNase I-seq and examined enriched TF binding motifs within sites changing after dnFOS induction. Within the 1416 lost DHSs upon dnFOS induction, we observed a significant enrichment of the AP-1 motif (Figure 3-11A) suggesting loss of AP-1 binding and chromatin closing as a consequence. An enrichment for CEBP was also seen in lost DHSs upon dnFOS induction which suggests that the cells are perhaps becoming less differentiated (Zhang et al., 1997a). Expression of dnFOS also gives rise to 698 new DHSs which were enriched for GATA motifs. *GATA2* in particular is the GATA family member upregulated by dnFOS (Figure 3-11D) and so enrichment of GATA motifs in dnFOS-specific DHSs suggests that the cells are becoming more immature (Tsai et al., 1994), consistent with finding of a loss of CEBP motifs. In addition, enriched RUNX motif were observed at gained and lost DHSs upon dnFOS induction, suggesting that RUNX1 or RUNX1-ETO may be re-distributed.

In order to confirm that motif enrichment changes correlated with changes in transcription factor binding, we performed ChIP-seq of Kasumi-1 cells with or

without induction of dnFOS and plotted the peaks alongside their DHSs (Figure 3-11B). Loss of JUND sites upon dnFOS induction confirmed that dnFOS indeed blocked AP-1 binding in chromatin. In cells derived from Embryonic Stem cells differentiated into haematopoietic progenitors, it has previously been shown that AP-1 binding occurs together with that of TEAD4 (Obier et al., 2016) and in Kasumi-1 cells, blocking AP-1 binding also reduced TEAD4 binding. C/EBPα and GATA2 ChIP-seq confirmed our motif analysis findings above. RUNX1 and RUNX1-ETO ChIP-seq showed that both TFs are redistributed upon dnFOS induction and that loss of these factors in particular led to a closing of chromatin sites. Lastly, LDB1 is a chromatin looping TF that is part of the core RUNX1-ETO complex (Abe et al., 2015) and LBD1 binding reduced at sites concordant with RUNX1-ETO upon dnFOS induction.

Of note, the DHS ranking correlated well with gene expression changes (Figure 3-11A-B) which suggests that AP-1 related changes in chromatin accessibility were happening at important cis-regulatory elements such as enhancers.

To exemplify dnFOS-related epigenetic and transcriptomic changes, Figure 3-11C shows a UCSC genome browser screenshot of all the data at the *GATA2* locus. Induction of dnFOS led to decreased JUND binding at two intragenic cis-regulatory elements (intron 2 and intron 5) and led to a slight decrease in chromatin accessibility at the intron 2 site. Induction of dnFOS also led to a decrease in C/EBPα, TEAD4 and RUNX1 binding and an increase in GATA2 and LMO2 binding at the intron 2 site. These changes correlated with an increase in *GATA2* mRNA expression (Figure 3-11D).

A

dnFOS DNasel-seq

no Dox | plus Dox

Motifs: AP-1, C/EBP, GATA, MAX, RUNX, CREB, NRF1, NFY, ETS, PU.1, CTCF, SP1

RNA-seq Gene expression
FC (+Dox/no Dox)

Ranked by fold-change

n=698 1
2
n=1416 3

+/-1kb

+/- 1kb

-1.00
-0.67
-0.33
0.00
0.33
0.67
1.00

B

dnFOS DNasel-seq

CTRL | +dnFOS

ChIP-seq

GATA2: CTRL | +dnFOS
JUND: CTRL | +dnFOS
TEAD4: CTRL | +dnFOS
CEBPA: CTRL | +dnFOS
LDB1: MM | +dnFOS
RUNX1: MM | +dnFOS
RUNX1-ETO: MM | +dnFOS

Gene expression
FC (+dox/-dox)

Ranked by fold-change

n=698 1
2
n=1416 3

+/-1kb

-1.0   1.0

167

C

D GATA2

Figure 3-11: Transcription factor binding is altered by the induction of dnFOS.

(A-B) Density plots of DNase I Hypersensitive Site (DHS) peaks in a 2kb window, ranked top to bottom by relative tag count of peaks in Kasumi-1 cells with doxycycline-induced dnFOS, compared to the uninduced control. (A) Enriched transcription factor binding motifs centred around the DHS summit or (B) ChIP-seq peaks are plotted alongside, as well as gene expression fold changes of the doxycycline-induced cells compared with the no doxycycline control. The bars on the side depict the number of specific and shared peaks. (C) UCSC genome browser screenshot of the indicated ChIP-seq, DNase I-seq or RNA-seq signals at the *GATA2* locus. (D) mRNA expression of *GATA2*, as measured by RNA-seq of Kasumi-1 cells with and without doxycycline-mediated induction of dnFOS. Some of the ChIP-seq experiments performed by Dr Anetta Ptasinska and Dr Paulynn Chin.

Putting together the transcriptomic and the epigenetic data, the induction of dnFOS induces a more immature stem-cell state in Kasumi-1 cells together with aberrant lymphocyte pathway activation.

Figure 3-12 shows the genes that are upregulated after dnFOS expression and are organised into overlapping network modules bound by C/EBPα, RUNX1 or RUNX1-ETO with some such as *BCL6* bound by all three factors. BCL6 is known to repress *CCND2* expression by binding together with STAT5 at a cis-regulatory element (Fernández de Mattos et al., 2004) and may represent another mechanism by with induction of dnFOS decreased *CCND2* expression.
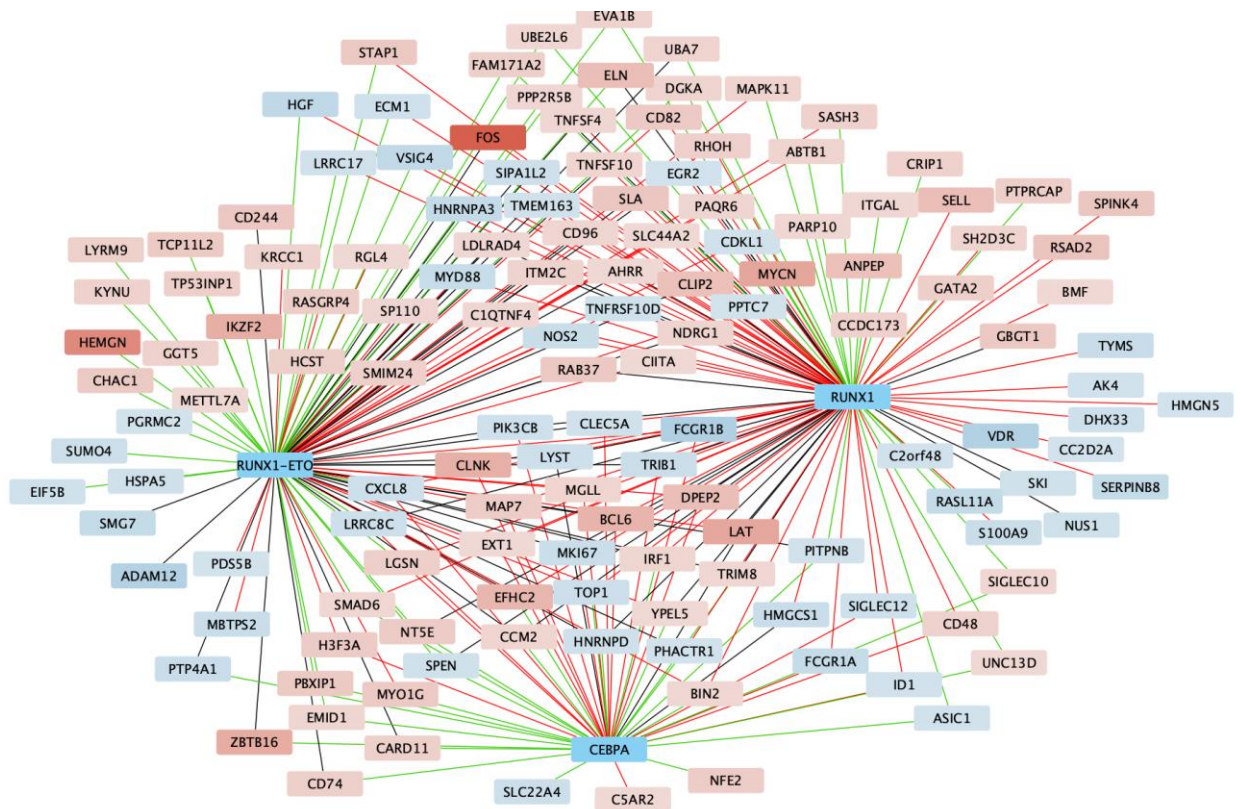
Figure 3-12: dnFOS responsive RUNX1-ETO, C/EBPα and RUNX1 bound specific gene network.

RUNX1-ETO, CEBPA and RUNX1 bound genes were determined by ChIP-seq in Kasumi-1 cells and then were filtered against genes upregulated by dnFOS expression. Gene expression indicated by the colour of the box ranked from high (red) to low (blue). Network constructed together with Dr Sophie Kellaway.

3.3.2 Induction of dnFOS leads to distinct epigenetic and transcriptional changes in primary t(8;21) and FLT3-ITD AML cells

In order to confirm whether blocking AP-1 binding was important in primary patient-derived cells as well, we transduced the doxycycline-inducible dnFOS or the Empty Vector control plasmid into these cells.

Starting with primary t(8;21) cells, we confirmed that doxycycline addition led to induction of dnFOS in dnFOS transduced cells but not in the Empty Vector control (Figure 3-13A). We next performed an analysis of accessible chromatin using ATAC-seq which showed that a few open chromatin sites enriched for AP-1 (and ATF) motifs were lost when dnFOS was induced (Figure 3-13B). Furthermore, an enrichment of RUNX and C/EBP motifs was seen in sites specific for control cells, similar to the changes seen in the Kasumi-1 cell line experiments (Section 3.3.1).

Next, we associated the dnFOS responsive sites with their respective genes and identified enriched KEGG pathways and observed a strong upregulation of many Zinc Finger genes (Figure 3-13C). Many Zinc Finger proteins are transcription factors but many also have roles in RNA binding or protein recruitment (Cassandri et al., 2017). Consequently, by upregulating many of these Zinc Finger factors, blocking AP-1 could have potent transcriptional, post-transcriptional and phenotypic changes. For example, blocking AP-1 with dnFOS led to an upregulation of RNA post-transcriptional modifiers such as *UPF2* that triggers nonsense mediated decay and may partially explain the downregulation of many genes. Similarly, within downregulated KEGG pathways, spliceosome, RNA transportation and ribosomal biogenesis genes were downregulated, which may further explain many of the transcription changes exerted by AP-1 in t(8;21) AML.

A dnFOS

B

	AP-1	92.81%	1e-942
	PU.1	34.12%	1e-96
	RUNX	27.34%	1e-36
	ATF	15.31%	1e-25
	C/EBP	10.59%	1e-37
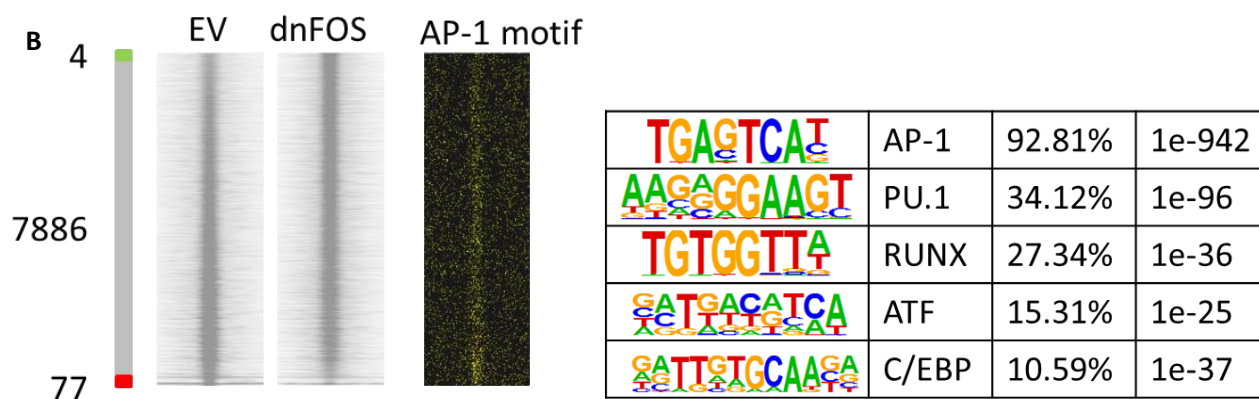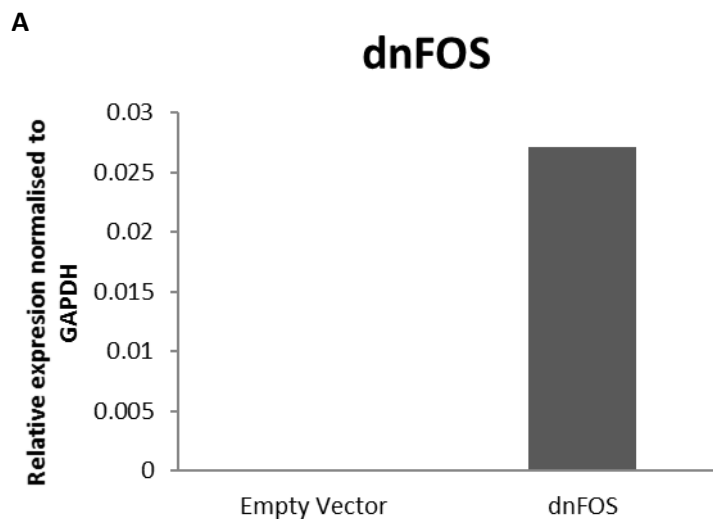
C Primary t(8;21) AML dnFOS upregulated KEGG pathways
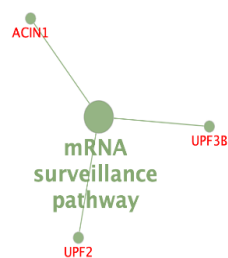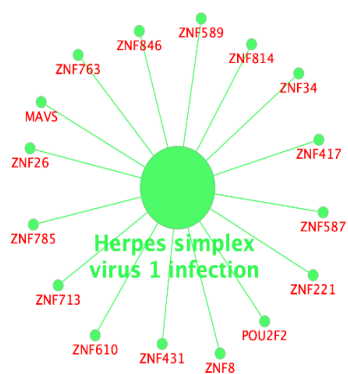
D Primary t(8;21) AML dnFOS downregulated KEGG pathways
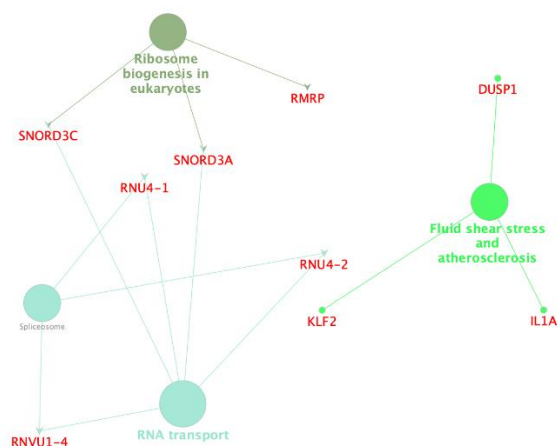
 (A) RT-qPCR of relative expression of dnFOS mRNA compared to *GAPDH* mRNA in primary t(8;21) AML. (B) Density plots of ATAC-seq peaks in a 2kb window, ranked top to bottom by relative tag count of peaks in doxycycline-induced *dnFOS* transduced cells relative to doxycycline-induced Empty Vector transduced cells in primary t(8;21) AML. Bars on the left indicate the number of peaks that are specific or shared between the cell types. The tables of enriched transcription factor binding motifs in the Empty Vector specific group is depicted alongside. (C-D) KEGG pathways derived from (C) upregulated or (D) downregulated enriched genes of primary t(8;21) AML transduced with doxycycline-induced dnFOS relative to an Empty Vector control.

We next repeated these experiments in primary FLT3-ITD AML. Transduction and induction of the dnFOS construct again increased dnFOS mRNA expression (Figure 3-14A) and led to a decreased in colony formation ability, relative to the Empty Vector control (Figure 3-14B). The analysis of chromatin accessibility by ATAC-seq showed that induction of dnFOS led to a loss of 651 open chromatin sites which were enriched for AP-1 motifs (Figure 3-14C). Induction of dnFOS also led to 942 new open chromatin sites that were enriched for GATA motifs suggesting that the cells were becoming more stem-cell like, similar to what was seen in the Kasumi-1 cell line experiments (Section 3.3.1). However, in contrast to t(8;21) AML where C/EBP enriched motifs were lost, in the case of FLT3-ITD AML dnFOS led to new open chromatin sites with C/EBP enriched motifs. This data therefore shows differences in the responses of the mutation-specific transcriptional networks to the abolition of AP-1 activity.

The KEGG pathway analysis of differentially expressed genes in FLT3-ITD AML cells after dnFOS induction showed an upregulation of the RAS signalling pathway (Figure 3-13D). Similar to what was seen in Kasumi-1 cells, this result suggests a negative feedback loop within the AP-1 signalling pathway whereupon blocking AP-1 binding de-represses the expression of upstream members of its signalling cascade (Section 3.3.1). Calcium signalling including *PLCB4* (phospholipase B4) was also found to be upregulated by dnFOS induction and represents a further mechanism by which negative feedback into the AP-1 signalling pathway occurs (Wu et al., 2019).

Further analysis of downregulated KEGG pathways showed that cell cycle genes were downregulated after induction of dnFOS (Figure 3-14E) including key components of the G1 cell cycle transition such as *CDC2* and *CCND1*. The importance of AP-1 in permitting the cell cycle transition was also seen in the Kasumi-1 cell experiments (Section 3.2.1).

C

	Control	dnFOS	AP-1	GATA	RUNX
942					
13912					
651					

dnFOS specific

Motif	Match	% of sites	p-value
AAAGAGGAAGTG	PU.1	48.6	1e-126
TTGCATAACC	C/EBP	31.7	1e-18
ATTGTGGTTT	RUNX	29.8	1e-63
AGATAAGA	GATA	14.3	1e-20

Empty Vector specific

Motif	Match	% of sites	p-value
AAGAGGAAGT	PU.1	37.7	1e-59
TGACTCA	AP-1	15.2	1e-39
CCACTAGGGGGC	CTCF	13.2	1e-57
TGGCAGTTAA	MYB	12.6	1e-14
TGTGGTTTCC	RUNX	10.0	1e-29

Primary FLT3-ITD AML dnFOS upregulated KEGG pathways

Primary FLT3-ITD AML dnFOS downregulated KEGG pathways

Figure 3-14: Induction of dnFOS in primary FLT3-ITD AML cells leads to epigenetic and transcriptional changes.

(A) RT-qPCR of relative expression of dnFOS mRNA compared to *GAPDH* mRNA in primary FLT3-ITD AML. (B) Relative colony formation ability of primary

FLT3-ITD AML cells transduced with doxycycline-induced Empty Vector or dnFOS vector. (C) Density plots of ATAC-seq peaks in a 2kb window, ranked top to bottom by relative tag count of peaks in doxycycline-induced dnFOS transduced cells relative to doxycycline-induced Empty Vector transduced cells in primary FLT3-ITD AML. Bars on the left indicate the number of peaks that are specific or shared between the cell types. The tables of enriched transcription factor binding motifs in the specific groups are depicted alongside. (D-E) KEGG pathways derived from (D) upregulated or (E) downregulated enriched genes from primary t(8;21) AML transduced with doxycycline-induced dnFOS relative to an Empty Vector control.

Lastly, in order to infer whether any potential AP-1 therapies would have toxicity in healthy cells, we transduced dnFOS into healthy CD34+ cells and confirmed induction of dnFOS by RT-qPCR (Figure 3-15A). Intriguingly, induction of dnFOS did not alter colony formation ability of the cells suggesting that AP-1 is an AML-specific target (Figure 3-15B). The analysis of ATAC-seq on these cells showed that a proportion of open chromatin sites changed upon induction of dnFOS with a loss of AP-1 and ATF motif enriched open chromatin sites along with a gain of open chromatin sites with no specific motif enrichment (Figure 3-15C). As outlined below, these small changes in open chromatin sites did not lead to many changes in the transcriptome and no enriched KEGG pathways were found.
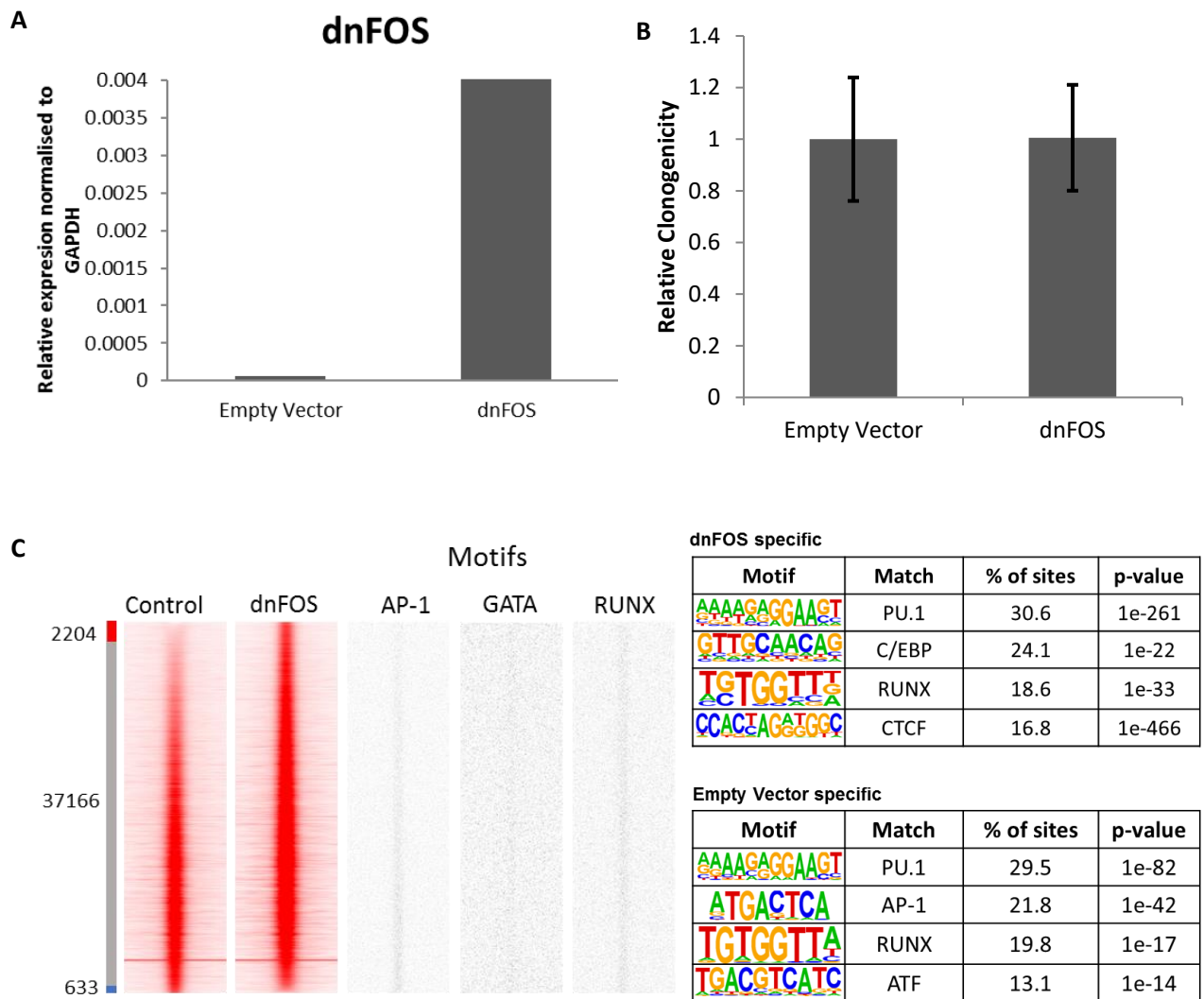
Figure 3-15: Induction of dnFOS in primary healthy CD34+ cells has little effect on cellular growth.

(A) RT-qPCR of relative expression of dnFOS mRNA compared to *GAPDH* mRNA in healthy CD34+ cells. (B) Relative colony formation ability of healthy CD34+ cells transduced with doxycycline-induced Empty Vector or dnFOS. (C) Density plots of ATAC-seq peaks in a 2kb window, ranked top to bottom by relative tag count of peaks in doxycycline-induced dnFOS transduced cells relative to doxycycline-induced Empty Vector transduced cells in healthy CD34+ cells. Bars on the left indicate the number of peaks that are specific or shared between the cell types. The tables of enriched transcription factor binding motifs in the specific groups are depicted alongside.

Next, we compared gene expression changes induced by dnFOS expression across the different primary cell types. Whist several hundred genes were differentially expressed in primary FLT3-ITD and t(8;21) AML upon dnFOS induction, only 9 genes were differentially expressed in healthy CD34$^+$ cells (Figure 3-16A). Furthermore, in a principle component analysis of gene expression data, healthy CD34$^+$ cells did not respond to dnFOS induction, in contrast to FLT3-ITD and t(8;21) AML cells which showed differences in the expression of a large number of genes (Figure 3-16B). Intriguingly, the variation upon dnFOS induction in the FLT3-ITD and t(8;21) AML cells was in distinct principle components again suggesting that the differentially expressed genes within the two AML subtypes were distinct. This result was confirmed by determining the overlap between differentially expressed genes in a Venn diagram, which showed minimal overlap (Figure 3-16C). This suggest that AP-1 has effects on the transcriptome in very distinct manners depending on the driver oncogene.

Upregulated Genes

Healthy   FLT3-ITD

4   0   210

0

0   8

362

t(8;21)

Downregulated Genes

Healthy   FLT3-ITD

5   0   297

0

0   0

69

t(8;21)
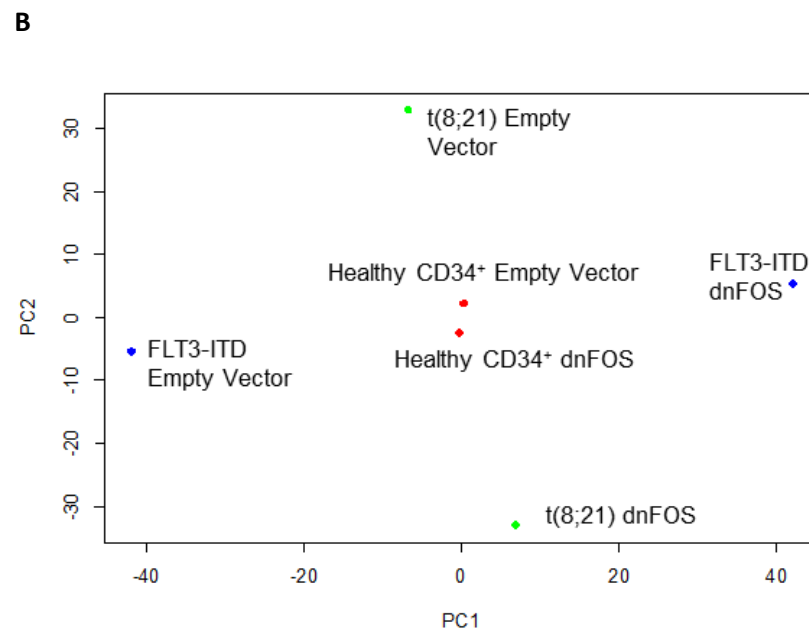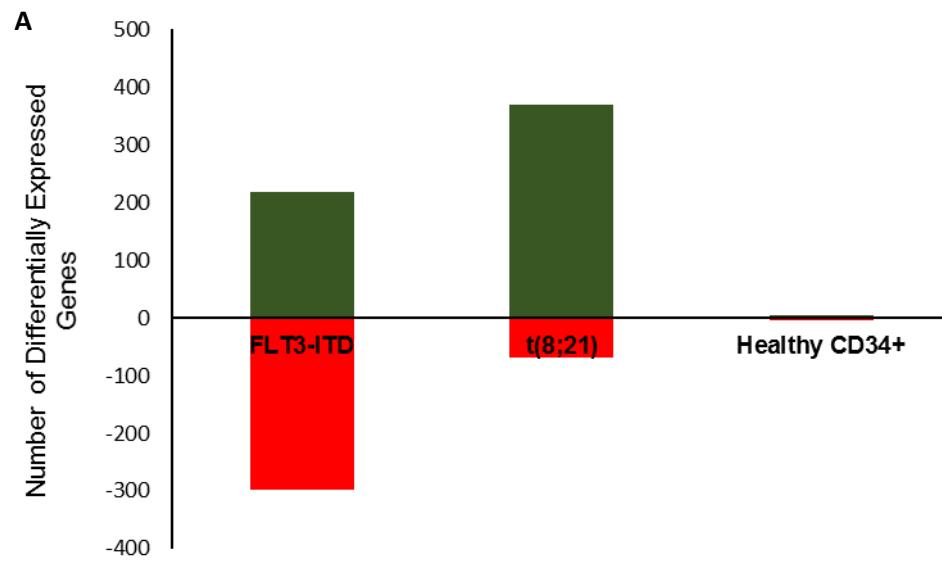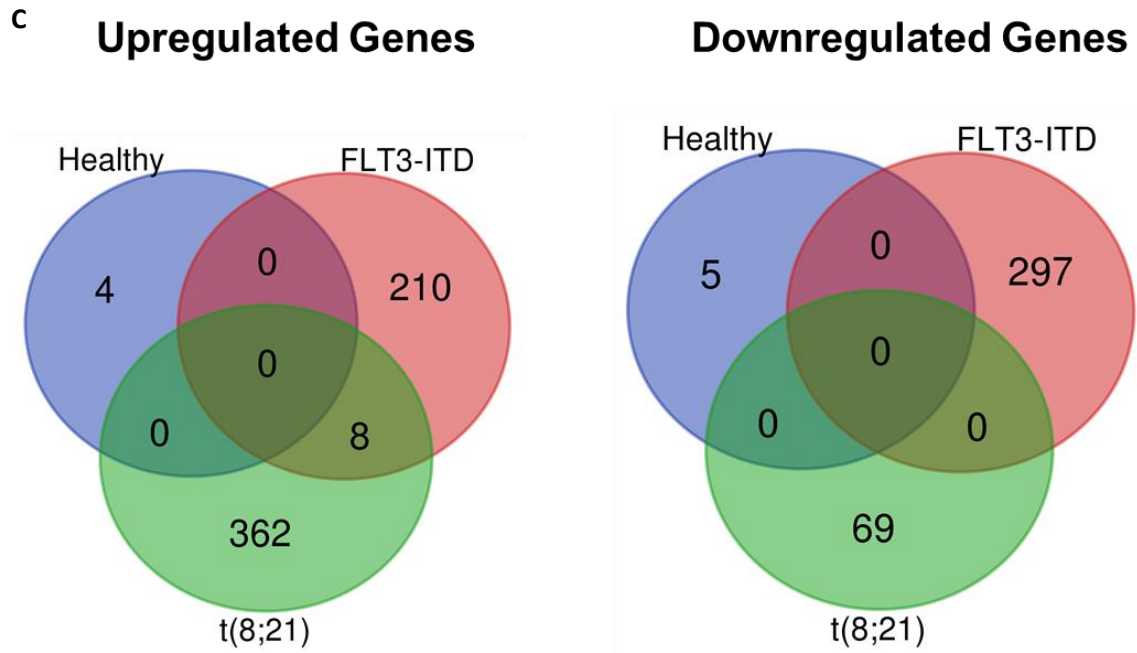
Figure 3-16: Different genes are dysregulated in different cell types upon dnFOS induction.

(A) Number of upregulated (above axis) and downregulated (below axis) genes in the RNA-seq data of the indicated primary cells transduced with dnFOS relative to Empty Vector control. (B) Principle component analysis of RNA-seq data from the indicated primary cells transduced with dnFOS or Empty Vector control. (C) Venn diagrams of upregulated and downregulated genes between each of the indicated primary cell types.

Chapter 4: A crucial role of WT1 in AML

4.1 Wilms Tumour 1 is Essential for Leukaemic Maintenance

4.1.1 Wilms Tumour 1 is upregulated in AML

Whilst *WT1* is known to be upregulated in AML in studies such at The Cancer Genome Atlas, for reasons discussed in Section 3.1.1, we assayed *WT1* gene expression in purified leukaemic blasts. Indeed, across different mutational and cytogenetic subtypes of AML, *WT1* was upregulated compared to healthy CD34[+] Peripheral Blood Stem Cells (PBSCs) (Figure 4-1).
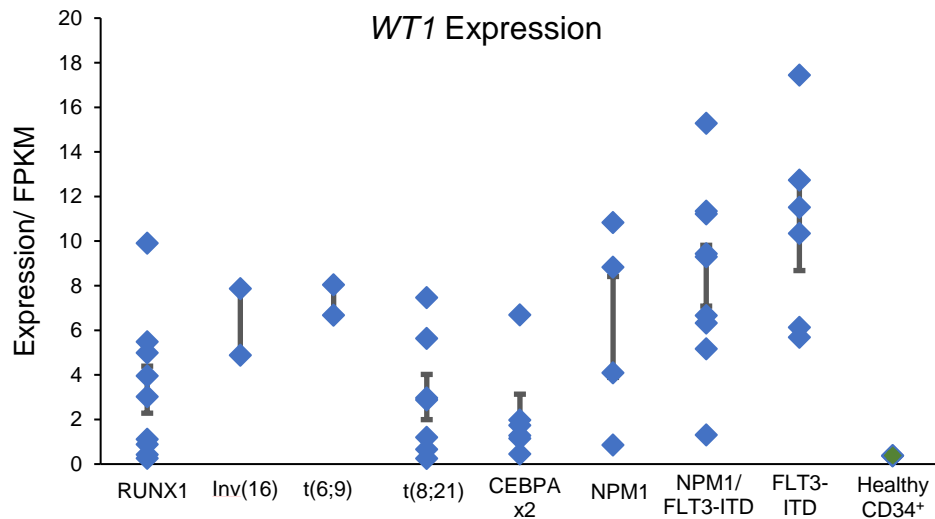
Figure 4-1: *WT1* is upregulated relative to healthy CD34[+] cells in all genetic subtypes of AML.

Expression of *WT1* mRNA in purified primary AML leukaemic blasts with different mutational subtypes. Error bars show standard error of the mean.

4.1.2 Wilms Tumour 1 is essential for leukaemic maintenance *in vivo* and *in vitro*

In order to establish the significance of the observed *WT1* upregulation, functional experiments are vital. As part of the shRNA depletion screen discussed earlier (see Section 3.1.2; (Martinez-Soria et al., 2018)) induction of shRNA against *WT1* in all three experimental conditions led to a significant depletion of *WT1* targeting shRNA barcodes, suggesting that WT1 is a key factor involved in leukaemic maintenance. Intriguingly, similar to what was seen for *JUN*, the most significant depletion was seen in the *in vivo* experiments suggesting that WT1 may have an important role in leukaemic blast survival within the bone marrow microenvironment.
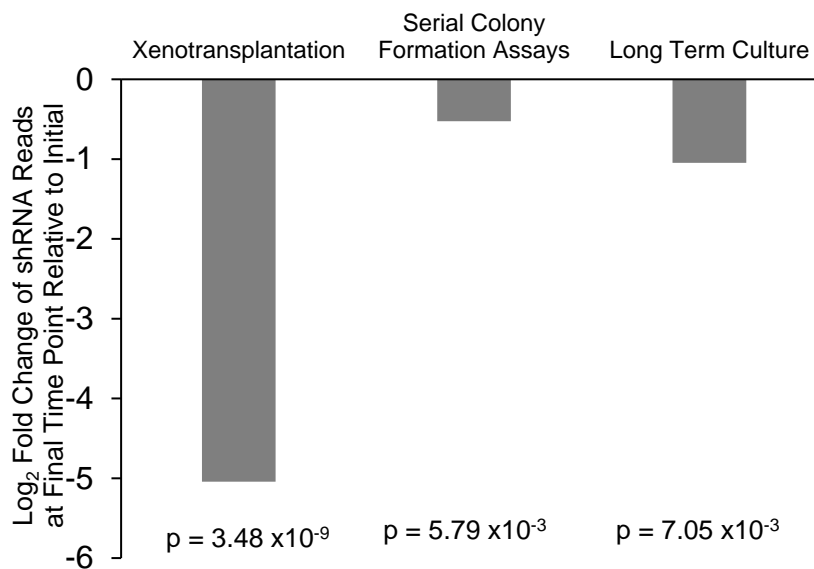
Figure 4-2: shRNA targeting *WT1* are depleted in a shRNA depletion screen.

Log_2 fold change of shRNA reads at the final time point relative to initial time point in xenotransplantation, serial colony formation assays or long term culture. Experiments performed by Dr Natalia Martinez-Soria, Newcastle University (Martinez-Soria et al., 2018).

4.1.3 Knockdown of *WT1* leads to decreased growth and colony formation in AML cells but not in healthy CD34+ cells

In order the confirm the important role of *WT1* in single gene perturbation experiments, two independent shRNA designs against *WT1* that were different to those in the shRNA depletion screens, were cloned into the pTRIPZ plasmid and transduced into Kasumi-1 cells. Induction of shRNA against *WT1* with the addition of doxycycline led to a reduction of WT1 protein expression (Figure 4-3B) and decreased cell growth, whilst induction of a non-targeting control had no effect upon growth (Figure 4-3A). Induction of shRNA against *WT1* also led to decreased colony formation ability (Figure 4-3C) and increased apoptosis (Figure 4-3D) but did not affect the cell cycle stage of leukaemic cells (Figure 4-3E).

A
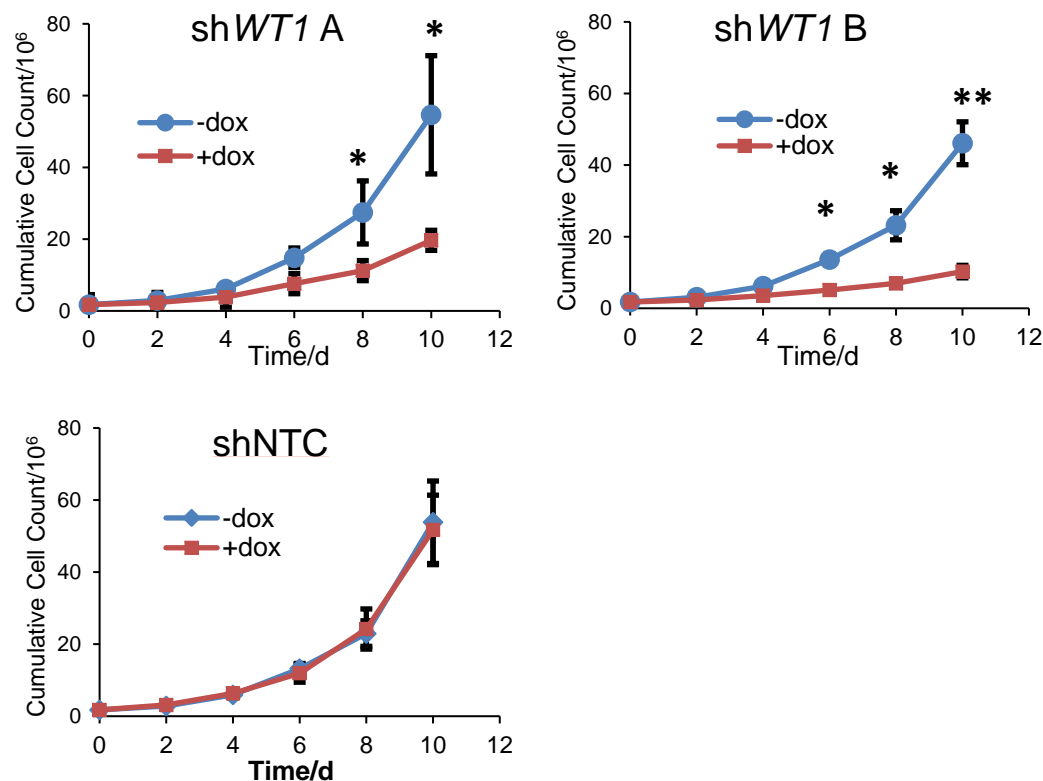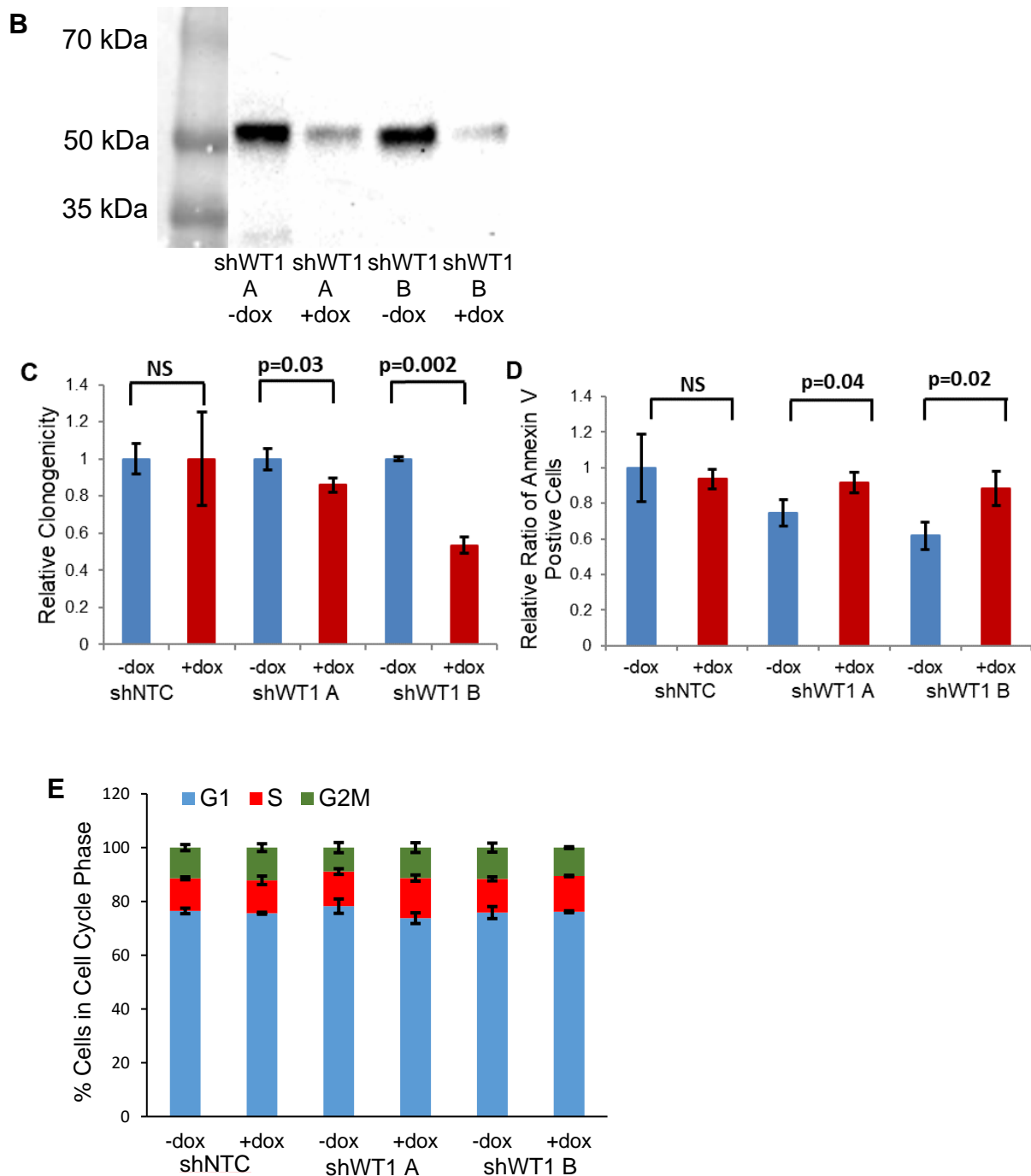
Figure 4-3: Knockdown of *WT1* in Kasumi-1 cells reduces growth and colony formation.

Kasumi-1 cells transduced with one of two different sh*WT1* or shNTC plasmids +/- doxycycline and phenotypic assessment by (A) measuring growth curves, (B) Western blot determining WT1 protein levels (C) colony formation ability of 2000 cell seeded into methylcellulose, (D) measuring annexin V positive cells and (E) measuring cell cycle phase distribution. n=3 and error bars show standard

deviation. Significance was determined using a 2 tailed student's t-test. * indicates p-value <0.05, ** indicates p-value <0.005.

Kasumi-1 cells are a well-studied model primary t(8;21) AML (Ptasinska et al., 2012). However, to show the relevance for primary AML, we performed similar experiments on patient derived t(8;21) AML and primary FLT3-ITD AML, together with healthy CD34$^+$ bone marrow cells. Unfortunately, primary t(8;21) cells do not form colonies *in vitro* on methylcellulose. However, primary FLT3-ITD AML did form colonies and fewer colonies were formed after cells were transduced with shRNA against *WT1* compared with a non-targeting control (Figure 4-4A). By contrast, the colony formation ability of CD34$^+$ cells derived from a healthy donor were unaffected by shRNA against *WT1* (Figure 4-4B).
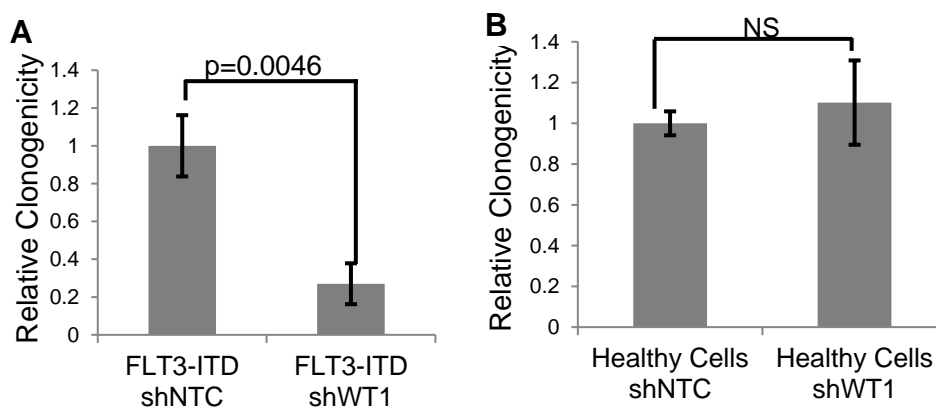
Figure 4-4: Colony formation of FLT3-ITD but not normal healthy CD34$^+$ cells is reduced by *WT1* knockdown.

Number of colonies formed per 2000 cells seeded in methylcellulose, normalised to shNTC (non-targeting control) using (A) primary FLT3-ITD AML or (B) primary CD34$^+$ bone marrow cells from a healthy donor transduced with doxycycline-inducible shNTC or sh*WT1* and induced with doxycycline. n=3 and error bars

show standard deviation. Significance was determined using a 2 tailed student's t-test.

4.1.4 Knockdown of *WT1* using shRNA led to distinct changes in the transcriptome

Examination of significantly upregulated and downregulated genes by RNA-seq and looking at KEGG pathways revealed different groups of genes regulated by WT1. Within genes upregulated by knockdown of *WT1*, we found an enrichment of the Glutathione Metabolism pathway (Figure 4-5A). Glutathione metabolism is a known therapeutic vulnerability within AML stem cells (Pei et al., 2013) and so further upregulation upon *WT1* knockdown may represent an opportunity for synthetic lethality. In addition, WT1 is thought to have a role in signalling and *WT1* mutations are found together with signalling gene mutations on a single cell level (Miles et al., 2020a). Consistent with that premise, within these knockdown experiments, the Rap1 signalling pathway is downregulated (Figure 4-5B).

sh*WT1* upregulated KEGG pathways

sh*WT1* downregulated KEGG pathways
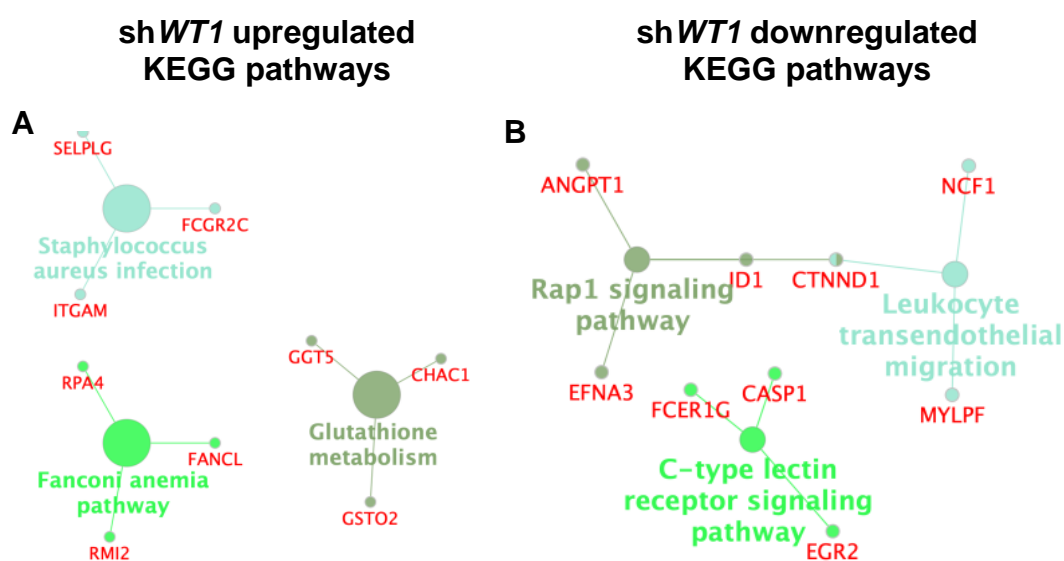
Figure 4-5: Distinct transcriptional changes are elicited by shRNA against *WT1.*

(A-B) KEGG pathways derived from RNA-seq data divided into (A) upregulated genes and (B) downregulated genes after Kasumi-1 cells were transduced with shWT1, compared with shNTC.

In order to confirm whether the findings in Kasumi-1 cells were representative of t(8;21) AML biology, the *WT1* knockdown experiments were repeated in patient-derived t(8;21) AML and over 50% knockdown was achieved (Figure 4-6A). Knockdown led to a loss of enrichment of RUNX motifs within open chromatin peaks as assessed by ATAC-seq (Figure 4-6B). In terms of gene expression, no enriched KEGG pathways could be found amongst upregulated genes. However, several signalling pathways including NFκB signalling were enriched amongst downregulated KEGG pathways.

B

Motifs in shWT1 specific peaks n = 389			
Motif	Match	p-value	% Targets
AAAGAGGAAGT	PU.1	1e-25	22.15
GCCACCTAGTGG	CTCF	1e-24	11.07
TGAATCATCCT	AP-1	1e-14	10.74

Motifs in shNTC specific peaks n = 1884			
Motif	Match	p-value	% Targets
TTACGCAA	C/EBP	1e-22	28.99
TGTGGTTTCG	RUNX	1e-39	16.10
AAAGAGGAAGTG	PU.1	1e-111	15.64
GATGACTCAT	AP-1	1e-74	13.61
CCACTAGGGGGC	CTCF	1e-119	13.22
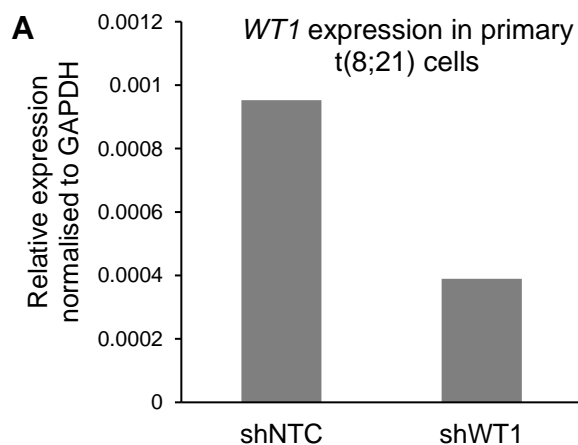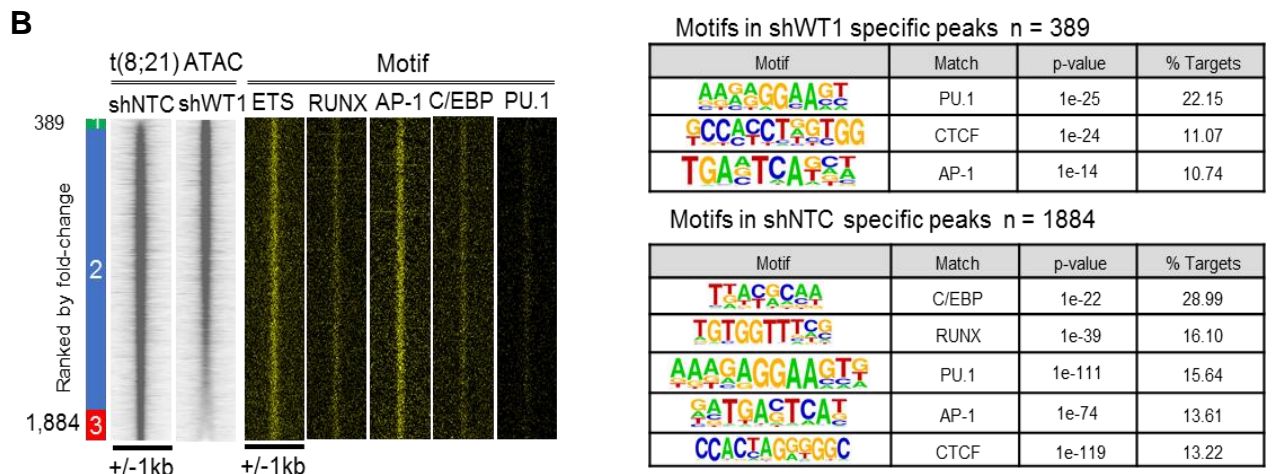
Figure 4-6: Distinct epigenetic changes are elicited by *WT1* knockdown in primary t(8;21) AML cells.

(A) Expression of *WT1* mRNA normalised to *GAPDH* mRNA in primary t(8;21) cells transduced with shNTC or sh*WT1* plasmids. (B) Density plots of ATAC-seq peaks in a 2kb window, ranked top to bottom by relative tag count of peaks in doxycycline-induced sh*WT1* transduced cells relative to doxycycline-induced shNTC transduced cells in t(8;21) primary AML cells. Tables of enriched motifs are shown alongside.

189
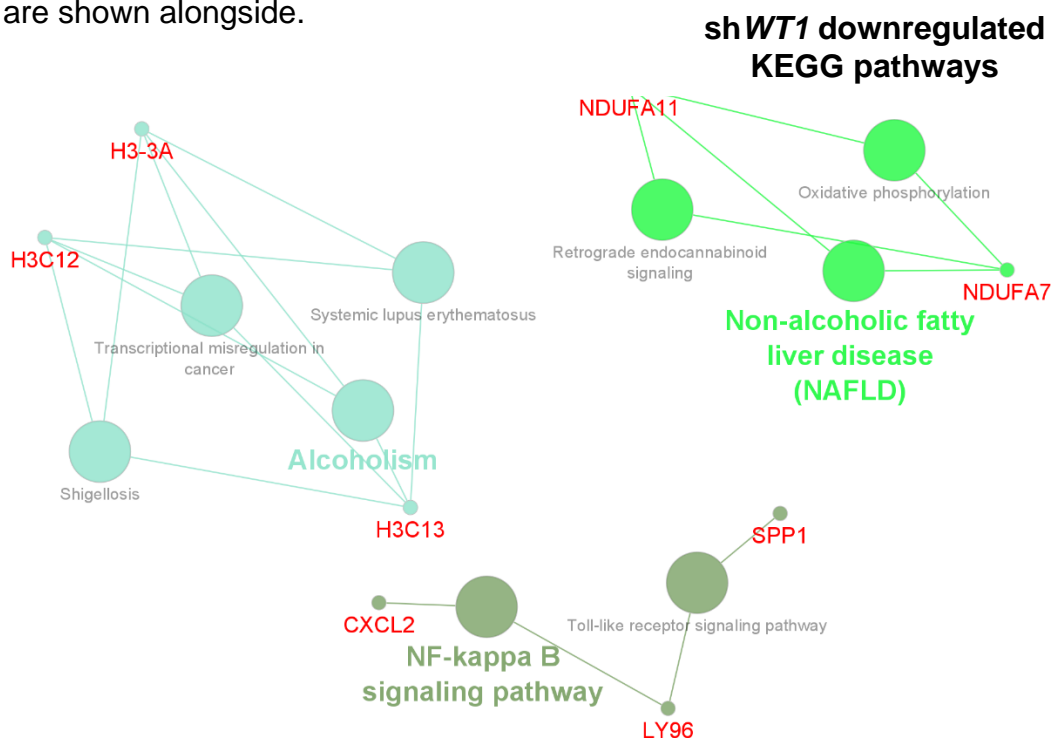
Figure 4-7: Distinct transcriptional changes are elicited by *WT1* knockdown in primary t(8;21) AML cells.

(A-B) Enriched KEGG pathways in genes (A) upregulated or (B) downregulated upon knockdown of *WT1* compared with an Empty Vector control, as determined by RNA-seq.

The transduction of *WT1* shRNA into primary FLT3-ITD AML led to over 80% knockdown of *WT1* mRNA (Figure 4-8A). However, we found few changes in chromatin accessibility with no significant differential enrichment of TF binding motifs (Figure 4-8B). However, *WT1* knockdown caused numerous changes in gene expression with 37 significantly upregulated genes and 286 downregulated genes upon knockdown of *WT1*. To the small number of upregulated genes, no enriched KEGG pathways could be found. Of note, a decrease in the glycerophospholipid and ether lipid metabolism was seen upon knockdown of *WT1*. These lipids are crucial for AML cell growth and survival (Ricciardi et al., 2015) and so WT1 may have a role in upregulating cellular metabolic pathways.
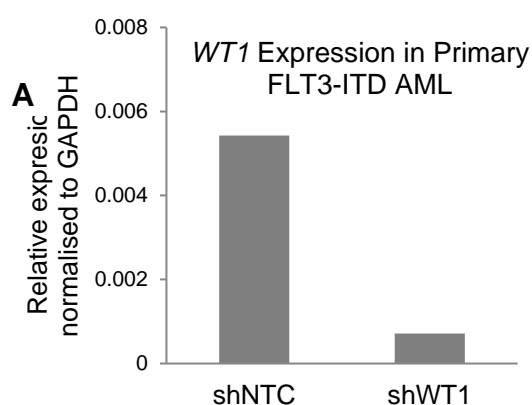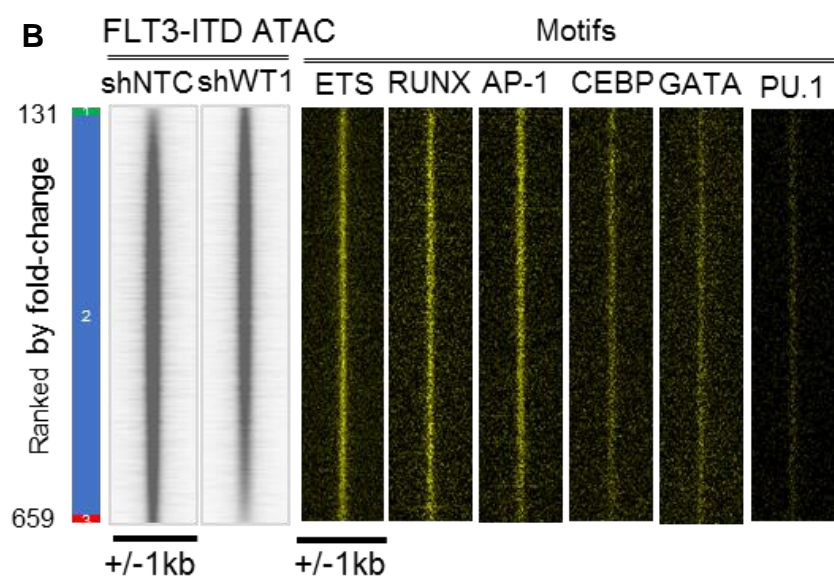
Figure 4-8: Distinct epigenetic changes are elicited by *WT1* knockdown in primary FLT3-ITD AML cells.

(A) Expression of *WT1* mRNA normalised to *GAPDH* mRNA in primary FLT3-ITD cells transduced with shNTC or sh*WT1* plasmids. (B) Density plots of ATAC-seq peaks in a 2kb window, ranked top to bottom by relative tag count of peaks in doxycycline-induced sh*WT1* transduced cells relative to doxycycline-induced shNTC transduced cells in primary FLT3-ITD AML cells. A table of enriched motifs is shown alongside.

sh*WT1* downregulated KEGG pathways
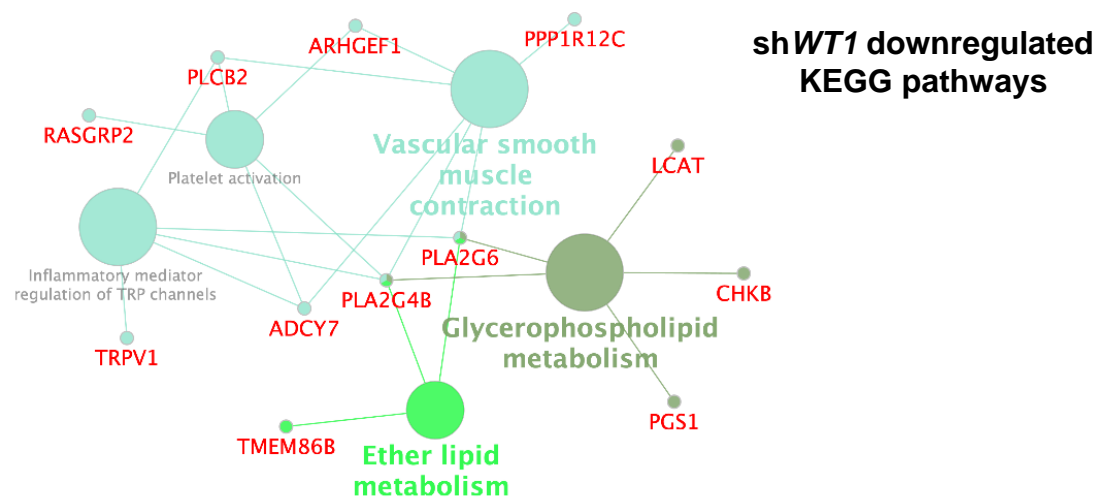
Figure 4-9: Distinct transcriptional changes are elicited by *WT1* knockdown in primary FLT3-ITD AML cells.

Enriched KEGG pathways derived from genes downregulated upon knockdown of *WT1* compared with an Empty Vector control, as determined by RNA-seq.

When a similar WT1 shRNA experiment was repeated on healthy patient-derived CD34+ cells, despite achieving over 50% *WT1* knockdown (Figure 4-10A) and we observed very few changes in chromatin accessibility (Figure 4-10B) and gene expression.
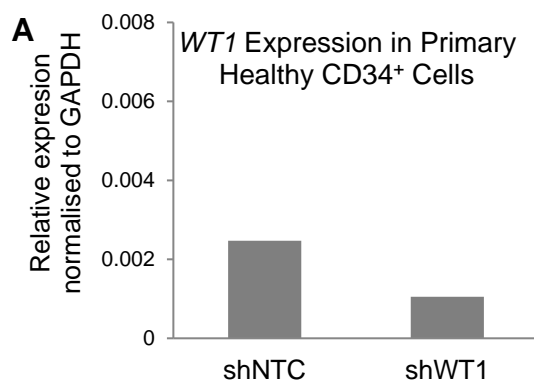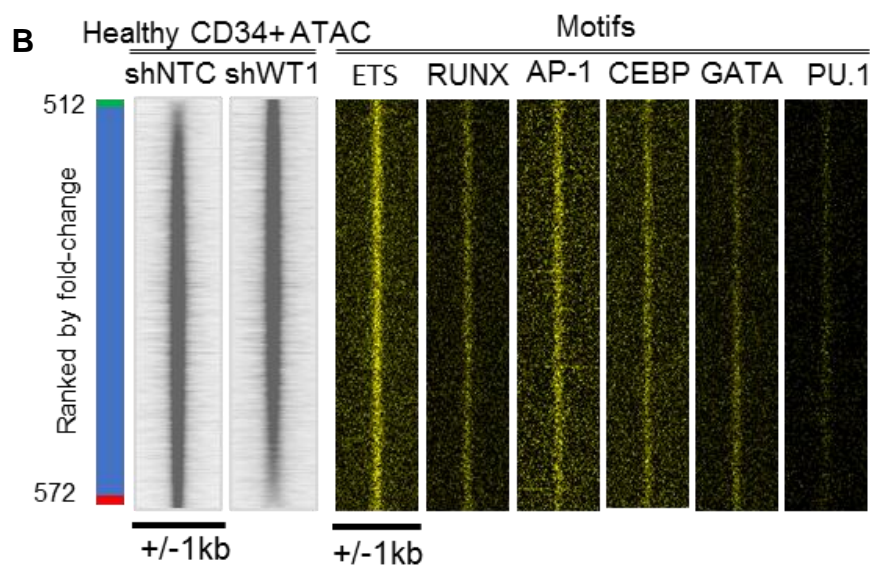
Figure 4-10: Minimal transcriptional or epigenetic changes are elicited by *WT1* knockdown in primary healthy CD34+ cells.

(A) Expression of *WT1* mRNA normalised to *GAPDH* mRNA in primary healthy CD34+ cells transduced with shNTC or sh*WT1* plasmids. (B) Density plots of ATAC-seq peaks in a 2kb window, ranked top to bottom by relative tag count of peaks in doxycycline-induced sh*WT1* transduced cells relative to doxycycline-induced shNTC transduced cells in primary healthy CD34+ AML cells.

4.2 The regulation of *WT1* gene expression

4.2.1 Four distal *WT1* cis-regulatory elements are present in all leukaemic blasts examined

We have previously systematically characterised the chromatin accessibility landscape using DNase I-seq on numerous patient samples (Assi et al., 2019) and in this study, we characterised further samples. Focusing on t(8;21) AML (Figure 4-11) but equally applicable to other subtypes of AML (data not shown), four distal cis-regulatory elements were found that were present in all patients examined. Superimposition of promoter capture HiC data that was previously

published by the Bonifer lab (Assi et al., 2019) along with the primary cell ATAC-seq experiment described here in section 4.1.3 showed that each of these cis-regulatory elements interacts with the *WT1* promoter.
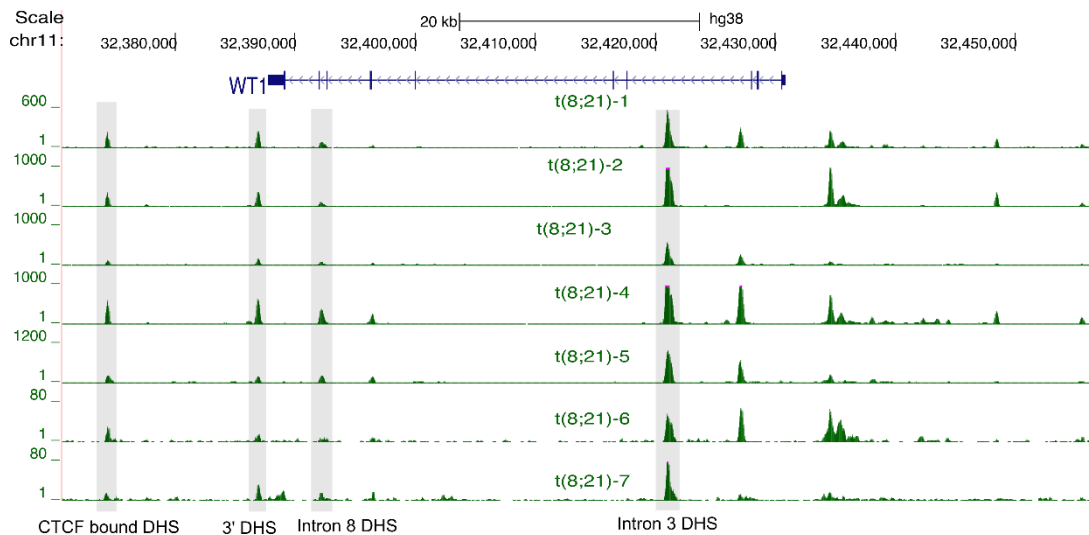
Figure 4-11: Four distal open chromatin sites around the *WT1* locus are present in all primary t(8;21) AML samples examined.

UCSC genome browser screenshot of DNase I Hypersensitive sites in 7 primary t(8;21) AML, centred at the *WT1* locus. Some of the experiments were performed by Dr Rosie Imperato and Dr Daniel Coleman.
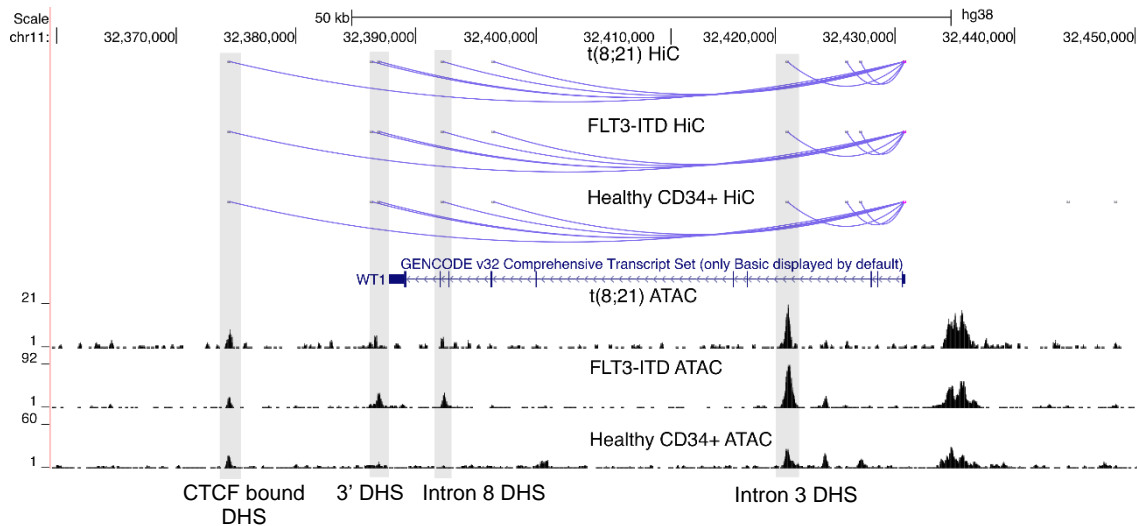
Figure 4-12: Four distal open chromatin sites around the *WT1* locus interact with the promoter in primary cells.

UCSC genome browser screenshot showing promoter Capture HiC interactions and ATAC-seq peaks at the *WT1* locus in primary t(8;21), primary FLT3-ITD and healthy CD34[+] bone marrow cells. Capture HiC experiments were performed by Dr Anna Pickin (Assi et al., 2019).

Further examination of these open chromatin sites with ChIP-seq data previously published by the Bonifer lab and others (Ptasinska et al., 2014, Ptasinska et al., 2019, Ben-Ami et al., 2013, Martens et al., 2012) revealed that the 3' DHS displayed almost no transcription factor binding or histone modification and that the CTCF bound DHS only had CTCF binding seen (Figure 4-13). The intron 3 DHS bound numerous transcription factors including RUNX1-ETO, JunD, LDB1, CEBPA, LMO2, RUNX1, PU.1 and AP4 and displayed both activation (p300, H3K27Ac) and repression (NCoR) histone modifications. The intron 8 DHS also displayed binding of numerous transcription factor binding including WT1, RUNX1-ETO, LMO2, RUNX1, PU.1 and AP4 but intriguingly did not show any of the histone modifications examined.
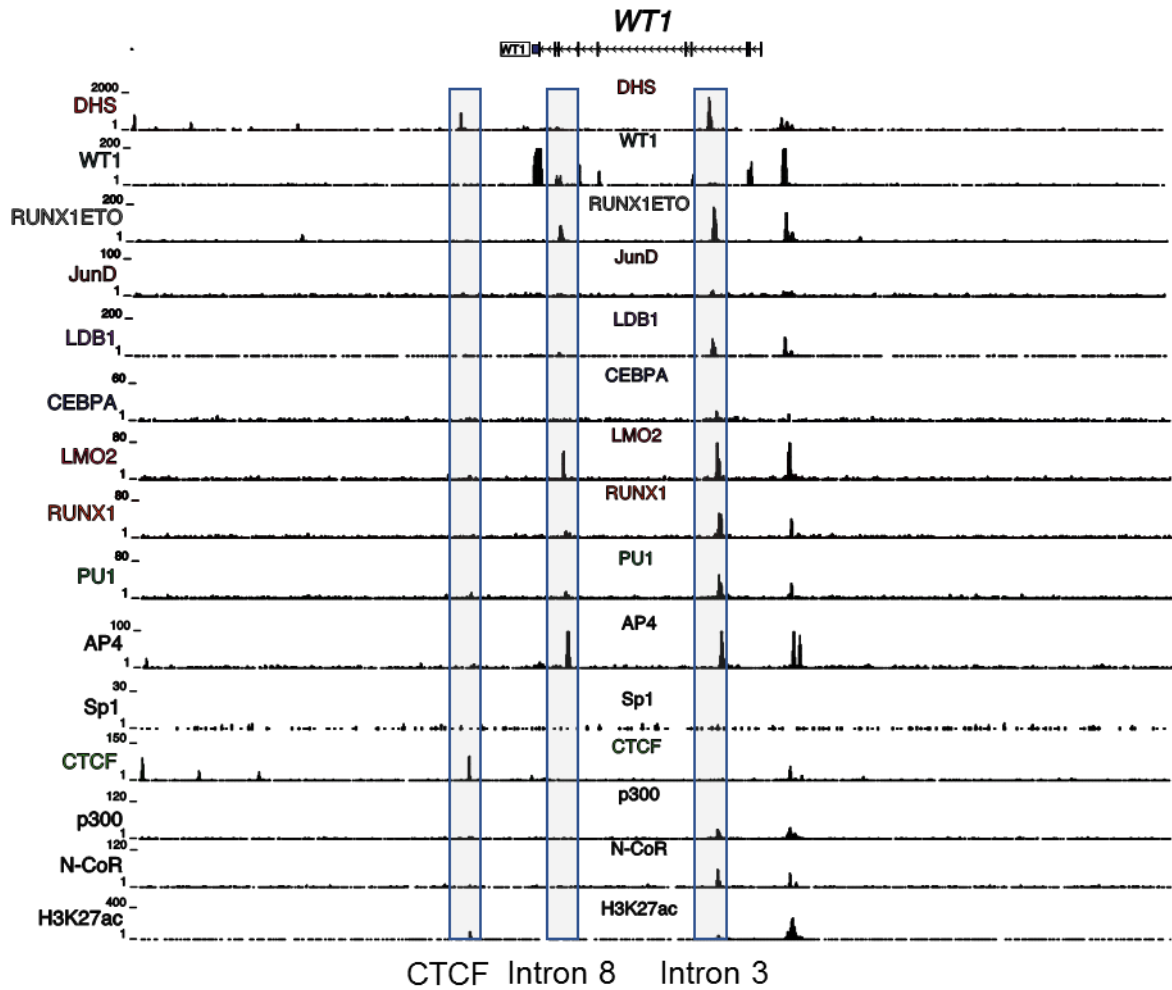
Figure 4-13: ChIP-seq of indicated transcription factor or histone modifications at the *WT1* locus in Kasumi-1 cells.

UCSC genome browser screenshot of DNase I Hypersensitive sites and ChIP-seq peaks of the indicated factors in Kasumi-1 cells, centred at the *WT1* locus (some ChIP-seq data derived from (Ptasinska et al., 2014, Ptasinska et al., 2019, Ben-Ami et al., 2013, Martens et al., 2012).

4.2.2 Activation of the intron 8 DHS by CRISPR-p300 reveals an enhancer element

We went on to examine the function of the distal cis-regulatory elements by using CRISPR activation (CRISPRa) which enables p300 to be targeted to a site of

interest and hence forcing histone H3K27 acetylation at the site. The 3' DHS and CTCF bound DHS CRISPRa were used as controls to assess the effect of forced acetylation at an open chromatin site in general and an empty vector control to assess for any non-targeted effects of the dCas9-p300 protein. CRISPRa at the intron 8 DHS led to a large upregulation of *WT1* gene expression compared with CRISPRa at other sites (Figure 4-14A) and also an increase in WT1 protein expression (Figure 4-14C). Performing ChIP-qPCR for H3K27Ac and normalising to a known positive region, the *SPI1* -14kb enhancer, CRISPR activation at the intron 8 DHS led to a significant in H3K27ac compared with the control sites (Figure 4-14B). This confirmed that changes in gene expression are likely to be due to the recruitment of p300 in the CRISPRa experiment.
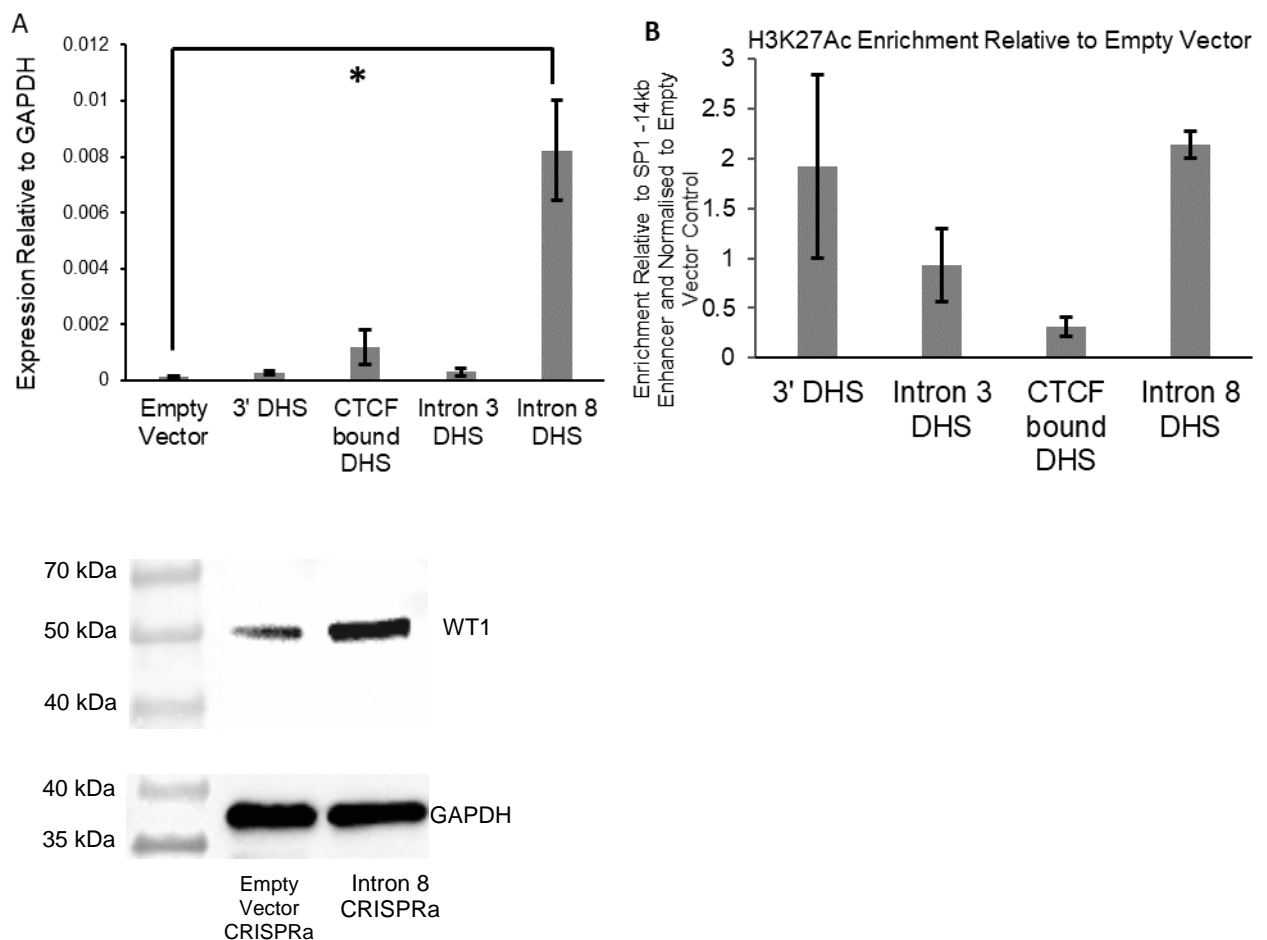
(A) Relative expression levels of *WT1* mRNA normalised to *GAPDH* mRNA in Kasumi-1 cells transduced with vectors expressing CRISPRa targeted to the indicated DHSs. (B) H3K27ac ChIP-qPCR enrichment analysis of Kasumi-1 cells transduced with a CRISPRa vector targeted to intron 8, enrichment levels were normalised to the empty vector control and to the PU.1 -14 kb Enhancer (positive control). (C) Western blot showing WT1 and GAPDH protein expression in Kasumi-1 cells transduced with an empty vector or with sgRNA targeted to intron 8 DHS.

4.2.3 CRISPRa at the intron 8 DHS leads to increased cell growth and colony formation and decreased apoptosis

Next, we examined the functional consequences of CRISPRa at four cis-regulatory elements. CRISPRa at the intron 8 DHS led to increased cell growth (Figure 4-15A) and a large increase in colony formation ability (Figure 4-15B) but CRISPRa at other sites had no effect. CRISPRa did not affect the cell cycle phase of cells but did lead to a decrease in apoptosis when targeted to the intron 8 DHS.
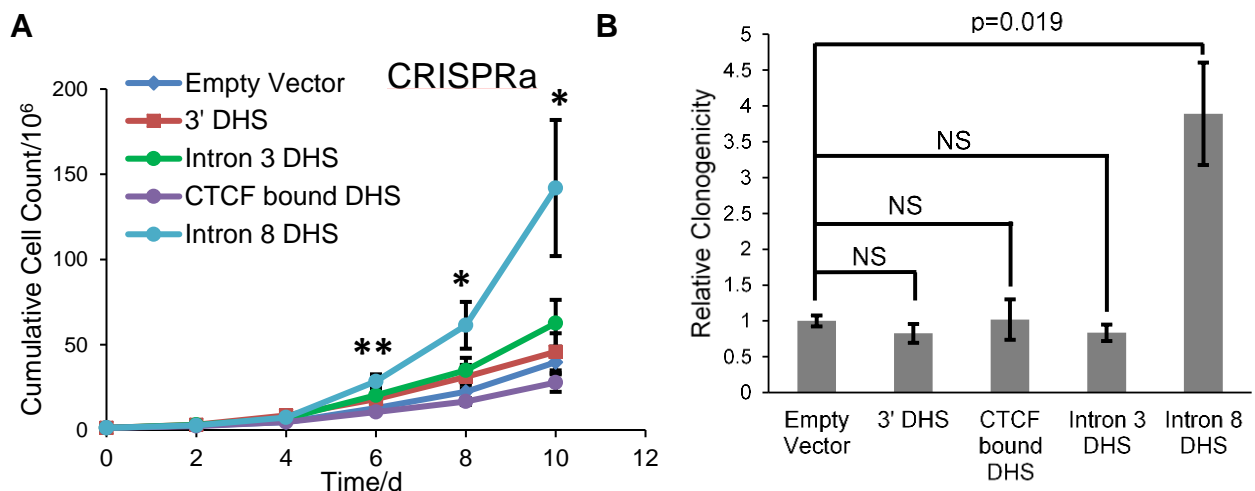
C
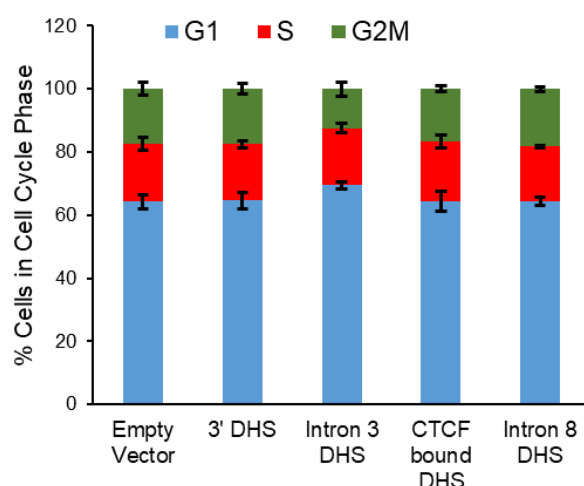
D
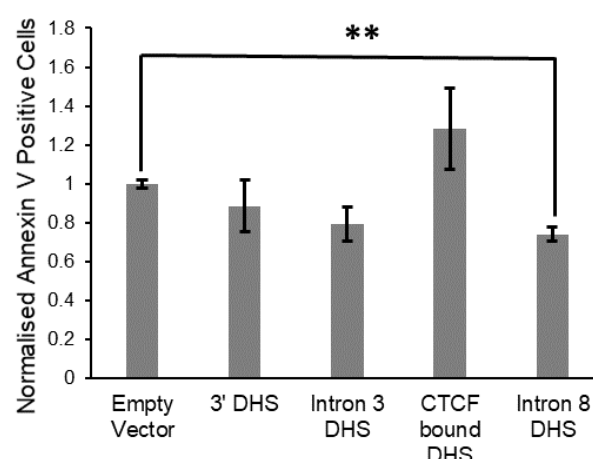
Figure 4-15: CRISPR activation at the intron 8 DHS increases growth and colony formation and decreases apoptosis.

(A) Time course of cumulative cell growth of Kasumi-1 cells expressing a CRISPRa protein targeted to the indicated *WT1* DHSs. (B) Colony formation ability in methylcellulose of 2000 Kasumi-1 cells expressing a CRISPRa targeted against the indicated *WT1* DHSs and normalised to the empty vector control. (C-D) Kasumi-1 cells transduced with CRISPRa targeted to the indicated *WT1* DHSs examined for (C) Cell cycle phase analysis and (D) analysis of annexin V positive cells, normalised to data obtained with the empty vector control. For all panels, n=3 and error bars show standard deviation. Significance was determined using a 2 tailed student's t-test. * indicates p-value <0.05, ** indicates p-value <0.005.

4.2.4 Several families of transcription factors can bind to the intron 8 DHS

Further examination of the intron 8 DHS revealed numerous motifs for transcription factor binding including five RUNX sites, 3 ETS sites and an AP-1 site (Figure 4-16). We have previously described that RUNX1 and RUNX1-ETO bind to the intron 8 DHS (section 4.2.1). A time-course of knockdown of *RUNX1-ETO* using siRNA against led to decreased *WT1* expression at the mRNA (Figure 4-17A) and protein level (Figure 4-17B).

Knockdown of RUNX1 however did not lead to changes in WT1 expression (Ben-Ami et al., 2013). The effect of blocking AP-1 binding upon WT1 will be discussed in greater detail in section 4.8.

Figure 4-16: ETS, RUNX and AP-1 motifs are present in the *WT1* intron 8 DHS.

DNA sequence at the intron 8 open chromatin site and with highlighted transcription factor binding motifs.
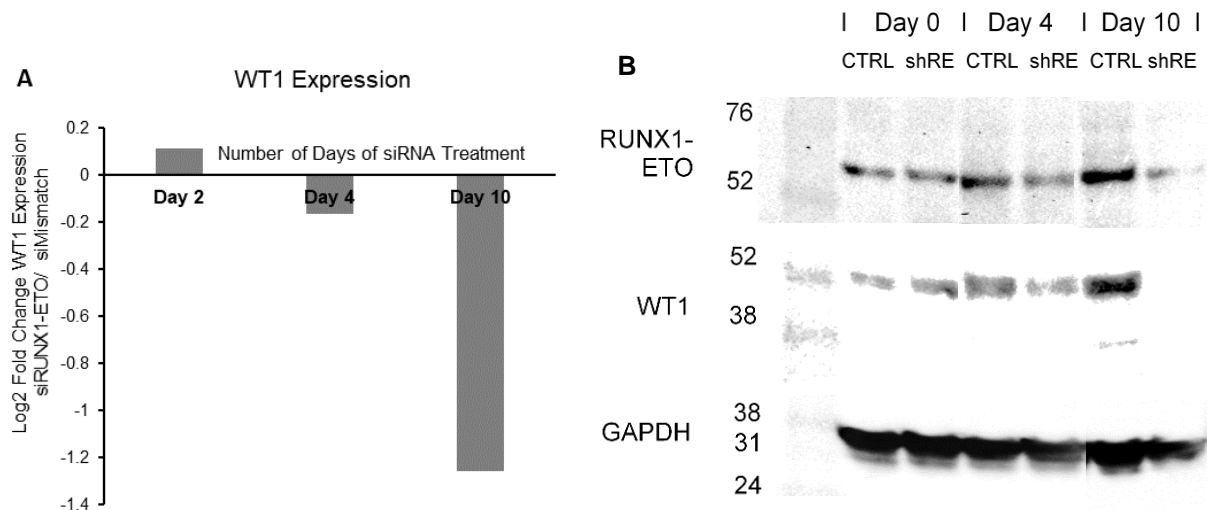
Figure 4-17: <u>*WT1* gene and protein expression are decreased with *RUNX1-ETO*</u>
<u>knockdown.</u>

Time course of Kasumi-1 cells transfected with siRNA against *RUNX1-ETO*
followed by (A) determination of Log2 Fold Change of mRNA levels before and
after RUNX1-ETO knock-down (FPKM of shRNA/control data from RNA-seq) and
B) protein level analysis by Western blot measuring RUNX1-ETO, WT1 and
GAPDH.

4.3 Expression of *WT1* -KTS and +KTS isoforms causes isoform-specific phenotypic changes in AML

WT1 produces as at least 12 isoforms, 8 of which are produced in haematopoietic
cells; Transcription start site (TSS) 0 is only employed in germline tissue e.g.
testes but TSS 1 and 2 are employed in haematopoietic cells producing long and
short mRNA isoforms respectively. In addition, alternative splicing occurs in the
51bp sequence in exon 5 coding for a 17 amino acid '17AA' sequence and in a
9bp sequence at the end of exon 9 coding for a 3 amino acid Lysine-Threonine-
Serine 'KTS' sequence. Figure 4-18 is a schematic of WT1 structure and function,
primarily based upon site-directed mutagenesis work (Wang et al., 1993).

Transcription Start Site	0	1	2		
Base#	-219	1	378	750-800	1224-1232
	CTG	ATG	ATG	17AA	KTS

| Amino Acid # | -73 | 1 | 127 | 250-266 | 408-410 |
| Protein Weight/KDa | 60-62 | 52-54 | 36-38 | | |

Self-Association Domain (1-180)

Zinc Fingers (323-449)
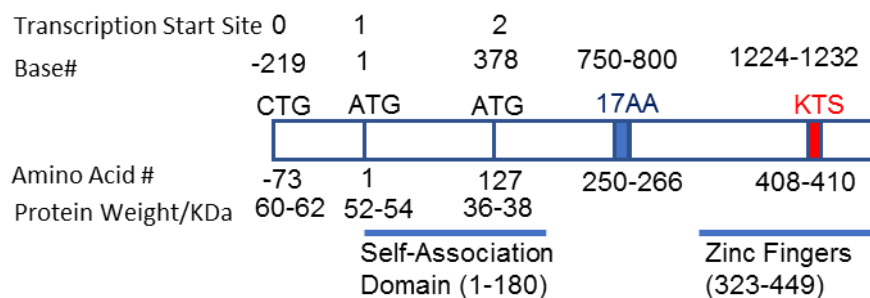
<u>Figure 4-18: *WT1* has three transcriptional start sites and two alternative splice sites.</u>

Schematic of WT1 structure and sites of alternative start sites and alternative splice sites.

Next we examined the ratio between *WT1* isoforms in AML by looking for the presence or absence of splice sequences within deeply sequenced RNA-seq reads from purified AML blasts. This analysis revealed that *WT1* isoforms containing the 9bp sequence coding for KTS sequence (*WT1* +KTS) were upregulated more than those lacking the 9bp sequence (*WT1* –KTS) in most genetic subtypes of AML whilst an equal ratio was seen in healthy peripheral blood stem cells (Figure 4-19). We also examined RNA-seq data from the Cancer Genome Atlas (TCGA, 2013) but did not find any differences in KTS isoform expression (data not shown); this once again shows the importance of sequencing purified blasts as we have done, rather than whole bone marrow with contaminating normal cells as the TCGA consortium has.
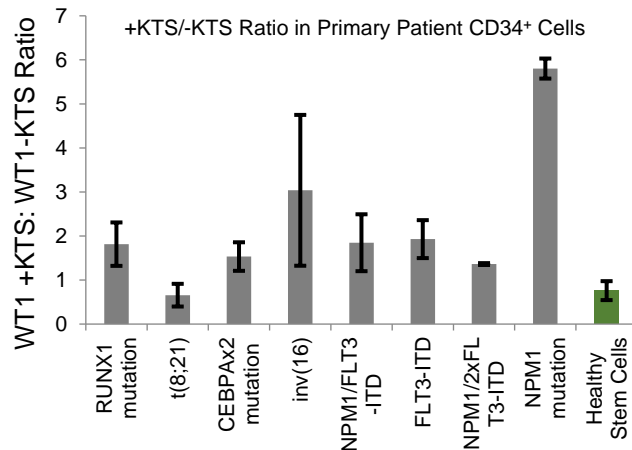
Figure 4-19: An increased *WT1* +KTS:*WT1* –KTS ratio is present in mutational subtypes of AML relative to healthy CD34+ cells.

Ratio between expression of mRNAs encoding the *WT1* +KTS and *WT1* –KTS isoforms in purified primary patient CD34+ cells.

4.3.1 Opposite effects of *WT1* –KTS isoforms and WT1 +KTS isoforms on leukaemic growth

Next we wanted to examine the functional roles of individual *WT1* isoforms in AML. We initially wanted to express these isoforms on a null background by knocking out *WT1* by CRISPR-Cas9 gene editing on Kasumi-1 cells. This knockout experiment was successful in generating nearly 100 single cell clones with a single allele *WT1* knockout, as assessed by PCR on genomic DNA and RT-qPCR on mRNA (Figure 4-20). However, no clones carried homozygous deletions of *WT1* suggesting that some *WT1* expression was necessary for leukaemic cell survival.

A Wildtype Allele PCR   Deletion Allele PCR

A B C D E F G H I J K L M N   A B C D E F G H I J K L M N

sgRNA cut   sgRNA cut

WT1 Deletion Allele
Forward Primer   WT1 Wildtype
                        Allele Primers

                                        WT1 Deletion Allele
                                        Reverse Primer

B WT1
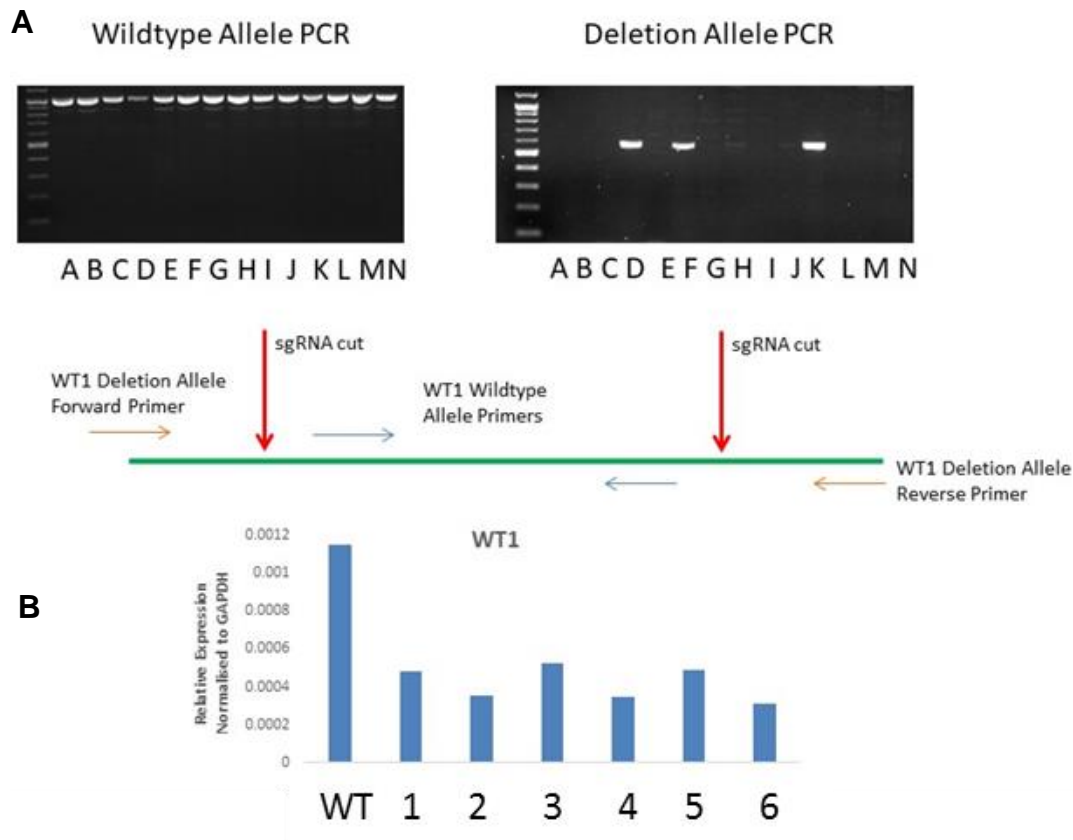
Figure 4-20: CRISPR-Cas9 gene editing of _WT1_ led to heterozygous but no homozygous clones.

(A) PCR and gel electrophoresis from DNA of 14 example single cell Kasumi-1 clones (A-N) with amplification of the wildtype allele or of the deletion allele, together with a schematic indicating the primer design. (B) _WT1_ mRNA expression relative to _GAPDH_ in wildtype or 6 example CRISPR-Cas9 gene edited single cell clones (1-6).

To obtain further insight into the role of WT1, we overexpressed individual _WT1_ isoforms to physiological level (up to 5 fold increased expression (Kramarzova et al., 2012) on a normal _WT1_ background. Figure 4-21A shows a schematic of the doxycycline-inducible experimental set-up used in experiments and a Western

blot of WT1 protein expression shows expression of individual isoforms over the background (empty vector control) (Figure 4-21B).

Expression of any of the four *WT1* –KTS isoforms led to decreased cell growth whereas expression of any of the four *WT1* +KTS isoforms led to increased growth (Figure 4-21C). Similarly expression of *WT1* –KTS isoforms led to decreased colony formation ability whilst expression of *WT1* +KTS isoforms did not affect colony formation (Figure 4-21D). Furthermore, expression of *WT1* – KTS isoforms increased apoptosis whilst expression of *WT1* +KTS isoforms had no effect (Figure 4-21E). Lastly, expression of *WT1* -KTS isoforms led to a G1 cell cycle arrest whilst expression of *WT1* +KTS isoforms did not affect cell cycle stage (Figure 4-21F). In all cases we did not see any effects that could be attributed to either the TSS employed nor to whether or not alternative splicing occurred at the 17AA splice sequence.

TSS	EV	2	2	2	2	1	1	1	1
17AA	-	-	+	+	+	-	+	-
KTS	-	+	-	+	+	-	-	+

C

TSS	EV	2	2	1	1	2	2	1	1
17AA		-	+	-	+	-	+	+	-
KTS		-	-	-	-	+	+	+	+

E

TSS	EV	2	2	1	1	2	2	1	1
17AA		-	+	-	+	-	+	+	-
KTS		-	-	-	-	+	+	+	+

F

TSS	EV	2	2	1	1	2	2	1	1
17AA		-	+	-	+	-	+	+	-
KTS		-	-	-	-	+	+	+	+
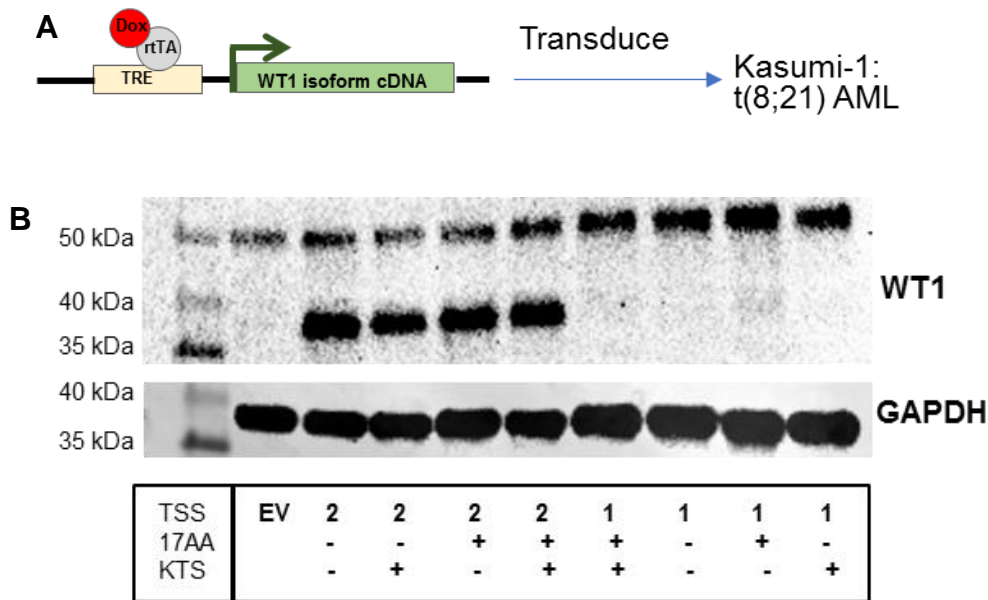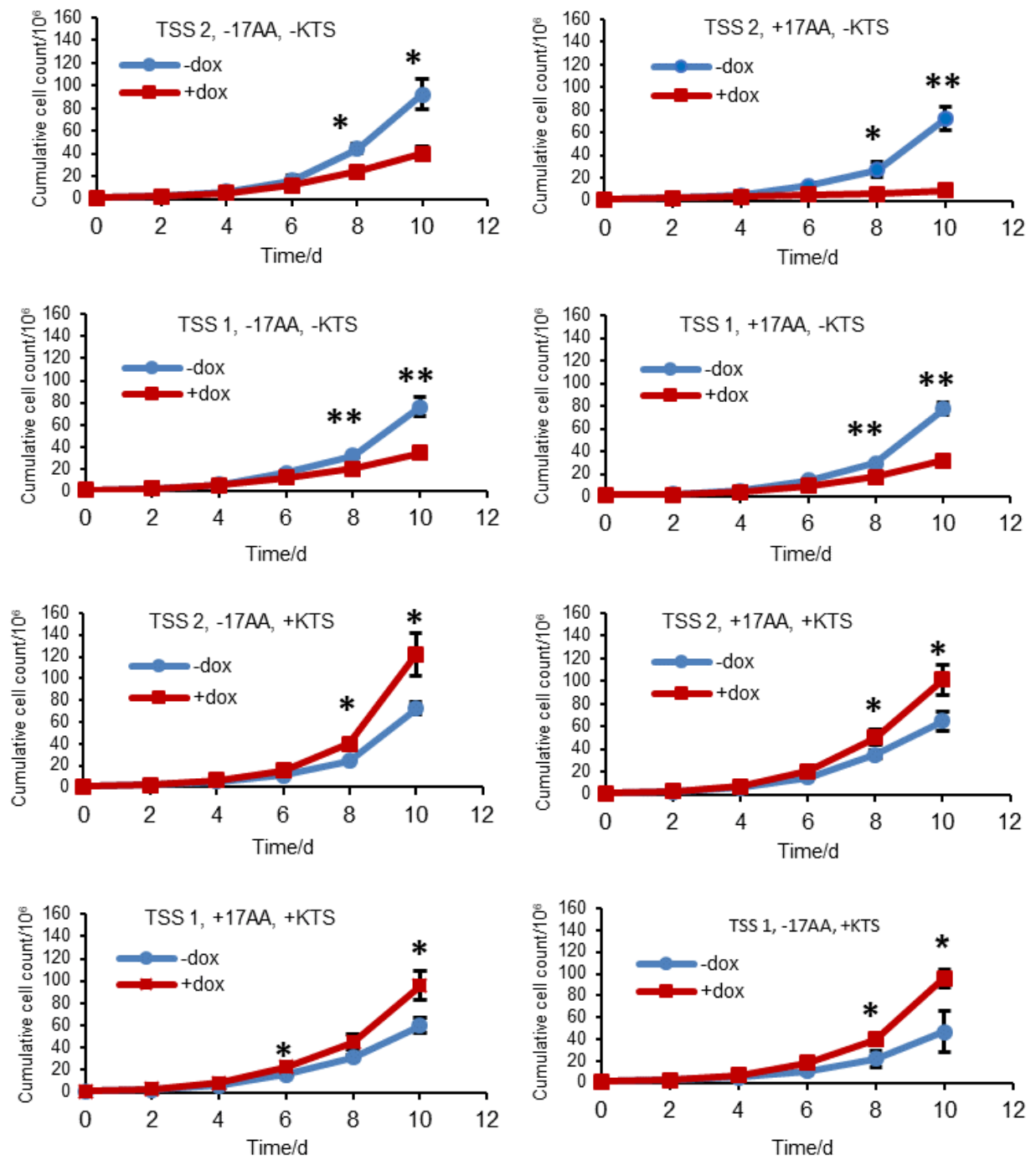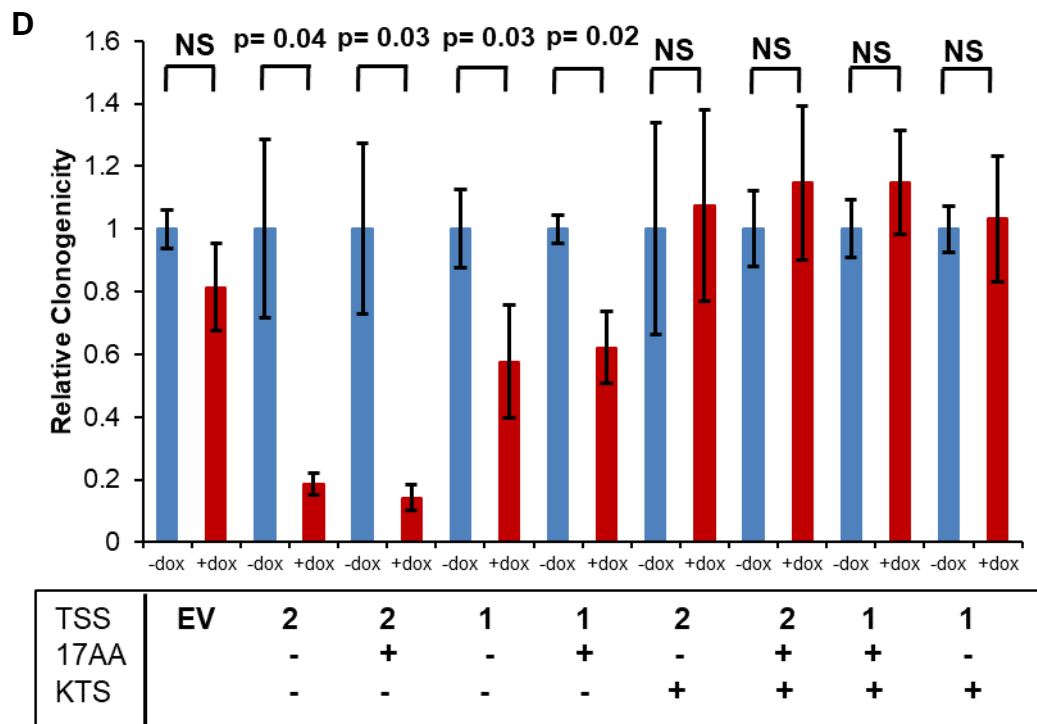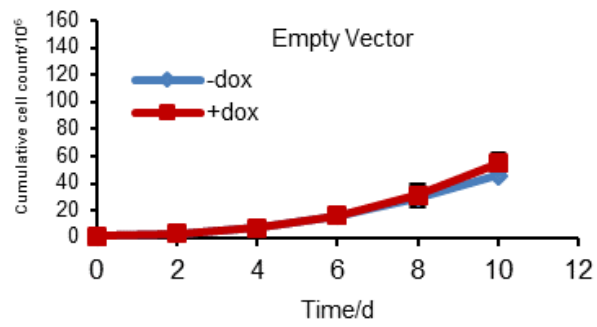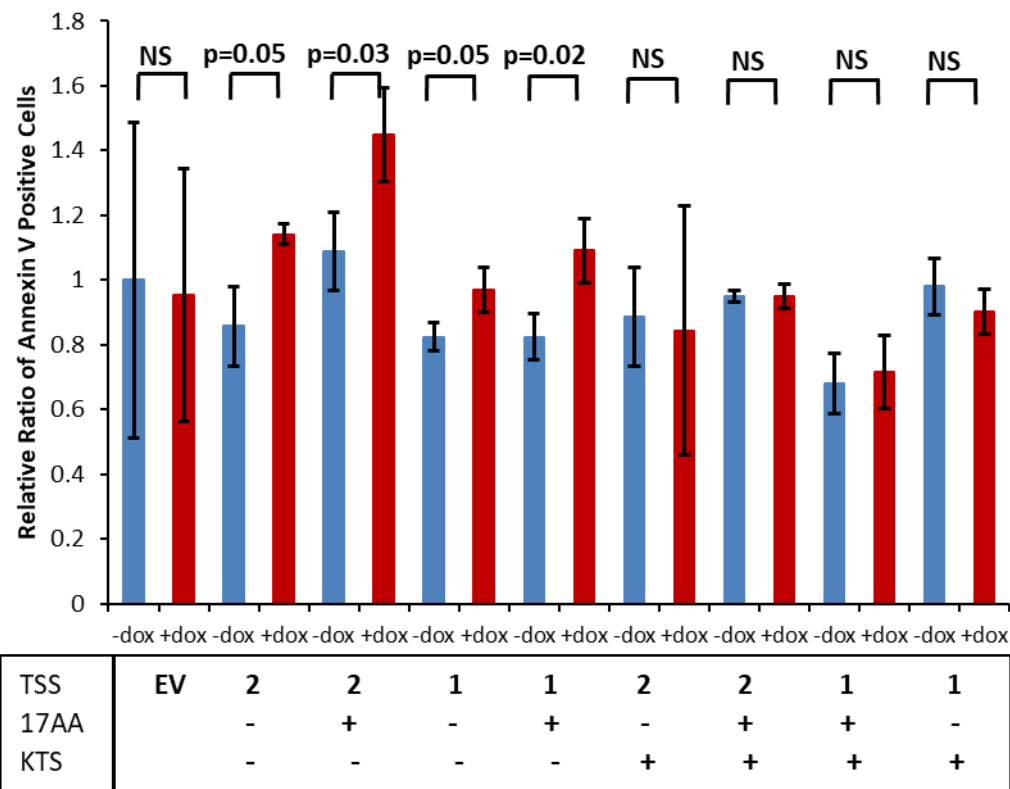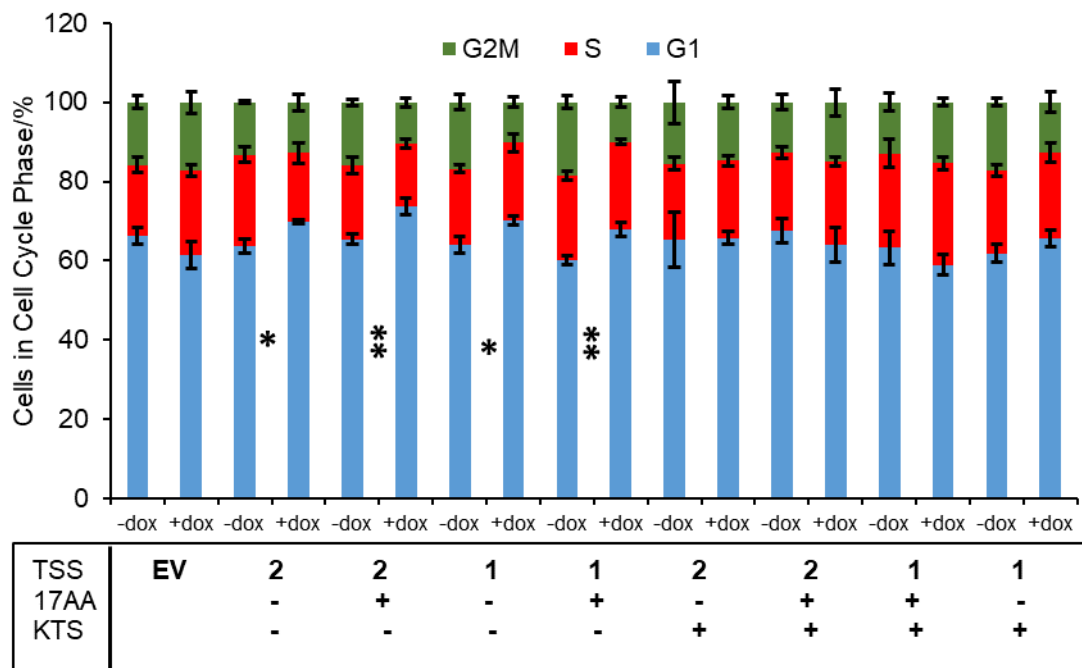
Figure 4-21: Expression of *WT1* –KTS isoforms decreases growth and colony formation of Kasumi-1 cells, increases apoptosis and induces a G1 cell arrest.

(A) Schematic showing the strategy of doxycycline induction strategy of *WT1* isoform expression in Kasumi-1 cells. (B) Western blot determining WT1 and GAPDH protein levels in Kasumi-1 cells transduced with each of the *WT1* isoforms. (C-F) Measurement of Kasumi-1 cells transduced with each of the *WT1* isoforms and assessing (C) time course of cumulative cell count (D) colony formation per 2000 cells seeded in methylcellulose and normalised to Empty Vector control (E) apoptosis relative to the empty vector control (F) cell cycle phase. For all panels, n=3 and error bars show standard deviation. Significance was determined using a 2 tailed student's t-test. * indicates p-value <0.05, ** indicates p-value <0.005.

Next, we correlated functional findings with gene expression. Global analysis of RNA-seq experiments will be discussed in Section 4.3.3. However, focusing on selected genes and pathways relevant to phenotypic findings, we found that that *CDKN2A* and *CASP9* were upregulated when *WT1* –KTS was induced (Figure 4-22). *CDKN2A* codes for the p14ARF and p16INK4A proteins which inhibit the cyclin-dependent kinase proteins CDK4 and CDK6 thus inhibiting the G1 cell cycle transition (Vidal and Koff, 2000). *CASP9* codes for the Caspase 9 protein which cleaves and activates effectors of apoptosis such as Caspase 3 (Fritsch et al., 2019).
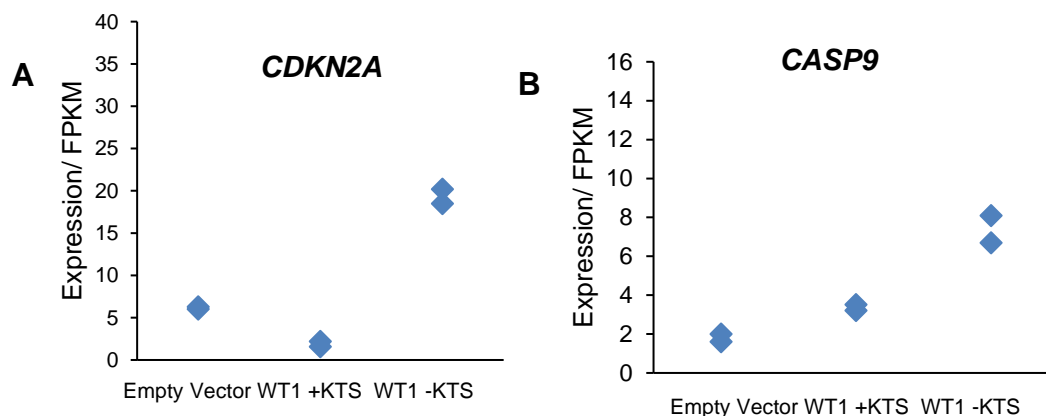
Figure 4-22: *CDKN2A* and *CASP9* expression are increased with *WT1* –KTS expression.

RT-qPCR of mRNA from Kasumi-1 cells transduced with doxycycline-induced *WT1* isoforms of (A) *CDKN2A* or (B) *CASP9.*

4.3.2 RUNX1-ETO and WT1 co-operate to block myeloid differentiation

Knockdown of the *RUNX1-ETO* driver oncogene in t(8;21) AML has been described to cause myeloid differentiation of t(8;21) cell lines and primary cells (Ptasinska et al., 2012). We therefore investigated whether WT1 co-operated with RUNX1-ETO to affect myeloid differentiation by assessing cell surface expression of the stem cell marker CD34 and the mature myeloid cell marker CD11b (Figure 4-23A-B). As expected, *RUNX1-ETO* knockdown with siRNA (siRE) and with expression of an empty vector control led to decreased CD34 and increased CD11b expression (myeloid differentiation) compared to cells receiving a mismatch control (siMM). Expression of *WT1* –KTS also led to myeloid differentiation, but of note, when *WT1* –KTS was expressed and *RUNX1-ETO* was knocked down, myeloid differentiation was escalated. Intriguingly, when *WT1* +KTS was expressed, not only did it not cause myeloid differentiation but it also prevented the myeloid differentiation that was expected after the *RUNX1-ETO* oncogene was knocked down.

Wright-Giemsa staining of cells showed differentiation to neutrophils when *WT1* –KTS was expressed whereas under the other conditions cells were maintained in the blast stage (Figure 4-23C). Similarly, RNA-seq showed downregulation of *CD34* and upregulation of *ITGAM* (which codes for CD11b) (Figures 4-23D-E).

A

CD34

B

CD11b

C

Empty Vector   WT1 +KTS   WT1 -KTS
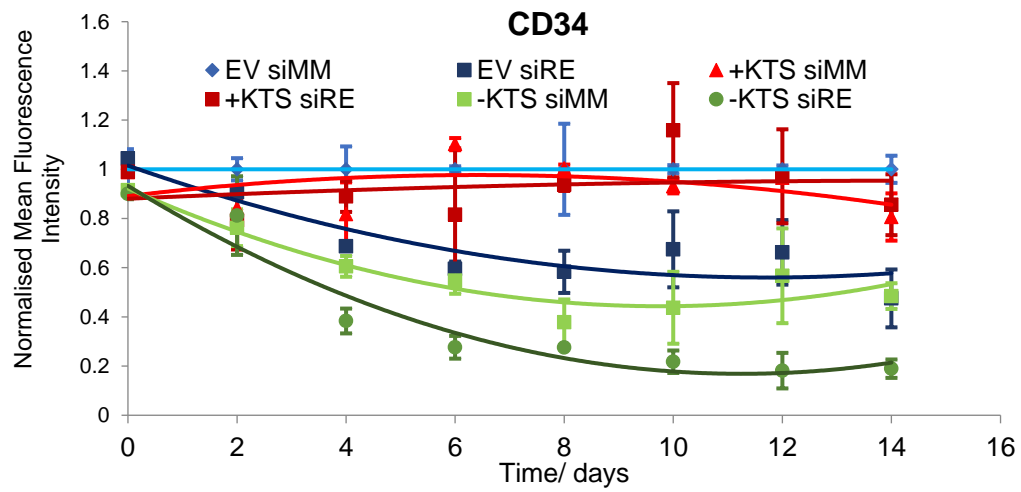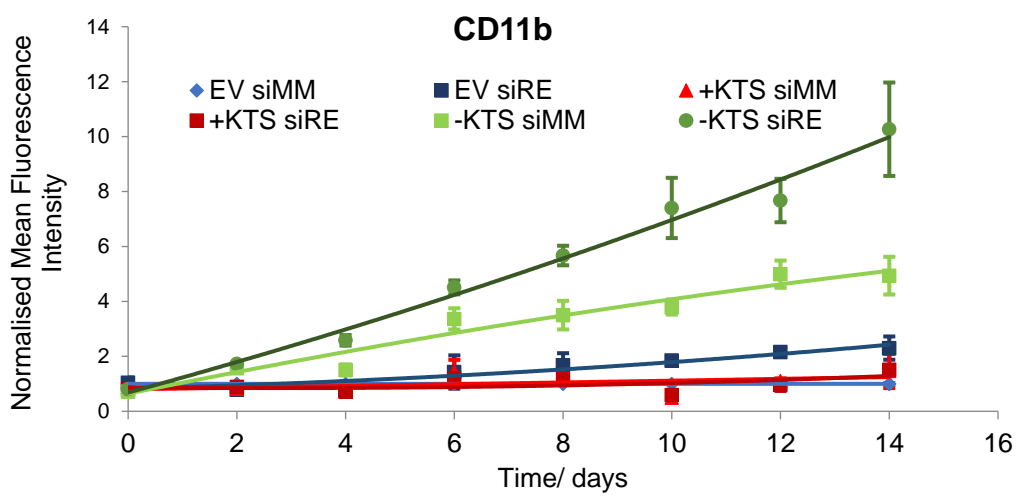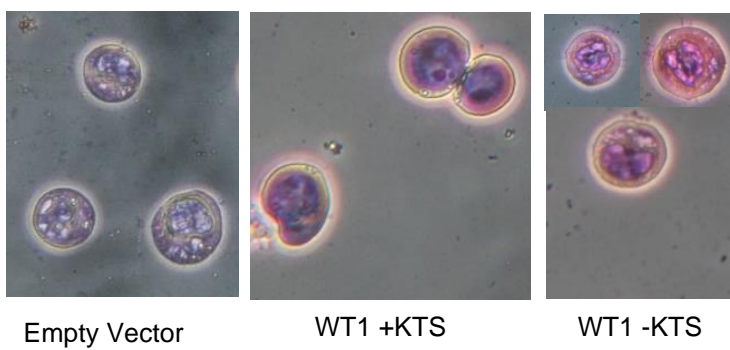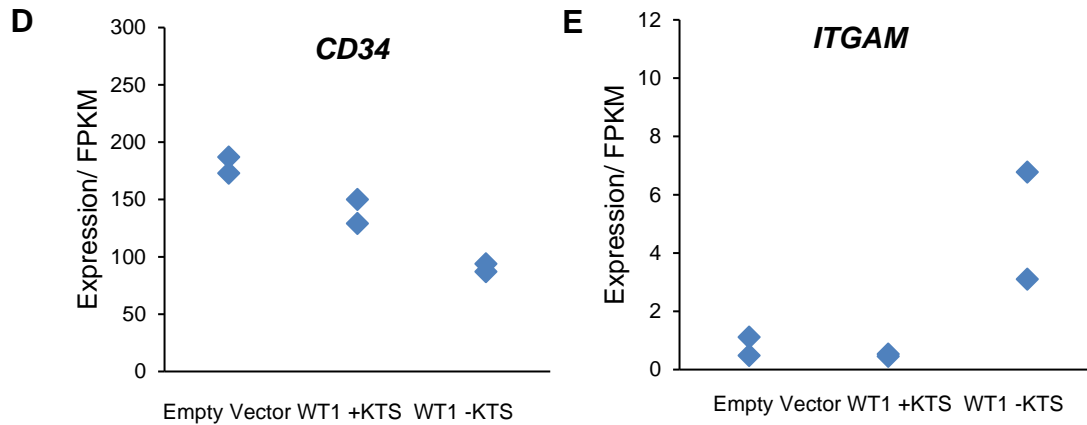
Figure 4-23: *RUNX1-ETO* knockdown co-operates with *WT1 –*KTS expression in permitting myeloid differentiation.

(A-B) Time course of mean fluorescence intensity changes of (A) CD34 and (B) CD11b surface marker expression in Kasumi-1 cells transduced with doxycycline-induced Empty Vector, *WT1* +KTS or *WT1* –KTS and transfected with siRNA against RUNX1-ETO (siRE) or a Mismatch (siMM) control. (C) Morphology of Kasumi-1 cells 10 days after transduction with doxycycline-induced Empty Vector, *WT1* +KTS or *WT1* –KTS. (D-E) Gene expression analysis from RNA-seq data from Kasumi-1 cells transduced with doxycycline-induced *WT1* isoforms measuring (D) *CD34* or (E) *ITGAM* (CD11b).

4.3.3 Expression of different *WT1* Isoforms leads to alterations in the chromatin landscape and in the transcriptome

Expression of *WT1* –KTS in Kasumi-1 cells led to the formation of 1742 new DHSs with enrichment for the WT1, PU.1 and RUNX motifs and a loss of ETS motifs in the 1345 lost peaks (Figure 4-24A). The change from ETS to PU.1 motifs in DHSs with expression of *WT1* -KTS is consistent with myeloid differentiation (Nerlov and Graf, 1998a). The enrichment of WT1 motifs within new DHSs is intriguing. WT1 is not a known pioneer factor and does not have the Fork head or NF-Y style heterotrimeric structure normally associated with such factors;

instead WT1 probably recruits other proteins which may remodel chromatin. Expression of *WT1* +KTS led to the formation of 973 new DHSs at sites with an enrichment for AP-1 and PU.1 motifs but notably no WT1 motif enrichment was seen in the 3131 lost DHSs (Figure 4-24B). The switch from PU.1 to ETS upon *WT1* +KTS expression perhaps suggests that the cells are becoming more stem cell like, consistent with the findings in Section 4.3.2. The lack of WT1 in lost motifs suggests that changes in chromatin conformation here were indirect changes and not directly as a consequence of WT1. We observed no particular change in the genomic distributions of DHSs seen when *WT1* isoforms were expressed (Figure 4-24C).

A
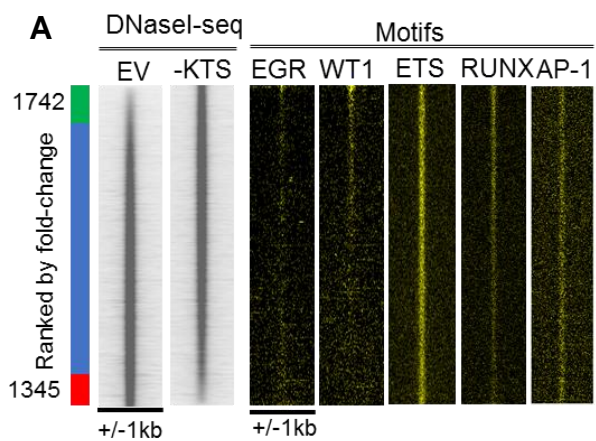
Motif in WT1 -KTS specific peaks n = 1742

Motif	Match	p value	% of Targets
AAAAAGGAAGT	PU.1	1e-234	31.5
GTGTGTGGGAGT	WT1	1e-189	27.0
AACTTGTGGTTA	RUNX	1e-117	28.4
ATCTATTTTTAG	MEF2	1e-33	10.72
TATGACTCAT	AP-1	1e-27	10.39

Motif	Match	p value	% of Targets
AACAGGAAGT	ETS	1e-106	31.83
CCACCAGGGGGC	CTCF	1e-92	13.20
TGTGGTTA	RUNX	1e-23	12.93
GCTGACTCAT	AP-1	1e-30	9.04

B
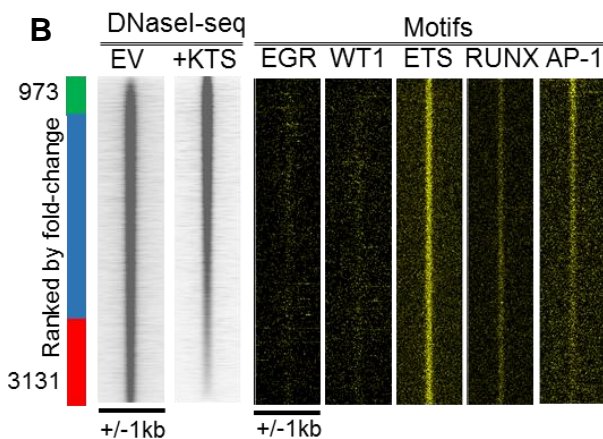
Motif	Match	p value	% of Targets
ACAGGAAGTA	ETS	1e-63	34.45
ATGTGGTT	RUNX	1e-20	24.13
CAATGACTCACC	AP-1	1e-47	17.0

Motif	Match	p value	% of Targets
AACAGGAAGT	PU.1	1e-319	46.89
ACTGTGGTTA	RUNX	1e-91	22.83
ATTTTGCAAT	CEBP	1e-77	14.19
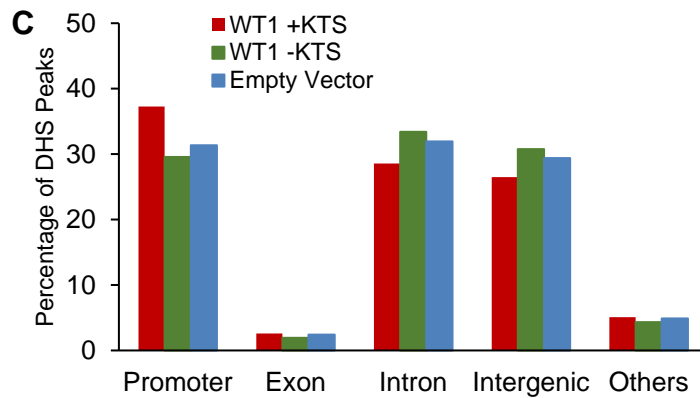
Figure 4-24: Induction of *WT1* isoforms leads to changes in open chromatin sites.

(A-B) Density plots of DNase I-seq peaks in a 2kb window, ranked top to bottom by relative tag count of peaks in (A) doxycycline-induced *WT1* -KTS or (B) doxycycline-induced *WT1* +KTS relative to Empty Vector control in Kasumi-1 cells. The bars on the side depict the number of specific and shared peaks. Plotted alongside are motifs occurring at the DHSs. The tables of enriched TF binding motifs within the specific peaks are shown alongside. (C) Histogram of the genomic distribution of DNase I Hypersensitive Sites in Kasumi-1 cells transduced with Empty Vector, *WT1* +KTS or *WT1* –KTS.

We next examined the global impact of *WT1* isoform overexpression using RNA-seq. *WT1* –KTS expression led to a distinct transcriptional signature which clusters away from other experimental conditions (*WT1* +KTS, Empty Vector control) when examined by hierarchical clustering of RNA-seq data (Figure 4-25A). A heatmap of differentially expressed genes showed that gene expression changes appear to be greatest with expression of *WT1* –KTS (Figure 4-25B). In section 4.3.2, we showed how WT1 and RUNX1-ETO co-operate to alter myeloid differentiation and so we therefore investigated differential gene expression after *WT1* was knocked down and in comparison to *RUNX1-ETO* knockdown. This analysis shows an alteration in expression of many common genes, however

some changes were distinct (Figure 4-25C). In order to assess which gene expression changes were likely to be directly due to WT1, we analysed how many differentially expressed genes were direct WT1 targets as assessed by WT1 ChIP-seq (global analysis of these experiments will be discussed in Section 4.4.2). Between 1/3 to 1/2 of all differentially expressed genes were direct WT1 targets (Figure 4.25D-E)
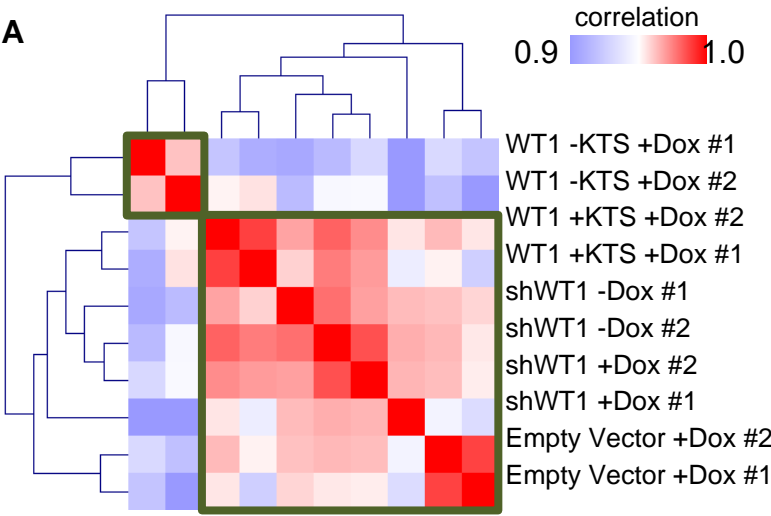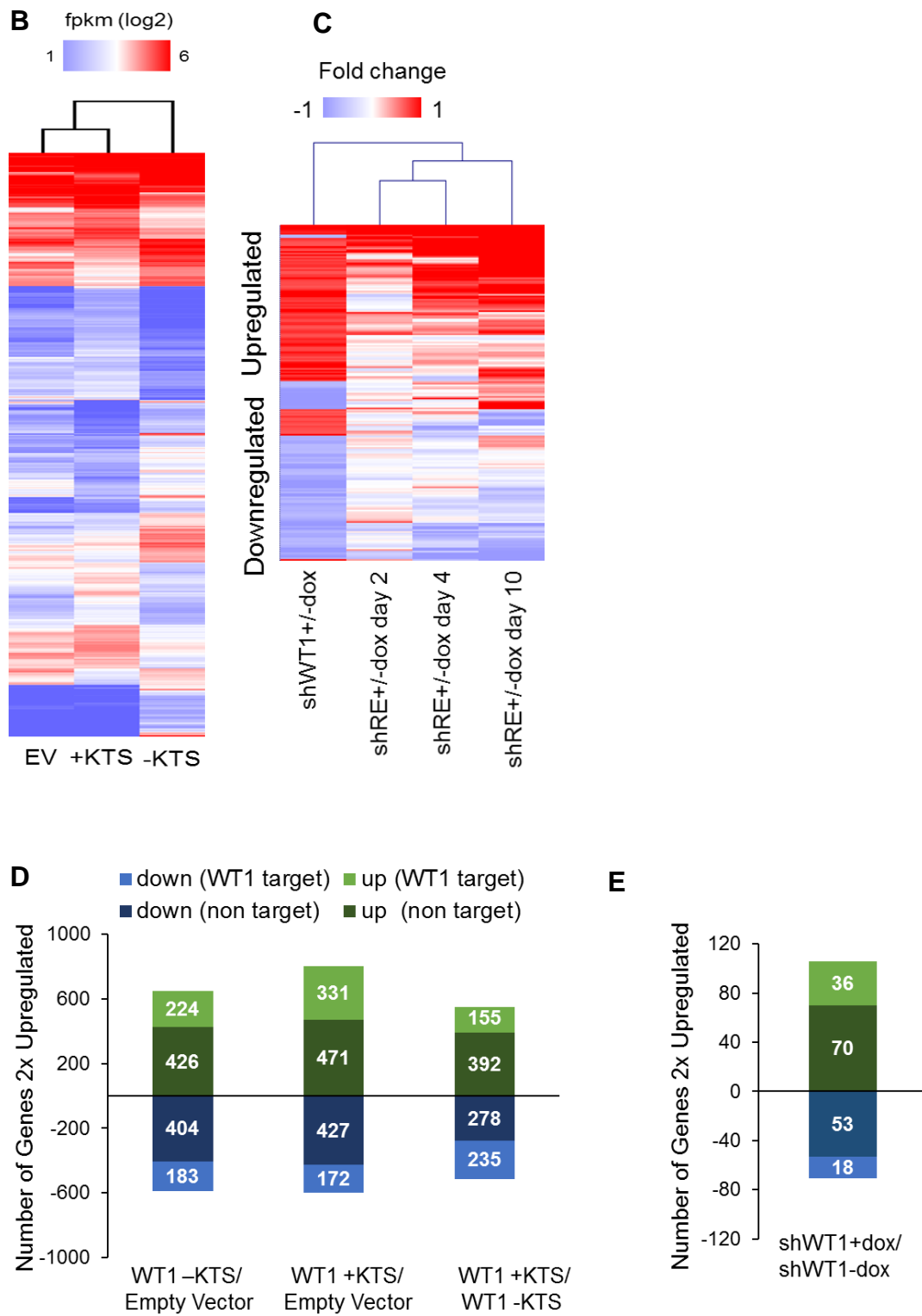
Figure 4-25: Distinct transcriptional changes are elicited with *WT1* isoform expression.

(A) Hierarchical clustering of Pearson correlation coefficients of RNA-seq experiments of Kasumi-1 cells transduced with Empty Vector, *WT1* +KTS, *WT1* –KTS or sh*WT1*. (B) Heatmap of hierarchical clustering of differentially expressed genes in Kasumi-1 cells transduced with Empty Vector, *WT1* +KTS or *WT1* – KTS. (C) Heatmap of hierarchical clustering of differentially expressed genes in Kasumi-1 cells transduced with sh*WT1* or sh*RUNX1-ETO* at day 2, 4 or 10 time points. (D-E) Number of differentially expressed genes and how many of which are direct WT1 targets with Kasumi-1 cells transduced with (D) Empty Vector, *WT1* +KTS or *WT1* –KTS or (E) sh*WT1*.

We performed a KEGG pathway analysis of differentially expressed genes, and focusing on the genes that are direct WT1 protein targets, revealed the cellular pathways being altered by the expression of the different isoforms of WT1. We found very few enriched upregulated pathways when *WT1* +KTS was expressed but the analysis of downregulated pathways revealed downregulation of genes normally involved in RNA degradation (Figure 4-26A-B), alluding to the role of WT1 in RNA processing (Duarte et al., 1998). Furthermore, downregulation of the p53 signalling pathway may permit the increased growth seen when WT1 +KTS was expressed (Section 4.3.1).

By contrast, expression of *WT1* –KTS led to the upregulation of many KEGG pathways. Notably, the upregulation of genes involved in many cellular signalling pathways including Jak-STAT, MAPK, NFκB and sphingolipid signalling (Figure 4-26C). However, most of these genes are negative regulators of these pathways such as the Dual Specificity Phosphatase (DUSP) 2 and 7 genes (Kim et al., 1999) which may explain the decreased growth of cells transduced with *WT1* – KTS. Downregulated pathways include cell cycle regulation, consistent with the cell cycle arrest seen in these cells (Section 4.3.1) as well as spliceosome genes,

which may represent another mechanism by which WT1 regulates the post-transcriptional modification of RNA (Figure 4-26D).

A *WT1 +KTS* upregulated KEGG pathways
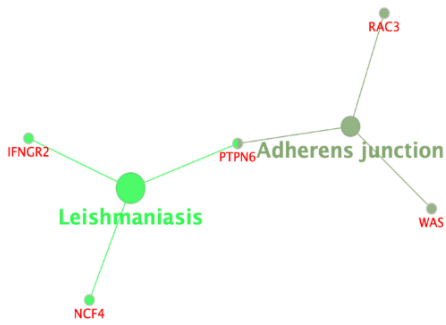
B *WT1 +KTS* downregulated KEGG pathways
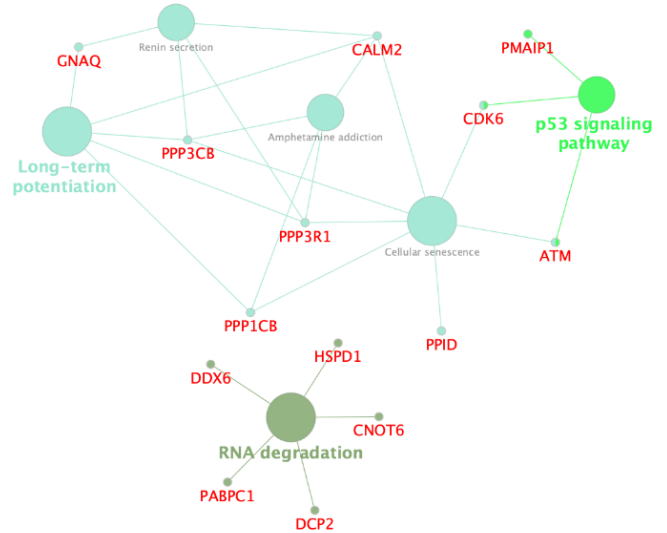
C *WT1 –KTS* upregulated KEGG pathways
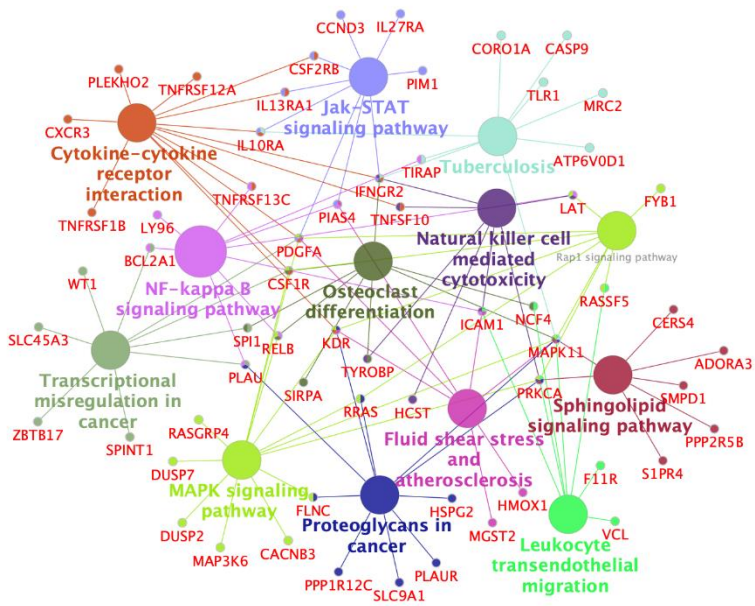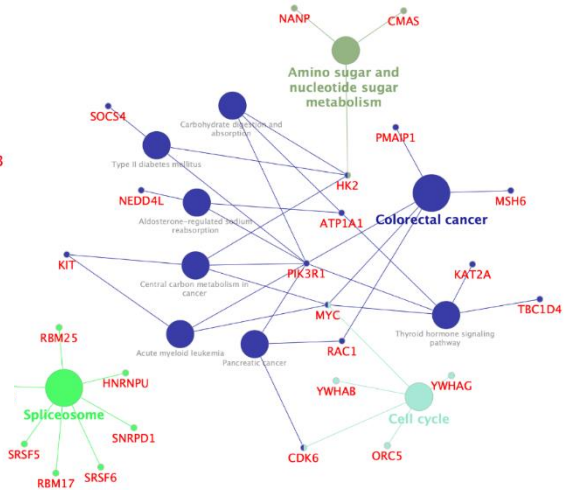
D *WT1 –KTS* downregulated KEGG pathways

Since *WT1* isoform expression appeared to affect splicing, we next analysed how splicing was altered, both globally and at specific genes. Globally, *WT1* +KTS expression led to the differential splicing of more genes compared with other experimental conditions (Figure 4-27A). Many of the genes that were differentially spliced were direct WT1 targets suggesting direct RNA processing by WT1 or its protein partners (Figure 4-27B) but some were not direct targets which may be explained by WT1 regulating the expression of genes coding for spliceosome proteins (Figure 4-26D). WT1 isoforms caused both increases and decreases in the Percentage of exonic reads Spliced In (PSI) in various genes (Figure 4-27C).

Focusing on the *WT1* locus, expression of the different isoforms led to differential splicing, as seen on a Sashimi plot (Figure 4-27D). Intriguingly, WT1 ChIP-seq showed peaks overlying each exon of *WT1* suggesting that WT1 may regulate its own splicing.

A

B

Number of WT1 targets that are differentially spliced

C
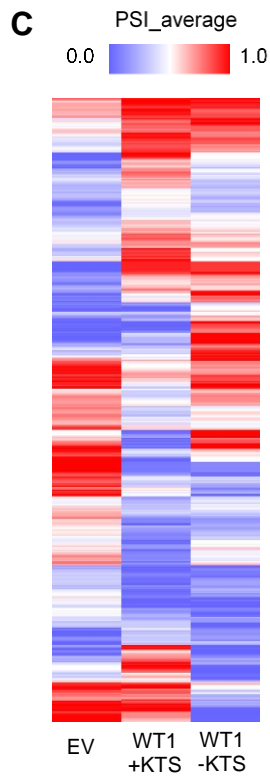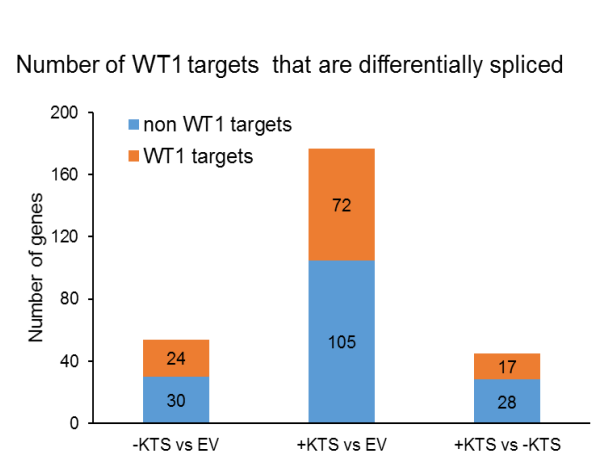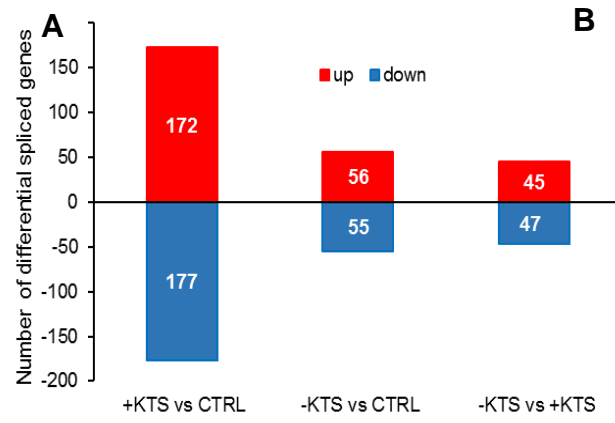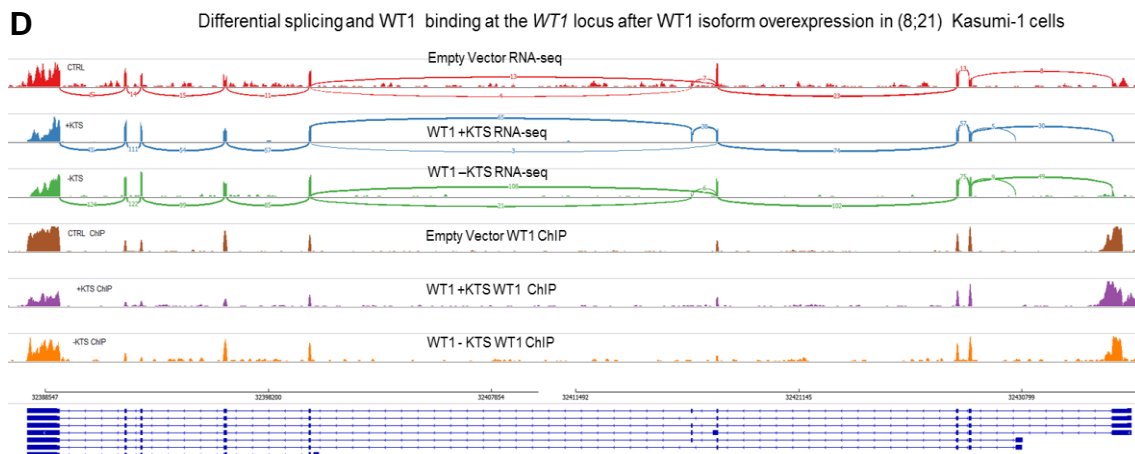
Figure 4-27: Expression of WT1 isoforms drives differential splicing.

(A-B) Number of differentially spliced genes, (A) all genes or (B) direct WT1 targets only, p-value <0.01, of Kasumi-1 cells transduced with doxycycline-induced Empty Vector, *WT1* +KTS or *WT1* –KTS. Increased PSI (Percentage of exonic reads Spliced In) shown above the axis or decreased PSI shown below the axis. (C) Heatmap showing hierarchical clustering of differentially spliced genes as determined by PSI, p-value <0.01 of Kasumi-1 cells transduced with doxycycline-induced Empty Vector, *WT1* +KTS or *WT1* –KTS. (D) Screenshot of splicing events in the *WT1* locus depicting the frequency of exon usage in control (empty vector transduced), *WT1* +KTS transduced and *WT1* -KTS transduced Kasumi-1 cells, together with ChIP data depicting the position of WT1 binding sites as indicated.

4.3.4 *WT1* expression alter the chromatin landscape and gene expression in primary t(8;21) AML

Having established that expression of WT1 isoforms leads to distinct transcriptional, epigenetic and phenotypic changes in Kasumi-1 cells, we wanted to see whether this was also true for primary patient derived cells, starting with t(8;21) AML. We transduced primary t(8;21) AML with doxycycline-induced Empty Vector, *WT1* +KTS or *WT1* –KTS and FACS sorted transduced cells after doxycycline induction based upon GFP expression and confirmed *WT1*

expression with RT-qPCR (Figure 4-28A). Analysis of ATAC-seq data from these FACS sorted cells comparing the Empty Vector and *WT1* –KTS cells showed distinct chromatin changes. Within the 1672 new open chromatin sites appearing after *WT1* –KTS was expressed, we found a gain in enrichment of PU.1 and C/EBP motifs, suggesting that these cells were undergoing myeloid differentiation (Zhang et al., 1997a) (Figure 4-28B).

The analysis of differentially expressed gene in the RNA-seq data from these FACS sorted primary cells showed that *WT1* –KTS expression caused distinct transcriptomic changes (Figure 4-28C); we found an upregulation of genes suggestive of myeloid differentiation such as *ITGAM* (which codes for CD11b) as well as pathways involved in IL-17 signalling suggesting a shift in the cells towards an inflammatory phenotype (Iwakura et al., 2011) (Figure 4-29A-B). Expression of *WT1* +KTS caused an upregulation of cytokine-cytokine receptor and toll-like receptor KEGG pathways (Figure 4-29C-D) which is consistent with increased signalling-related cell growth. In addition further dysregulation in Il-17 signalling was seen.

A

B

t(8;21) ATAC

	EV	-KTS	+KTS
1672			
1			
2			
653	3		

+/-1kb

Motifs

ETS RUNX AP-1 CEBP PU.1

+/-1kb

Motifs in WT1 -KTS specific peaks n = 1672

Motif	Match	p-value	% Targets
TTGTGGTT	RUNX	1e-27	32.61
AAGAGGAAGT	PU.1	1e-97	26.92
TATGAGTCAT	AP-1	1e-43	10.18
GCCCTCTAGTGG	CTCF	1e-73	6.33
TTGCGCAA	C/EBP	1e-8	3.93

Motifs in Empty Vector specific peaks n = 653

Motif	Match	p-value	% Targets
CAGTAGGGGG	CTCF	1e-14	19.35
ATGACTCA	AP-1	1e-27	14.35
TGTGGTTT	RUNX	1e-9	8.04

C

fpkm (log2)

0   3

EV   WT1 +KTS   WT1 -KTS

Figure 4-28: Expression of *WT1* isoforms leads to epigenetic and transcriptional changes in primary t(8;21) AML cells.

(A) Relative colony formation ability of primary t(8;21) AML cells transduced with doxycycline-induced Empty Vector, *WT1* +KTS or *WT1* –KTS. (B) Density plots of ATAC-seq peaks in a 2kb window, ranked top to bottom by relative tag count of peaks from doxycycline-induced *WT1* -KTS transduced cells relative to doxycycline-induced Empty Vector transduced cells in primary t(8;21) AML. Bars on the left indicate the number of peaks that are specific or shared between the cell types. The tables of enriched transcription factor binding motifs in the specific groups are depicted alongside. (C) Heatmap of differentially expressed genes of primary t(8;21) AML cells transduced with doxycycline-induced Empty Vector, *WT1* +KTS or *WT1* –KTS.

A Primary t(8;21) *WT1* –KTS upregulated KEGG pathways

B Primary t(8;21) *WT1* –KTS downregulated KEGG pathways

C Primary t(8;21) *WT1* +KTS upregulated KEGG pathways

D Primary t(8;21) *WT1* +KTS downregulated KEGG pathways

Figure 4-29: Expression of *WT1* isoforms leads to alterations in KEGG pathways in primary t(8;21) AML.

(A-D) KEGG pathways derived from enriched genes of primary t(8;21) AML transduced with doxycycline-induced Empty Vector, *WT1* +KTS or *WT1* –KTS.

We next integrated gene expression, open chromatin data and motif enrichment data to construct transcription factor (Gene Regulatory) networks in t(8;21) AML, which was facilitated by our previous capture HiC data which allowed us to annotate binding motifs to their rightful promoter (Assi et al., 2019). Here, we focused specifically on those transcription factors that were differentially expressed compared to healthy CD34[+] cells (Figure 4-30).

A large degree of mutual binding between various transcription factors and their respective genes can be seen and the AP-1 family of TFs stand out as a major node. Having established the WT1 motif through ChIP-seq (see Section 4.4.2), we found that WT1 was strongly regulated by the ETS family of TFs and that WT1 could in turn regulate the AP-1 and NFAT families. Thus WT1 forms an important node within the t(8;21) AML TF network.
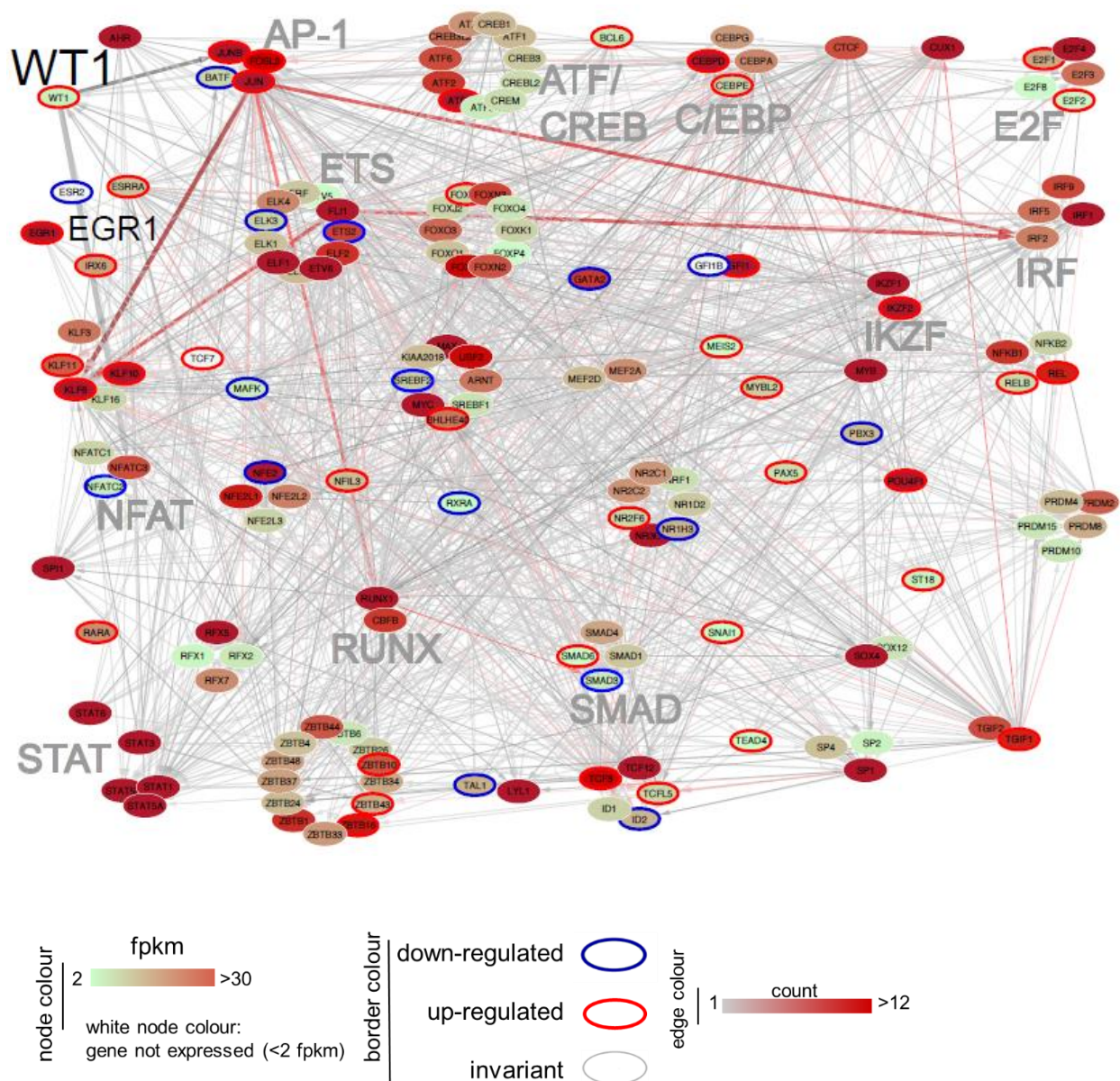
Node and edge attributes

Figure 4-30: AML-specific transcription factor network in t(8;21) AML.

Transcription Factor families binding to the same motif form a node, highlighted with large grey or black text with the family name. Arrows pointing outward from entire TF family node highlight footprinted motifs in individual genes generated by any member of this factor family whereby the footprint was annotated to the gene using the CHiC data where possible, or otherwise to the nearest gene. The colour of the edge (arrow) indicated the number of footprinted transcription factor binding motifs found in the target transcription factor gene. The expression level (FKPM) of the individual genes is depicted in white (low) or red (high) color. The

colour of the border indicated whether the gene is upregulated or downregulated compared to healthy CD34[+] cells. The network was constructed by Dr Salam Assi.

4.3.5 *WT1* expression alters the chromatin landscape and gene expression in primary FLT3-ITD AML

We next used the same strategy to examine the role of WT1 in FLT3-ITD AML, a second major subgroup of AML. We transduced and FACS sorted GFP positive cells and confirmed *WT1* overexpression through RT-qPCR (Figure 4-31A). Expression of *WT1* –KTS reduced colony formation from primary FLT3-ITD cells whereas expression of *WT1* +KTS did not affect it (Figure 4-31B). The analysis of ATAC-seq data from these cells showed 1111 new open chromatin sites after the expression of *WT1* –KTS (Figure 4-31C). Motif enrichment analysis of these new open chromatin sites showed a specific enrichment for PU.1 motifs, suggesting that as with t(8;21) AML, these cells may be undergoing myeloid differentiation.

A WT1 Expression in Primary FLT3-ITD

Relative Expression Normalised to GAPDH

Empty Vector   WT1 +KTS   WT1 -KTS

B

p=0.00012

NS

Relative Clonogenicity

Empty Vector   WT1 +KTS   WT1 -KTS

C

Motif	Match	p-value	% Targets
AAAAGAGGAAGT	PU.1	1e-78	29.98
TGTGGTTAGT	RUNX	1e-37	14.19
ATGACTCATC	AP-1	1e-30	12.13
GCCCTCTAGTGG	CTCF	1e-80	12.13

Motifs in WT1 -KTS specific peaks n = 1111

Motif	Match	p-value	% Targets
TTGTGGTTAT	RUNX	1e-49	21.92
TGATGACTCA	AP-1	1e-51	13.27
CCACTAGGGGGC	CTCF	1e-73	10.19

Motifs in Empty Vector specific peaks n =1315

Figure 4-31: Expression of the *WT1* –KTS isoforms leads decreased colony formation and distinct epigenetic changes in primary FLT3-ITD AML cells.

(A) RT-qPCR of relative expression of *WT1* mRNA compared to *GAPDH* mRNA in primary FLT3-ITD AML. (B) Relative colony formation ability of primary FLT3-ITD AML cells transduced with doxycycline-induced Empty Vector, *WT1* +KTS or *WT1* –KTS. (C) Density plots of ATAC-seq peaks in a 2kb window, ranked top to bottom by relative tag count of peaks in doxycycline-induced *WT1* -KTS transduced cells relative to doxycycline-induced Empty Vector transduced cells in primary FLT3-ITD AML. Bars on the left indicate the number of peaks that are specific or shared between the cell types. The tables of enriched transcription factor binding motifs in the specific groups are depicted alongside.

Analysis of differentially expressed genes from RNA-seq data showed that once again expression of *WT1* –KTS led to changes in gene expression distinct from the other experimental conditions (Figure 4-32). Analysis of KEGG pathways revealed that expression of *WT1* –KTS led to an upregulation of genes involved in apoptosis (Figure 4-33A), which may partially explain why colony formation was impaired. Downregulated genes included those involved in regulation of the cell cycle such as *CCND1, CCNE1* and *E2F1* (Figure 4-33B). By contrast, when

the *WT1* +KTS was expressed, cell cycle genes such as *CCNE1*, *CCNE2* and *E2F1* were upregulated, consistent with the idea that WT1 +KTS produces an increased growth phenotype (Figure 4-33C).
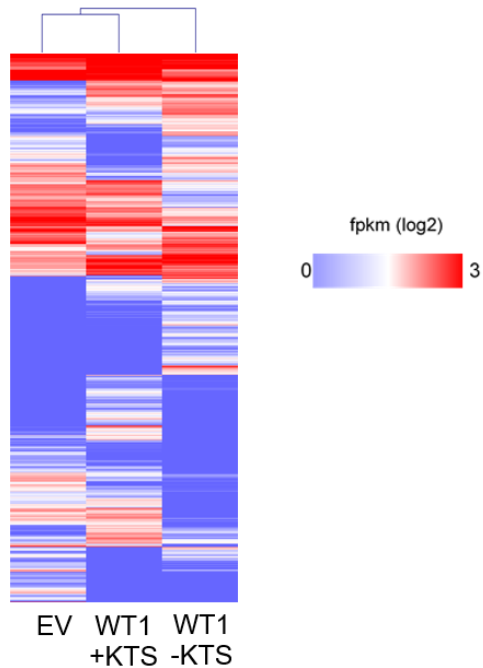
Figure 4-32: Expression of *WT1* isoforms leads to distinct transcriptional changes in primary FLT3-ITD AML cells.

Heatmap of differentially expressed genes of primary FLT-ITD AML cells transduced with doxycycline-induced Empty Vector, *WT1* +KTS or *WT1* –KTS.

**Primary FLT3-ITD *WT1* –KTS
upregulated KEGG pathways**

A

**Primary FLT3-ITD *WT1* –KTS
downregulated KEGG pathways**
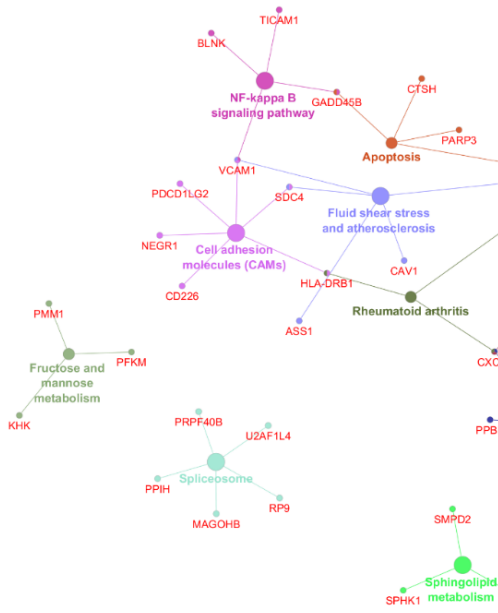
B

**Primary FLT3-ITD *WT1* +KTS
upregulated KEGG pathways**
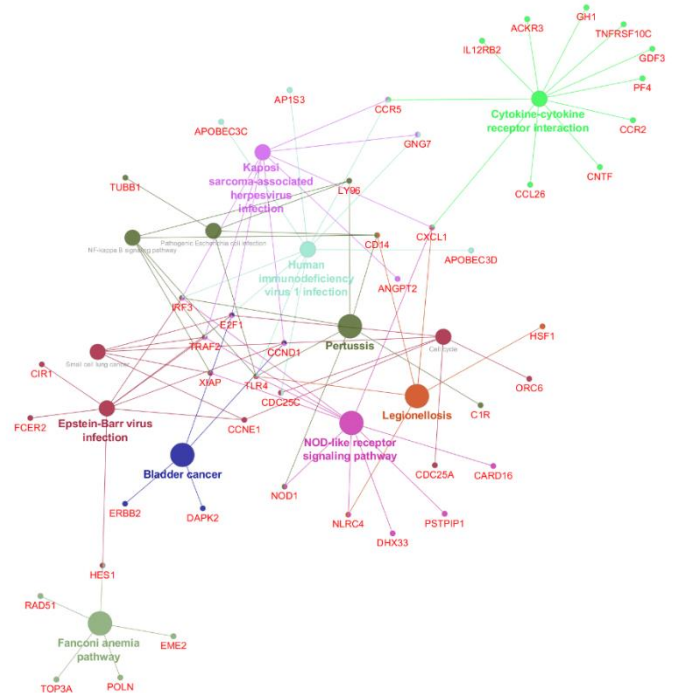
C

**Primary FLT3-ITD *WT1* +KTS
downregulated KEGG pathways**

D

230
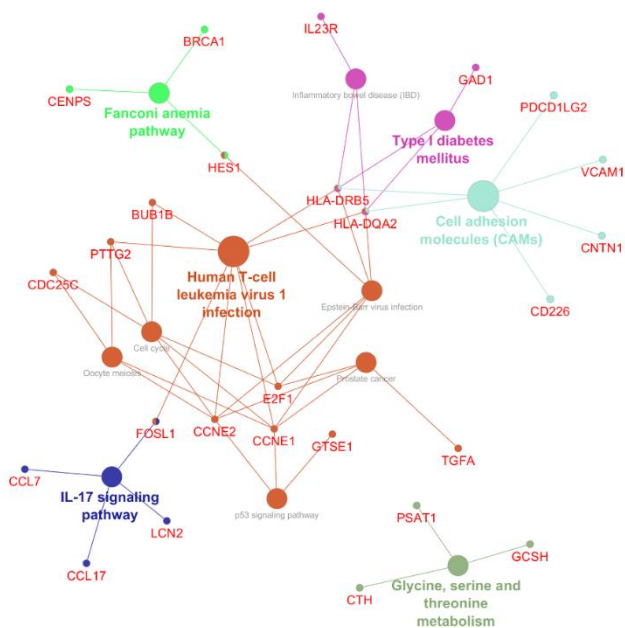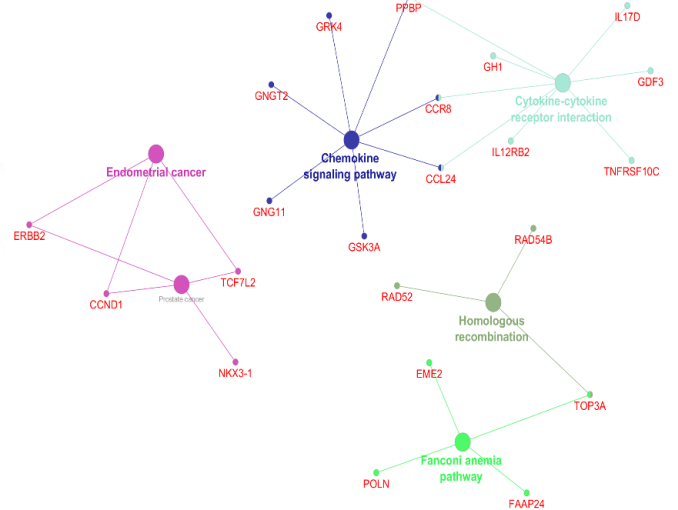
Figure 4-33: Expression of *WT1* isoforms leads to alterations in different KEGG pathways in primary FLT3-ITD AML.

(A-D) KEGG pathways derived from enriched genes of primary FLT3-ITD AML transduced with doxycycline-induced Empty Vector, *WT1* +KTS or *WT1* –KTS.

Integrating all of the multi-omics data, we determined the Transcription Factor Network for primary FLT3-ITD AML. Similar to t(8;21) AML, AP-1 and ETS factors were found to be major nodes within this network (Figure 4-34). Furthermore, WT1 once again emerged as an important regulator, which could regulate the AP-1 and NFAT families of TFs. However, we also found key differences such as the prominence of the HOX cluster and transcription factors such as NFIX and FOXC1 that was specific to the FLT3-ITD AML Transcription Factor Network.
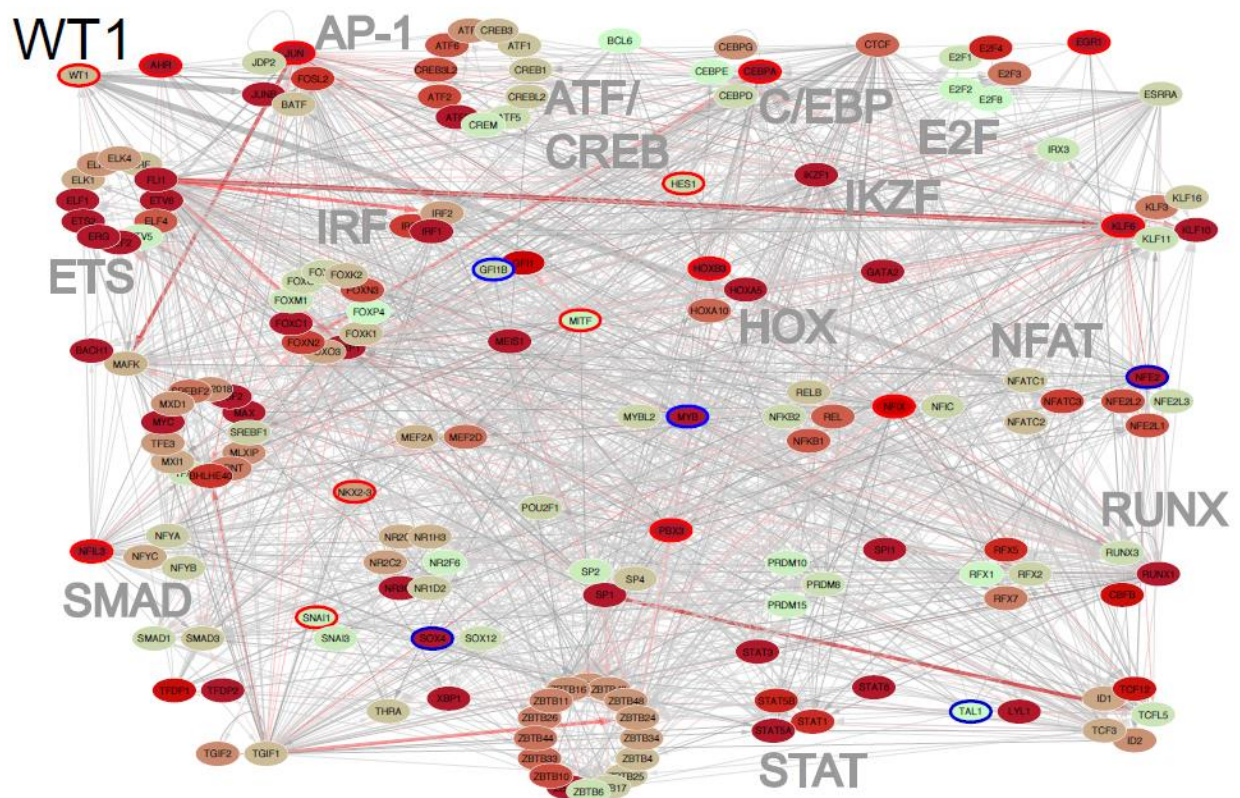
Node and edge attributes

Figure 4-34: AML-specific transcription factor network in FLT3-ITD AML.

Transcription Factor families binding to the same motif form a node, highlighted with large grey or black text with the family name. Arrows pointing outward from entire TF family node highlight footprinted motifs in individual genes generated by any member of this factor family whereby the footprint was annotated to the gene using the CHiC data where possible, or otherwise to the nearest gene. The colour of the edge (arrow) indicated the number of footprinted transcription factor binding motifs found in the target transcription factor gene. The expression level (FKPM) of the individual genes is depicted in white (low) or red (high) color. The colour of the border indicated whether the gene is upregulated or downregulated compared to healthy CD34⁺ cells. The network was constructed by Dr Salam Assi.

4.3.6 *WT1* expression causes minimal change in the chromatin landscape or transcriptome of healthy patient cells

When searching for therapeutic targets, it is important their perturbation does not affect normal cells, thus reducing toxicities for patients. In order to address this issue, we transduced the different *WT1* isoforms affected into CD34⁺ cells derived from the bone marrow of healthy patients.

We confirmed successful induction of *WT1* isoforms within these cells through RT-qPCR (Figure 4-35A). Reassuringly, altering expression of *WT1* isoforms did

not affect the colony formation ability of healthy cells (Figure 4-35B). ATAC-seq analysis showed very few changes in open chromatin sites (Figure 4-35C). In addition, whilst there were a small number of differentially expressed genes after *WT1* isoform expression (Figure 4-35D), no enriched KEGG pathways were found suggesting minimal transcriptomic change.

Motif	Match	p-value	% Targets
GCCCCCTSGTGG	CTCF	1e-116	24.23
AAAASAGGAAGT	ETS	1e-42	20.10

Motifs in EV specific peaks n = 259

Motif	Match	p-value	% Targets
GATGACTCAT	AP-1	1e-11	23.30
AAGAGGAAGT	ETS	1e-9	16.56

D
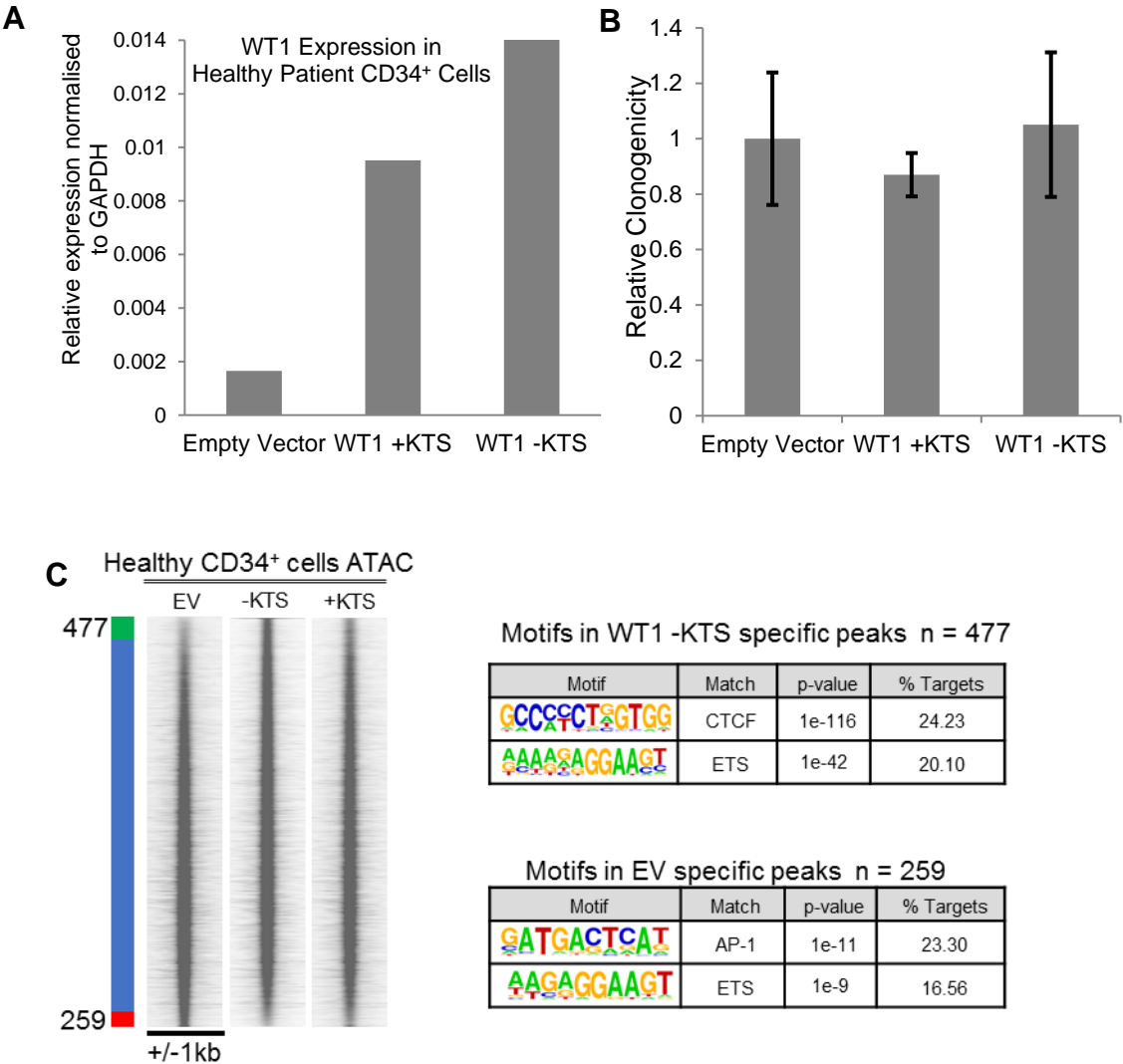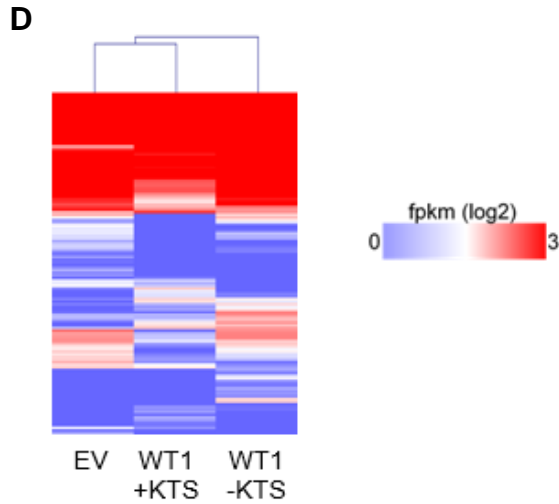
<u>Figure 4-35: Expression of *WT1* isoforms leads to limited epigenetic or transcriptional change in primary healthy CD34⁺ cells.</u>

(A) RT-qPCR of relative expression of *WT1* mRNA compared to *GAPDH* mRNA in primary FLT3-ITD AML. (B) Relative colony formation ability of primary healthy CD34⁺ cells transduced with doxycycline-induced Empty Vector, *WT1* +KTS or *WT1* –KTS. (C) Density plots of ATAC-seq peaks in a 2kb window, ranked top to bottom by relative tag count of peaks in doxycycline-induced *WT1* -KTS transduced cells relative to doxycycline-induced Empty Vector transduced cells in primary healthy CD34⁺ cells. Bars on the left indicate the number of peaks that are specific or shared between the cell types. The tables of enriched transcription factor binding motifs in the specific groups are depicted alongside. (D) Heatmap of differentially expressed genes of primary healthy CD34⁺ cells transduced with doxycycline-induced Empty Vector, *WT1* +KTS or *WT1* –KTS.

4.3.7 WT1 expression leads to AML subtype-specific changes in gene expression and in chromatin landscapes

Next, we performed a global analysis of RNA-seq (Figure 4-36A) and ATAC-seq (Figure 4-36B) patterns incorporating all primary cell experiments. This analysis found that cells predominantly clustered according to their parent cell type. Therefore, whilst WT1 expression or depletion does give rise to important

transcriptional and epigenetic alterations, these changes are relatively small compared to AML subtype-specific differences.

A

RNA-seq

correlation
0.6 — 1.0

Healthy CD34⁺ WT1 -KTS
Healthy CD34⁺ Empty Vector
Healthy CD34⁺ WT1 +KTS
Healthy CD34⁺ shWT1
Healthy CD34⁺ shNTC
FLT3-ITD shNTC
FLT3-ITD shWT1
FLT3-ITD Empty Vector
FLT3-ITD WT1 +KTS
FLT3-ITD WT1 –KTS
t(8;21) WT1 +KTS
t(8;21) Empty Vector
t(8;21) WT1 -KTS

B

ATAC-seq

correlation
0.0 1.0

Healthy CD34⁺ EV
Healthy CD34⁺ -KTS
Healthy CD34⁺ +KTS
Healthy CD34⁺ shNTC
Healthy CD34⁺ shWT1
FLT3-ITD shWT1
FLT3-ITD shNTC
FLT3-ITD WT1 +KTS
FLT3-ITD Empty Vector
FLT3-ITD WT1 –KTS
t(8;21) Empty Vector
t(8;21) WT1 -KTS
t(8;21) WT1 +KTS

Figure 4-36: Differences in the transcriptome and epigenome elicited by *WT1* perturbations occur within an AML subtype-specific cistrome.

(A-B) Heatmap showing hierarchical clustering of Pearson Correlation Coefficients of the indicated primary cell (A) RNA-seq experiments or (B) ATAC-seq experiments. Primary t(8;21) AML, primary FLT3-ITD AML or healthy CD34⁺ cells were transduced with doxycycline-induced Empty Vector, *WT1* +KTS, *WT1* –KTS, shNTC or shWT1.

Figure 4-37 shows the number of differentially expressed gene within the primary cell experiments. Of interest, the most differentially expressed genes occurred within the context of FLT3-ITD AML and of note, this is the AML subtype within which *WT1* mutations are most commonly found (Papaemmanuil et al., 2016).
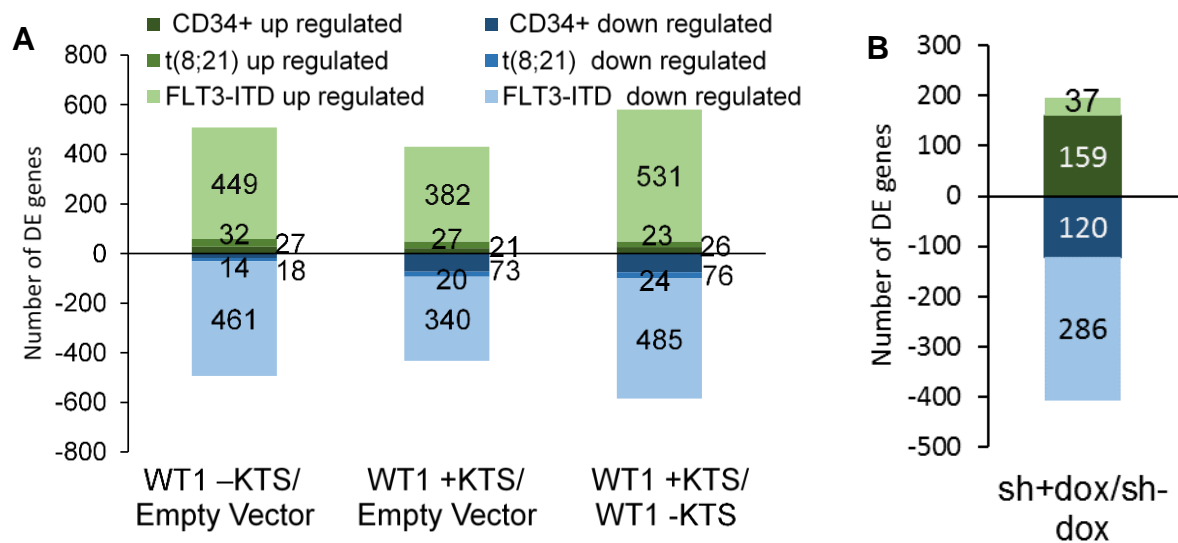
Figure 4-37: Number of genes dysregulated by *WT1* perturbation in primary t(8;21) AML, FLT3-ITD AML and healthy CD34+ cells.

(A-B) Number of differentially expressed genes from RNA-seq experiments of primary t(8;21) AML, primary FLT3-ITD AML or healthy CD34+ cells that were transduced with doxycycline-induced (A) Empty Vector, *WT1* +KTS, *WT1* –KTS or (B) shNTC or sh*WT1*.

4.4 WT1 binds to specific sites in an isoform specific manner

In order to understand the actions of transcription factors, it is essential to know to which genes they bind. Before this study, the only published ChIP-seq experiments in leukaemia examining WT1 binding were performed on a CML blast crisis cell line, K562 (Ullmark et al., 2017). However, after we downloaded this data, we discovered that the quality of data was extremely poor and did not allow any meaningful conclusions to be drawn. We therefore performed WT1 ChIP-seq after expression of several *WT1* isoforms in order to understand genomic binding sites in detail.

4.4.1 WT1 binds to sites distinct from the core transcription factor network characterising t(8;21) AML

In t(8;21) AML, a core set of co-localising transcription factors, which includes the components of the RUNX1-ETO complex, HEB, LDB1, LMO2 as well as PU.1, RUNX1, JUND and C/EBPα make up the t(8;21) specific transcription factor network (Ptasinska et al., 2014). To examine how WT1 binding relates to this network, we performed WT1 ChIP-seq after expression of six of the different WT1 isoforms or an Empty Vector control in Kasumi-1 cells. Hierarchical clustering of the binding sites in these WT1 ChIP-seq experiments showed that they cluster separately to those of the other transcription factors (Figure 4-38). Within the WT1 subcluster, the ChIP seq data clustered according to which KTS isoforms was expressed and independently of alternative splicing of the 17AA exon 5 site or the transcriptional start site used which give rise to long or short isoforms. Since the KTS amino acids are within the Zinc Finger DNA binding domain of WT1 (Laity et al., 2000), this results is therefore consistent with the finding that their presence or absence is the principle determinant of where they bind.
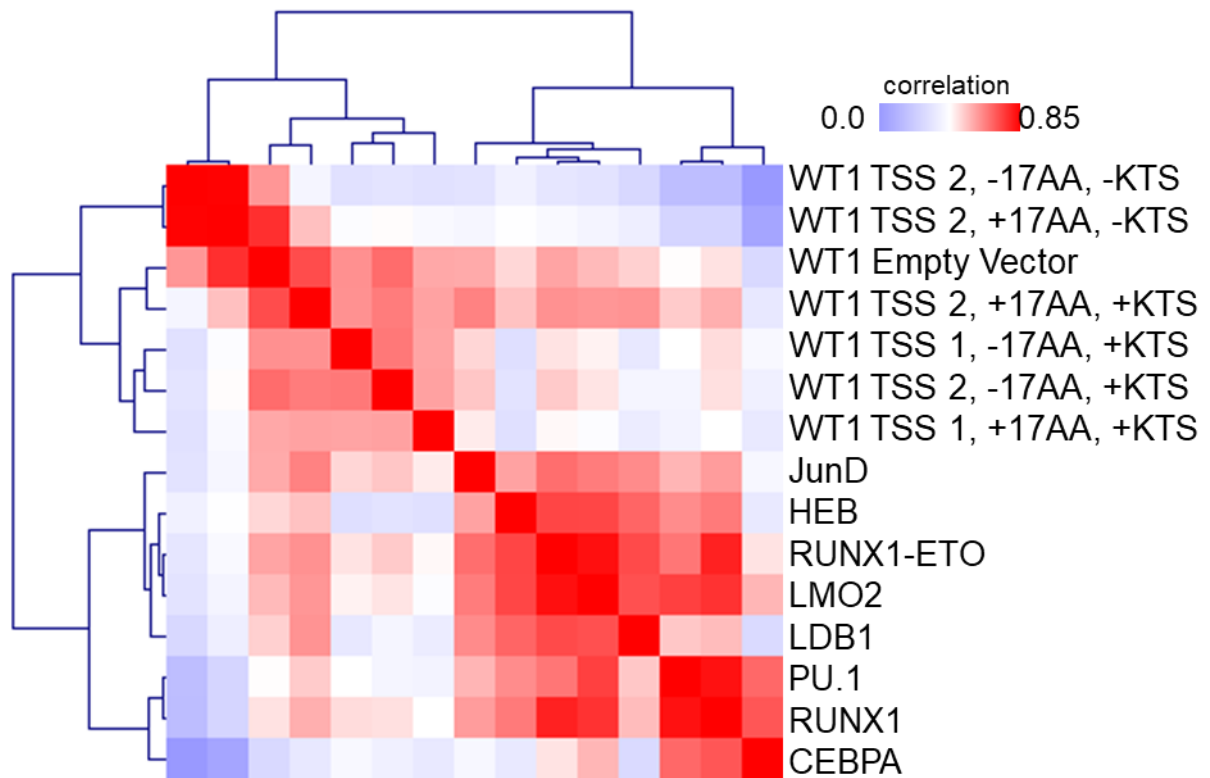
Figure 4-38: WT1 ChIP-seq binding sites cluster in an isoform-specific fashion and away from the RUNX1-ETO complex sites.

Heatmap showing hierarchical clustering of Pearson Correlation Coefficients of WT1 isoform ChIP-seq sequence profiles and the indicated transcription factor ChIP-seq profiles. Some ChIP-seq data from (Ptasinska et al., 2014).

4.4.2 WT1 binds in an isoform specific manner

We next determined the number of overlapping binding sites between the endogenous *WT1* (Empty Vector), the *WT1* –KTS and *WT1* +KTS isoform transduced cells. This analysis shows that *WT1* –KTS transduced cells contain many more and different binding sites compare with *WT1* +KTS transduced cells. Upon expression of the *WT1* –KTS isoform, 9744 new WT1 binding sites were present compared to the Empty Vector control, whilst expression of the *WT1*

+KTS isoform gave rise to 1648 new binding sites (Figure 4-39A). The general

genomic distribution of these WT1 binding sites was not greatly altered after *WT1*

isoform induction except for a slight redistribution of intronic WT1 binding sites to

intergenic binding sites upon *WT1* +KTS expression (Figure 4-39B).
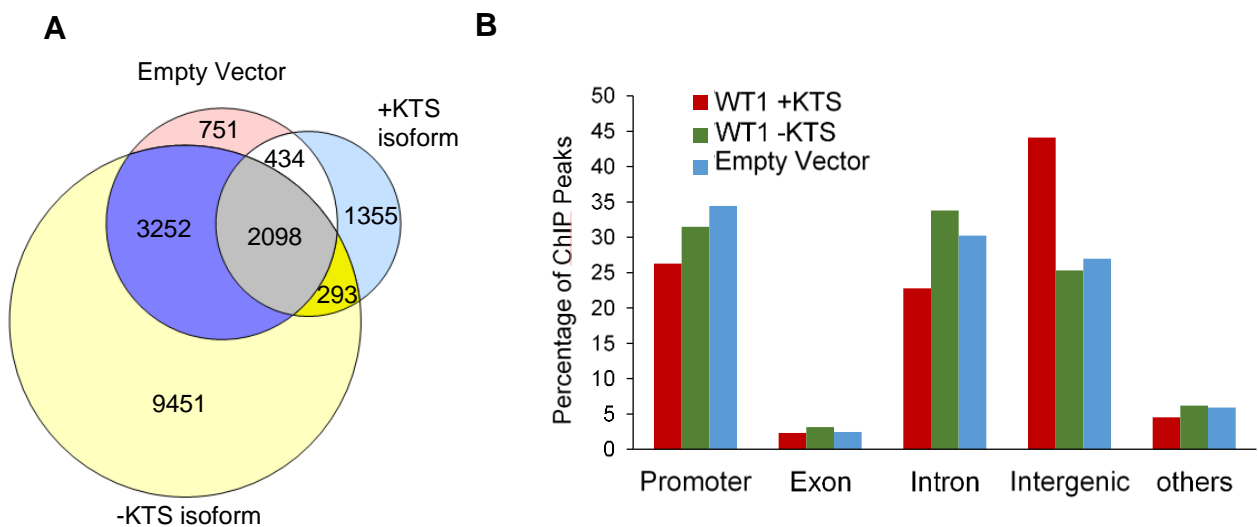
Figure 4-39: Distinct and overlapping binding sites of *WT1* isoforms.

(A) Venn diagram showing the overlap between WT1 ChIP-seq binding sites in
Kasumi-1 cells transduced with doxycycline-induced Empty Vector, *WT1* –KTS
or *WT1* +KTS. (B) Histogram of the genomic distribution of WT1 binding sites.

The quality of our data allowed us to determine a precise position weight matrix

determining WT1 binding sites. Motif enrichment analysis of WT1 ChIP-seq

peaks of all three cell populations highlighted an EGR (Early Growth Response)

transcription factor like Guanine-rich motif together with ETS, RUNX and CTCF

motifs. Expression of *WT1* +KTS led to a lower motif density of an EGR-like WT1

motif and motif enrichment compared with the Empty Vector control or *WT1* -

KTS. Furthermore, similar to what was seen with the previous DNase I-seq analysis (Section 4.3.3), expression of *WT1* +KTS led to enrichment of ETS motifs rather than PU.1 motifs within WT1 binding sites; this also suggest that *WT1* +KTS may confer a more stem-like phenotype (Figure 4-40 A-C). Analysis of motif enrichment showed that within the Empty Vector control experiments measuring the binding of the endogenous WT1, a WT1 motif was found within 46.77% of binding sites, which rose to 63.35% of sites when *WT1* –KTS was expressed but fell to 27.50% of sites when *WT1* +KTS was expressed. This is consistent with NMR studies of WT1 –KTS being a stronger DNA binder than WT1 +KTS (Laity et al., 2000) (Figure 4-40 A-C).

Detailed examination of the position weight matrix of the WT1 motif by comparison with the EGR motif showed that the WT1 motif contained an invariant Adenine base whilst the EGR motif would contain a Cytosine or Thymine (Figure 4-40D). This suggests that some but not all EGR motifs could be bound by WT1 as well. Intriguingly, the position weight matrix for the WT1 motif did not differ between the WT1 –KTS and WT1 +KTS isoforms. This result suggests that the DNA binding specificity of these WT1 isoforms is not purely determined by the DNA sequence.

A
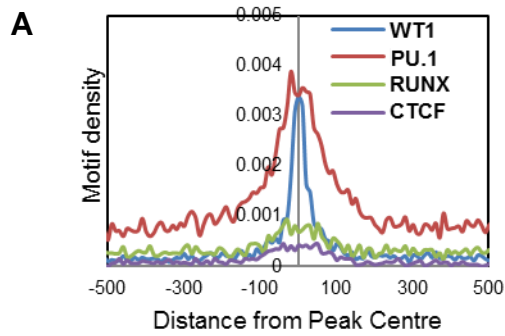
Motif	Match	p-value	% targets
	WT1	1e-641	46.77
	PU.1	1e-379	47.23
TGTGGTTA	RUNX	1e-106	28.83
CACCAGGGGGCG	CTCF	1e-77	6.68
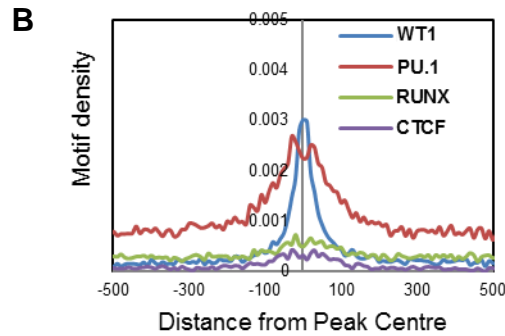
Empty Vector

B

Motif	Match	p-value	% targets
TGTGTGGGAGG	WT1	1e-1701	63.35
ACAGGAAGT	PU.1	1e-466	41.00
TGTGGTTT	RUNX	1e-125	17.27
GCCCCCTGTGG	CTCF	1e-142	5.71

WT1 -KTS

C

Motif	Match	p-value	% targets
AACAGGAAGT	ETS	1e-187	46.89
TGTGGTTA	RUNX	1e-45	27.57
TGCGTGGGAG	WT1	1e-159	27.50
CGCCCCCTGTG	CTCF	1e-60	8.05

WT1 +KTS

D

TGCGTGGGCGG EGR1 Motif

TGCGTGGGAGG WT1 Motif
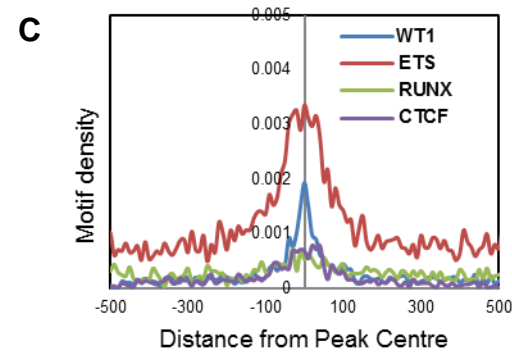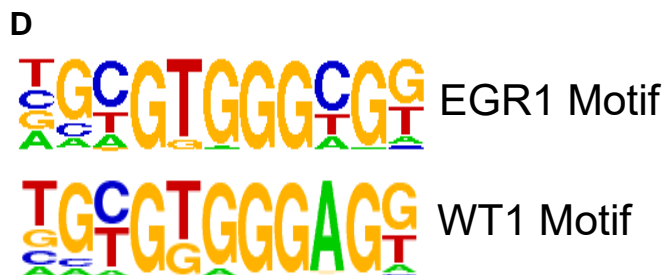
Figure 4-40: Identification of a precise WT1 Position Weight Matrix that differs from EGR.

(A-C) Average profile plots of motif density relative to distance from ChIP peak centre in Kasumi-1 cells transduced with doxycycline-induced (A) Empty Vector, (B) *WT1* –KTS or (C) *WT1* +KTS. Table of the frequencies of enriched TF binding motifs is shown alongside. (D) De novo generated Position Weight Matrix

underlying WT1 binding together with that of EGR1 based upon ChIP-seq experiments.

4.4.3 WT1 is associated with activating histone modifications

In order to determine what action WT1 was having upon the chromatin landscape, we ranked the Empty Vector WT1 ChIP binding sites from the biggest to the smallest peak and correlated it to Histone modification ChIP-seq (Figure 4-41A). The strongest WT1 ChIP peaks appeared to correlate with greater p300 and RNA polymerase II binding as well as the activating histone modifications H3K9Ac and H3K27Ac but not the repressive H3K27me3 histone mark.

Furthermore, the strongest WT1 binding sites correlated well with other Zinc Finger transcription factor ChIP-seq binding sites including EGR1 and Sp1 (see also Section 4.5) but not the core transcription factors of t(8;21) AML (Ptasinska et al., 2014) (Figure 4-41A). Examination on an isoform specific level show that the core transcription factors binding is correlated better with WT1 +KTS compared with WT1 –KTS (Figure 4-41B). This may be because cells with *WT1* –KTS isoform expression are more differentiated as discussed in Section 4.3.2. Further confirmatory evidence comes from the presence of moderately increased CEBPA binding with WT1 –KTS binding sites as well as the more obvious WT1 +KTS correlation (Figure 4-41B). CEBPA is involved in both stem cell and differentiated cell states (Zhang et al., 1997a).
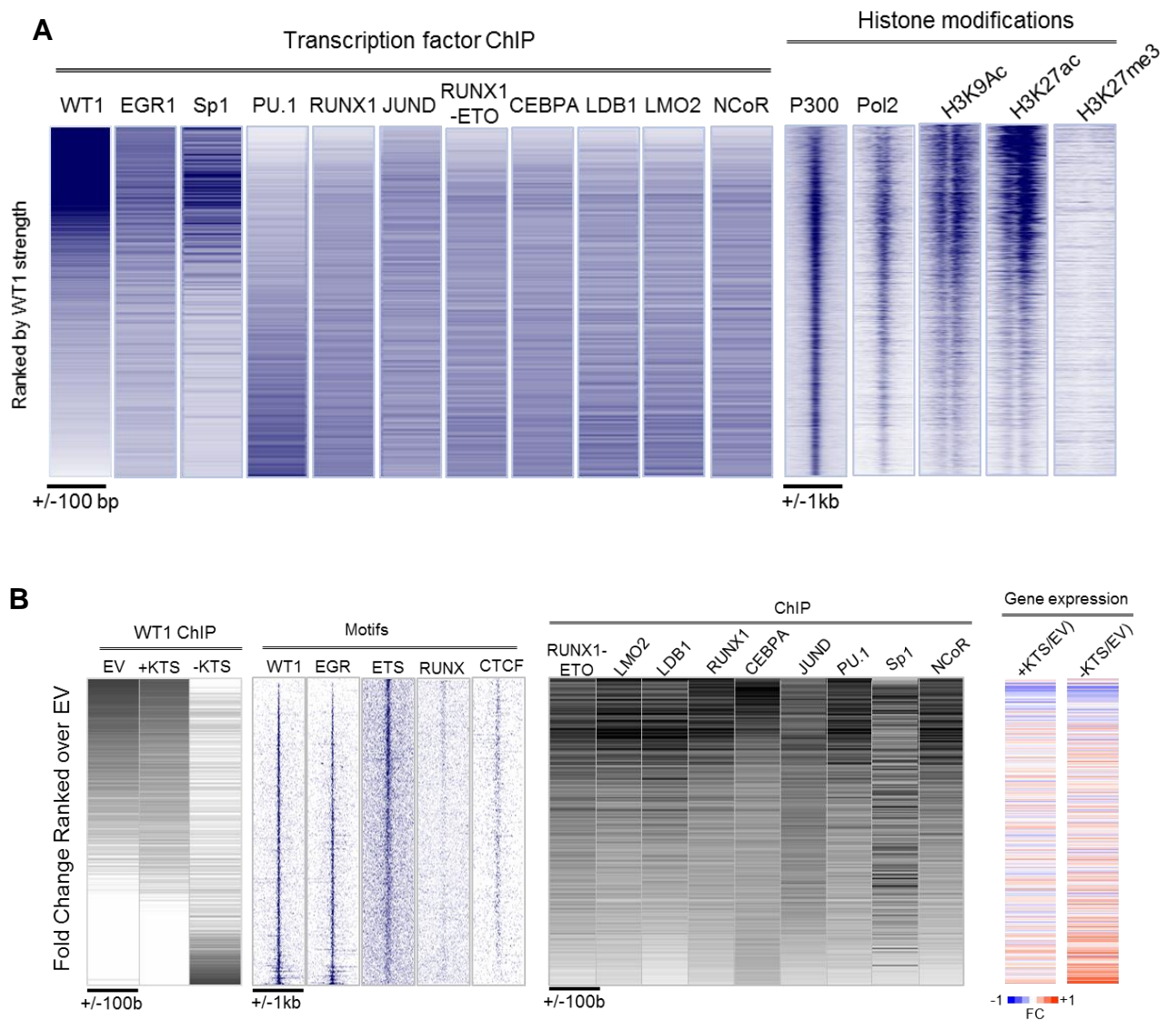
Figure 4-41: WT1 binding sites are associated with activation histone marks and are associated increased gene expression.

(A) Density plots of the indicated transcription factor and histone modification ChIP-seq peaks in Kasumi-1 cells within a 200 bp window, ranked by tag count of WT1 peak strength going from top to bottom. (B) Comparison of density plots of WT1 ChIP-seq peaks with the indicated transcription factor binding motifs or ChIP-seq peaks in Kasumi-1 cells transduced with doxycycline-induced EV, *WT1* +KTS or *WT1* –KTS within a 200 bp window, ranked by tag count of WT1 peak strength going from top to bottom. Plotted alongside is gene expression relative to the EV control. Some ChIP-seq data from (Ptasinska et al., 2014).

4.4.4 The expression of *WT1* isoforms leads to the establishment of distinct gene regulatory networks

Next, we generated Gene Regulatory Networks (GRNs) from data derived from cells transduced with *WT1* +KTS or *WT1* –KTS isoforms using the methodology discussed in Section 4.3.4 (Figures 4-42 and 4-43). Here, the bottom half of the networks shows the interactions between transcription factor encoding genes based on motif occupancy, whereas the top half shows non-transcription factor genes linked to the transcription factor network.

The WT1 –KTS GRN shows a specific downregulation of *MYC* and upregulation of *IRF1* and *CEBPE* relative to the Empty Vector control, which is not present in the WT1 +KTS GRN. *MYC* is a potent oncogene in leukaemia (Hoffman et al., 2002) and its downregulation is consistent with the abrogation of the leukaemic phenotype with *WT1* –KTS expression. *IRF1* downregulation or mutation is also important in leukaemogenesis (Preisler et al., 2001) and so it upregulation in the WT1 –KTS GRN also suggest reduction in leukaemic behaviour. *CEBPE* is a lineage-specifying transcription factor whereby its upregulation suggests neutrophil differentiation (Yamanaka et al., 1997).

In the WT1 +KTS GRN, AP-1 family members including *FOSL1* and *JDP2* are upregulated and such activated signalling may contribute the growth phenotype seen. In the GRNs of WT1 +KTS and WT1 –KTS, *CBFB* is downregulated; since CBFβ is a vital chaperone protein promoting RUNX1 protein binding (Tahirov et al., 2001).

Furthermore, looking at downstream upregulated genes from a therapeutics perspective, the WT1 +KTS GRN contains CD69 and PIM (Proto-oncogene serine/threonine protein) kinases such as PIM1 and the WT1 –KTS GRN contains the FES kinase, all of which are 'druggable' targets.
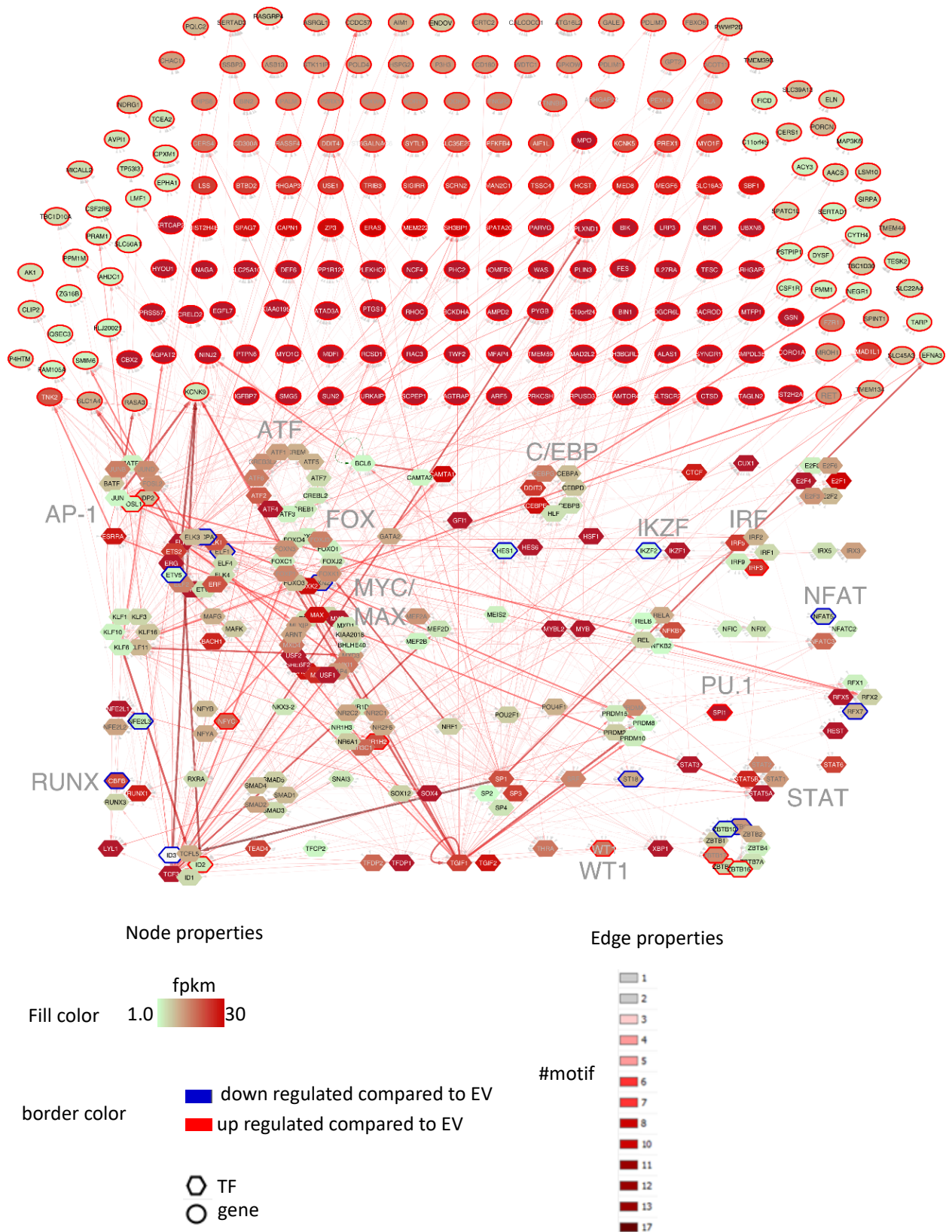
Figure 4-42: Gene regulatory network connected to upregulated WT1 +KTS
targets.

Transcription factor families binding to the same motif form a node, highlighted with large grey text with the family name. Arrows pointing outward from entire node highlight footprinted motifs in individual genes generated by any member of this factor family whereby the footprint was annotated to the gene using the CHiC data where possible, or otherwise to the nearest gene. The colour of the Edge (arrow) indicates the number of footprinted transcription factor motifs present in the target transcription factor gene. The expression level (FKPM) of the individual genes is depicted in white (low) or red (high) color. The border colour indicates whether the gene is upregulated or downregulated compared to the Empty Vector (EV) control. Network constructed by Dr Salam Assi.

Node properties

fpkm

Fill color   1.0 ▭ 30

border color   ▮ down regulated compared to EV
                ▮ up regulated compared to EV

                ⬡ TF
                ◯ gene

Edge properties

#motif

▭ 1
▭ 2
▭ 3
▭ 4
▭ 5
▭ 6
▭ 7
▭ 8
▭ 10
▭ 11
▭ 12
▭ 13
▭ 17

Transcription factor families binding to the same motif form a node, highlighted with large grey text with the family name. Arrows pointing outward from entire node highlight footprinted motifs in individual genes generated by any member of this factor family whereby the footprint was annotated to the gene using the CHiC data where possible, or otherwise to the nearest gene. The colour of the Edge (arrow) indicates the number of footprinted transcription factor motifs present in the target transcription factor gene. The expression level (FKPM) of the individual genes is depicted in white (low) or red (high) color. The border colour indicates whether the gene is upregulated or downregulated compared to the Empty Vector (EV) control. Network constructed by Dr Salam Assi.

4.4.5 RUNX1-ETO binding is altered by *WT1* isoform expression

In Section 4.3.2, we discussed how RUNX1-ETO and WT1 isoforms may co-operate to alter myeloid differentiation. In order to understand the mechanism by which this may occur, we performed RUNX1-ETO ChIP-seq upon expression of *WT1* isoforms. When the RUNX1-ETO ChIP-seq peaks were ranked by the Empty Vector against the *WT1* +KTS isoform, a redistribution of RUNX1-ETO was seen to sites with stronger RUNX, AP-1 and CEBP motif presence and away from TCF3 (Figure 4-44A). This shows that WT1 may be altering the chromatin landscape upon which RUNX1-ETO binds and hence changes RUNX1-ETO mediated gene expression. ChIP-qPCR analysis of RUNX1-ETO binding at several loci shows that in RUNX1-ETO binding site such as the S*PI1* (PU.1) enhancer, *LAT2* enhancer, *IGFBP7* promoter, *CSF1R* enhancer and *WT1* intron 3 enhancer, *WT1* +KTS expression increases RUNX1-ETO binding (Figure 4-44B). Enhanced RUNX1-ETO binding at such target sites may explain why *WT1* +KTS expression maintains a stem cell phenotype.
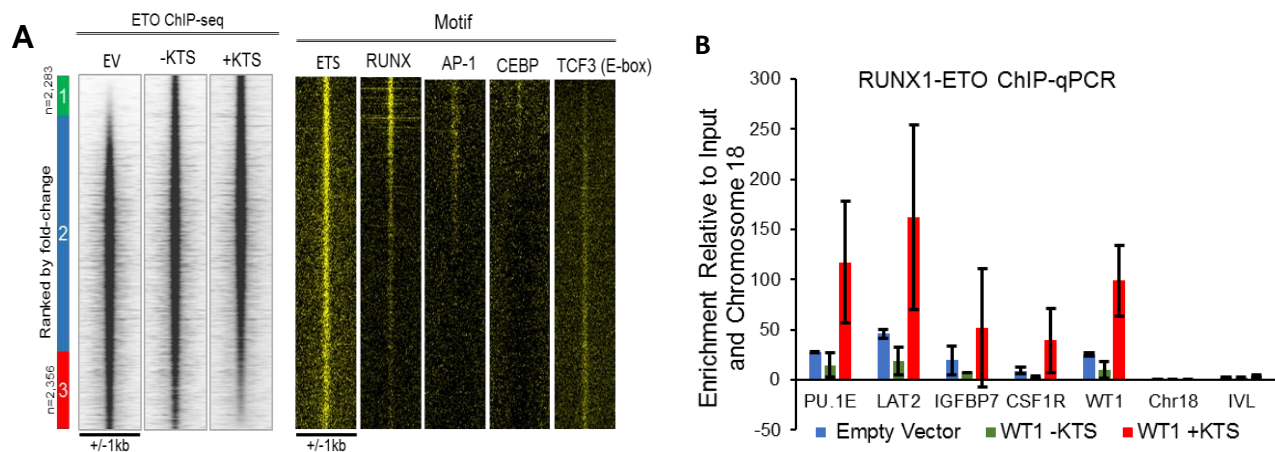
Figure 4-44: WT1 isoform expression alters RUNX1-ETO binding.

(A) Density plots of RUNX1-ETO ChIP-seq peaks in a 2kb window, ranked top to bottom by relative tag count of peaks in *WT1* +KTS relative to Empty Vector control in Kasumi-1 cells, together with TF binding motifs occurring at these sites. The bars on the side depict the number of specific and shared peaks. (B) RUNX1-ETO ChIP-qPCR of Kasumi-1 cells transduced with doxycycline-induced Empty Vector, *WT1* –KTS and *WT1* +KTS. n=3 and error bars signify standard deviation. ChIP-seq performed by Dr Paulynn Chin.

In order to assess what effects RUNX1-ETO may reciprocally have on *WT1*, we performed WT1 ChIP-seq with and without knockdown of *RUNX1-ETO* using doxycycline-induced shRNA. Density plots comparing the two experimental conditions showed that upon *RUNX1-ETO* knockdown, WT1 was redistributed away from RUNX1 sites but that WT1 motif enrichment at these sites was reduced and that there might be more indirect binding occurring through other factors (Figure 4-45).
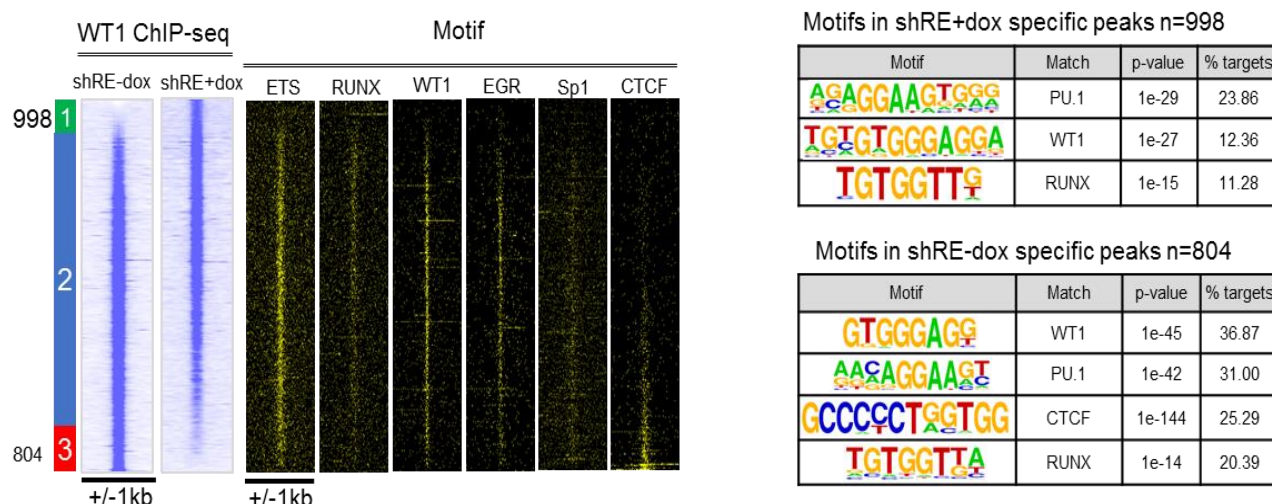
Figure 4-45: *RUNX1-ETO* knockdown alters WT1 binding.

Density plots of WT1 ChIP-seq peaks in a 2kb window, ranked top to bottom by relative tag count of peaks in doxycycline-induced Kasumi-1 cells expressing shRNA against *RUNX1-ETO* relative to the non-induced control in Kasumi-1 cells. Plotted alongside are enriched transcription factor binding motifs and tables of these enriched motifs.

4.5 Interaction between Zinc finger transcription factors

WT1 is a Zinc Finger transcription factor but several others Zinc Finger TFs are also important in haematopoiesis including EGR1 (Min et al., 2008), KLFs (Bialkowska et al., 2017) and Sp1 (Gilmour et al., 2019). Moreover, their binding sites share large similarities raising the possibility that they regulate each other's binding.

4.5.1 Similarities and differences in Sp1, WT1 and EGR1 binding

In order to compare the binding sites of these Zinc Finger TFs, we performed motif co-localisation analysis within 50bp of WT1 ChIP-seq sites. We found that the WT1 motif was most co-localised with EGR which is unsurprising given the

similarity of their Position Weight Matrices (Section 4.4.2). WT1 peaks also clustered closely to the other Zinc Finger factor motifs for KLF and Sp family members (Figure 4-46A). Increased co-localisation was also seen with ETS motifs but not with any other TF motifs.

An extended motif search within WT1 and EGR1 ChIP-seq peaks showed the presence of a composite motif between the Sp family and WT1/EGR1 (Figure 4-46B). This may explain the high degree of co-localisation seen between these factors and they may potentially co-operate or co-bind as a complex. Furthermore, analysis looking at either peaks from WT1, Sp1 and EGR1 ChIP-seq experiments (Figure 4-46C) or looking at motif density in open chromatin sites (Figure 4-46D) shows good co-localisation of all of these factors at WT1 binding sites. Quantification of the degree of overlap between these TF binding sites showed that there were 971 common binding sites between WT1 and Sp1 (Figure 4-46E) whilst there were 2477 common binding sites between WT1 and EGR1 (Figure 4-46F).

A

B

C

D
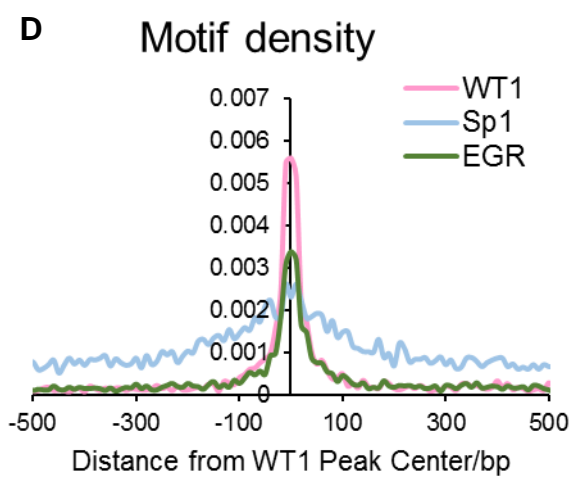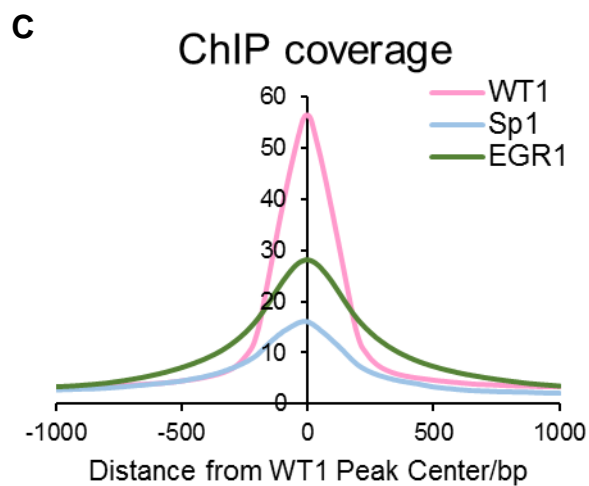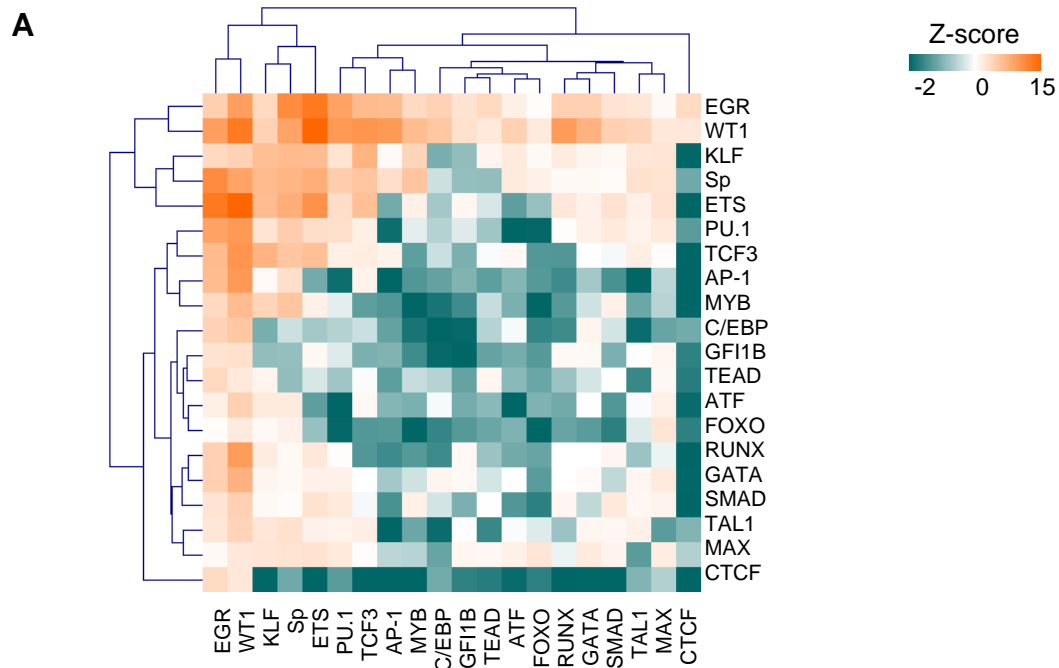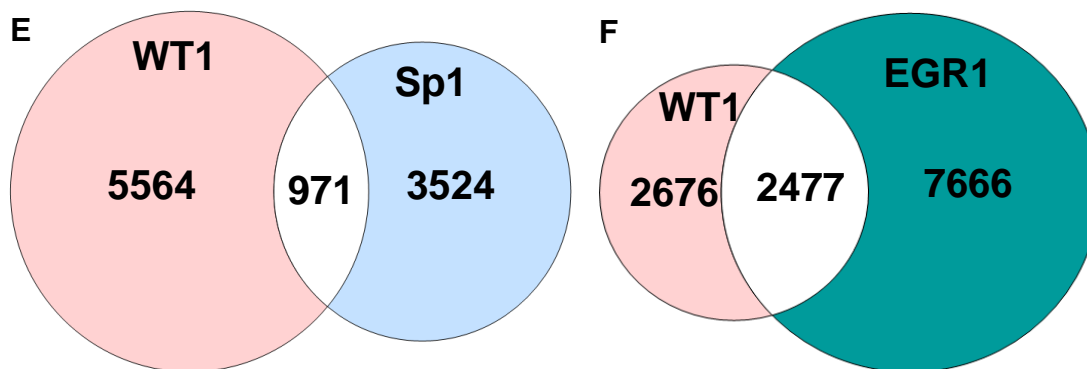
Figure 4-46: The Zinc Fingers transcription factors WT1, Sp1 and EGR1 have common and distinct binding sites.

(A) Bootstrapping analysis clustering the significance of co-localisation of the indicated TF binding motifs within 50 bp in WT1 ChIP-seq peaks in Kasumi-1 cells. (B) Extended HOMER motif search in WT1 or EGR1 ChIP-seq experiments showing a composite Sp:WT1/EGR1 motif. (C-D) Average profile plot of (C) ChIP coverage and (D) motif density relative to distance from the WT1 peak centre in Kasumi-1 cells. (E) Venn diagram showing the overlap between WT1 and Sp1 ChIP-seq binding sites in Kasumi-1 cells. (F) Venn diagram showing the overlap between WT1 and EGR1 ChIP-seq binding sites in Kasumi-1 cells. Sp1 ChIP-seq only performed by Dr Anna Pickin.

4.5.2 Strand specific Sp1 binding around EGR1 but not WT1 binding sites

In order to further understand how these Zinc Finger TFs were binding in proximity, we performed strand specific ChIP-seq analyses. The motif frequency of Sp1 on EGR1 ChIP-seq positive strand reads was greater compared to the negative strand (Figure 4-47A), which suggests that both the Sp1 and EGR1 zinc fingers preferentially wrap around the same strand of DNA. This preferential strand-specific co-localisation was not seen for other co-localised transcription factors such as ETS (Figure 4-47B).

When performing the same analysis for WT1 ChIP-seq experiments, WT1 and Sp1 did not appear to co-localise in a strand-specific manner (Figure 4-47C) nor did WT1 co-localise with ETS in a strand specific manner (Figure 4-47D). Therefore, strand specific binding appeared to be a unique to Sp1 and EGR1.
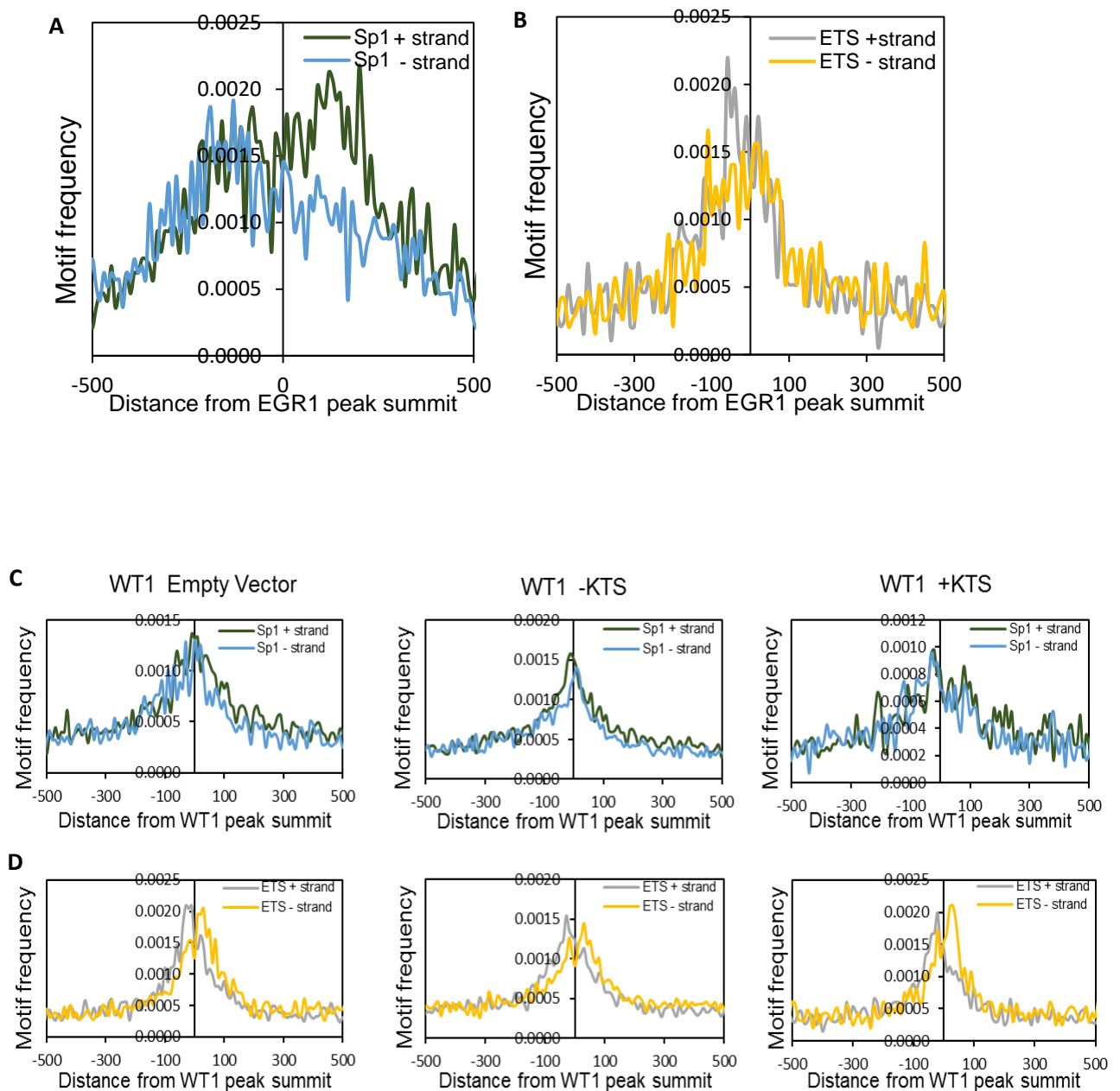
Figure 4-47: Sp1 binds in a strands specific manner around EGR1 but not WT1.

(A-B) Average profile plot of strand-specific (A) Sp1 and (B) ETS motif frequency relative to the distance from the EGR1 ChIP-seq peak summit in Kasumi-1 cells. (C-D) Average profile plot of strand-specific (C) Sp1 and (D) ETS motif frequency relative to distance from the WT1 ChIP-seq peak summit in Kasumi-1 cells transduced with Empty Vector, *WT1* –KTS or *WT1* +KTS.

4.5.3 Interplay between WT1 and EGR1 binding

Next, we wanted to understand how WT1 and EGR1 binding may affect one another given that their Position Weight Matrices are similar. We performed EGR1 ChIP-seq on Kasumi-1 cells transduced with an Empty Vector control or *WT1* isoforms and ranked tag count of the *WT1* –KTS against the Empty Vector (Figure 4-48A). Induction of either *WT1* isoform led to loss of a large number (14569) of EGR1 binding sites, i.e. more than 1/3rd of all sites. The sites that were lost appeared to be those at which Sp1 binding was strongest which suggests that WT1 may be able to displace the EGR1-Sp1 complex alluded to in the earlier co-localisation analysis (Figure 4-47A).

We also performed the reciprocal experiment whereby *WT1* was knocked down using shRNA and EGR1 ChIP-seq was performed. Knockdown of *WT1* led to 1533 new EGR1 binding sites as expected (Figure 4-48B). However, a loss of 1930 EGR1 binding sites was also seen and in particular at strong Sp1 binding sites. This result suggests that the relationship between WT1 and EGR1 is complex. Further to this finding, average profile analysis showed that whilst WT1 isoform expression did reduce EGR1 binding globally (Figure 4-48C), *WT1* knockdown also reduced EGR1 binding globally (Figure 4-48D) which again suggests a complex relationship between EGR1 and WT1.

A

EGR1 ChIP-seq

EV -KTS +KTS

Motif

ETS RUNX EGR Sp1 ChIP

Ranked by fold-change

289
1
2
3
14,569

+/-1kb   +/-1kb

Motifs in -KTS specific peaks n = 289

Motif	Match	p-value	% of Targets
(logo)	EGR	1e-37	48.15
(logo)	PU.1	1e-21	18.98

Motifs in EV specific peaks n = 14569

Motif	Match	p-value	% of Targets
(logo)	ETS	1e-416	21.54
(logo)	RUNX	1e-61	3.55
(logo)	EGR	1e-81	10.93

B

EGR1 ChIP

shWT1 -dox shWT1 +dox

Motifs

ETS RUNX Sp1 EGR

ChIP-seq

WT1 Sp1 RUNX1 PU.1

Ranked by fold-change

1533
1930

+/-1kb   +/-1kb   +/-1kb

258

Figure 4-48: _WT1_ isoform expression or knockdown affects EGR1 binding in a complex manner.

(A-B) Density plots of EGR1 ChIP-seq peaks within a 2kb window, ranked top to bottom by relative tag count of peaks in (A) doxycycline-induced Kasumi-1 cells transduces with the _WT1_ –KTS construct relative to Empty Vector control or (B) doxycycline-induced shRNA against WT1 relative to non-induced control. The bars on the side depict the number of specific and shared peaks. Plotted alongside are enriched motifs. (C-D) Average profile plots of EGR1 ChIP coverage in (C) Kasumi-1 cells transduced with Empty Vector, _WT1_ –KTS or _WT1_ +KTS or (D) Kasumi-1 cells transduced with sh_WT1_ +/- doxycycline. EGR1 ChIP-seq performed by Dr Paulynn Chin.

4.5.4 Example sites demonstrating the relationship between WT1 and EGR1 binding

When we examined several loci for example an _ELK3_ upstream cis-regulatory element (Figure 4-49A) and the _ART1_ intragenic element (Figure 4-49B), _WT1_ knockdown led to increased EGR1 binding. Upon expression of the _WT1_ –KTS isoform, a loss of EGR1 binding was seen at several loci including the _SOCS3_ downstream cis-regulatory element (Figure 4-50A) and the _CASP9_ downstream element (Figure 4-50B).
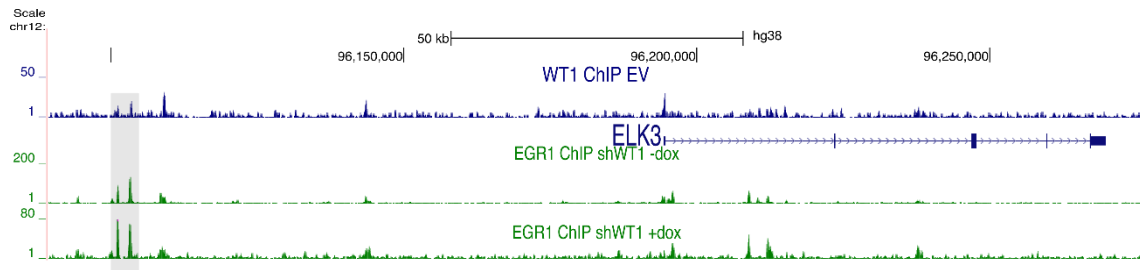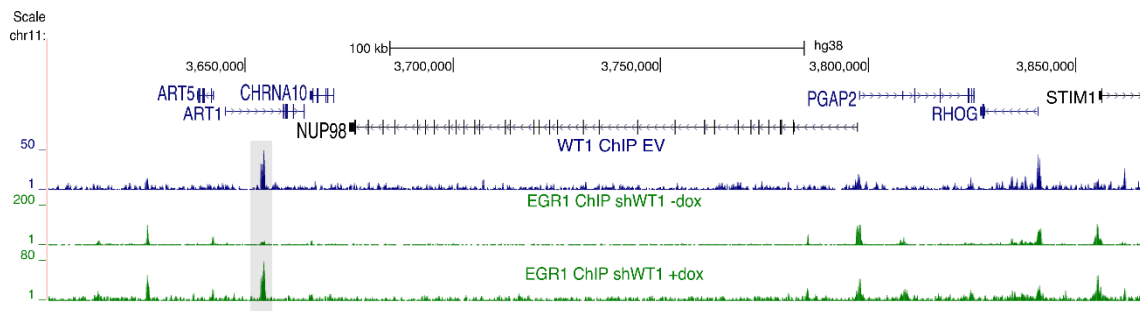
Figure 4-49: Knockdown of *WT1* increases EGR1 binding at several cis-regulatory elements.

(A-B) UCSC genome browser screenshots depicting WT1 ChIP and EGR1 ChIP peaks at the (A) *ELK3* locus or (B) *ART1* locus with and without *WT1* knockdown.
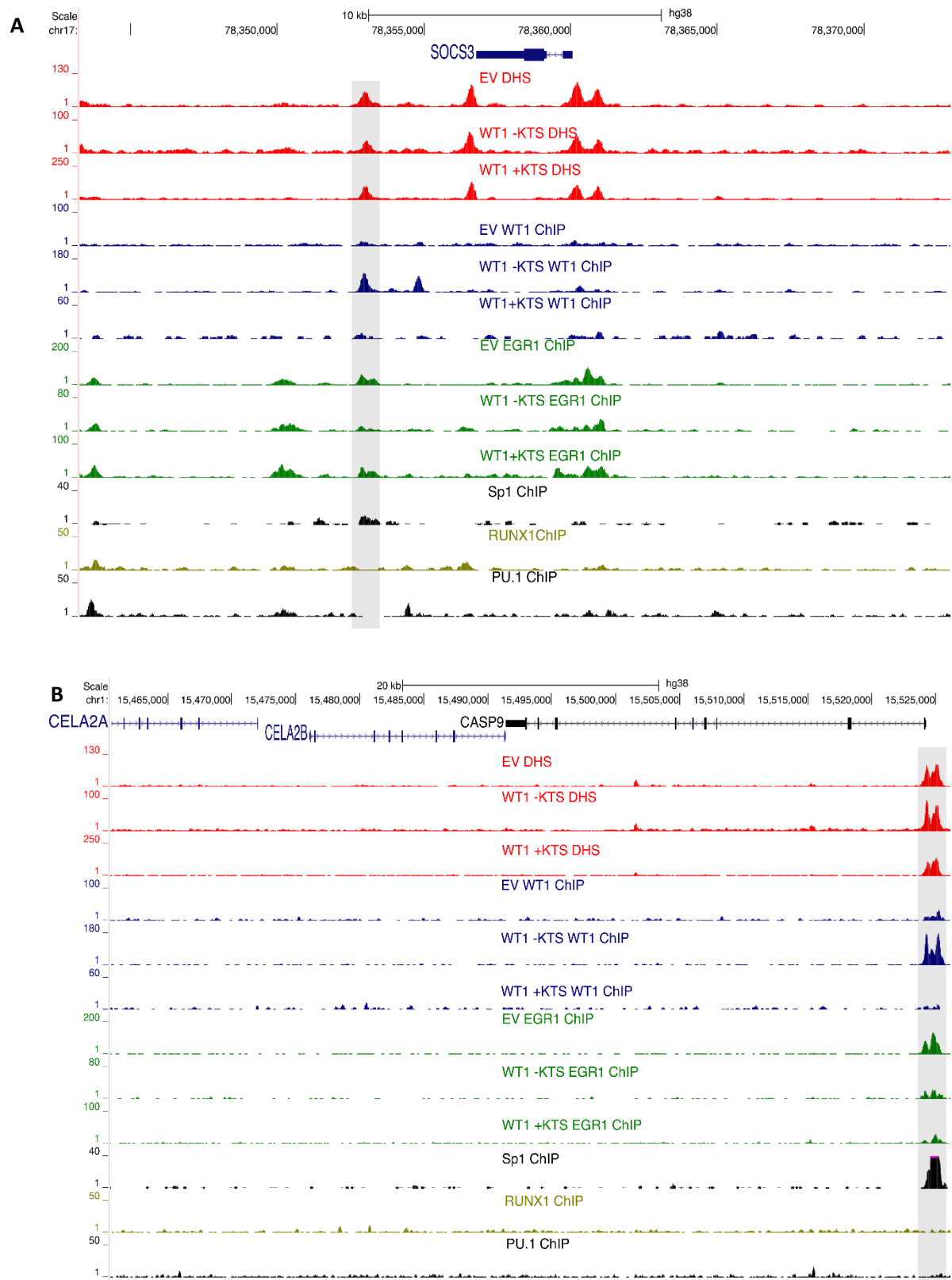
Figure 4-50: Expression of the *WT1* –KTS isoforms displaces EGR1 at some cis-regulatory elements.

(A-B) UCSC genome browser screenshots depicting DNase I-seq, WT1 ChIP, EGR1 ChIP, Sp1 ChIP, RUNX1 ChIP and PU.1 ChIP peaks at the (A) *SOCS3* locus and (B) *CASP9* locus.

4.5.5 WT1 and EGR1 modules

Using the specific motifs derived from WT1 and EGR1 ChIP-seq experiments together with our digital footprinting and Capture HiC data, we examined their specific regulatory modules in our primary AML data. Modules were created identifying WT1 or EGR1 transcription factor binding motifs in distal cis-regulatory elements of genes upregulated in primary AML compared to healthy CD34⁺ control cells.

Initially focusing on t(8;21) AML, the presence of occupied WT1 binding sites is correlated with the upregulation of AP-1 family members such as *JUNB*, *FOSL2* and *JDP2* as well as growth receptors such as *FGFR1* suggesting activated signalling (Figure 4-51A). The presence of EGR1 occupied motifs is correlated with the upregulation of signalling genes such as *MAPKAPK2* and other signalling including NFκB signalling (*NFKBIB)*. Both WT1 and EGR1 also regulate key t(8;21) genes such as *UBASH3B*, an ubiquitin ligase (Goyama et al., 2016).

Next, we identified the WT1 and EGR1 bound modules in FLT3-ITD (and NPM1) mutated AML. The presence of occupied WT1 motifs correlated with upregulation of *JUND* and *FOSL2* (Figure 4-51B). The presence of occupies EGR1 motifs correlated with the upregulation of key transcription factors genes in FLT3-ITD leukaemogenesis including *NFIX* and *FOXC1* (Assi et al., 2019).

Figure 4-51: Upregulated Gene Regulatory Network linked to WT1/ EGR1 binding sites in primary AML cells.

WT1 and EGR1 linked to their AML-specific upregulated gene targets as determined by the presence of occupied WT1 motifs in (A) primary t(8;21) AML and (B) primary FLT3-ITD AML (Assi et al., 2019). Network constructed together with Dr Daniel Coleman.

4.6 Functional analysis of recurrent *WT1* mutations

4.6.1 *WT1* exon 8 mutations increase leukaemic growth and self-renewal

WT1 mutations occur in around 10% of all AML (King-Underwood et al., 1996) with mutations most commonly in exon 7, 8 or 9 which harbour different Zinc Finger domains. In order to investigate what the role of *WT1* mutations may be, we created doxycycline-inducible *WT1* lentiviral constructs carrying frameshift mutations in exons 7, 8 and 9 and transduced them into Kasumi-1 cells. Figure 4-52A shows a schematic of where these frameshift mutations were created.

Induction of the *WT1* exon 8 mutation led to increased cell growth whereas induction of the *WT1* exon 7 or exon 9 mutations or of the Empty Vector control had no growth altering effect (Figure 4-52B). The *WT1* exon 8 mutation also led to increased colony formation ability (Figure 4-52C) and decreased apoptosis (Figure 4-52D) whereas the other mutations or Empty Vector control did not have an effect. Lastly, none of the mutations had an effect upon the cell cycle (Figure 4-52E).

A

Ex7 Ex7 Ex8 Ex9
mut mut mut mut
A B

Transcription Start Site: 0　1　2

Base #:
-219 CTG　1 ATG　378 ATG　750-800　801-896　897-1049　1050-1139　1140-1223　1233-1346

E6　E7　E8　E9　E10

Amino Acid #:
-73　1　127　250-266　267-298　299-349　350-379　380-407　411-448

Protein Weight/ KDa:
60-62　52-54　36-38

B

Empty Vector
Exon 7 Mutation
Exon 8 Mutation *
Exon 9 Mutation

-dox
+dox

Cumulative Cell Count/10^6

Time/ days

265
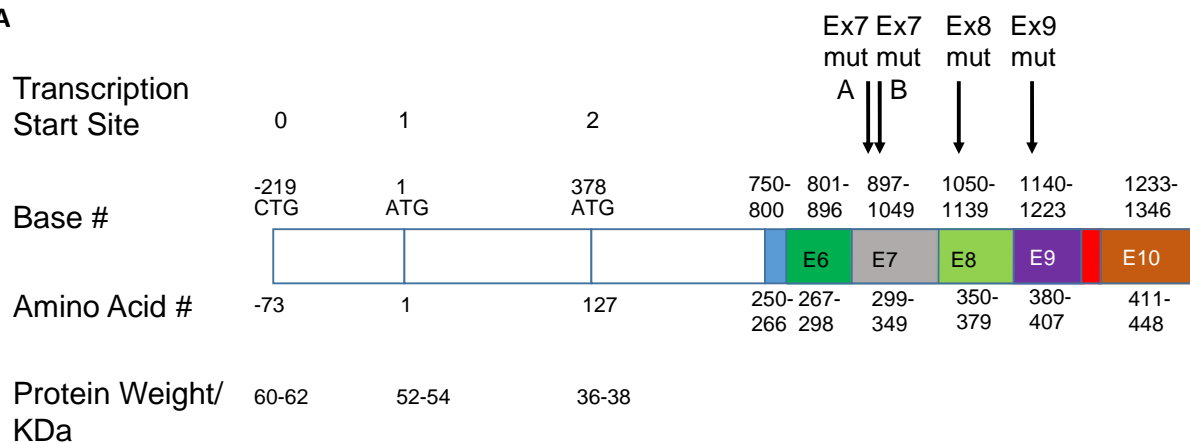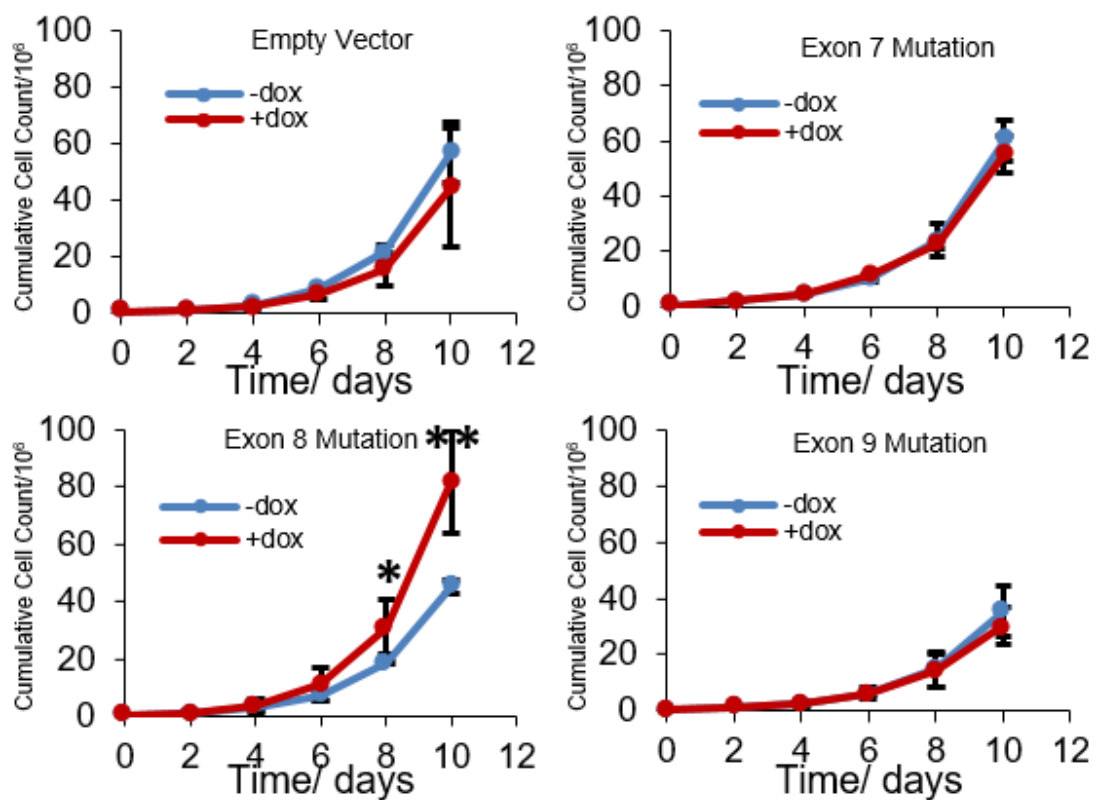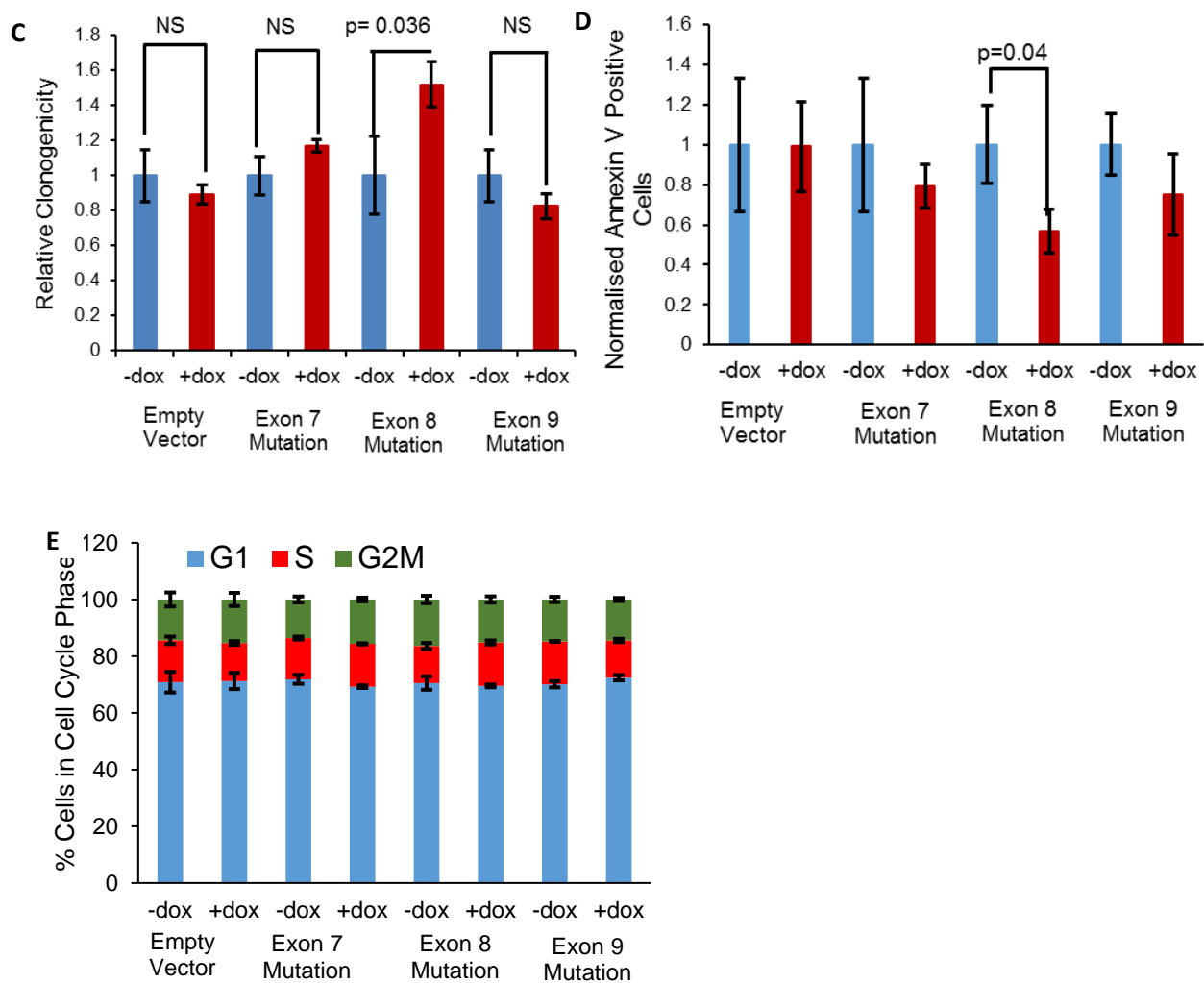
Figure 4-52: *WT1* mutations alter cell growth, colony formation and apoptosis in a mutation site-specific manner.

(A) Schematic showing where frameshift mutations were introduced in *WT1*. (B) Time course of cumulative cell growth of Kasumi-1 cells with or without doxycycline induction of *WT1* carrying frameshift mutations in different exons. (C-E) Kasumi-1 cells transduced with doxycycline-induced *WT1* carrying frameshift mutations in different exons examined for (C) colony formation ability of 2000 cells seeded in methylcellulose, with data normalised to the Empty Vector control, (D) number of annexin V positive cells (data normalised to Empty Vector control) or (E) cell cycle phase. For all panels, n=3 and error bars show standard deviation. Significance was determined using a 2 tailed student's t-test. * indicates p-value <0.05, ** indicates p-value <0.005.

4.6.2 Binding activity of *WT1* exonic mutations

We were intrigued by the finding that only induction of the exon 8 frameshift mutation had a phenotypic effect and not the other mutations. We performed WT1 ChIP-seq on Kasumi-1 cells after induction of the exon 8 mutated *WT1*, which gave very few peaks whilst an Empty Vector control WT1 ChIP performed in parallel gave good enrichment at expected genomic sites (Figure 4-53); this result suggested that the exon 8 mutation could act as a dominant negative to WT1 and perhaps other Zinc Finger TFs which have very similar binding sites (Section 4.5.3).

We had expected that the exon 7 mutation would also have a phenotypic effect but were surprised to find that induction of this construct did not lead to the formation of a mutant protein in either of two different constructs with *WT1* mutations in different places in exon 7, whereas induction of exon 8 and exon 9 mutants did give rise to a truncated protein (Figure 4-54A). In order to confirm that protein could be produced, we transfected HEK 293T cells with the plasmids and a mutant protein could be seen (Figure 4-54B). Furthermore, in primary patient AML cells with a heterozygous exon 7 *WT1* mutation, again a mutant protein could not be seen. This suggested that the exon 7 mutation is unstable within leukaemic cells and we hypothesise that either the mutant *WT1* RNA is degraded by nonsense-mediated decay or that the mutant protein is degraded by the proteasome.
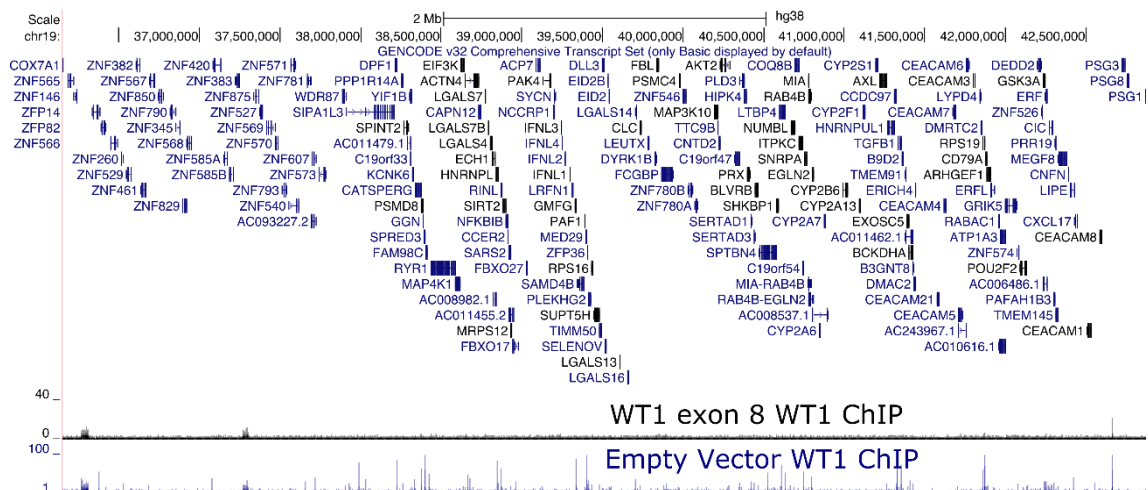
Figure 4-53: The WT1 exon 8 mutant binds to fewer sites than wildtype WT1.

UCSC genome browser screenshot depicting WT1 ChIP-seq on Kasumi-1 cells transduced with either doxycycline-induced *WT1* exon 8 mutant or an Empty Vector control.
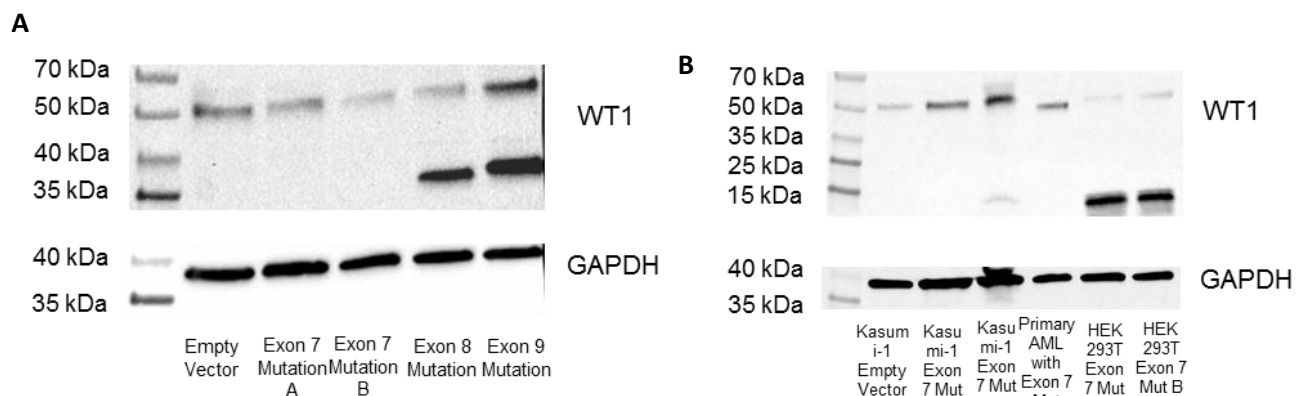
Figure 4-54: *WT1* exon 7 mutants do not give rise to a stable protein in Kasumi-1 cells.

(A-B) Western blots examining WT1 and GAPDH protein levels from (A) Kasumi-1 cells transduced with doxycycline-induced *WT1* carrying mutations in different exons or (B) Kasumi-1, primary AML or HEK 293T cells transduced with Empty Vector or exon 7 mutant *WT1*.

4.7 WT1 and EGR1 differentially respond to oncogenic signalling and Gene Regulatory Network perturbations

In FLT3-ITD AML, activation of signalling is derived from the constitutively active FLT3 receptor which imparts a specific chromatin pattern common to all cells with this mutation (Cauchy et al., 2015) (Assi et al., 2019). We therefore examined how *WT1* expression responded to the perturbation of the signalling and Gene Regulatory Network. To this end, we treated primary FLT3-ITD cells with the FLT3 inhibitor, AC220 (Quizartinib) which led to decreased *WT1* and *EGR1* expression (Figure 4-55A).

RUNX1 is master transcription factor in AML and whilst it is difficult to directly target, inhibiting the CBFβ-RUNX1 complex with a CBFβ inhibitor reduces RUNX1 binding to DNA (Illendula et al., 2016). Using the CBFβ inhibitor on primary FLT3-ITD cells led to a decrease in *WT1* expression but did not affect *EGR1* expression (Figure 4-55B).

As previously discussed, the AP-1 family of transcription factors are key transducers of signalling. In order to block AP-1 binding, we expressed dnFOS in primary FLT3-ITD cells in a doxycycline-induced manner. Induction of the dnFOS peptide led to decreased *WT1* expression but did not alter *EGR1* expression (Figure 4-55C).
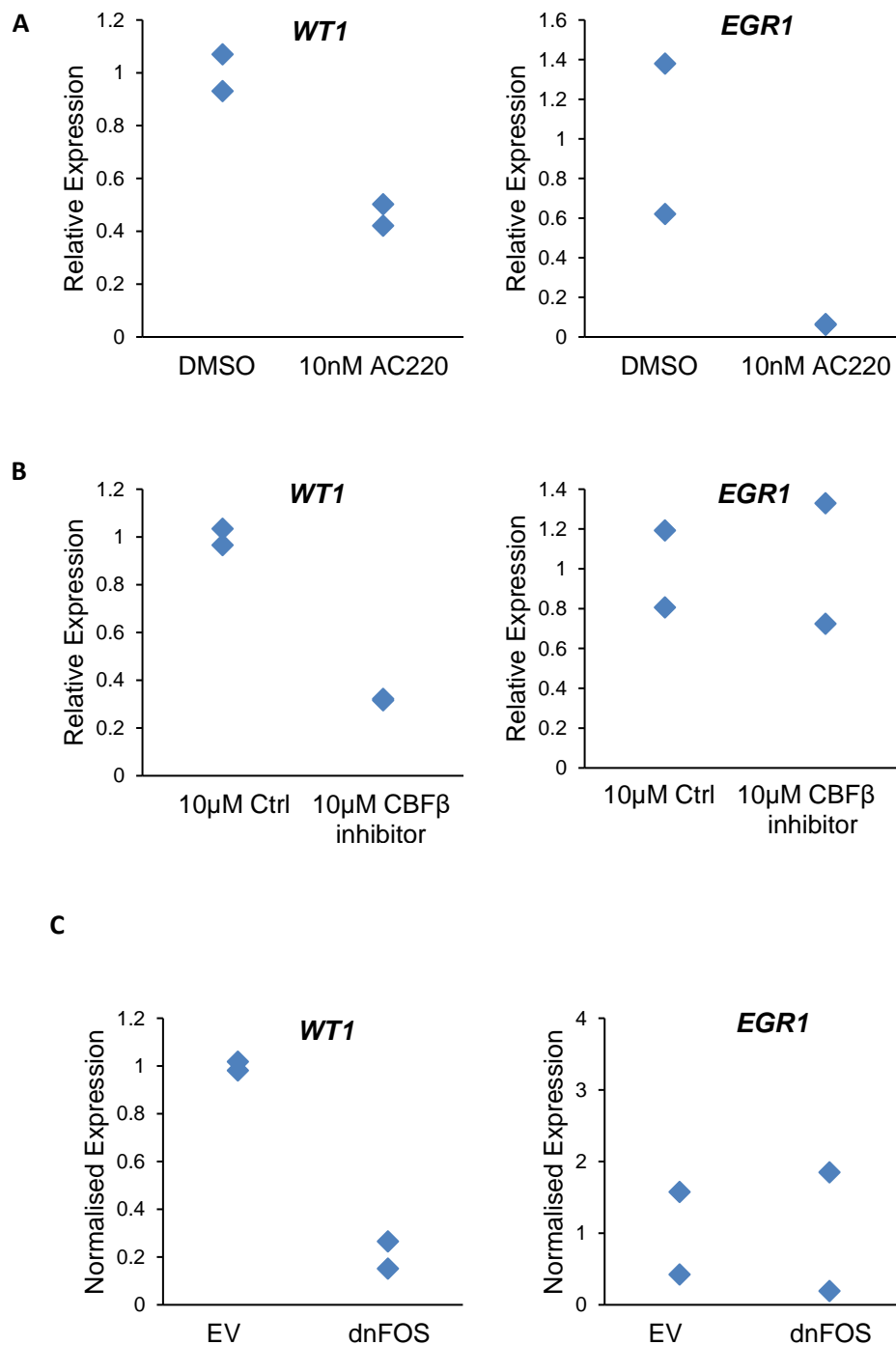
Figure 4-55: Perturbations in signalling affect WT1 and EGR1 expression.

(A) Dot plot of relative gene expression values as determined by RNA-seq of *WT1* and *EGR1* in primary FLT3-ITD cells with treatment with DMSO or the FLT3-ITD inhibitor Quizartinib (AC220). (B) Dot plot of relative gene expression values as determined by RNA-seq of *WT1* and *EGR1* in cultured primary FLT3-ITD cells after treatment with a control compound or the RUNX1-CBFβ inhibitor (Illendula

et al., 2016). (C) Dot plot of relative gene expression of *WT1* and *EGR1* in primary FLT3-ITD cells transduced with doxycycline-induced Empty Vector or Dominant Negative FOS (dnFOS) as measured by RT-PCR. AC220 and RUNX inhibitor experiments performed by Dr Daniel Coleman.

4.8 AP-1 and WT1 Regulate Each Other

The fact that WT1 responds to oncogenic signalling raised the possibility that it may be regulated by AP-1 and that these factors are intricately related and can regulate each other's Gene Regulatory Networks. Induction of dnFOS led to downregulation of *WT1* in primary FLT3-ITD cells but not in primary t(8;21) or healthy CD34[+] cells (Figure 4-56). Similarly, overexpression of either *WT1* +KTS or *WT1* –KTS isoforms led to an increase in *FOSL1* expression in primary FLT3-ITD cells but not in primary t(8;21) or healthy CD34[+] cells (Figure 4-57); this finding again demonstrates that the different subtypes of AML differ in the response of their Gene Regulatory networks to AP-1 and WT1 perturbation (Sections 4.3.4 and 4.3.5).
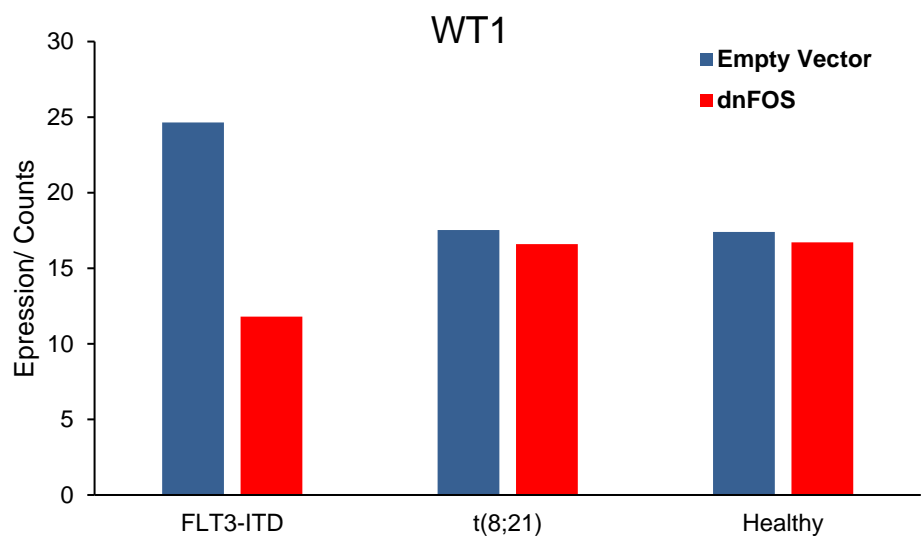
Figure 4-56: *WT1* expression responds to dnFOS induction in FLT3-ITD AML.

WT1 read counts as measured by RNA-seq of primary FLT3-ITD, t(8;21) and healthy CD34$^+$ cells induced and transduced with doxycycline-induced Empty Vector control or dnFOS.
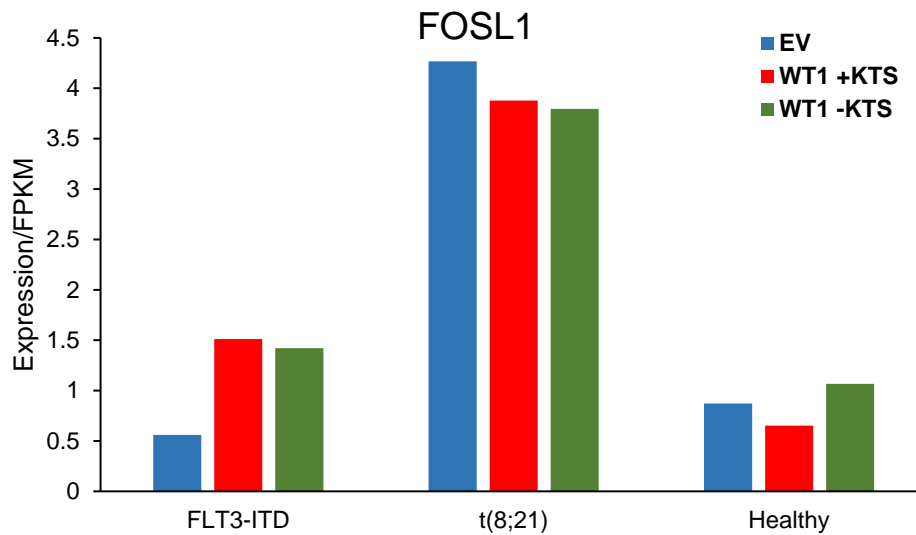
Figure 4-57: *FOSL1* is upregulated by *WT1* expression in primary FLT3-ITD AML.

(A) *FOSL1* expression as measured by RNA-seq of primary FLT3-ITD, t(8;21) and healthy CD34$^+$ cells induced and transduced with doxycycline-induced Empty Vector control or *WT1* +KTS or *WT1* –KTS isoforms.

4.9 Summary

From our WT1 work, our most important conclusions are that:

(i)   *WT1* is upregulated and an important component of AML maintenance within the Gene Regulatory Networks of multiple subtypes of AML but not in healthy stem cells.

(ii)   Overexpression of endogenous *WT1* leads to increased leukaemic growth and clonogenicity whilst knockdown of *WT1* leads to decreased leukaemic growth.

(iii)   WT1 acts in an isoform-specific fashion with WT1 +KTS isoforms increasing leukaemic growth and clonogenicity in contrast to WT1 – KTS isoforms which decrease leukaemic growth through cell cycle inhibition, apoptosis and differentiation.

(iv)   WT1 +KTS and WT1 –KTS isoforms have very distinct binding sites and also differentially regulate RNA splicing, allowing for distinct alterations to the epigenome and transcriptome.

(v)   Certain WT1 mutants leading to alterations in DNA binding can lead to increased leukaemic growth.

(vi)   Our ChIP and gene expression data allow us to propose models of AML sub-type-specific up-regulation of these transcription factors in response to external signals and GRN alterations (see Section 5.1.1-5.1.2).

From our AP-1 work, our most important conclusions are that:

i)   AP-1 is a central node in the Gene Regulatory Networks of multiple subtypes of AML and many of its family members are upregulated in AML compared with healthy stem cells.

ii)   Blocking AP-1 binding through use of a dominant negative FOS peptide led to decreased growth by a G1 cell cycle arrest through regulation of Cyclin D2 as well as leading to aberrant differentiation.

iii)   Our understanding of AP-1 biology may be exploited therapeutically to increase survival of mice in xenotransplantation experiments by either blocking AP-1 binding or through pharmacologically inhibiting one of its downstream genes, Cyclin D2.

Chapter 5: Discussion

5.1 Oncogenic signalling

In this work, we gained novel insights into how several specific transcription factors facilitate oncogenic signalling. We describe their actions at a molecular level as to how they contribute to tumourigenesis and AML fitness as well as identify the target genes that they bind to. We have explored the role of two distinct classes of transcription factors:

1) The AP-1 family of TFs, within the Basic Domain/Leucine Zipper class of TFs.

2) Zinc Finger TFs, focusing mainly on WT1 but also on how WT1 co-operates and competes with two other Zinc Finger TFs EGR1 and Sp1.

5.1.1 WT1 as part of an oncogenic signalling hub

WT1 overexpression is a marker for an aggressive type of AML independent of the AML sub-type. *WT1* is upregulated after AML relapse and is a biomarker for relapse prediction as it can be detected in minimal residual disease (MRD) following treatments such as chemotherapy (Inoue et al., 1994). *WT1* expression in MRD is also a biomarker for AML relapse post-allogeneic stem cell transplantation (Nomdedéu et al., 2018). This work points to elevated growth factor signalling in leukemic cells as a major reason for *WT1* upregulation and also provides an explanation for the worse prognosis in such patients as WT1 itself modulates the gene expression patterns of leukemic cells towards enhanced growth.

Indeed, activation of signalling plays a key role in growth and increasing evolutionary fitness in AML. In an examination of apparently healthy individuals, the detection of mutations in signalling genes such as *KRAS*, *NRAS*, *NF1*, *JAK2* and *CBL* could predict the development of AML around 6 years later (Abelson et al., 2018)*.* In another example, activating *KRAS* mutations have been demonstrated to confer increased competitive survival in xenotransplantation experiments (Burgess et al., 2017). Furthermore, the acquisition of a FLT3-ITD mutation in a *NPM1* mutant background accelerates the relapse time as compared to the *NPM1* mutation alone (Höllein et al., 2018). Finally, *WT1* mutations have been described to co-operate with FLT3-ITD mutations *in vivo* in the development of a fully penetrant AML (Pronier et al., 2018) which again alludes to the clonal advantages conferred by activated signalling.

Integrating all the available ChIP-seq data from this study and others (Ben-Ami et al., 2013, Martinez-Soria et al., 2018, Martens et al., 2012, Ptasinska et al., 2012, Ptasinska et al., 2019) as well as RNA-seq data, we can construct a WT1 and EGR1 centred signalling network in t(8;21) AML (Figure 5-1). In response to signals from cytokines binding to growth receptors, from constitutively active signalling due to signalling mutations e.g. RAS or KIT or due to the RUNX1-ETO driver oncoprotein (Martinez-Soria et al., 2018), the AP-1 family of transcription factors is activated. AP-1 binds to *EGR1* cis-regulatory elements together with WT1, Sp1, RUNX1-ETO, AP4 (E-box factor) and EGR1 (autoregulation). The experiments presented here show that *WT1* knockdown reduces *EGR1* expression (Section 4.1.4). In addition, *EGR1* is an Early Growth Factor gene and

so can respond to signals directly through its Serum Response Element (Clarkson et al., 1999).

WT1 cis-regulatory elements are bound by RUNX1-ETO, AP4, AP-1, RUNX1, PU.1, LMO2, C/EBPα and WT1 itself (some data from (Ptasinska et al., 2014) RUNX1-ETO knockdown leads to decreased WT1 expression (Section 4.2.4). WT1 and EGR1 can compete or otherwise alter each other's binding globally in order to determine the transcriptome of a leukaemic cell (Section 4.5.3).
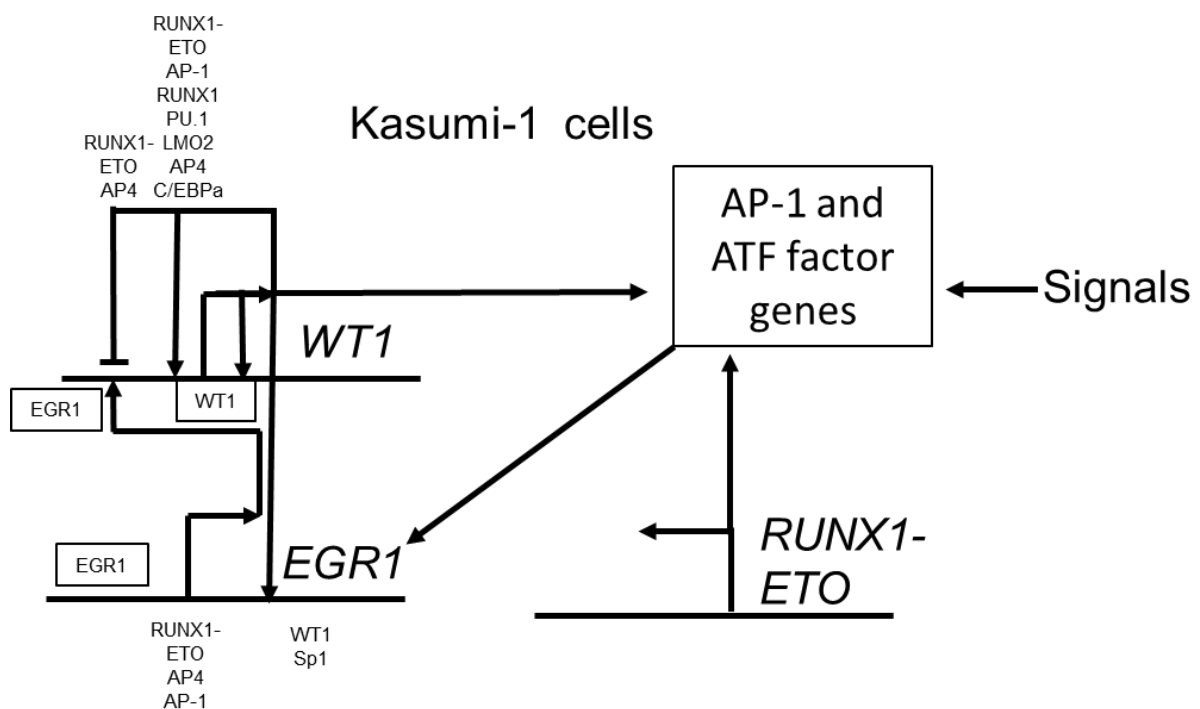
Figure 5-1: Kasumi-1 signalling network

WT1 and EGR1 regulatory sub-network specific for Kasumi-1 cells based on the ChIP from our work and elsewhere (Ben-Ami et al., 2013; Martens et al., 2012; Ptasinska et al., 2014) as well as our perturbation experiments (dnFOS expression, *RUNX1-ETO* knockdown, *WT1* knockdown, *RUNX1* knockdown).

The signalling network feeding into *WT1* and *EGR1* in FLT3-ITD AML is different. Here, *RUNX1* is part of the FLT3-ITD AML gene regulatory network and its inhibition leads to *WT1* down-regulation (see Section 4.7). The expression and activity of *RUNX1* is modulated by the FLT3-ITD signal as it has been shown that FLT3-ITD signalling can activate RUNX1 phosphorylation, which can co-operate with the FLT3-ITD signal to induce leukaemia (Behrens et al., 2017).

Incorporating the perturbation data and ChIP-seq data from (Cauchy et al., 2015), we constructed a FLT3-ITD signalling network (Figure 5-2). FLT3-ITD signalling activates AP-1, STAT5 and EGR1 signalling (Cauchy et al., 2015). EGR1, together with WT1, Sp1, STAT5, AP-1, RUNX1 and itself (autoregulation) can bind to *EGR1* cis-regulatory elements. RUNX1, EGR1, STAT5, AP-1 and WT1 itself can also bind to *WT1* cis-regulatory elements. Consequently, the WT1 and EGR1 factors can regulate gene expression in response to signalling (Section 4.3.5).

FLT3-ITD/NPM1 AML
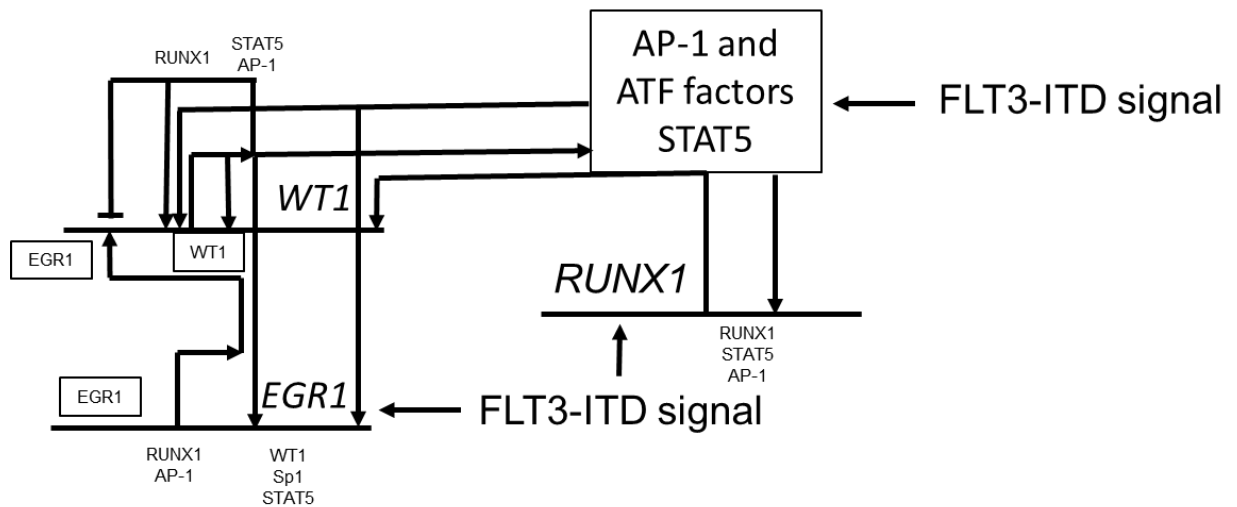
Figure 5-2: Primary FLT3-ITD AML signalling network.

WT1 and EGR1 regulatory sub-network specific for FLT3-ITD/NPM1 mutated AML. Based partly on data from (Assi et al., 2019) and STAT5 ChIP data from (Cauchy et al., 2015).

WT1 appears to act as a transcriptional activator as the strongest WT1 binding sites were correlated with activating histone modifications such as H3K9Ac and H3K27Ac and anti-correlated with repressive histone marks such as H3K27me3. Furthermore, WT1 binding sites were correlated with increased gene expression (see Section 4.4.3). Consequently WT1 is likely to be propagating oncogenic signalling.

On an isoform-specific level, we showed that WT1 +KTS expression increased leukaemic growth whilst WT1 –KTS expression decreased it. Neither the presence of the 17AA splice sequence nor the length of the isoform by employing the first transcriptional start site appeared to make any difference for cellular

growth (see Section 4.3.1) or the distribution of DNA binding sites (see Section 4.4.1). Whilst the extra sequences in the longer TSS 1 isoforms have been previously described as a 'repression domain' because of their interaction with transcriptional repression-associated proteins such as BASP1 (Carpenter et al., 2004) we didn't see any functional evidence of changes.

EGR1 is another Zinc Finger TF which we have shown as having a binding motif overlapping with WT1 and which can compete with WT1 for its binding sites (see Section 4.5.3). *EGR1* can also bind to and regulate *WT1* but WT1 can also regulate EGR1 as seen in knockdown experiments (see Section 4.1.4). EGR1 is bound by multiple signalling responsive factors such as AP-1 and STAT5 (Cauchy et al., 2015) and RUNX1-ETO and so can respond to oncogenic signalling. But equally, EGR1 can also be modulated through its five Serum Response Elements in response to general cellular stress.

Oncogenic signalling is not the only signalling mechanism that may elicit transcriptional changes. In murine models of AML, inflammation of mesenchymal stem cells promoted the transition of pre-leukaemia to leukaemia (Zambetti et al., 2016). In another murine study, high IL1 levels promoted myeloid skewing of HSCs and accelerated leukaemogenesis (Chavez et al., 2019). Therefore these transcription factors that are driven by oncogenic signalling may potentially also be driven/ initiated from exogenous signalling and it would be very interesting to see whether *WT1* expression or binding are altered by inflammatory signals, thus promoting oncogenesis.

5.1.2 AP-1 as part of an oncogenic signalling hub

AP-1 has been known to be part of a tumour-specific oncogenic signalling process for decades (Angel et al., 1988). Whilst an appreciation of the role of AP-1 within AML existed for several years (Steidl et al., 2006b), its precise role in leukaemogenesis has not been followed up in detail. Part of the difficulty has been the large degree of redundancy of AP-1 family members, consequently knockdown of one or even more AP-1 genes does not always affect leukaemic cells. Furthermore, even current published inhibitors of AP-1 (Brennan et al., 2020) do not work when tested in our Lab (James Griffin, Daniel Coleman, unpublished results). Whilst other labs have looked for simple and non-specific phenotypic changes such as cell death, our lab have created an AP-1 reporter cellular assay, which has clearly shown that none of the trialled inhibitors work (unpublished data). Therefore, due to redundancy and lack of inhibitor, an understanding of AP-1 on a molecular level has been poor, until this work which used a dominant negative FOS peptide that blocked the binding of all AP-1 proteins.

AP-1 is activated through growth factor receptor signalling for example by Stem Cell Factor binding to the KIT receptor (Martin et al., 1990) or the FLT3 ligand binding to the FLT3 receptor (Lyman et al., 1993a). Binding of these growth factors to their receptors leads to autophosphorylation of their tyrosine kinase domains, which recruits adaptor proteins such as GRB2 which in turn recruit Guanine nucleotide exchange factors such as SOS which can promote RAS phosphorylation and hence activate RAS signalling (Egan et al., 1993). Activated RAS subsequently activates downstream members of it signalling pathway

including RAF, MEK, ERK, MAPK and JNK through subsequent serine and threonine moiety phosphorylation (Campbell et al., 1998). AP-1 family members are then activated through MAPK signalling, as exemplified by ATF2 which is activated by p38 MAPK (Raingeaud et al., 1996) and JNK (Gupta et al., 1995) and JUN which is activated by JNK (Dérijard et al., 1994).

In AML, Class I mutations such as *KIT*, *FLT3*, *NRAS* and *KRAS* mutations are common (Papaemmanuil et al., 2016) and drive constitutive AP-1 signalling independent of growth factor binding. Further support of this concept comes from FLT3 knockdown studies in AML which led to both a reduction in MAPK pathway member phosphorylation and a reduction in AP-1 binding, thus demonstrating that AP-1 is a vital transducer of oncogenic signalling in the context of leukaemia (Cauchy et al., 2015). However, Class II mutations such as RUNX1-ETO can also activate AP-1 signalling (Elsässer et al., 2003) (Frank et al., 1999) and we have shown that *RUNX1-ETO* knockdown reduces the expression of AP-1 family members (see Section 3.1.2), although the precise molecular mechanism linking RUNX1-ETO and AP-1 is currently a subject of active investigation.

Whilst inhibiting AP-1 binding led to genome-wide changes (see Section 3.3), we uncovered an important link between AP-1 binding and the cell cycle and hence a possible molecular mechanism for how oncogenic signalling actually leads to increased cell growth. AP-1 binds to the Cyclin D2 (*CCND2)* promoter and increases *CCND2* mRNA and protein expression (see Section 3.2.1). Cyclin D2 is an essential protein that activates Cyclin-Dependent Kinases 4 and 6 (CDK4, CDK6) and is vital for the G1 cell cycle transition (Kato et al., 1993). Either knocking down *CCND2* or expressing dnFOS led to a G1 cell cycle arrest

(Martinez-Soria et al., 2018); this finding provides confirmation as to how signalling feeds into the cell cycle and also highlights a therapeutic vulnerability whereby the CDK4/6 inhibitor Palbociclib can inhibit the growth effects of oncogenic signalling.

5.1.3 Maintenance of oncogenic signalling through regulatory loops

In our work, we have shown that both AP-1 and *WT1* are upregulated as a consequence of the expression of the RUNX1-ETO driver oncoproteins. Upon knockdown of the driver oncoprotein, both AP-1 and *WT1* are eventually downregulated (see Sections 3.2.2 and 4.2.4) albeit 10 days after knockdown. This delay in AP-1 or *WT1* downregulation may be due to feedback loops maintaining AP-1 and WT1 oncogenic signalling.

In the case of AP-1, when dnFOS was induced, genes coding for members of the RAS signalling pathway, that are upstream of AP-1, were upregulated (see Section 3.3.1). Furthermore, *MYCN* was upregulated nearly 7 fold upon dnFOS induction and MYCN has previously been shown to contribute to leukaemic proliferation in mouse models (Kawagoe et al., 2007) through increased JUN signalling. These two mechanisms may therefore be responsible for the maintenance of AP-1 signalling when AP-1 binding/expression (driven by the driver oncoprotein) is reduced.

In the case of WT1, WT1 was seen to bind to its own intron 8 cis-regulatory element thus establishing a classical feed-forward loop. This element was found to interact with the *WT1* promoter and CRISPR activation led to increased *WT1*

expression (see Section 4.2). Consequently, this is a mechanism by which WT1 can maintain its own expression in order to maintain enhanced expression following oncogenic signalling.

5.1.4 Alterations in oncogenic signalling through WT1 mutations

Whilst mutations in AP-1 family members are not recurrent in AML, *WT1* mutations occur in around 10% of AML (King-Underwood et al., 1996). *WT1* mutations occur most commonly in exon 7 but can also be found in exon 8 or 9 (Krauth et al., 2015). *WT1* mutations also confer a poor prognosis (Hou et al., 2010).

In our work, we found that overexpressing the exon 8 mutant increased leukaemic growth whilst exon 7 and exon 9 mutations did not (see Section 4.6). Furthermore, whilst expression of exon 8 or exon 9 mutants produced truncated proteins, the exon 7 mutation did not give rise to a stable protein, presumably due to nonsense-mediated decay (Abbas et al., 2010) or protein degradation through the proteasome, thus resembling a null allele. Examination of a primary patient sample with an exon 7 mutation confirmed that a truncated WT1 protein was not detectable in these cells (see Section 4.6.2).

Consequently, exon 7 mutations are likely to give rise to any phenotypic effects through haploinsufficiency, which we could not test in our overexpression experiments. By contrast, exon 8 mutations did give rise to a functional protein which led to the loss of even normal WT1 binding, as seen by ChIP-seq (see Section 4.6.2) suggesting a dominant negative phenotype.

Therefore, whilst wildtype WT1 can act as an oncogenic signal, our mutation experiments suggest that it also has a role as a tumour suppressor and either haploinsufficiency or loss of function leads to a leukaemic phenotype.

5.1.5 Oncogenic signalling interacts with the driver oncoprotein transcriptional program

Intriguingly, we found surprisingly little overlap between different mutational subtypes of AML with regards to epigenetic or transcriptional changes after the expression of dnFOS or *WT1* isoforms (see Sections 3.3.2 and 4.3.7). We found that RNA-seq or ATAC-seq experiments clustered according to the driver mutation rather than the nature of the genetic perturbation performed, with different pathways being affected. Consequently, whilst similar phenotypic effects were elicited by the perturbation including the stunting of growth, the genome-wide effects of the AP-1 and WT1 transcription factors appear to be contingent upon the chromatin landscape or transcriptional program on which they act.

This finding that these transcription factors behave so differently in different cells is of therapeutic advantage. For both AP-1 and WT1 we found that healthy CD34$^+$ stem cells were unaffected by our genetic perturbations experiments. Therefore we would expect healthy cells to be unaffected by any inhibitors targeting these transcription factors, that may be developed in the future. This would be a significant advantage over chemotherapy which kills healthy and leukaemic stem cells alike and renders patients pancytopenic for extended periods of time.

We explored this interaction between changes elicited by the driver oncoprotein and transcription factor expression in the case of t(8;21) AML with *WT1* isoform

expression. Expression of *WT1 –KTS* cooperated with *RUNX1-ETO* knockdown to promote myeloid differentiation whilst *WT1* +KTS expression blocked myeloid differentiation even when *RUNX1-ETO* was knocked down (see Section 4.3.2). This results agrees with the finding that *WT1* is a target of RUNX1-ETO (see Section 4.2.4). However, we also found that *WT1* isoform expression altered RUNX1-ETO binding sites (see Section 4.4.5). Therefore the driver oncoprotein and downstream transcription factor actually regulate one another in a complex manner; it is not just the case that the driver oncoprotein determines where the downstream transcription factor binds.

5.1.6 Oncogenic signalling may affect cellular metabolism

It has long been known that cancer cells alter their metabolism to adapt to their environment, for example, the Warburg effect describes a normoxic increase in lactate production due to glycolytic activity in proliferative cells (Warburg et al., 1927). Many further metabolic alterations in cancer, to compensate for restrictions in the availability of nutrients in the tumour microenvironment, are now understood (Finicle et al., 2018) including scavenging extracellular nutrients, autophagy to utilise intracellular nutrients and upregulation of metabolic enzymes (Rosario et al., 2018).

As rapidly growing cells require more nutrients and metabolites, growth driving signalling appears to be intricately connected with cellular metabolism pathways including the Phosphoinositide 3-kinase, PTEN (phosphatase and tensin homolog) and tyrosine pathways (Ward and Thompson, 2012).

A recurrent theme within our work is that oncogenic signalling through transcription factor expression elicits changes in gene expression of genes in metabolic pathways. When *WT1* was knocked down in primary FLT3-ITD cells, downregulation of genes in the glycerophospholipid and ether lipid metabolism pathways was seen (see Section 4.1.4). These lipids are crucial for AML cell growth and survival (Ricciardi et al., 2015) demonstrating that WT1 may have a role in upregulating cellular metabolic pathways.

Similarly, when AP-1 binding was blocked using a dominant negative FOS, downregulation of genes in the steroid biosynthesis, butanoate, pyruvate, tryptophan, arginine and proline metabolism pathway occurred (see Section 3.3.1). Most notably within these metabolic pathways, AML cells are addicted to arginine and cannot replicate or engraft well without it (Mussai et al., 2015).

5.2 Determinants of binding in structurally similar transcription factors and isoforms

5.2.1 Sequence specific determinants of transcription factor binding

In our work, we found that the DNA consensus sequence for WT1 binding differed by just one invariant Adenine base; the position weight matrix for EGR1 was SYGTGGGYGK (IUPAC nomenclature) compared with SYGTGGAGK for WT1 (Section 4.4.2). The change in just one base was enough to lead to considerably different binding sites with only 2477 binding sites shared between the two factors and with 2676 WT1-specific binding sites and 7666 EGR1-specific binding site (section 4.5.1).

The similarity between the WT1 and EGR1 binding sequences is likely because EGR1 has 3 zinc fingers which are similar to zinc fingers 2-4 of WT1. A crystallographic study showed that only zinc fingers 2-4 insert deeply into the DNA major grove to make base-specific contacts (Stoll et al., 2007). The invariant Adenine base in position 8 of the position weight matrix of WT1 is the key difference from EGR1 where position 8 has a Cytosine or Thymine and NMR studies suggest that the closest contact with this base is the Zinc Finger 1 which is not present in EGR1 (Yengo et al., 2018). Consequently, in contrast to the crystallography study showing that zinc finger 1 does not insert deeply into the DNA major grove, it nevertheless appears very important in conferring specificity as to whether the transcription factor binds to the site or not.

5.2.2 Non-sequence specific determinants of transcription factor binding

A further intriguing finding from our work was that the WT1 +KTS and WT1 –KTS isoforms bind to the same consensus sequence yet their genomic binding sites were quite distinct with WT1 +KTS having 9451 unique binding sites and WT1 –KTS having 1355 unique binding sites (Section 4.4.3). This finding agrees with EMSA studies where different shifts have been seen with the isoforms (Bickmore et al., 1992). Whilst some of these changes may simply be explained by the differences in the number of open chromatin accessible to the WT1 protein (Section 4.3.3), there were not enough sites changing in accessibility to account for all of the differential binding seen.

The +KTS splice site affects Zinc fingers 3 and 4 (Yengo et al., 2018) which may be sites of protein-protein interaction. With this knowledge, other possibilities which can explain differences in binding outside of sequence specificity include:

i) Altered binding through other transcription factors which are recruited via zinc finger 3 or 4 of WT1. TF complexes such as AP-1 and the p50-RelA NFκB complexes (Chen et al., 1998) rely on distinct TF co-operating in order to bind. It is also known that TFs changing binding partner can alter binding sites, for example the Retinoid X receptor binds to 1,2 or 5bp spaced repeats of the hormone response element when it heterodimerises with the Retinoic acid receptor (Näär et al., 1991) but binds to 3bp spaced repeats of that element when it heterodimerises with the Vitamin D receptor (Towers et al., 1993), hence conferring differential binding. Whilst we have found an Sp1:WT1 composite motif suggesting co-operative binding with Sp1, this didn't appear to be the cause of isoform specific binding (Section 3.5.1) and perhaps other proteins can co-bind with WT1 in an isoform-specific manner in order to affect its binding. Indeed, a high throughput study *in vitro* revealed over 300 TF co-operative binding pairs in a colonic cell line, many of which were previously unknown (Jolma et al., 2015).

ii) Local DNA modifications such as methylation differentially affecting how each transcription factor binds, for example NRF1 can only bind to unmethylated DNA (Domcke et al., 2015) and KLF4 preferentially binds to methylated DNA (Hu et al., 2013). WT1 is known to recruit TET2 and hence modify DNA methylation (Rampal et al., 2014). Therefore, WT1 isoform-specific changes in methylation can potentially alter their genomic binding.

iii) Differences in GC composition surrounding the binding site (Dror et al., 2015).

iv) Differences in DNA shape features within the motif such as minor groove width and rotational parameters such as helix twist, propeller twist and roll. This feature has been recognised for the Hox family transcription factor SCR (Abe et al., 2015) but also in basic Helix-Loop-Helix factors such as CBF1 and TYE1 in yeast whereby nucleotides outside of the DNA binding motif were found to influence the 3D structure of DNA binding sites and hence confer differential binding (Gordân et al., 2013).

v) Differences in the strength of binding. Even if the same binding sequence exists, alterations in the strength of transcription factor binding can affect binding sites. Indeed, nuclear magnetic resonance studies showed that the WT1 KTS insertion increased the movement of a linker between zinc fingers 3 and 4 and so reduced the stability of the 4th zinc finger in the major groove of DNA (Laity et al., 2000).

5.3 Regulation of transcription factor genes

By understanding how expression of transcription factor genes is regulated by other transcription factors, we can derive gene regulatory networks (see Sections 4.3.4 and 4.3.5). By examining these networks, it is possible to understand how a gene mutation or overexpression, perhaps through oncogenic signalling, can lead to large cellular changes in the transcriptome and epigenome.

We have profiled open chromatin sites around the *WT1* locus in primary AML cells and identified an intron 3 DHS, intron 8 DHS, 3' DHS and CTCF-bound DHS. In addition, we have confirmed interactions of all of these sites with the promoter through HiC and investigated the effects of Lysine 27 acetylation in histone H3 through CRISPR activation (see Section 4.2).

Out of these open chromatin sites, the 3' DHS has been previously suggested to be an enhancer, which can bind GATA1 and GATA 2 as seen in Electrophoretic Mobility Shift Assays (EMSA) (Furuhata et al., 2009). The intron 3 DHS has also been described as an enhancer which can bind to GATA1 and Myb (Zhang et al., 1997b).

Enhancers typically carry H3K27ac marks and p300 binding and contain numerous transcription factor binding sites (Ernst et al., 2011). The 3' DHS did not have any of these features but the intron 3 DHS had all of them (see Section 4.2.1). The intron 8 DHS did not display H3K27 acetylation or p300 binding but several transcription factors including RUNX1, RUNX1-ETO, WT1 and PU.1 bound there. CRISPR activation experiments showed that H3K27 acetylation at the intron 8 DHS increased *WT1* expression but forced acetylation at the other open chromatin sites had no effect. Putting this altogether, we did not find any evidence of any enhancer activity of the 3' DHS or CTCF bound DHS in leukaemia cells. The intron 3 DHS appears to be a constitutively active enhancer in leukaemia. However, the intron 8 DHS may be a poised enhancer site (Creyghton et al., 2010) which upon H3K27 acetylation can change it into an active enhancer (Rada-Iglesias et al., 2011). Given the responsiveness of *WT1* to cellular signalling, the intron 8 site is likely to be a signalling responsive site.

We speculate that depending on the particular genetic subtype of AML and its transcription factor network, this enhancer may be active to different extents, which may also explain the difference in *WT1* expression in different types of AML (see Section 4.1.1).

AP-1 has classically been thought to be regulated through post-translational modifications such as signalling; JNK can phosphorylate and hence activate JUN at positions 63 and 73 (Pulverer et al., 1993). However, we have also described how AP-1 factors can also be regulated at a gene expression level, for example *RUNX1-ETO* knockdown in t(8;21) AML reduces gene expression of several AP-1 family members including *JUN*, *FOSL1* and *FOSL2* (see Section 3.1.2); RUNX1-ETO binds to the promoter and several open chromatin sites near each of these AP-1 genes. Therefore, we demonstrate a further mechanism of the regulation of AP-1 activity.

5.4 Targeting transcription factors

We have studied in detail several transcription factors which clearly have major impacts upon leukaemia maintenance. However, to take this work further, one needs to consider the therapeutic potential of targeting these factors. Whilst transcription factors were long considered "undruggable", recent studies have pointed towards several potential therapeutic strategies targeting their activities.

5.4.1 Direct inhibition of transcription factor complexes

Some transcription factors do not act alone and form complexes with other factors on DNA. This notion is exemplified by the CBFβ-RUNX1 complex; whilst RUNX1

can by itself bind to its DNA consensus sequence through its RUNT domain, CBFβ heterodimerising with RUNX1 increases DNA binding strength around 10 fold by restricting mobile parts of the RUNT domain (Tahirov et al., 2001). Consequently, inhibiting the CBFβ-RUNX1 complex, as confirmed by Fluorescence Resonance Energy Transfer (FRET) and co-immunoprecipitation, can decrease RUNX1 binding as seen through ChIP-seq (Illendula et al., 2016). We have used this inhibitor in our work to perturb the WT1 centred signalling network in FLT3-ITD AML (see Section 4.7.2). Similarly, in inv(16) AML, an inhibitor of the CBFβ-SMMHC oncoprotein to RUNX1 complex could also be used to restore normal RUNX1 binding and kill leukaemic cells (Illendula et al., 2015).

5.4.2 Promotion of direct protein degradation

The phthalimide class of drug have long been used to treat multiple myeloma and del(5q) myelodysplasia. However, their mechanism of action was later discovered to be through protein degradation of the Ikaros transcription factors IKZF1 and IKZF3 (Lu et al., 2014). By conjugating protein ligands to a phthalimide drugs, target proteins can also be targeted for degradation by the proteasome (Winter et al., 2015). Such a strategy could theoretically be used to degrade any transcription factor.

5.4.3 Target upstream regulators or downstream effector genes

In some instances, key upstream regulators or downstream targets may be more druggable than the transcription factor itself, which is why we went to great lengths to understand the Gene Regulatory Networks (see Section 4.4.4). We found that the WT1 +KTS responsive GRN contains CD69 and PIM (Proto-

oncogene serine/threonine protein) kinases such as PIM1 and the WT1 –KTS responsive GRN contains the FES kinase, all of which are druggable targets.

Similarly, we showed that AP-1 upregulated Cyclin D2 which highlighted a therapeutic vulnerability that could be targeted with Palbociclib to great effect *in vitro* and *in vivo* (see Section 3.2.2).

5.4.4 Targeting the spliceosome

Mutations in genes coding for the spliceosome such as *SRSF2*, *SF3B1*, *U2AF1* and *ZRSF2* are known to promote leukaemogenesis (Yoshimi et al., 2019) and such AMLs are vulnerable to spliceosome inhibitors (Lee et al., 2016). We have shown in this work that alternative splicing of *WT1* particularly at the exon 9 KTS splice site was key to leukaemic behaviour but also that WT1 can regulate splicing directly. Therefore, use of spliceosome inhibitors such as the SF3b inhibitor Spliceostatin A (Kaida et al., 2007) or E7107 (Kotake et al., 2007) may alter this behaviour.

5.4.5 Targeting Epigenetic regulators

Numerous inhibitors of epigenetic regulators are being developed and undergoing clinical trials. Whilst none of these epigenetic regulators selectively targets a particular transcription factor, histone modification of the cis-regulatory elements of the transcription factor may affect its expression. The following table summarises some of the drugs currently being developed:

Epigenetic Regulator	Nomen-clature	Mechanism	Inhibitor	Clinical Trial
Lysine Specific Demethylase 1	LSD1	Demethylates H3K4 and H3K9 mono-	Tranylcypromine	NCT02717884

		and di- methyl marks		
Enhancer of Zeste Homolog 2	EZH2	H3K27 Trimethylation	Tazemetostat	NCT01897571
Disruptor of Telomeric Silencing 1-like	DOT1L	H3K79 methylation	Pinometostat	NCT01684150
Lysine Acetyltransferases	EP300, CREBBP	Histone acetylation e.g. H3K27	CCS1477	NCT04068597
Arginine Methyltransferase 5	PRMT5	Arginine methylation of H4R3, H2AR3 and H3R8.	GSK3326595	NCT03614728
Bromodomain containing protein 4	BRD4	Epigenetic reader	GSK525762	NCT01943851

Table 5-1: List of epigenetic regulators with drugs undergoing clinical trials

5.4.6 Immune therapies

Since *WT1* is upregulated within minimal residual disease, immune therapies have been trialled in this context. Vaccination trials against WT1 have produced limited effects (Maslak et al., 2018). However, T Cell Receptor therapies against WT1 have been more encouraging (Chapuis et al., 2019).

Our findings that different WT1 isoforms have contrasting biological activities, some of which normally decrease leukaemic maintenance highlights the need to take caution when designing therapeutic approaches that may shift the isoform ratio. Our results also highlights the need to account for the AML sub-type as the genomic response to *WT1* overexpression differs between AML sub-types and *WT1* may respond differently to whatever oncogenic signalling is present in the

different subtypes, meriting efforts to understand how their signalling network is rewired to keep AML cells alive and how this feeds into the WT1/EGR1 axis.

5.5 Summary

Through the study of several transcription factors upregulated in AML such as the AP-1 family and Zinc Finger family of TFs, we have shown how leukaemic maintenance may be orchestrated at a molecular level resulting in corrupt transcription factor networks. Targeting these AML-specific corrupt networks will be an avenue to future therapeutic interventions.

5.6 Future Work

5.6.1 Splicing and RNA modification

In this work we have shown that the phenotype of the leukaemia cell is strongly dependent upon which of *WT1* +KTS or *WT1* –KTS is expressed and so splicing would be expected to alter the ratios of *WT1* isoforms expressed. Furthermore, WT1 isoforms themselves affect splicing of other genes through direct binding at intron-exon boundaries as well as through altering gene expression of splicing factors (see Section 4.3.3). Consequently, investigating how spliceosome inhibitors, such as the ones discussed in Section 5.4.4, affect *WT1* isoform expression and how they affect splicing of genes regulated by WT1 would be of great interest.

Furthermore, a greater understanding of splicing alterations induced by WT1 would be important as alternative splicing events can lead to alterations in the protein translated (Kalsotra and Cooper, 2011). Whilst in this work we have described how alterations in the Percentage Spliced In (PSI) of exons is affected by WT1 expression, we have not determined the precise splicing alteration from the numerous described possibilities including intron retention, cassette exons, mutually exclusive exons, exon skipping, alternative 5' splice site usage and alternative 3' splice site usage (Dvinge et al., 2016). Such an analysis would be carried out both in the cell line as well as the primary cell RNA-seq data. We would also perform experiments to determine precisely which RNA that WT1 isoforms bind to by UV crosslinking protein-RNA complexes, performing immunoprecipitation of WT1 bound RNA and then performing RNA sequencing 'ClIP-seq' (Bharathavikru et al., 2017).

5.6.2 Tagging approaches

Throughout our work, we have used WT1 antibodies raised against a sequence that is present in all WT1 isoforms and in ectopically expressed WT1. Since we believe that it is the ratio between different WT1 isoforms that is crucial for leukaemic behaviour (Gu et al., 2010), using a WT1 antibody against all types of WT1 is advantageous as it allows us to understand where all WT1 binding is occurring when certain ratios are expressed. However, a limitation of this approach is that it becomes difficult to attribute a specific binding site to a specific isoform when multiple isoforms could be binding. Consequently, tagging WT1 transgene with a HA-tag and then using a HA antibody can allow us to specifically understand where the transgene is binding. It is unclear whether tagging at the

N-terminal or C-terminal end would work and certainly a concern with tagging at the C-terminal end would be that DNA binding might be altered, given that just the three amino acid KTS sequence caused such drastic alterations in DNA binding.

Another important use of tagging would be for performing Immunoprecipitation when no suitable antibodies are available. An estimate by the ENCODE consortium is that less than 10% of supposedly specific commercially available antibodies are suitable for ChIP-seq. In this work, we were able to show that the dominant negative FOS could reduce binding of JUND to DNA. However, we were not able to show this as being the case for any other AP-1 family members due to a lack of specific antibodies. Tagging endogenous genes using CRISPR-Cas9 gene editing would allow for ChIP-seq where antibodies are not available to the native protein. Two current approaches involve utilise either Homology Directed Repair (Zhang et al., 2008a) or Microhomology Mediated End Joining (Sakuma et al., 2016).

5.6.3 Enhancer assays

We have discussed that the intron 3 DHS in *WT1* is likely to be an active enhancer and that the intron 8 DHS is likely to be a poised DHS (see Section 5.3) based upon their histone modification and CRISPR activation responses. However, this hypothesis needs to be confirmed with functional assays; classically this is done through a luciferase assay. In this assay, putative enhancers are cloned in front of a minimal promoter in a plasmid to drive firefly luciferase expression. When this plasmid is transfected into the cell of interest and a sequence with enhancer

activity is present in front of the promoter, a luciferin signal that can be detected and quantified is emitted (de Wet et al., 1987).

Other confirmatory evidence could come through a CRISPR interference assay; by targeting a catalytically dead Cas9 fused to an inhibitory Kruppel associated box (KRAB) domain to each putative enhancer, where gene expression decreases, the targeted site is a likely enhancer (Qi et al., 2013). Thus, if CRISPRi at the intron 3 DHS reduced *WT1* gene expression, it would suggest that the intron 3 site is an enhancer.

5.6.4 Protein-protein interactions

In our work to understand corrupt transcription factor networks in leukaemia, we have assumed that TFs bind to their motifs directly but we know that factors may bind at other sites through protein-protein interactions or alter the epigenome by recruiting chromatin modifiers.

It is currently known that WT1 binds to chaperone proteins such as Hsp70 (Maheswaran et al., 1998), the transcriptional repressor BASP1 (Carpenter et al., 2004), the tumour suppressor protein TP53 (Maheswaran et al., 1993a), the transcriptional co-activator CREB Binding Protein (Wang et al., 2001), the DNA methylation modifier TET2 (Wang et al., 2015) (Rampal et al., 2014) (see also Section 1.3.2) and the RNA modifier Wilms Tumour Associating Protein (WTAP) (Little et al., 2000). However, many other protein partners are yet to be established.

Immunoprecipitation followed by Mass Spectrometry may allow us to understand which proteins WT1 binds to (Free et al., 2009). However, such a technique is

prone to false positive results and so it would be important to validate any findings through co-immunoprecipitation experiments (Bharathavikru and von Kriegsheim, 2016).

5.6.5 Xenotransplantation experiments

With the dnFOS experiments, we found a larger effect of dnFOS induction *in vivo* with a longer latency to tumour engraftment (Section 3.1.4) than might otherwise be expected from just a 40% reduction in colony formation ability as seen in the *in vitro* assays (Section 3.1.2). Some of the genes that were downregulated with dnFOS induction included those coding for matrix metalloproteinase which are needed to create a niche within the bone marrow microenvironment (Saw et al., 2019).

Similarly, it would be very interesting to see how changes in *WT1* gene expression interacts with the bone marrow microenvironment and whether induction of *WT1* +KTS increases leukaemic growth *in vivo*. Such experiments could be performed in a similar manner to the dnFOS experiments by xenotransplantation.

5.6.6 Metabolic Studies

In order to correlate whether dysregulation in gene expression of any of the metabolic pathways identified translates to metabolic changes within the leukaemic cell, more functional studies are necessary. Mass spectrometry with isotope tracing is a powerful method of identifying metabolic changes and has been used successfully to identify metabolic vulnerabilities in leukaemic stem cells that may be targeted with small molecule inhibitors (Kuntz et al., 2017). More

recently, the Bonifer lab has collaborated to set up live cell Nuclear Magnetic Resonance studies of the extracellular milieu of leukaemic cells in culture (Roberts, 2019). These techniques could be employed to trace specific metabolic changes.

References

ABBAS, S., ERPELINCK-VERSCHUEREN, C. A., GOUDSWAARD, C. S., LÖWENBERG, B. & VALK, P. J. 2010. Mutant Wilms' tumor 1 (WT1) mRNA with premature termination codons in acute myeloid leukemia (AML) is sensitive to nonsense-mediated RNA decay (NMD). *Leukemia,* 24**,** 660-3.

ABE, N., DROR, I., YANG, L., SLATTERY, M., ZHOU, T., BUSSEMAKER, HARMEN J., ROHS, R. & MANN, RICHARD S. 2015. Deconvolving the Recognition of DNA Shape from Sequence. *Cell,* 161**,** 307-318.

ABELSON, S., COLLORD, G., NG, S. W. K., WEISSBROD, O., MENDELSON COHEN, N., NIEMEYER, E., BARDA, N., ZUZARTE, P. C., HEISLER, L., SUNDARAVADANAM, Y., LUBEN, R., HAYAT, S., WANG, T. T., ZHAO, Z., CIRLAN, I., PUGH, T. J., SOAVE, D., NG, K., LATIMER, C., HARDY, C., RAINE, K., JONES, D., HOULT, D., BRITTEN, A., MCPHERSON, J. D., JOHANSSON, M., MBABAALI, F., EAGLES, J., MILLER, J. K., PASTERNACK, D., TIMMS, L., KRZYZANOWSKI, P., AWADALLA, P., COSTA, R., SEGAL, E., BRATMAN, S. V., BEER, P., BEHJATI, S., MARTINCORENA, I., WANG, J. C. Y., BOWLES, K. M., QUIRÓS, J. R., KARAKATSANI, A., LA VECCHIA, C., TRICHOPOULOU, A., SALAMANCA-FERNÁNDEZ, E., HUERTA, J. M., BARRICARTE, A., TRAVIS, R. C., TUMINO, R., MASALA, G., BOEING, H., PANICO, S., KAAKS, R., KRÄMER, A., SIERI, S., RIBOLI, E., VINEIS, P., FOLL, M., MCKAY, J., POLIDORO, S., SALA, N., KHAW, K.-T., VERMEULEN, R., CAMPBELL, P. J., PAPAEMMANUIL, E., MINDEN, M. D., TANAY, A., BALICER, R. D., WAREHAM, N. J., GERSTUNG, M., DICK, J. E., BRENNAN, P., VASSILIOU, G. S. & SHLUSH, L. I. 2018. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature,* 559**,** 400-404.

AGARWAL, S. K., GURU, S. C., HEPPNER, C., ERDOS, M. R., COLLINS, R. M., PARK, S. Y., SAGGAR, S., CHANDRASEKHARAPPA, S. C., COLLINS, F. S., SPIEGEL, A. M., MARX, S. J. & BURNS, A. L. 1999. Menin interacts with the AP1 transcription factor JunD and represses JunD-activated transcription. *Cell,* 96**,** 143-52.

AKASHI, K., TRAVER, D., MIYAMOTO, T. & WEISSMAN, I. L. 2000. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature,* 404**,** 193-7.

AL SERAIHI, A. F., RIO-MACHIN, A., TAWANA, K., BÖDÖR, C., WANG, J., NAGANO, A., HEWARD, J. A., IQBAL, S., BEST, S., LEA, N., MCLORNAN, D., KOZYRA, E. J., WLODARSKI, M. W., NIEMEYER, C. M., SCOTT, H., HAHN, C., ELLISON, A., TUMMALA, H., CARDOSO, S. R., VULLIAMY, T., DOKAL, I., BUTLER, T., SMITH, M., CAVENAGH, J. & FITZGIBBON, J. 2018. GATA2 monoallelic expression underlies reduced penetrance in inherited GATA2-mutated MDS/AML. *Leukemia,* 32**,** 2502-2507.

ALBERTA, J. A., SPRINGETT, G. M., RAYBURN, H., NATOLI, T. A., LORING, J., KREIDBERG, J. A. & HOUSMAN, D. 2003. Role of the WT1 tumor suppressor in murine hematopoiesis. *Blood,* 101**,** 2570.

ANDERSSON, R., REFSING ANDERSEN, P., VALEN, E., CORE, L. J., BORNHOLDT, J., BOYD, M., HEICK JENSEN, T. & SANDELIN, A. 2014. Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nature Communications,* 5**,** 5336.

ANGEL, P., HATTORI, K., SMEAL, T. & KARIN, M. 1988. The jun proto-oncogene is positively autoregulated by its product, Jun/AP-1. *Cell,* 55**,** 875-85.

ANGELOV, D., MOLLA, A., PERCHE, P. Y., HANS, F., COTE, J., KHOCHBIN, S., BOUVET, P. & DIMITROV, S. 2003. The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. *Mol Cell,* 11**,** 1033-41.

302

ASOU, H., TASHIRO, S., HAMAMOTO, K., OTSUJI, A., KITA, K. & KAMADA, N. 1991. Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation. *Blood,* 77**,** 2031.

ASSI, S. A., IMPERATO, M. R., COLEMAN, D. J. L., PICKIN, A., POTLURI, S., PTASINSKA, A., CHIN, P. S., BLAIR, H., CAUCHY, P., JAMES, S. R., ZACARIAS-CABEZA, J., GILDING, L. N., BEGGS, A., CLOKIE, S., LOKE, J. C., JENKIN, P., UDDIN, A., DELWEL, R., RICHARDS, S. J., RAGHAVAN, M., GRIFFITHS, M. J., HEIDENREICH, O., COCKERILL, P. N. & BONIFER, C. 2019. Subtype-specific regulatory network rewiring in acute myeloid leukemia. *Nature Genetics,* 51**,** 151-162.

BAIRD, P. N. & SIMMONS, P. J. 1997. Expression of the Wilms' tumor gene (WT1) in normal hemopoiesis. *Exp Hematol,* 25**,** 312-20.

BANERJI, J., RUSCONI, S. & SCHAFFNER, W. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell,* 27**,** 299-308.

BANNISTER, A. J., ZEGERMAN, P., PARTRIDGE, J. F., MISKA, E. A., THOMAS, J. O., ALLSHIRE, R. C. & KOUZARIDES, T. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature,* 410**,** 120-4.

BARBIERI, I., TZELEPIS, K., PANDOLFINI, L., SHI, J., MILLÁN-ZAMBRANO, G., ROBSON, S. C., ASPRIS, D., MIGLIORI, V., BANNISTER, A. J., HAN, N., DE BRAEKELEER, E., PONSTINGL, H., HENDRICK, A., VAKOC, C. R., VASSILIOU, G. S. & KOUZARIDES, T. 2017. Promoter-bound METTL3 maintains myeloid leukaemia by m6A-dependent translation control. *Nature,* 552**,** 126.

BARRAGAN, E., CERVERA, J., BOLUFER, P., BALLESTER, S., MARTIN, G., FERNANDEZ, P., COLLADO, R., SAYAS, M. J. & SANZ, M. A. 2004. Prognostic implications of Wilms tumor gene (WT1) expression in patients with de novo acute myeloid leukemia. *Haematologica,* 89**,** 926.

BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., SCHONES, D. E., WANG, Z., WEI, G., CHEPELEV, I. & ZHAO, K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell,* 129**,** 823-37.

BATEMAN, C. M., COLMAN, S. M., CHAPLIN, T., YOUNG, B. D., EDEN, T. O., BHAKTA, M., GRATIAS, E. J., VAN WERING, E. R., CAZZANIGA, G., HARRISON, C. J., HAIN, R., ANCLIFF, P., FORD, A. M., KEARNEY, L. & GREAVES, M. 2010. Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia. *Blood,* 115**,** 3553-8.

BECKER, H., MARCUCCI, G., MAHARRY, K., RADMACHER, M. D., MRÓZEK, K., MARGESON, D., WHITMAN, S. P., PASCHKA, P., HOLLAND, K. B., SCHWIND, S., WU, Y.-Z., POWELL, B. L., CARTER, T. H., KOLITZ, J. E., WETZLER, M., CARROLL, A. J., BAER, M. R., MOORE, J. O., CALIGIURI, M. A., LARSON, R. A. & BLOOMFIELD, C. D. 2010. Mutations of the Wilms tumor 1 gene (WT1) in older patients with primary cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood,* 116**,** 788.

BEHRENS, A., JOCHUM, W., SIBILIA, M. & WAGNER, E. F. 2000. Oncogenic transformation by ras and fos is mediated by c-Jun N-terminal phosphorylation. *Oncogene,* 19**,** 2657-2663.

BEHRENS, K., MAUL, K., TEKIN, N., KRIEBITZSCH, N., INDENBIRKEN, D., PRASSOLOV, V., MÜLLER, U., SERVE, H., CAMMENGA, J. & STOCKING, C. 2017. RUNX1 cooperates with FLT3-ITD to induce leukemia. *J Exp Med,* 214**,** 737-752.

BEN-AMI, O., FRIEDMAN, D., LESHKOWITZ, D., GOLDENBERG, D., ORLOVSKY, K., PENCOVICH, N., LOTEM, J., TANAY, A. & GRONER, Y. 2013. Addiction of t(8;21) and inv(16) acute myeloid leukemia to native RUNX1. *Cell Rep,* 4**,** 1131-43.

BENNETT, J. M., CATOVSKY, D., DANIEL, M.-T., FLANDRIN, G., GALTON, D. A. G., GRALNICK, H. R. & SULTAN, C. 1976. Proposals for the Classification of the Acute Leukaemias French-American-British (FAB) Co-operative Group. *British Journal of Haematology,* 33**,** 451-458.

BERNSTEIN, B. E., MIKKELSEN, T. S., XIE, X., KAMAL, M., HUEBERT, D. J., CUFF, J., FRY, B., MEISSNER, A., WERNIG, M., PLATH, K., JAENISCH, R., WAGSCHAL, A., FEIL, R., SCHREIBER, S. L. & LANDER, E. S. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell,* 125**,** 315-26.

BHARATHAVIKRU, R., DUDNAKOVA, T., AITKEN, S., SLIGHT, J., ARTIBANI, M., HOHENSTEIN, P., TOLLERVEY, D. & HASTIE, N. 2017. Transcription factor Wilms' tumor 1 regulates developmental RNAs through 3' UTR interaction. *Genes Dev,* 31**,** 347-352.

BHARATHAVIKRU, R. & VON KRIEGSHEIM, A. 2016. WT1-Associated Protein-Protein Interaction Networks. *Methods Mol Biol,* 1467**,** 189-96.

BHATTACHARYA, S. K., RAMCHANDANI, S., CERVONI, N. & SZYF, M. 1999. A mammalian protein with specific demethylase activity for mCpG DNA. *Nature,* 397**,** 579-83.

BI, W., WU, L., COUSTRY, F., DE CROMBRUGGHE, B. & MAITY, S. N. 1997. DNA Binding Specificity of the CCAAT-binding Factor CBF/NF-Y. *Journal of Biological Chemistry,* 272**,** 26562-26572.

BIALKOWSKA, A. B., YANG, V. W. & MALLIPATTU, S. K. 2017. Krüppel-like factors in mammalian stem cells and development. *Development (Cambridge, England),* 144**,** 737-754.

BICKMORE, W. A., OGHENE, K., LITTLE, M. H., SEAWRIGHT, A., VAN HEYNINGEN, V. & HASTIE, N. D. 1992. Modulation of DNA binding specificity by alternative splicing of the Wilms tumor wt1 gene transcript. *Science,* 257**,** 235.

BINDEA, G., MLECNIK, B., HACKL, H., CHAROENTONG, P., TOSOLINI, M., KIRILOVSKY, A., FRIDMAN, W. H., PAGÈS, F., TRAJANOSKI, Z. & GALON, J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics,* 25**,** 1091-3.

BLOM VAN ASSENDELFT, G., HANSCOMBE, O., GROSVELD, F. & GREAVES, D. R. 1989. The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell,* 56**,** 969-77.

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-20.

BOLOTIN, A., QUINQUIS, B., SOROKIN, A. & EHRLICH, S. D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology,* 151**,** 2551-2561.

BONNET, D. & DICK, J. E. 1997. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Medicine,* 3**,** 730.

BOUILLOUX, F., JUBAN, G., COHET, N., BUET, D., GUYOT, B., VAINCHENKER, W., LOUACHE, F. & MORLE, F. 2008. EKLF restricts megakaryocytic differentiation at the benefit of erythrocytic differentiation. *Blood,* 112**,** 576-84.

BOYD, A. L., ASLOSTOVAR, L., REID, J., YE, W., TANASIJEVIC, B., PORRAS, D. P., SHAPOVALOVA, Z., ALMAKADI, M., FOLEY, R., LEBER, B., XENOCOSTAS, A. & BHATIA, M. 2018. Identification of Chemotherapy-Induced Leukemic-Regenerating Cells Reveals a Transient Vulnerability of Human AML Recurrence. *Cancer Cell,* 34**,** 483-498.e5.

BRAHMA, S., UDUGAMA, M. I., KIM, J., HADA, A., BHARDWAJ, S. K., HAILU, S. G., LEE, T.-H. & BARTHOLOMEW, B. 2017. INO80 exchanges H2A.Z for H2A by translocating on DNA proximal to histone dimers. *Nature Communications,* 8**,** 15616.

BRAUN, T. P., OKHOVAT, M., COBLENTZ, C., CARRATT, S. A., FOLEY, A., SCHONROCK, Z., SMITH, B. M., NEVONEN, K., DAVIS, B., GARCIA, B., LATOCHA, D., WEEDER, B. R., GRZADKOWSKI, M. R., ESTABROOK, J. C., MANNING, H. G., WATANABE-SMITH, K., JENG, S., SMITH, J. L., LEONTI, A. R., RIES, R. E., MCWEENEY, S., DI GENUA, C., DRISSEN, R., NERLOV, C., MESHINCHI, S., CARBONE, L., DRUKER, B. J. & MAXSON, J. E. 2019. Myeloid lineage enhancers drive oncogene synergy in CEBPA/CSF3R mutant acute myeloid leukemia. *Nature Communications,* 10**,** 5455.

BRENNAN, A., LEECH, J. T., KAD, N. M. & MASON, J. M. 2020. Selective antagonism of cJun for cancer therapy. *Journal of Experimental & Clinical Cancer Research,* 39**,** 184.

BROUNS, S. J., JORE, M. M., LUNDGREN, M., WESTRA, E. R., SLIJKHUIS, R. J., SNIJDERS, A. P., DICKMAN, M. J., MAKAROVA, K. S., KOONIN, E. V. & VAN DER OOST, J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science,* 321**,** 960-4.

BRUSSELBACH, S., MOHLE-STEINLEIN, U., WANG, Z. Q., SCHREIBER, M., LUCIBELLO, F. C., MULLER, R. & WAGNER, E. F. 1995. Cell proliferation and cell cycle progression are not impaired in fibroblasts and ES cells lacking c-Fos. *Oncogene,* 10**,** 79-86.

BUENROSTRO, J. D., CORCES, M. R., LAREAU, C. A., WU, B., SCHEP, A. N., ARYEE, M. J., MAJETI, R., CHANG, H. Y. & GREENLEAF, W. J. 2018. Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell,* 173**,** 1535-1548.e16.

BUENROSTRO, J. D., GIRESI, P. G., ZABA, L. C., CHANG, H. Y. & GREENLEAF, W. J. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods,* 10**,** 1213-8.

BUREL, S. A., HARAKAWA, N., ZHOU, L., PABST, T., TENEN, D. G. & ZHANG, D.-E. 2001. Dichotomy of AML1-ETO Functions: Growth Arrest versus Block of Differentiation. *Molecular and Cellular Biology,* 21**,** 5577-5590.

BURGESS, M. R., HWANG, E., MROUE, R., BIELSKI, C. M., WANDLER, A. M., HUANG, B. J., FIRESTONE, A. J., YOUNG, A., LACAP, J. A., CROCKER, L., ASTHANA, S., DAVIS, E. M., XU, J., AKAGI, K., LE BEAU, M. M., LI, Q., HALEY, B., STOKOE, D., SAMPATH, D., TAYLOR, B. S., EVANGELISTA, M. & SHANNON, K. 2017. KRAS Allelic Imbalance Enhances Fitness and Modulates MAP Kinase Dependence in Cancer. *Cell,* 168**,** 817-829.e15.

BUSCH, K., KLAPPROTH, K., BARILE, M., FLOSSDORF, M., HOLLAND-LETZ, T., SCHLENNER, S. M., RETH, M., HOFER, T. & RODEWALD, H. R. 2015. Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature,* 518**,** 542-6.

BUSHNELL, D. A., WESTOVER, K. D., DAVIS, R. E. & KORNBERG, R. D. 2004. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science,* 303**,** 983-8.

CALL, K. M., GLASER, T., ITO, C. Y., BUCKLER, A. J., PELLETIER, J., HABER, D. A., ROSE, E. A., KRAL, A., YEGER, H., LEWIS, W. H., JONES, C. & HOUSMAN, D. E. 1990. Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. *Cell,* 60**,** 509-520.

CAMPBELL, S. L., KHOSRAVI-FAR, R., ROSSMAN, K. L., CLARK, G. J. & DER, C. J. 1998. Increasing complexity of Ras signaling. *Oncogene,* 17**,** 1395-1413.

CARICASOLE, A., DUARTE, A., LARSSON, S. H., HASTIE, N. D., LITTLE, M., HOLMES, G., TODOROV, I. & WARD, A. 1996. RNA binding by the Wilms tumor suppressor zinc finger proteins. *Proc Natl Acad Sci U S A,* 93**,** 7562-6.

CARPENTER, B., HILL, K. J., CHARALAMBOUS, M., WAGNER, K. J., LAHIRI, D., JAMES, D. I., ANDERSEN, J. S., SCHUMACHER, V., ROYER-POKORA, B., MANN, M., WARD, A. & ROBERTS, S. G. E. 2004. BASP1 Is a Transcriptional Cosuppressor for the Wilms' Tumor Suppressor Protein WT1. *Molecular and Cellular Biology,* 24**,** 537-549.

CARULLO, V. N. N. & DAY, J. J. 2019. Genomic Enhancers in Brain Health and Disease. *Genes,* 10.

CASSANDRI, M., SMIRNOV, A., NOVELLI, F., PITOLLI, C., AGOSTINI, M., MALEWICZ, M., MELINO, G. & RASCHELLÀ, G. 2017. Zinc-finger proteins in health and disease. *Cell Death Discovery,* 3**,** 17071.

CAUCHY, P., JAMES, SALLY R., ZACARIAS-CABEZA, J., PTASINSKA, A., IMPERATO, MARIA R., ASSI, SALAM A., PIPER, J., CANESTRARO, M., HOOGENKAMP, M., RAGHAVAN, M., LOKE, J., AKIKI, S., CLOKIE, SAMUEL J., RICHARDS, STEPHEN J., WESTHEAD, DAVID R., GRIFFITHS, MICHAEL J., OTT, S., BONIFER, C. & COCKERILL, PETER N. 2015. Chronic FLT3-ITD

Signaling in Acute Myeloid Leukemia Is Connected to a Specific Chromatin Signature. *Cell Reports,* 12**,** 821-836.

CAZZANIGA, G., VAN DELFT, F. W., LO NIGRO, L., FORD, A. M., SCORE, J., IACOBUCCI, I., MIRABILE, E., TAJ, M., COLMAN, S. M., BIONDI, A. & GREAVES, M. 2011. Developmental origins and impact of BCR-ABL1 fusion and IKZF1 deletions in monozygotic twins with Ph+ acute lymphoblastic leukemia. *Blood,* 118**,** 5559-64.

CHALLEN, G. A., SUN, D., JEONG, M., LUO, M., JELINEK, J., BERG, J. S., BOCK, C., VASANTHAKUMAR, A., GU, H., XI, Y., LIANG, S., LU, Y., DARLINGTON, G. J., MEISSNER, A., ISSA, J. P., GODLEY, L. A., LI, W. & GOODELL, M. A. 2011. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet,* 44**,** 23-31.

CHAPUIS, A. G., EGAN, D. N., BAR, M., SCHMITT, T. M., MCAFEE, M. S., PAULSON, K. G., VOILLET, V., GOTTARDO, R., RAGNARSSON, G. B., BLEAKLEY, M., YEUNG, C. C., MUHLHAUSER, P., NGUYEN, H. N., KROPP, L. A., CASTELLI, L., WAGENER, F., HUNTER, D., LINDBERG, M., COHEN, K., SEESE, A., MCELRATH, M. J., DUERKOPP, N., GOOLEY, T. A. & GREENBERG, P. D. 2019. T cell receptor gene therapy targeting WT1 prevents acute myeloid leukemia relapse post-transplant. *Nature Medicine,* 25**,** 1064-1072.

CHARROT, S., ARMES, H., RIO-MACHIN, A. & FITZGIBBON, J. 2020. AML through the prism of molecular genetics. *British Journal of Haematology,* 188**,** 49-62.

CHAU, Y.-Y., BROWNSTEIN, D., MJOSENG, H., LEE, W.-C., BUZA-VIDAS, N., NERLOV, C., JACOBSEN, S. E., PERRY, P., BERRY, R., THORNBURN, A., SEXTON, D., MORTON, N., HOHENSTEIN, P., FREYER, E., SAMUEL, K., VAN'T HOF, R. & HASTIE, N. 2011. Acute Multiple Organ Failure in Adult Mice Deleted for the Developmental Regulator Wt1. *PLOS Genetics,* 7**,** e1002404.

CHAVEZ, J., RABE, J. L., HIGA, K., LOEFFLER, D., NOURAIZ, A., MILLS, T., IDLER, B. M., STEVENS, B. M., FLEENOR, C., NERLOV, C., NAKAJIMA, H., HAGMAN, J., SCHROEDER, T., KIM, H. M., JORDAN, C. T., DEGREGORI, J. & PIETRAS, E. 2019. PU.1 Enforces Hematopoietic Stem Cell Quiescence during Chronic Inflammation. *Blood,* 134**,** 822-822.

CHEN, F. E., HUANG, D.-B., CHEN, Y.-Q. & GHOSH, G. 1998. Crystal structure of p50/p65 heterodimer of transcription factor NF-κB bound to DNA. *Nature,* 391**,** 410-413.

CHEUNG, A. C. & CRAMER, P. 2011. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature,* 471**,** 249-53.

CHIOCCHETTI, A., TOLOSANO, E., HIRSCH, E., SILENGO, L. & ALTRUDA, F. 1997. Green fluorescent protein as a reporter of gene expression in transgenic mice. *Biochimica et biophysica acta,* 1352**,** 193-202.

CHOU, F.-S., GRIESINGER, A., WUNDERLICH, M., LIN, S., LINK, K. A., SHRESTHA, M., GOYAMA, S., MIZUKAWA, B., SHEN, S., MARCUCCI, G. & MULLOY, J. C. 2012. The thrombopoietin/MPL/Bcl-xL pathway is essential for survival and self-renewal in human preleukemia induced by AML1-ETO. *Blood,* 120**,** 709.

CHOU, F.-S., WUNDERLICH, M., GRIESINGER, A. & MULLOY, J. C. 2011. N-Ras<sup>G12D</sup> induces features of stepwise transformation in preleukemic human umbilical cord blood cultures expressing the AML1-ETO fusion gene. *Blood,* 117**,** 2237.

CIRILLO, L. A. & ZARET, K. S. 1999. An Early Developmental Transcription Factor Complex that Is More Stable on Nucleosome Core Particles Than on Free DNA. *Molecular Cell,* 4**,** 961-969.

CLAPIER, C. R. & CAIRNS, B. R. 2012. Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature,* 492**,** 280-4.

CLARKSON, R. W., SHANG, C. A., LEVITT, L. K., HOWARD, T. & WATERS, M. J. 1999. Ternary complex factors Elk-1 and Sap-1a mediate growth hormone-induced transcription of

egr-1 (early growth response factor-1) in 3T3-F442A preadipocytes. *Mol Endocrinol,* 13**,** 619-31.

COCKERILL, P. N. 2011. Structure and function of active chromatin and DNase I hypersensitive sites. *The FEBS Journal,* 278**,** 2182-2210.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

CORCES-ZIMMERMAN, M. R., HONG, W.-J., WEISSMAN, I. L., MEDEIROS, B. C. & MAJETI, R. 2014. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proceedings of the National Academy of Sciences,* 111**,** 2548.

CORCES, M. R., BUENROSTRO, J. D., WU, B., GREENSIDE, P. G., CHAN, S. M., KOENIG, J. L., SNYDER, M. P., PRITCHARD, J. K., KUNDAJE, A., GREENLEAF, W. J., MAJETI, R. & CHANG, H. Y. 2016. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet,* 48**,** 1193-203.

CORE, L. J., MARTINS, A. L., DANKO, C. G., WATERS, C. T., SIEPEL, A. & LIS, J. T. 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet,* 46**,** 1311-20.

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. & JAENISCH, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A,* 107**,** 21931-6.

CURRAN, T. & TEICH, N. M. 1982. Identification of a 39,000-dalton protein in cells transformed by the FBJ murine osteosarcoma virus. *Virology,* 116**,** 221-35.

DAO, L. T. M., GALINDO-ALBARRAN, A. O., CASTRO-MONDRAGON, J. A., ANDRIEU-SOLER, C., MEDINA-RIVERA, A., SOUAID, C., CHARBONNIER, G., GRIFFON, A., VANHILLE, L., STEPHEN, T., ALOMAIRI, J., MARTIN, D., TORRES, M., FERNANDEZ, N., SOLER, E., VAN HELDEN, J., PUTHIER, D. & SPICUGLIA, S. 2017. Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat Genet,* 49**,** 1073-1081.

DAVIES, R. C., CALVIO, C., BRATT, E., LARSSON, S. H., LAMOND, A. I. & HASTIE, N. D. 1998. WT1 interacts with the splicing factor U2AF65 in an isoform-dependent manner and can be incorporated into spliceosomes. *Genes & Development,* 12**,** 3217-3225.

DE BOER, B., PRICK, J., PRUIS, M. G., KEANE, P., IMPERATO, M. R., JAQUES, J., BROUWERS-VOS, A. Z., HOGELING, S. M., WOOLTHUIS, C. M., NIJK, M. T., DIEPSTRA, A., WANDINGER, S., VERSELE, M., ATTAR, R. M., COCKERILL, P. N., HULS, G., VELLENGA, E., MULDER, A. B., BONIFER, C. & SCHURINGA, J. J. 2018. Prospective Isolation and Characterization of Genetically and Functionally Distinct AML Subclones. *Cancer Cell,* 34**,** 674-689.e8.

DE WET, J. R., WOOD, K. V., DELUCA, M., HELINSKI, D. R. & SUBRAMANI, S. 1987. Firefly luciferase gene: structure and expression in mammalian cells. *Mol Cell Biol,* 7**,** 725-37.

DELTCHEVA, E., CHYLINSKI, K., SHARMA, C. M., GONZALES, K., CHAO, Y., PIRZADA, Z. A., ECKERT, M. R., VOGEL, J. & CHARPENTIER, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature,* 471**,** 602-7.

DÉRIJARD, B., HIBI, M., WU, I. H., BARRETT, T., SU, B., DENG, T., KARIN, M. & DAVIS, R. J. 1994. JNK1: a protein kinase stimulated by UV light and Ha-Ras that binds and phosphorylates the c-Jun activation domain. *Cell,* 76**,** 1025-37.

DEVARY, Y., GOTTLIEB, R. A., SMEAL, T. & KARIN, M. 1992. The mammalian ultraviolet response is triggered by activation of src tyrosine kinases. *Cell,* 71**,** 1081-1091.

DIAO, Y., FANG, R., LI, B., MENG, Z., YU, J., QIU, Y., LIN, K. C., HUANG, H., LIU, T., MARINA, R. J., JUNG, I., SHEN, Y., GUAN, K. L. & REN, B. 2017. A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods,* 14**,** 629-635.

DING, L., LEY, T. J., LARSON, D. E., MILLER, C. A., KOBOLDT, D. C., WELCH, J. S., RITCHEY, J. K., YOUNG, M. A., LAMPRECHT, T., MCLELLAN, M. D., MCMICHAEL, J. F., WALLIS, J. W., LU, C., SHEN, D., HARRIS, C. C., DOOLING, D. J., FULTON, R. S., FULTON, L. L., CHEN, K., SCHMIDT, H., KALICKI-VEIZER, J., MAGRINI, V. J., COOK, L., MCGRATH, S. D., VICKERY, T. L., WENDL, M. C., HEATH, S., WATSON, M. A., LINK, D. C., TOMASSON, M. H., SHANNON, W. D., PAYTON, J. E., KULKARNI, S., WESTERVELT, P., WALTER, M. J., GRAUBERT, T. A., MARDIS, E. R., WILSON, R. K. & DIPERSIO, J. F. 2012. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature,* 481**,** 506-10.

DIXON, J. R., SELVARAJ, S., YUE, F., KIM, A., LI, Y., SHEN, Y., HU, M., LIU, J. S. & REN, B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature,* 485**,** 376-80.

DOENCH, J. G., FUSI, N., SULLENDER, M., HEGDE, M., VAIMBERG, E. W., DONOVAN, K. F., SMITH, I., TOTHOVA, Z., WILEN, C., ORCHARD, R., VIRGIN, H. W., LISTGARTEN, J. & ROOT, D. E. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology,* 34**,** 184.

DOMCKE, S., BARDET, A. F., ADRIAN GINNO, P., HARTL, D., BURGER, L. & SCHÜBELER, D. 2015. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature,* 528**,** 575.

DONI JAYAVELU, N., JAJODIA, A., MISHRA, A. & HAWKINS, R. D. 2020. Candidate silencer elements for the human and mouse genomes. *Nature Communications,* 11**,** 1061.

DOULATOV, S., NOTTA, F., EPPERT, K., NGUYEN, L. T., OHASHI, P. S. & DICK, J. E. 2010. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat Immunol,* 11**,** 585-93.

DOVER, J., SCHNEIDER, J., TAWIAH-BOATENG, M. A., WOOD, A., DEAN, K., JOHNSTON, M. & SHILATIFARD, A. 2002. Methylation of Histone H3 by COMPASS Requires Ubiquitination of Histone H2B by Rad6. *Journal of Biological Chemistry,* 277**,** 28368-28371.

DROR, I., GOLAN, T., LEVY, C., ROHS, R. & MANDEL-GUTFREUND, Y. 2015. A widespread role of the motif environment in transcription factor binding across diverse protein families. *Genome Research*.

DUARTE, A., CARICASOLE, A., GRAHAM, C. F. & WARD, A. 1998. Wilms' tumour-suppressor protein isoforms have opposite effects on Igf2 expression in primary embryonic cells, independently of p53 genotype. *British Journal of Cancer,* 77**,** 253-259.

DUPLOYEZ, N., MARCEAU-RENAUT, A., BOISSEL, N., PETIT, A., BUCCI, M., GEFFROY, S., LAPILLONNE, H., RENNEVILLE, A., RAGU, C., FIGEAC, M., CELLI-LEBRAS, K., LACOMBE, C., MICOL, J.-B., ABDEL-WAHAB, O., CORNILLET, P., IFRAH, N., DOMBRET, H., LEVERGER, G., JOURDAN, E. & PREUDHOMME, C. 2016. Comprehensive mutational profiling of core binding factor acute myeloid leukemia. *Blood,* 127**,** 2451.

DVINGE, H., KIM, E., ABDEL-WAHAB, O. & BRADLEY, R. K. 2016. RNA splicing factors as oncoproteins and tumour suppressors. *Nature Reviews Cancer,* 16**,** 413-430.

DYKSTRA, B., KENT, D., BOWIE, M., MCCAFFREY, L., HAMILTON, M., LYONS, K., LEE, S. J., BRINKMAN, R. & EAVES, C. 2007. Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell,* 1**,** 218-29.

EDIRIWICKREMA, A., ALESHIN, A., REITER, J. G., CORCES, M. R., KÖHNKE, T., STAFFORD, M., LIEDTKE, M., MEDEIROS, B. C. & MAJETI, R. 2020. Single-cell mutational profiling enhances the clinical evaluation of AML MRD. *Blood Advances,* 4**,** 943-952.

EFERL, R., SIBILIA, M., HILBERG, F., FUCHSBICHLER, A., KUFFERATH, I., GUERTL, B., ZENZ, R., WAGNER, E. F. & ZATLOUKAL, K. 1999. Functions of c-Jun in Liver and Heart Development. *The Journal of Cell Biology,* 145**,** 1049.

EGAN, S. E., GIDDINGS, B. W., BROOKS, M. W., BUDAY, L., SIZELAND, A. M. & WEINBERG, R. A. 1993. Association of Sos Ras exchange protein with Grb2 is implicated in tyrosine kinase signal transduction and transformation. *Nature,* 363**,** 45-51.

ELLISEN, L. W., CARLESSO, N., CHENG, T., SCADDEN, D. T. & HABER, D. A. 2001. The Wilms tumor suppressor WT1 directs stage-specific quiescence and differentiation of human hematopoietic progenitor cells. *The EMBO Journal,* 20**,** 1897.

ELSÄSSER, A., FRANZEN, M., KOHLMANN, A., WEISSER, M., SCHNITTGER, S., SCHOCH, C., REDDY, V. A., BUREL, S., ZHANG, D.-E., UEFFING, M., TENEN, D. G., HIDDEMANN, W. & BEHRE, G. 2003. The fusion protein AML1-ETO in acute myeloid leukemia with translocation t(8;21) induces c-jun protein expression via the proximal AP-1 site of the c-jun promoter in an indirect, JNK-dependent manner. *Oncogene,* 22**,** 5646-5657.

ENGEL, C., GUBBEY, T., NEYER, S., SAINSBURY, S., OBERTHUER, C., BAEJEN, C., BERNECKY, C. & CRAMER, P. 2017. Structural Basis of RNA Polymerase I Transcription Initiation. *Cell,* 169**,** 120-131.e22.

ERICKSON, P., GAO, J., CHANG, K. S., LOOK, T., WHISENANT, E., RAIMONDI, S., LASHER, R., TRUJILLO, J., ROWLEY, J. & DRABKIN, H. 1992. Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to Drosophila segmentation gene, runt. *Blood,* 80**,** 1825.

ERNST, J., KHERADPOUR, P., MIKKELSEN, T. S., SHORESH, N., WARD, L. D., EPSTEIN, C. B., ZHANG, X., WANG, L., ISSNER, R., COYNE, M., KU, M., DURHAM, T., KELLIS, M. & BERNSTEIN, B. E. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature,* 473**,** 43-9.

FEI, J., TORIGOE, S. E., BROWN, C. R., KHUONG, M. T., KASSAVETIS, G. A., BOEGER, H. & KADONAGA, J. T. 2015. The prenucleosome, a stable conformational isomer of the nucleosome. *Genes & Development,* 29**,** 2563-2575.

FERNÁNDEZ DE MATTOS, S., ESSAFI, A., SOEIRO, I., PIETERSEN, A. M., BIRKENKAMP, K. U., EDWARDS, C. S., MARTINO, A., NELSON, B. H., FRANCIS, J. M., JONES, M. C., BROSENS, J. J., COFFER, P. J. & LAM, E. W. 2004. FoxO3a and BCR-ABL regulate cyclin D2 transcription through a STAT5/BCL6-dependent mechanism. *Mol Cell Biol,* 24**,** 10058-71.

FIGUEROA, M. E., ABDEL-WAHAB, O., LU, C., WARD, P. S., PATEL, J., SHIH, A., LI, Y., BHAGWAT, N., VASANTHAKUMAR, A., FERNANDEZ, H. F., TALLMAN, M. S., SUN, Z., WOLNIAK, K., PEETERS, J. K., LIU, W., CHOE, S. E., FANTIN, V. R., PAIETTA, E., LOWENBERG, B., LICHT, J. D., GODLEY, L. A., DELWEL, R., VALK, P. J., THOMPSON, C. B., LEVINE, R. L. & MELNICK, A. 2010. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell,* 18**,** 553-67.

FINICLE, B. T., JAYASHANKAR, V. & EDINGER, A. L. 2018. Nutrient scavenging in cancer. *Nat Rev Cancer,* 18**,** 619-633.

FINN, R. S., CROWN, J. P., LANG, I., BOER, K., BONDARENKO, I. M., KULYK, S. O., ETTL, J., PATEL, R., PINTER, T., SCHMIDT, M., SHPARYK, Y., THUMMALA, A. R., VOYTKO, N. L., FOWST, C., HUANG, X., KIM, S. T., RANDOLPH, S. & SLAMON, D. J. 2015. The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. *Lancet Oncol,* 16**,** 25-35.

FIRE, A., ALBERTSON, D., HARRISON, S. W. & MOERMAN, D. G. 1991. Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. *Development,* 113**,** 503.

FORRESTER, W. C., THOMPSON, C., ELDER, J. T. & GROUDINE, M. 1986. A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proceedings of the National Academy of Sciences,* 83**,** 1359.

FORSTER, V. J., NAHARI, M. H., MARTINEZ-SORIA, N., BRADBURN, A. K., PTASINSKA, A., ASSI, S. A., FORDHAM, S. E., MCNEIL, H., BONIFER, C., HEIDENREICH, O. & ALLAN, J. M. 2015. The leukemia-associated RUNX1/ETO oncoprotein confers a mutator phenotype. *Leukemia,* 30**,** 251.

FRANK, R. C., SUN, X., BERGUIDO, F. J., JAKUBOWIAK, A. & NIMER, S. D. 1999. The t(8;21) fusion protein, AML1/ETO, transforms NIH3T3 cells and activates AP-1. *Oncogene,* 18**,** 1701-1710.

FRANZA, B. R., RAUSCHER, F. J., JOSEPHS, S. F. & CURRAN, T. 1988. The Fos complex and Fos-related antigens recognize sequence elements that contain AP-1 binding sites. *Science,* 239**,** 1150.

FREE, R. B., HAZELWOOD, L. A. & SIBLEY, D. R. 2009. Identifying novel protein-protein interactions using co-immunoprecipitation and mass spectroscopy. *Current protocols in neuroscience,* Chapter 5**,** Unit-5.28.

FRITSCH, M., GÜNTHER, S. D., SCHWARZER, R., ALBERT, M.-C., SCHORN, F., WERTHENBACH, J. P., SCHIFFMANN, L. M., STAIR, N., STOCKS, H., SEEGER, J. M., LAMKANFI, M., KRÖNKE, M., PASPARAKIS, M. & KASHKAR, H. 2019. Caspase-8 is the molecular switch for apoptosis, necroptosis and pyroptosis. *Nature,* 575**,** 683-687.

FU, Y., SANDER, J. D., REYON, D., CASCIO, V. M. & JOUNG, J. K. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology,* 32**,** 279.

FURUHATA, A., MURAKAMI, M., ITO, H., GAO, S., YOSHIDA, K., SOBUE, S., KIKUCHI, R., IWASAKI, T., TAKAGI, A., KOJIMA, T., SUZUKI, M., ABE, A., NAOE, T. & MURATE, T. 2009. GATA-1 and GATA-2 binding to 3' enhancer of WT1 gene is essential for its transcription in acute leukemia and solid tumor cell lines. *Leukemia,* 23**,** 1270-1277.

GARNEAU, J. E., DUPUIS, M. E., VILLION, M., ROMERO, D. A., BARRANGOU, R., BOYAVAL, P., FREMAUX, C., HORVATH, P., MAGADAN, A. H. & MOINEAU, S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature,* 468**,** 67-71.

GELMETTI, V., ZHANG, J., FANELLI, M., MINUCCI, S., PELICCI, P. G. & LAZAR, M. A. 1998. Aberrant Recruitment of the Nuclear Receptor Corepressor-Histone Deacetylase Complex by the Acute Myeloid Leukemia Fusion Partner ETO. *Molecular and Cellular Biology,* 18**,** 7185-7191.

GILHAM, C., PETO, J., SIMPSON, J., ROMAN, E., EDEN, T. O., GREAVES, M. F. & ALEXANDER, F. E. 2005. Day care in infancy and risk of childhood acute lymphoblastic leukaemia: findings from UK case-control study. *Bmj,* 330**,** 1294.

GILLESPIE, M. A., PALII, C. G., SANCHEZ-TALTAVULL, D., SHANNON, P., LONGABAUGH, W. J. R., DOWNES, D. J., SIVARAMAN, K., HUGHES, J. R., PRICE, N. D., PERKINS, T. J., RANISH, J. A. & BRAND, M. 2019. Absolute quantification of transcription factors reveals principles of gene regulation in erythropoiesis. *bioRxiv***,** 812123.

GILMOUR, J., O'CONNOR, L., MIDDLETON, C. P., KEANE, P., GILLEMANS, N., CAZIER, J.-B., PHILIPSEN, S. & BONIFER, C. 2019. Robust hematopoietic specification requires the ubiquitous Sp1 and Sp3 transcription factors. *Epigenetics & chromatin,* 12**,** 33-33.

GLISOVIC, T., BACHORIK, J. L., YONG, J. & DREYFUSS, G. 2008. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters,* 582**,** 1977-1986.

GOARDON, N., MARCHI, E., ATZBERGER, A., QUEK, L., SCHUH, A., SONEJI, S., WOLL, P., MEAD, A., ALFORD, K. A., ROUT, R., CHAUDHURY, S., GILKES, A., KNAPPER, S., BELDJORD, K., BEGUM, S., ROSE, S., GEDDES, N., GRIFFITHS, M., STANDEN, G., STERNBERG, A., CAVENAGH, J., HUNTER, H., BOWEN, D., KILLICK, S., ROBINSON, L., PRICE, A., MACINTYRE, E., VIRGO, P., BURNETT, A., CRADDOCK, C., ENVER, T., JACOBSEN, S. E. W., PORCHER, C. & VYAS, P. 2011. Coexistence of LMPP-like and GMP-like Leukemia Stem Cells in Acute Myeloid Leukemia. *Cancer Cell,* 19**,** 138-152.

GORDÂN, R., SHEN, N., DROR, I., ZHOU, T., HORTON, J., ROHS, R. & BULYK, MARTHA L. 2013. Genomic Regions Flanking E-Box Binding Sites Influence DNA Binding Specificity of bHLH Transcription Factors through DNA Shape. *Cell Reports,* 3**,** 1093-1104.

GOYAMA, S., SCHIBLER, J., GASILINA, A., SHRESTHA, M., LIN, S., LINK, K. A., CHEN, J., WHITMAN, S. P., BLOOMFIELD, C. D., NICOLET, D., ASSI, S. A., PTASINSKA, A., HEIDENREICH, O., BONIFER, C., KITAMURA, T., NASSAR, N. N. & MULLOY, J. C. 2016. UBASH3B/Sts-1-CBL axis regulates myeloid proliferation in human preleukemia induced by AML1-ETO. *Leukemia,* 30**,** 728-39.

GRUNBERG, S., WARFIELD, L. & HAHN, S. 2012. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat Struct Mol Biol,* 19**,** 788-96.

GU, W., HU, S., CHEN, Z., QIU, G., CEN, J., HE, B., HE, J. & WU, W. 2010. High expression of WT1 gene in acute myeloid leukemias with more predominant WT1+17AA isoforms at relapse. *Leukemia Research,* 34**,** 46-49.

GU, X., HU, Z., EBRAHEM, Q., CRABB, J. S., MAHFOUZ, R. Z., RADIVOYEVITCH, T., CRABB, J. W. & SAUNTHARARAJAH, Y. 2014. Runx1 Regulation of Pu.1 Corepressor/Coactivator Exchange Identifies Specific Molecular Targets for Leukemia Differentiation Therapy. *Journal of Biological Chemistry,* 289**,** 14881-14895.

GUALDI, R., BOSSARD, P., ZHENG, M., HAMADA, Y., COLEMAN, J. R. & ZARET, K. S. 1996. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev,* 10**,** 1670-82.

GUCCIONE, E., BASSI, C., CASADIO, F., MARTINATO, F., CESARONI, M., SCHUCHLAUTZ, H., LUSCHER, B. & AMATI, B. 2007. Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. *Nature,* 449**,** 933-7.

GUELEN, L., PAGIE, L., BRASSET, E., MEULEMAN, W., FAZA, M. B., TALHOUT, W., EUSSEN, B. H., DE KLEIN, A., WESSELS, L., DE LAAT, W. & VAN STEENSEL, B. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature,* 453**,** 948.

GUPTA, S., CAMPBELL, D., DERIJARD, B. & DAVIS, R. J. 1995. Transcription factor ATF2 regulation by the JNK signal transduction pathway. *Science,* 267**,** 389.

HABER, D. A., SOHN, R. L., BUCKLER, A. J., PELLETIER, J., CALL, K. M. & HOUSMAN, D. E. 1991. Alternative splicing and genomic structure of the Wilms tumor gene WT1. *Proceedings of the National Academy of Sciences,* 88**,** 9618.

HABERLE, V. & STARK, A. 2018. Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology,* 19**,** 621-637.

HAI, T. & CURRAN, T. 1991. Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. *Proceedings of the National Academy of Sciences,* 88**,** 3720.

HAMMES, A., GUO, J.-K., LUTSCH, G., LEHESTE, J.-R., LANDROCK, D., ZIEGLER, U., GUBLER, M.-C. & SCHEDL, A. 2001. Two Splice Variants of the Wilms' Tumor 1 Gene Have Distinct Functions during Sex Determination and Nephron Formation. *Cell,* 106**,** 319-329.

HAN, M. & GRUNSTEIN, M. 1988. Nucleosome loss activates yeast downstream promoters in vivo. *Cell,* 55**,** 1137-1145.

HANNON, G. J. & ROSSI, J. J. 2004. Unlocking the potential of the human genome with RNA interference. *Nature,* 431**,** 371-378.

HASSA, P. O., HAENNI, S. S., ELSER, M. & HOTTIGER, M. O. 2006. Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? *Microbiol Mol Biol Rev,* 70**,** 789-829.

HEBBES, T. R., THORNE, A. W. & CRANE-ROBINSON, C. 1988. A direct link between core histone acetylation and transcriptionally active chromatin. *Embo j,* 7**,** 1395-402.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell,* 38**,** 576-589.

HEITZ, E. 1928. Das heterochromatin der Moose. Jb. Wiss. Bot. 69: 728-818. *View Article PubMed/NCBI Google Scholar*.

HEMMATI, S., HAQUE, T. & GRITSMAN, K. 2017. Inflammatory Signaling Pathways in Preleukemic and Leukemic Stem Cells. *Frontiers in Oncology,* 7**,** 265.

HEYWORTH, C., PEARSON, S., MAY, G. & ENVER, T. 2002. Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *The EMBO journal,* 21**,** 3770-3781.

HIGA, K. C., GOODSPEED, A., CHAVEZ, J. S., ZABEREZHNYY, V., RABE, J. L., TENEN, D. G., PIETRAS, E. M. & DEGREGORI, J. 2020. Chronic Interleukin-1 exposure triggers selective expansion of Cebpa-deficient multipotent hematopoietic progenitors. *bioRxiv***,** 2020.03.25.008250.

HIGUCHI, M., O'BRIEN, D., KUMARAVELU, P., LENNY, N., YEOH, E.-J. & DOWNING, J. R. 2002. Expression of a conditional AML1-ETO oncogene bypasses embryonic lethality and establishes a murine model of human t(8;21) acute myeloid leukemia. *Cancer Cell,* 1**,** 63-74.

HILTON, I. B., D'IPPOLITO, A. M., VOCKLEY, C. M., THAKORE, P. I., CRAWFORD, G. E., REDDY, T. E. & GERSBACH, C. A. 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol,* 33**,** 510-7.

HIRSCH, P., ZHANG, Y., TANG, R., JOULIN, V., BOUTROUX, H., PRONIER, E., MOATTI, H., FLANDRIN, P., MARZAC, C., BORIES, D., FAVA, F., MOKRANI, H., BETEMS, A., LORRE, F., FAVIER, R., FÉGER, F., MOHTY, M., DOUAY, L., LEGRAND, O., BILHOU-NABERA, C., LOUACHE, F. & DELHOMMEAU, F. 2016. Genetic hierarchy and temporal variegation in the clonal history of acute myeloid leukaemia. *Nature Communications,* 7**,** 12475.

HNISZ, D., ABRAHAM, BRIAN J., LEE, TONG I., LAU, A., SAINT-ANDRÉ, V., SIGOVA, ALLA A., HOKE, HEATHER A. & YOUNG, RICHARD A. 2013. Super-Enhancers in the Control of Cell Identity and Disease. *Cell,* 155**,** 934-947.

HNISZ, D., WEINTRAUB, A. S., DAY, D. S., VALTON, A.-L., BAK, R. O., LI, C. H., GOLDMANN, J., LAJOIE, B. R., FAN, Z. P., SIGOVA, A. A., REDDY, J., BORGES-RIVERA, D., LEE, T. I., JAENISCH, R., PORTEUS, M. H., DEKKER, J. & YOUNG, R. A. 2016. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science,* 351**,** 1454.

HOFFMAN, B., AMANULLAH, A., SHAFARENKO, M. & LIEBERMANN, D. A. 2002. The proto-oncogene c-myc in hematopoietic development and leukemogenesis. *Oncogene,* 21**,** 3414-21.

HÖLLEIN, A., MEGGENDORFER, M., DICKER, F., JEROMIN, S., NADARAJAH, N., KERN, W., HAFERLACH, C. & HAFERLACH, T. 2018. NPM1 mutated AML can relapse with wild-type NPM1: persistent clonal hematopoiesis can drive relapse. *Blood advances,* 2**,** 3118-3125.

HOPE, I. A. & STRUHL, K. 1985. GCN4 protein, synthesized in vitro, binds HIS3 regulatory sequences: implications for general control of amino acid biosynthetic genes in yeast. *Cell,* 43**,** 177-88.

HOPE, I. A. & STRUHL, K. 1987. GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *Embo j,* 6**,** 2781-4.

HOSEN, N., SHIRAKATA, T., NISHIDA, S., YANAGIHARA, M., TSUBOI, A., KAWAKAMI, M., OJI, Y., OKA, Y., OKABE, M., TAN, B., SUGIYAMA, H. & WEISSMAN, I. L. 2007. The Wilms' tumor gene WT1-GFP knock-in mouse reveals the dynamic regulation of WT1 expression in normal and leukemic hematopoiesis. *Leukemia,* 21**,** 1783.

HOU, H.-A., HUANG, T.-C., LIN, L.-I., LIU, C.-Y., CHEN, C.-Y., CHOU, W.-C., TANG, J.-L., TSENG, M.-H., HUANG, C.-F., CHIANG, Y.-C., LEE, F.-Y., LIU, M.-C., YAO, M., HUANG, S.-Y., KO, B.-S.,

HSU, S.-C., WU, S.-J., TSAY, W., CHEN, Y.-C. & TIEN, H.-F. 2010. WT1 mutation in 470 adult patients with acute myeloid leukemia: stability during disease evolution and implication of its incorporation into a survival scoring system. *Blood,* 115**,** 5222.

HU, S., WAN, J., SU, Y., SONG, Q., ZENG, Y., NGUYEN, H. N., SHIN, J., COX, E., RHO, H. S., WOODARD, C., XIA, S., LIU, S., LYU, H., MING, G.-L., WADE, H., SONG, H., QIAN, J. & ZHU, H. 2013. DNA methylation presents distinct binding sites for human transcription factors. *eLife,* 2**,** e00726.

ILLENDULA, A., GILMOUR, J., GREMBECKA, J., TIRUMALA, V. S. S., BOULTON, A., KUNTIMADDI, A., SCHMIDT, C., WANG, L., PULIKKAN, J. A., ZONG, H., PARLAK, M., KUSCU, C., PICKIN, A., ZHOU, Y., GAO, Y., MISHRA, L., ADLI, M., CASTILLA, L. H., RAJEWSKI, R. A., JANES, K. A., GUZMAN, M. L., BONIFER, C. & BUSHWELLER, J. H. 2016. Small Molecule Inhibitor of CBFβ-RUNX Binding for RUNX Transcription Factor Driven Cancers. *EBioMedicine,* 8**,** 117-131.

ILLENDULA, A., PULIKKAN, J. A., ZONG, H., GREMBECKA, J., XUE, L., SEN, S., ZHOU, Y., BOULTON, A., KUNTIMADDI, A., GAO, Y., RAJEWSKI, R. A., GUZMAN, M. L., CASTILLA, L. H. & BUSHWELLER, J. H. 2015. A small-molecule inhibitor of the aberrant transcription factor CBFβ-SMMHC delays leukemia in mice. *Science,* 347**,** 779.

INOUE, K., SUGIYAMA, H., OGAWA, H., NAKAGAWA, M., YAMAGAMI, T., MIWA, H., KITA, K., HIRAOKA, A., MASAOKA, T. & NASU, K. 1994. WT1 as a new prognostic factor and a new marker for the detection of minimal residual disease in acute leukemia. *Blood,* 84**,** 3071.

ITO, S., SHEN, L., DAI, Q., WU, S. C., COLLINS, L. B., SWENBERG, J. A., HE, C. & ZHANG, Y. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science,* 333**,** 1300-3.

ITO, T., IKEHARA, T., NAKAGAWA, T., KRAUS, W. L. & MURAMATSU, M. 2000. p300-Mediated acetylation facilitates the transfer of histone H2A–H2B dimers from nucleosomes to a histone chaperone. *Genes & Development,* 14**,** 1899-1907.

IWAKURA, Y., ISHIGAME, H., SAIJO, S. & NAKAE, S. 2011. Functional specialization of interleukin-17 family members. *Immunity,* 34**,** 149-62.

IWASAKI, H., SOMOZA, C., SHIGEMATSU, H., DUPREZ, E. A., IWASAKI-ARAI, J., MIZUNO, S.-I., ARINOBU, Y., GEARY, K., ZHANG, P., DAYARAM, T., FENYUS, M. L., ELF, S., CHAN, S., KASTNER, P., HUETTNER, C. S., MURRAY, R., TENEN, D. G. & AKASHI, K. 2005. Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. *Blood,* 106**,** 1590.

IZZO, F., LEE, S. C., PORAN, A., CHALIGNE, R., GAITI, F., GROSS, B., MURALI, R. R., DEOCHAND, S. D., ANG, C., JONES, P. W., NAM, A. S., KIM, K.-T., KOTHEN-HILL, S., SCHULMAN, R. C., KI, M., LHOUMAUD, P., SKOK, J. A., VINY, A. D., LEVINE, R. L., KENIGSBERG, E., ABDEL-WAHAB, O. & LANDAU, D. A. 2020. DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nature Genetics*.

JAISWAL, S., FONTANILLAS, P., FLANNICK, J., MANNING, A., GRAUMAN, P. V., MAR, B. G., LINDSLEY, R. C., MERMEL, C. H., BURTT, N., CHAVEZ, A., HIGGINS, J. M., MOLTCHANOV, V., KUO, F. C., KLUK, M. J., HENDERSON, B., KINNUNEN, L., KOISTINEN, H. A., LADENVALL, C., GETZ, G., CORREA, A., BANAHAN, B. F., GABRIEL, S., KATHIRESAN, S., STRINGHAM, H. M., MCCARTHY, M. I., BOEHNKE, M., TUOMILEHTO, J., HAIMAN, C., GROOP, L., ATZMON, G., WILSON, J. G., NEUBERG, D., ALTSHULER, D. & EBERT, B. L. 2014. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine,* 371**,** 2488-2498.

JAN, M., SNYDER, T. M., CORCES-ZIMMERMAN, M. R., VYAS, P., WEISSMAN, I. L., QUAKE, S. R. & MAJETI, R. 2012. Clonal Evolution of Preleukemic Hematopoietic Stem Cells Precedes Human Acute Myeloid Leukemia. *Science Translational Medicine,* 4**,** 149ra118.

JAVIERRE, B. M., BURREN, O. S., WILDER, S. P., KREUZHUBER, R., HILL, S. M., SEWITZ, S., CAIRNS, J., WINGETT, S. W., VÁRNAI, C., THIECKE, M. J., BURDEN, F., FARROW, S., CUTLER, A. J., REHNSTRÖM, K., DOWNES, K., GRASSI, L., KOSTADIMA, M., FREIRE-PRITCHETT, P., WANG, F., CONSORTIUM, B., STUNNENBERG, H. G., TODD, J. A., ZERBINO, D. R., STEGLE, O., OUWEHAND, W. H., FRONTINI, M., WALLACE, C., SPIVAKOV, M. & FRASER, P. 2016. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell,* 167**,** 1369-1384.e19.

JI, H., EHRLICH, L. I. R., SEITA, J., MURAKAMI, P., DOI, A., LINDAU, P., LEE, H., ARYEE, M. J., IRIZARRY, R. A., KIM, K., ROSSI, D. J., INLAY, M. A., SERWOLD, T., KARSUNKY, H., HO, L., DALEY, G. Q., WEISSMAN, I. L. & FEINBERG, A. P. 2010. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature,* 467**,** 338-342.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337**,** 816-21.

JING, H., VAKOC, C. R., YING, L., MANDAT, S., WANG, H., ZHENG, X. & BLOBEL, G. A. 2008. Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Mol Cell,* 29**,** 232-42.

JOCHUM, W., DAVID, J. P., ELLIOTT, C., WUTZ, A., PLENK, H., JR., MATSUO, K. & WAGNER, E. F. 2000. Increased bone formation and osteosclerosis in mice overexpressing the transcription factor Fra-1. *Nat Med,* 6**,** 980-4.

JOLMA, A., YIN, Y., NITTA, K. R., DAVE, K., POPOV, A., TAIPALE, M., ENGE, M., KIVIOJA, T., MORGUNOVA, E. & TAIPALE, J. 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature,* 527**,** 384-388.

JONES, L. K., NEAT, M. J., VAN DELFT, F. W., MITCHELL, M. P., ADAMAKI, M., STONEHAM, S. J., PATEL, N. & SAHA, V. 2003. Cryptic rearrangement involving MLL and AF10 occurring in utero. *Leukemia,* 17**,** 1667-1669.

KAGEY, M. H., NEWMAN, J. J., BILODEAU, S., ZHAN, Y., ORLANDO, D. A., VAN BERKUM, N. L., EBMEIER, C. C., GOOSSENS, J., RAHL, P. B., LEVINE, S. S., TAATJES, D. J., DEKKER, J. & YOUNG, R. A. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature,* 467**,** 430-5.

KAIDA, D., MOTOYOSHI, H., TASHIRO, E., NOJIMA, T., HAGIWARA, M., ISHIGAMI, K., WATANABE, H., KITAHARA, T., YOSHIDA, T., NAKAJIMA, H., TANI, T., HORINOUCHI, S. & YOSHIDA, M. 2007. Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nature Chemical Biology,* 3**,** 576-583.

KALSOTRA, A. & COOPER, T. A. 2011. Functional consequences of developmentally regulated alternative splicing. *Nature Reviews Genetics,* 12**,** 715-729.

KAMEI, Y., XU, L., HEINZEL, T., TORCHIA, J., KUROKAWA, R., GLOSS, B., LIN, S.-C., HEYMAN, R. A., ROSE, D. W., GLASS, C. K. & ROSENFELD, M. G. 1996. A CBP Integrator Complex Mediates Transcriptional Activation and AP-1 Inhibition by Nuclear Receptors. *Cell,* 85**,** 403-414.

KAMPEN, K. R. 2012. The discovery and early understanding of leukemia. *Leukemia Research,* 36**,** 6-13.

KANN, M., ETTOU, S., JUNG, Y. L., LENZ, M. O., TAGLIENTI, M. E., PARK, P. J., SCHERMER, B., BENZING, T. & KREIDBERG, J. A. 2015. Genome-Wide Analysis of Wilms' Tumor 1-Controlled Gene Expression in Podocytes Reveals Key Regulatory Mechanisms. *J Am Soc Nephrol,* 26**,** 2097-104.

KAROLCHIK, D., HINRICHS, A. S., FUREY, T. S., ROSKIN, K. M., SUGNET, C. W., HAUSSLER, D. & KENT, W. J. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Research,* 32**,** D493-D496.

KATO, J., MATSUSHIME, H., HIEBERT, S. W., EWEN, M. E. & SHERR, C. J. 1993. Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4. *Genes Dev,* 7**,** 331-42.

KATSUMURA, K. R. & BRESNICK, E. H. 2017. The GATA factor revolution in hematology. *Blood,* 129**,** 2092-2102.

KAWAGOE, H., KANDILCI, A., KRANENBURG, T. A. & GROSVELD, G. C. 2007. Overexpression of N-Myc Rapidly Causes Acute Myeloid Leukemia in Mice. *Cancer Research,* 67**,** 10677.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome Res,* 12**,** 996-1006.

KESARWANI, M., KINCAID, Z., GOMAA, A., HUBER, E., ROHRABAUGH, S., SIDDIQUI, Z., BOUSO, M. F., LATIF, T., XU, M., KOMUROV, K., MULLOY, J. C., CANCELAS, J. A., GRIMES, H. L. & AZAM, M. 2017. Targeting c-FOS and DUSP1 abrogates intrinsic resistance to tyrosine-kinase inhibitor therapy in BCR-ABL-induced leukemia. *Nat Med,* 23**,** 472-482.

KHVOROVA, A., REYNOLDS, A. & JAYASENA, S. D. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell,* 115**,** 209-16.

KIM, D., LANGMEAD, B. & SALZBERG, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods,* 12**,** 357-360.

KIM, M.-S., PINTO, S. M., GETNET, D., NIRUJOGI, R. S., MANDA, S. S., CHAERKADY, R., MADUGUNDU, A. K., KELKAR, D. S., ISSERLIN, R., JAIN, S., THOMAS, J. K., MUTHUSAMY, B., LEAL-ROJAS, P., KUMAR, P., SAHASRABUDDHE, N. A., BALAKRISHNAN, L., ADVANI, J., GEORGE, B., RENUSE, S., SELVAN, L. D. N., PATIL, A. H., NANJAPPA, V., RADHAKRISHNAN, A., PRASAD, S., SUBBANNAYYA, T., RAJU, R., KUMAR, M., SREENIVASAMURTHY, S. K., MARIMUTHU, A., SATHE, G. J., CHAVAN, S., DATTA, K. K., SUBBANNAYYA, Y., SAHU, A., YELAMANCHI, S. D., JAYARAM, S., RAJAGOPALAN, P., SHARMA, J., MURTHY, K. R., SYED, N., GOEL, R., KHAN, A. A., AHMAD, S., DEY, G., MUDGAL, K., CHATTERJEE, A., HUANG, T.-C., ZHONG, J., WU, X., SHAW, P. G., FREED, D., ZAHARI, M. S., MUKHERJEE, K. K., SHANKAR, S., MAHADEVAN, A., LAM, H., MITCHELL, C. J., SHANKAR, S. K., SATISHCHANDRA, P., SCHROEDER, J. T., SIRDESHMUKH, R., MAITRA, A., LEACH, S. D., DRAKE, C. G., HALUSHKA, M. K., PRASAD, T. S. K., HRUBAN, R. H., KERR, C. L., BADER, G. D., IACOBUZIO-DONAHUE, C. A., GOWDA, H. & PANDEY, A. 2014. A draft map of the human proteome. *Nature,* 509**,** 575-581.

KIM, S.-C., HAHN, J.-S., MIN, Y.-H., YOO, N.-C., KO, Y.-W. & LEE, W.-J. 1999. Constitutive Activation of Extracellular Signal-Regulated Kinase in Human Acute Leukemias: Combined Role of Activation of MEK, Hyperexpression of Extracellular Signal-Regulated Kinase, and Downregulation of a Phosphatase, PAC1. *Blood,* 93**,** 3893-3899.

KIM, T. K. & SHIEKHATTAR, R. 2015. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell,* 162**,** 948-59.

KING-UNDERWOOD, L., LITTLE, S., BAKER, M., CLUTTERBUCK, R., DELASSUS, S., ENVER, T., LEBOZER, C., MIN, T., MOORE, A., SCHEDL, A. & PRITCHARD-JONES, K. 2005. Wt1 is not essential for hematopoiesis in the mouse. *Leuk Res,* 29**,** 803-12.

KING-UNDERWOOD, L., RENSHAW, J. & PRITCHARD-JONES, K. 1996. Mutations in the Wilms' tumor gene WT1 in leukemias. *Blood,* 87**,** 2171.

KNEZETIC, J. A. & LUSE, D. S. 1986. The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell,* 45**,** 95-104.

KO, M., HUANG, Y., JANKOWSKA, A. M., PAPE, U. J., TAHILIANI, M., BANDUKWALA, H. S., AN, J., LAMPERTI, E. D., KOH, K. P., GANETZKY, R., LIU, X. S., ARAVIND, L., AGARWAL, S., MACIEJEWSKI, J. P. & RAO, A. 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature,* 468**,** 839-43.

KOLOVOS, P., KNOCH, T. A., GROSVELD, F. G., COOK, P. R. & PAPANTONIS, A. 2012. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics & Chromatin,* 5**,** 1.

KOMOR, A. C., KIM, Y. B., PACKER, M. S., ZURIS, J. A. & LIU, D. R. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature,* 533**,** 420-4.

KONDO, M., WEISSMAN, I. L. & AKASHI, K. 1997. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell,* 91**,** 661-72.

KORNBERG, R. D. 2005. Mediator and the mechanism of transcriptional activation. *Trends in Biochemical Sciences,* 30**,** 235-239.

KOTAKE, Y., SAGANE, K., OWA, T., MIMORI-KIYOSUE, Y., SHIMIZU, H., UESUGI, M., ISHIHAMA, Y., IWATA, M. & MIZUI, Y. 2007. Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nature Chemical Biology,* 3**,** 570-575.

KRAMARZOVA, K., STUCHLY, J., WILLASCH, A., GRUHN, B., SCHWARZ, J., CERMAK, J., MACHOVA-POLAKOVA, K., FUCHS, O., STARY, J., TRKA, J. & BOUBLIKOVA, L. 2012. Real-time PCR quantification of major Wilms' tumor gene 1 (WT1) isoforms in acute myeloid leukemia, their characteristic expression patterns and possible functional consequences. *Leukemia,* 26**,** 2086-2095.

KRAUTH, M. T., ALPERMANN, T., BACHER, U., EDER, C., DICKER, F., ULKE, M., KUZNIA, S., NADARAJAH, N., KERN, W., HAFERLACH, C., HAFERLACH, T. & SCHNITTGER, S. 2015. WT1 mutations are secondary events in AML, show varying frequencies and impact on prognosis between genetic subgroups. *Leukemia,* 29**,** 660-7.

KREIDBERG, J. A., SARIOLA, H., LORING, J. M., MAEDA, M., PELLETIER, J., HOUSMAN, D. & JAENISCH, R. 1993. WT-1 is required for early kidney development. *Cell,* 74**,** 679-91.

KRIVEGA, I., DALE, R. K. & DEAN, A. 2014. Role of LDB1 in the transition from chromatin looping to transcription activation. *Genes & Development,* 28**,** 1278-1290.

KUNTZ, E. M., BAQUERO, P., MICHIE, A. M., DUNN, K., TARDITO, S., HOLYOAKE, T. L., HELGASON, G. V. & GOTTLIEB, E. 2017. Targeting mitochondrial oxidative phosphorylation eradicates therapy-resistant chronic myeloid leukemia stem cells. *Nature medicine,* 23**,** 1234-1240.

KUWAHARA, J., YONEZAWA, A., FUTAMURA, M. & SUGIURA, Y. 1993. Binding of transcription factor Sp1 to GC box DNA revealed by footprinting analysis: different contact of three zinc fingers and sequence recognition mode. *Biochemistry,* 32**,** 5994-6001.

LAI, F., OROM, U. A., CESARONI, M., BERINGER, M., TAATJES, D. J., BLOBEL, G. A. & SHIEKHATTAR, R. 2013. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature,* 494**,** 497.

LAIMINS, L., HOLMGREN-KÖNIG, M. & KHOURY, G. 1986. Transcriptional "silencer" element in rat repetitive sequences associated with the rat insulin 1 gene locus. *Proceedings of the National Academy of Sciences,* 83**,** 3151.

LAITY, J. H., CHUNG, J., DYSON, H. J. & WRIGHT, P. E. 2000. Alternative Splicing of Wilms' Tumor Suppressor Protein Modulates DNA Binding Activity through Isoform-Specific DNA-Induced Conformational Changes. *Biochemistry,* 39**,** 5341-5348.

LAMBERT, J., LAMBERT, J., NIBOUREL, O., PAUTAS, C., HAYETTE, S., CAYUELA, J.-M., TERRÉ, C., ROUSSELOT, P., DOMBRET, H., CHEVRET, S., PREUDHOMME, C., CASTAIGNE, S. & RENNEVILLE, A. 2014. MRD assessed by WT1 and NPM1 transcript levels identifies distinct outcomes in AML patients and is influenced by gemtuzumab ozogamicin. *Oncotarget,* 5**,** 6280-6288.

LANDSCHULZ, W. H., JOHNSON, P. F. & MCKNIGHT, S. L. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science,* 240**,** 1759-64.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods,* 9**,** 357-359.

LARSSON, S. H., CHARLIEU, J. P., MIYAGAWA, K., ENGELKAMP, D., RASSOULZADEGAN, M., ROSS, A., CUZIN, F., VAN HEYNINGEN, V. & HASTIE, N. D. 1995. Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing. *Cell,* 81**,** 391-401.

LE GUEZENNEC, X., VERMEULEN, M., BRINKMAN, A. B., HOEIJMAKERS, W. A., COHEN, A., LASONDER, E. & STUNNENBERG, H. G. 2006. MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol Cell Biol,* 26**,** 843-51.

LEE, S. C., DVINGE, H., KIM, E., CHO, H., MICOL, J. B., CHUNG, Y. R., DURHAM, B. H., YOSHIMI, A., KIM, Y. J., THOMAS, M., LOBRY, C., CHEN, C. W., PASTORE, A., TAYLOR, J., WANG, X., KRIVTSOV, A., ARMSTRONG, S. A., PALACINO, J., BUONAMICI, S., SMITH, P. G., BRADLEY, R. K. & ABDEL-WAHAB, O. 2016. Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. *Nat Med,* 22**,** 672-8.

LEE, T. I. & YOUNG, R. A. 2013. Transcriptional regulation and its misregulation in disease. *Cell,* 152**,** 1237-51.

LEE, W., HASLINGER, A., KARIN, M. & TJIAN, R. 1987. Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. *Nature,* 325**,** 368-372.

LEE, Y., HUR, I., PARK, S. Y., KIM, Y. K., SUH, M. R. & KIM, V. N. 2006. The role of PACT in the RNA silencing pathway. *Embo j,* 25**,** 522-32.

LEY, T. J., MILLER, C., DING, L., RAPHAEL, B. J., MUNGALL, A. J., ROBERTSON, A., HOADLEY, K., TRICHE, T. J., JR., LAIRD, P. W., BATY, J. D., FULTON, L. L., FULTON, R., HEATH, S. E., KALICKI-VEIZER, J., KANDOTH, C., KLCO, J. M., KOBOLDT, D. C., KANCHI, K. L., KULKARNI, S., LAMPRECHT, T. L., LARSON, D. E., LIN, L., LU, C., MCLELLAN, M. D., MCMICHAEL, J. F., PAYTON, J., SCHMIDT, H., SPENCER, D. H., TOMASSON, M. H., WALLIS, J. W., WARTMAN, L. D., WATSON, M. A., WELCH, J., WENDL, M. C., ALLY, A., BALASUNDARAM, M., BIROL, I., BUTTERFIELD, Y., CHIU, R., CHU, A., CHUAH, E., CHUN, H. J., CORBETT, R., DHALLA, N., GUIN, R., HE, A., HIRST, C., HIRST, M., HOLT, R. A., JONES, S., KARSAN, A., LEE, D., LI, H. I., MARRA, M. A., MAYO, M., MOORE, R. A., MUNGALL, K., PARKER, J., PLEASANCE, E., PLETTNER, P., SCHEIN, J., STOLL, D., SWANSON, L., TAM, A., THIESSEN, N., VARHOL, R., WYE, N., ZHAO, Y., GABRIEL, S., GETZ, G., SOUGNEZ, C., ZOU, L., LEISERSON, M. D., VANDIN, F., WU, H. T., APPLEBAUM, F., BAYLIN, S. B., AKBANI, R., BROOM, B. M., CHEN, K., MOTTER, T. C., NGUYEN, K., WEINSTEIN, J. N., ZHANG, N., FERGUSON, M. L., ADAMS, C., BLACK, A., BOWEN, J., GASTIER-FOSTER, J., GROSSMAN, T., LICHTENBERG, T., WISE, L., DAVIDSEN, T., DEMCHOK, J. A., SHAW, K. R., SHETH, M., SOFIA, H. J., YANG, L., DOWNING, J. R. & ELEY, G. 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med,* 368**,** 2059-74.

LICHTINGER, M., INGRAM, R., HANNAH, R., MÜLLER, D., CLARKE, D., ASSI, S. A., LIE-A-LING, M., NOAILLES, L., VIJAYABASKAR, M. S., WU, M., TENEN, D. G., WESTHEAD, D. R., KOUSKOFF, V., LACAUD, G., GÖTTGENS, B. & BONIFER, C. 2012. RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *The EMBO Journal,* 31**,** 4318.

LIN, S., PTASINSKA, A., CHEN, X., SHRESTHA, M., ASSI, S. A., CHIN, P. S., IMPERATO, M. R., ARONOW, B. J., ZHANG, J., WEIRAUCH, M. T., BONIFER, C. & MULLOY, J. C. 2017. A FOXO1-induced oncogenic network defines the AML1-ETO preleukemic program. *Blood,* 130**,** 1213.

317

LIN, S., WEI, J., WUNDERLICH, M., CHOU, F.-S. & MULLOY, J. C. 2016. Immortalization of human AE pre-leukemia cells by hTERT allows leukemic transformation. *Oncotarget,* 7**,** 55939-55950.

LINK, K. A., LIN, S., SHRESTHA, M., BOWMAN, M., WUNDERLICH, M., BLOOMFIELD, C. D., HUANG, G. & MULLOY, J. C. 2016. Supraphysiologic levels of the AML1-ETO isoform AE9a are essential for transformation. *Proceedings of the National Academy of Sciences,* 113**,** 9075.

LITTLE, M. H., WILLIAMSON, K. A., MANNENS, M., KELSEY, A., GOSDEN, C., HASTIE, N. D. & VAN HEYNINGEN, V. 1993. Evidence that WT1 mutations in Denys-Drash syndrome patients may act in a dominant-negative fashion. *Hum Mol Genet,* 2**,** 259-64.

LITTLE, N. A., HASTIE, N. D. & DAVIES, R. C. 2000. Identification of WTAP, a novel Wilms' tumour 1-associating protein. *Human Molecular Genetics,* 9**,** 2231-2239.

LIU, Y., CHENEY, M. D., GAUDET, J. J., CHRUSZCZ, M., LUKASIK, S. M., SUGIYAMA, D., LARY, J., COLE, J., DAUTER, Z., MINOR, W., SPECK, N. A. & BUSHWELLER, J. H. 2006. The tetramer structure of the Nervy homology two domain, NHR2, is critical for AML1/ETO's activity. *Cancer Cell,* 9**,** 249-260.

LOKE, J., CHIN, P. S., KEANE, P., PICKIN, A., ASSI, S. A., PTASINSKA, A., IMPERATO, M. R., COCKERILL, P. N. & BONIFER, C. 2018. C/EBPα overrides epigenetic reprogramming by oncogenic transcription factors in acute myeloid leukemia. *Blood Advances,* 2**,** 271-284.

LORCH, Y., MAIER-DAVIS, B. & KORNBERG, R. D. 2006. Chromatin remodeling by nucleosome disassembly in vitro. *Proc Natl Acad Sci U S A,* 103**,** 3090-3.

LU, G., MIDDLETON, R. E., SUN, H., NANIONG, M., OTT, C. J., MITSIADES, C. S., WONG, K.-K., BRADNER, J. E. & KAELIN, W. G. 2014. The Myeloma Drug Lenalidomide Promotes the Cereblon-Dependent Destruction of Ikaros Proteins. *Science,* 343**,** 305.

LU, R., NEFF, N. F., QUAKE, S. R. & WEISSMAN, I. L. 2011. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat Biotechnol,* 29**,** 928-33.

LUGER, K., MÄDER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature,* 389**,** 251.

LUNDE, B. M., MOORE, C. & VARANI, G. 2007. RNA-binding proteins: modular design for efficient function. *Nature Reviews Molecular Cell Biology,* 8**,** 479-490.

LUTTERBACH, B., WESTENDORF, J. J., LINGGI, B., ISAAC, S., SETO, E. & HIEBERT, S. W. 2000. A Mechanism of Repression by Acute Myeloid Leukemia-1, the Target of Multiple Chromosomal Translocations in Acute Leukemia. *Journal of Biological Chemistry,* 275**,** 651-656.

LYMAN, S. D., JAMES, L., BOS, T. V., DE VRIES, P., BRASEL, K., GLINIAK, B., HOLLINGSWORTH, L. T., PICHA, K. S., MCKENNA, H. J., SPLETT, R. R., FLETCHER, F. A., MARASKOVSKY, E., FARRAH, T., FOXWORTHE, D., WILLIAMS, D. E. & BECKMANN, M. P. 1993a. Molecular cloning of a ligand for the flt3flk-2 tyrosine kinase receptor: A proliferative factor for primitive hematopoietic cells. *Cell,* 75**,** 1157-1167.

LYMAN, S. D., JAMES, L., VANDEN BOS, T., DE VRIES, P., BRASEL, K., GLINIAK, B., HOLLINGSWORTH, L. T., PICHA, K. S., MCKENNA, H. J., SPLETT, R. R. & ET AL. 1993b. Molecular cloning of a ligand for the flt3/flk-2 tyrosine kinase receptor: a proliferative factor for primitive hematopoietic cells. *Cell,* 75**,** 1157-67.

MAEDER, M. L., LINDER, S. J., CASCIO, V. M., FU, Y., HO, Q. H. & JOUNG, J. K. 2013. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods,* 10**,** 977-9.

MAHESWARAN, S., ENGLERT, C., ZHENG, G., LEE, S. B., WONG, J., HARKIN, D. P., BEAN, J., EZZELL, R., GARVIN, A. J., MCCLUSKEY, R. T., DECAPRIO, J. A. & HABER, D. A. 1998. Inhibition of cellular proliferation by the Wilms tumor suppressor WT1 requires association with the inducible chaperone Hsp70. *Genes & development,* 12**,** 1108-1120.

MAHESWARAN, S., PARK, S., BERNARD, A., MORRIS, J. F., RAUSCHER, F. J., 3RD, HILL, D. E. & HABER, D. A. 1993a. Physical and functional interaction between WT1 and p53 proteins. *Proceedings of the National Academy of Sciences of the United States of America,* 90**,** 5100-5104.

MAHESWARAN, S., PARK, S., BERNARD, A., MORRIS, J. F., RAUSCHER, F. J., HILL, D. E. & HABER, D. A. 1993b. Physical and functional interaction between WT1 and p53 proteins. *Proceedings of the National Academy of Sciences,* 90**,** 5100.

MAKAROVA, K. S., ARAVIND, L., GRISHIN, N. V., ROGOZIN, I. B. & KOONIN, E. V. 2002. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic acids research,* 30**,** 482-496.

MANDEL, C. R., KANEKO, S., ZHANG, H., GEBAUER, D., VETHANTHAM, V., MANLEY, J. L. & TONG, L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature,* 444**,** 953-956.

MARSHALL, N. F. & PRICE, D. H. 1995. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J Biol Chem,* 270**,** 12335-8.

MARTENS, J. H. A., MANDOLI, A., SIMMER, F., WIERENGA, B.-J., SAEED, S., SINGH, A. A., ALTUCCI, L., VELLENGA, E. & STUNNENBERG, H. G. 2012. ERG and FLI1 binding sites demarcate targets for aberrant epigenetic regulation by AML1-ETO in acute myeloid leukemia. *Blood,* 120**,** 4038-4048.

MARTIN, F. H., SUGGS, S. V., LANGLEY, K. E., LU, H. S., TING, J., OKINO, K. H., MORRIS, C. F., MCNIECE, I. K., JACOBSEN, F. W., MENDLAZ, E. A., BIRKETT, N. C., SMITH, K. A., JOHNSON, M. J., PARKER, V. P., FLORES, J. C., PATEL, A. C., FISHER, E. F., ERJAVEC, H. O., HERRERA, C. J., WYPYCH, J., SACHDEV, R. K., POPE, J. A., LESLIE, I., WEN, D., LIN, C.-H., CUPPLES, R. L. & ZSEBO, K. M. 1990. Primary structure and functional expression of rat and human stem cell factor DNAs. *Cell,* 63**,** 203-211.

MARTINEZ-SORIA, N., MCKENZIE, L., DRAPER, J., PTASINSKA, A., ISSA, H., POTLURI, S., BLAIR, H. J., PICKIN, A., ISA, A., CHIN, P. S., TIRTAKUSUMA, R., COLEMAN, D., NAKJANG, S., ASSI, S., FORSTER, V., REZA, M., LAW, E., BERRY, P., MUELLER, D., OSBORNE, C., ELDER, A., BOMKEN, S. N., PAL, D., ALLAN, J. M., VEAL, G. J., COCKERILL, P. N., WICHMANN, C., VORMOOR, J., LACAUD, G., BONIFER, C. & HEIDENREICH, O. 2018. The Oncogenic Transcription Factor RUNX1/ETO Corrupts Cell Cycle Regulation to Drive Leukemic Transformation. *Cancer Cell,* 34**,** 626-642.e8.

MASLAK, P. G., DAO, T., BERNAL, Y., CHANEL, S. M., ZHANG, R., FRATTINI, M., ROSENBLAT, T., JURCIC, J. G., BRENTJENS, R. J., ARCILA, M. E., RAMPAL, R., PARK, J. H., DOUER, D., KATZ, L., SARLIS, N., TALLMAN, M. S. & SCHEINBERG, D. A. 2018. Phase 2 trial of a multivalent WT1 peptide vaccine (galinpepimut-S) in acute myeloid leukemia. *Blood advances,* 2**,** 224-234.

MATHENY, C. J., SPECK, M. E., CUSHING, P. R., ZHOU, Y., CORPORA, T., REGAN, M., NEWMAN, M., ROUDAIA, L., SPECK, C. L., GU, T. L., GRIFFEY, S. M., BUSHWELLER, J. H. & SPECK, N. A. 2007. Disease mutations in RUNX1 and RUNX2 create nonfunctional, dominant-negative, or hypomorphic alleles. *Embo j,* 26**,** 1163-75.

MAZUMDAR, C., SHEN, Y., XAVY, S., ZHAO, F., REINISCH, A., LI, R., CORCES, M. R., FLYNN, R. A., BUENROSTRO, J. D., CHAN, S. M., THOMAS, D., KOENIG, J. L., HONG, W. J., CHANG, H. Y. & MAJETI, R. 2015. Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation. *Cell Stem Cell,* 17**,** 675-688.

MCHALE, C. M., WIEMELS, J. L., ZHANG, L., MA, X., BUFFLER, P. A., GUO, W., LOH, M. L. & SMITH, M. T. 2003. Prenatal origin of TEL-AML1-positive acute lymphoblastic leukemia in children born in California. *Genes Chromosomes Cancer,* 37**,** 36-43.

MCKAY, L. M., CARPENTER, B. & ROBERTS, S. G. E. 1999. Regulation of the Wilms' tumour suppressor protein transcriptional activation domain. *Oncogene,* 18**,** 6546.

MECHTA-GRIGORIOU, F., GERALD, D. & YANIV, M. 2001. The mammalian Jun proteins: redundancy and specificity. *Oncogene,* 20**,** 2378-89.

MECHTA, F., LALLEMAND, D., PFARR, C. M. & YANIV, M. 1997. Transformation by ras modifies AP1 composition and activity. *Oncogene,* 14**,** 837-847.

MIFSUD, B., TAVARES-CADETE, F., YOUNG, A. N., SUGAR, R., SCHOENFELDER, S., FERREIRA, L., WINGETT, S. W., ANDREWS, S., GREY, W., EWELS, P. A., HERMAN, B., HAPPE, S., HIGGS, A., LEPROUST, E., FOLLOWS, G. A., FRASER, P., LUSCOMBE, N. M. & OSBORNE, C. S. 2015. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet,* 47**,** 598-606.

MILES, L. A., BOWMAN, R. L., MERLINSKY, T. R., CSETE, I. S., OOI, A., DURRUTHY-DURRUTHY, R., BOWMAN, M., FAMULARE, C., PATEL, M. A., MENDEZ, P., AINALI, C., MANIVANNAN, M., SAHU, S., GOLDBERG, A. D., BOLTON, K., ZEHIR, A., RAMPAL, R., CARROLL, M. P., MEYER, S. E., VINY, A. D. & LEVINE, R. L. 2020a. Single cell mutational profiling delineates clonal trajectories in myeloid malignancies. *bioRxiv,* 2020.02.07.938860.

MILES, L. A., BOWMAN, R. L., MERLINSKY, T. R., CSETE, I. S., OOI, A. T., DURRUTHY-DURRUTHY, R., BOWMAN, M., FAMULARE, C., PATEL, M. A., MENDEZ, P., AINALI, C., DEMAREE, B., DELLEY, C. L., ABATE, A. R., MANIVANNAN, M., SAHU, S., GOLDBERG, A. D., BOLTON, K. L., ZEHIR, A., RAMPAL, R., CARROLL, M. P., MEYER, S. E., VINY, A. D. & LEVINE, R. L. 2020b. Single-cell mutation analysis of clonal evolution in myeloid malignancies. *Nature*.

MIN, I. M., PIETRAMAGGIORI, G., KIM, F. S., PASSEGUÉ, E., STEVENSON, K. E. & WAGERS, A. J. 2008. The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. *Cell Stem Cell,* 2**,** 380-91.

MITCHELL, P. J. & TJIAN, R. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science,* 245**,** 371-8.

MIYAMOTO, T., WEISSMAN, I. L. & AKASHI, K. 2000. AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. *Proceedings of the National Academy of Sciences,* 97**,** 7521.

MOJICA, F. J., DIEZ-VILLASENOR, C., GARCIA-MARTINEZ, J. & SORIA, E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol,* 60**,** 174-82.

MOJICA, F. J. M., DIEZ-VILLASENOR, C., GARCIA-MARTINEZ, J. & ALMENDROS, C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology,* 155**,** 733-740.

MORIYA, S., TAKIGUCHI, M., & SEKI, N. 2008. Expression of the WT1 gene -KTS domain isoforms suppresses the invasive ability of human lung squamous cell carcinoma cells. *International Journal of Oncology***,** 349-356.

MOSTOSLAVSKY, G., KOTTON, D. N., FABIAN, A. J., GRAY, J. T., LEE, J.-S. & MULLIGAN, R. C. 2005. Efficiency of transduction of highly purified murine hematopoietic stem cells by lentiviral and oncoretroviral vectors under conditions of minimal in vitro manipulation. *Molecular Therapy,* 11**,** 932-940.

MULLER-SIEBURG, C. E., CHO, R. H., KARLSSON, L., HUANG, J. F. & SIEBURG, H. B. 2004. Myeloid-biased hematopoietic stem cells have extensive self-renewal capacity but generate diminished lymphoid progeny with impaired IL-7 responsiveness. *Blood,* 103**,** 4111-8.

MULLOY, J. C., CAMMENGA, J., BERGUIDO, F. J., WU, K., ZHOU, P., COMENZO, R. L., JHANWAR, S., MOORE, M. A. S. & NIMER, S. D. 2003. Maintaining the self-renewal and differentiation potential of human CD34<sup>+</sup> hematopoietic cells using a single genetic element. *Blood,* 102**,** 4369.

MUSSAI, F., EGAN, S., HIGGINBOTHAM-JONES, J., PERRY, T., BEGGS, A., ODINTSOVA, E., LOKE, J., PRATT, G., U, K. P., LO, A., NG, M., KEARNS, P., CHENG, P. & DE SANTO, C. 2015. Arginine dependence of acute myeloid leukemia blast proliferation: a novel therapeutic target. *Blood,* 125**,** 2386.

NÄÄR, A. M., BOUTIN, J.-M., LIPKIN, S. M., YU, V. C., HOLLOWAY, J. M., GLASS, C. K. & ROSENFELD, M. G. 1991. The orientation and spacing of core DNA-binding motifs dictate selective transcriptional responses to three nuclear receptors. *Cell,* 65**,** 1267-1279.

NAKAGAMA, H., HEINRICH, G., PELLETIER, J. & HOUSMAN, D. E. 1995. Sequence and structural requirements for high-affinity DNA binding by the WT1 gene product. *Molecular and cellular biology,* 15**,** 1489-1498.

NATOLI, T. A., LIU, J., EREMINA, V., HODGENS, K., LI, C., HAMANO, Y., MUNDEL, P., KALLURI, R., MINER, J. H., QUAGGIN, S. E. & KREIDBERG, J. A. 2002. A Mutant Form of the Wilms' Tumor Suppressor Gene WT1 Observed in Denys-Drash Syndrome Interferes with Glomerular Capillary Development. *Journal of the American Society of Nephrology,* 13**,** 2058-2067.

NERLOV, C. & GRAF, T. 1998a. PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes & development,* 12**,** 2403-2412.

NERLOV, C. & GRAF, T. 1998b. PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes & Development,* 12**,** 2403-2412.

NGUYEN, T. A., JONES, R. D., SNAVELY, A. R., PFENNING, A. R., KIRCHNER, R., HEMBERG, M. & GRAY, J. M. 2016. High-throughput functional comparison of promoter and enhancer activities. *Genome Research*.

NISHIDA, S., HOSEN, N., SHIRAKATA, T., KANATO, K., YANAGIHARA, M., NAKATSUKA, S.-I., HOSHIDA, Y., NAKAZAWA, T., HARADA, Y., TATSUMI, N., TSUBOI, A., KAWAKAMI, M., OKA, Y., OJI, Y., AOZASA, K., KAWASE, I. & SUGIYAMA, H. 2006. AML1-ETO rapidly induces acute myeloblastic leukemia in cooperation with the Wilms tumor gene, WT1. *Blood,* 107**,** 3303.

NOMDEDÉU, J. F., ESQUIROL, A., CARRICONDO, M., PRATCORONA, M., HOYOS, M., GARRIDO, A., RUBIO, M., BUSSAGLIA, E., GARCÍA-CADENAS, I., ESTIVILL, C., BRUNET, S., MARTINO, R. & SIERRA, J. 2018. Bone Marrow WT1 Levels in Allogeneic Hematopoietic Stem Cell Transplantation for Acute Myelogenous Leukemia and Myelodysplasia: Clinically Relevant Time Points and 100 Copies Threshold Value. *Biology of Blood and Marrow Transplantation,* 24**,** 55-63.

NOVERSHTERN, N., SUBRAMANIAN, A., LAWTON, L. N., MAK, R. H., HAINING, W. N., MCCONKEY, M. E., HABIB, N., YOSEF, N., CHANG, C. Y., SHAY, T., FRAMPTON, G. M., DRAKE, A. C. B., LESKOV, I., NILSSON, B., PREFFER, F., DOMBKOWSKI, D., EVANS, J. W., LIEFELD, T., SMUTKO, J. S., CHEN, J., FRIEDMAN, N., YOUNG, R. A., GOLUB, T. R., REGEV, A. & EBERT, B. L. 2011. Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis. *Cell,* 144**,** 296-309.

O'LEARY, N. A., WRIGHT, M. W., BRISTER, J. R., CIUFO, S., HADDAD, D., MCVEIGH, R., RAJPUT, B., ROBBERTSE, B., SMITH-WHITE, B., AKO-ADJEI, D., ASTASHYN, A., BADRETDIN, A., BAO, Y., BLINKOVA, O., BROVER, V., CHETVERNIN, V., CHOI, J., COX, E., ERMOLAEVA, O., FARRELL, C. M., GOLDFARB, T., GUPTA, T., HAFT, D., HATCHER, E., HLAVINA, W., JOARDAR, V. S., KODALI, V. K., LI, W., MAGLOTT, D., MASTERSON, P., MCGARVEY, K. M., MURPHY, M. R., O'NEILL, K., PUJAR, S., RANGWALA, S. H., RAUSCH, D., RIDDICK, L. D., SCHOCH, C., SHKEDA, A., STORZ, S. S., SUN, H., THIBAUD-NISSEN, F., TOLSTOY, I., TULLY, R. E., VATSAN, A. R., WALLIN, C., WEBB, D., WU, W., LANDRUM, M. J., KIMCHI, A., TATUSOVA, T., DICUCCIO, M., KITTS, P., MURPHY, T. D. & PRUITT, K. D. 2015. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research,* 44**,** D733-D745.

O'SHEA, E. K., RUTKOWSKI, R. & KIM, P. S. 1989. Evidence that the leucine zipper is a coiled coil. *Science,* 243**,** 538-42.

OBIER, N., CAUCHY, P., ASSI, S. A., GILMOUR, J., LIE-A-LING, M., LICHTINGER, M., HOOGENKAMP, M., NOAILLES, L., COCKERILL, P. N., LACAUD, G., KOUSKOFF, V. & BONIFER, C. 2016. Cooperative binding of AP-1 and TEAD4 modulates the balance between vascular smooth muscle and hemogenic cell fate. *Development,* 143**,** 4324.

OGBOURNE, S. & ANTALIS, T. M. 1998. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J,* 331 (Pt 1)**,** 1-14.

OKUDA, T., CAI, Z., YANG, S., LENNY, N., LYU, C.-J., VAN DEURSEN, J. M. A., HARADA, H. & DOWNING, J. R. 1998. Expression of a Knocked-In AML1-ETO Leukemia Gene Inhibits the Establishment of Normal Definitive Hematopoiesis and Directly Generates Dysplastic Hematopoietic Progenitors. *Blood,* 91**,** 3134.

OKUDA, T., VAN DEURSEN, J., HIEBERT, S. W., GROSVELD, G. & DOWNING, J. R. 1996. AML1, the Target of Multiple Chromosomal Translocations in Human Leukemia, Is Essential for Normal Fetal Liver Hematopoiesis. *Cell,* 84**,** 321-330.

OLINS, D. E. & OLINS, A. L. 2003. Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol,* 4**,** 809-14.

OLIVE, M., KRYLOV, D., ECHLIN, D. R., GARDNER, K., TAPAROWSKY, E. & VINSON, C. 1997. A Dominant Negative to Activation Protein-1 (AP1) That Abolishes DNA Binding and Inhibits Oncogenesis. *Journal of Biological Chemistry,* 272**,** 18586-18594.

ONG, C.-T. & CORCES, V. G. 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics,* 12**,** 283.

PALII, C. G., CHENG, Q., GILLESPIE, M. A., SHANNON, P., MAZURCZYK, M., NAPOLITANI, G., PRICE, N. D., RANISH, J. A., MORRISSEY, E., HIGGS, D. R. & BRAND, M. 2019. Single-Cell Proteomics Reveal that Quantitative Changes in Co-expressed Lineage-Specific Transcription Factors Determine Cell Fate. *Cell Stem Cell,* 24**,** 812-820.e5.

PAN, Q., SHAI, O., LEE, L. J., FREY, B. J. & BLENCOWE, B. J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics,* 40**,** 1413-1415.

PANIGRAHI, A. K. & PATI, D. 2012. Higher-order orchestration of hematopoiesis: is cohesin a new player? *Experimental hematology,* 40**,** 967-973.

PAPAEMMANUIL, E., GERSTUNG, M., BULLINGER, L., GAIDZIK, V. I., PASCHKA, P., ROBERTS, N. D., POTTER, N. E., HEUSER, M., THOL, F., BOLLI, N., GUNDEM, G., VAN LOO, P., MARTINCORENA, I., GANLY, P., MUDIE, L., MCLAREN, S., O'MEARA, S., RAINE, K., JONES, D. R., TEAGUE, J. W., BUTLER, A. P., GREAVES, M. F., GANSER, A., DÖHNER, K., SCHLENK, R. F., DÖHNER, H. & CAMPBELL, P. J. 2016. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *New England Journal of Medicine,* 374**,** 2209-2221.

PARK, S. M., CHO, H., THORNTON, A. M., BARLOWE, T. S., CHOU, T., CHHANGAWALA, S., FAIRCHILD, L., TAGGART, J., CHOW, A., SCHURER, A., GRUET, A., WITKIN, M. D., KIM, J. H., SHEVACH, E. M., KRIVTSOV, A., ARMSTRONG, S. A., LESLIE, C. & KHARAS, M. G. 2019. IKZF2 Drives Leukemia Stem Cell Self-Renewal and Inhibits Myeloid Differentiation. *Cell Stem Cell,* 24**,** 153-165.e7.

PASSEGUÉ, E., JOCHUM, W., SCHORPP-KISTNER, M., MÖHLE-STEINLEIN, U. & WAGNER, E. F. 2001. Chronic Myeloid Leukemia with Increased Granulocyte Progenitors in Mice Lacking JunB Expression in the Myeloid Lineage. *Cell,* 104**,** 21-32.

PASSEGUE, E. & WAGNER, E. F. 2000. JunB suppresses cell proliferation by transcriptional activation of p16(INK4a) expression. *Embo j,* 19**,** 2969-79.

PASSEGUÉ, E., WAGNER, E. F. & WEISSMAN, I. L. 2004. JunB Deficiency Leads to a Myeloproliferative Disorder Arising from Hematopoietic Stem Cells. *Cell,* 119**,** 431-443.

PAVLETICH, N. P. & PABO, C. O. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science,* 252**,** 809-17.

PEI, S., MINHAJUDDIN, M., CALLAHAN, K. P., BALYS, M., ASHTON, J. M., NEERING, S. J., LAGADINOU, E. D., CORBETT, C., YE, H., LIESVELD, J. L., O'DWYER, K. M., LI, Z., SHI, L., GRENINGER, P., SETTLEMAN, J., BENES, C., HAGEN, F. K., MUNGER, J., CROOKS, P. A., BECKER, M. W. & JORDAN, C. T. 2013. Targeting aberrant glutathione metabolism to eradicate human acute myelogenous leukemia cells. *The Journal of biological chemistry,* 288**,** 33542-33558.

PELLETIER, J., BRUENING, W., LI, F. P., HABER, D. A., GLASER, T. & HOUSMAN, D. E. 1991. WT1 mutations contribute to abnormal genital system development and hereditary Wilms' tumour. *Nature,* 353**,** 431.

PEREIRA, C.-F., CHANG, B., QIU, J., NIU, X., PAPATSENKO, D., HENDRY, CAROLINE E., CLARK, NEIL R., NOMURA-KITABAYASHI, A., KOVACIC, JASON C., MA'AYAN, A., SCHANIEL, C., LEMISCHKA, IHOR R. & MOORE, K. 2013. Induction of a Hemogenic Program in Mouse Fibroblasts. *Cell Stem Cell,* 13**,** 205-218.

PERTEA, M., PERTEA, G. M., ANTONESCU, C. M., CHANG, T.-C., MENDELL, J. T. & SALZBERG, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology,* 33**,** 290-295.

PEVNY, L., SIMON, M. C., ROBERTSON, E., KLEIN, W. H., TSAI, S. F., D'AGATI, V., ORKIN, S. H. & COSTANTINI, F. 1991. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature,* 349**,** 257-60.

PFARR, C. M., MECHTA, F., SPYROU, G., LALLEMAND, D., CARILLO, S. & YANIV, M. 1994. Mouse JunD negatively regulates fibroblast growth and antagonizes transformation by ras. *Cell,* 76**,** 747-60.

POLLARD, J. A., ALONZO, T. A., GERBING, R. B., HO, P. A., ZENG, R., RAVINDRANATH, Y., DAHL, G., LACAYO, N. J., BECTON, D., CHANG, M., WEINSTEIN, H. J., HIRSCH, B., RAIMONDI, S. C., HEEREMA, N. A., WOODS, W. G., LANGE, B. J., HURWITZ, C., ARCECI, R. J., RADICH, J. P., BERNSTEIN, I. D., HEINRICH, M. C. & MESHINCHI, S. 2010. Prevalence and prognostic significance of KIT mutations in pediatric patients with core binding factor AML enrolled on serial pediatric cooperative trials for de novo AML. *Blood,* 115**,** 2372.

PREISLER, H. D., PERAMBAKAM, S., LI, B., HSU, W. T., VENUGOPAL, P., CREECH, S., SIVARAMAN, S. & TANAKA, N. 2001. Alterations in IRF1/IRF2 expression in acute myelogenous leukemia. *Am J Hematol,* 68**,** 23-31.

PREKER, P., NIELSEN, J., KAMMLER, S., LYKKE-ANDERSEN, S., CHRISTENSEN, M. S., MAPENDANO, C. K., SCHIERUP, M. H. & JENSEN, T. H. 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science,* 322**,** 1851-4.

PRONIER, E., BOWMAN, R. L., AHN, J., GLASS, J., KANDOTH, C., MERLINSKY, T. R., WHITFIELD, J. T., DURHAM, B. H., GRUET, A., HANASOGE SOMASUNDARA, A. V., RAMPAL, R., MELNICK, A., KOCHE, R. P., TAYLOR, B. S. & LEVINE, R. L. 2018. Genetic and epigenetic evolution as a contributor to WT1-mutant leukemogenesis. *Blood,* 132**,** 1265-1278.

PTASHNE, M. 1967. Specific Binding of the λ Phage Repressor to λ DNA. *Nature,* 214**,** 232-234.

PTASINSKA, A., ASSI, S. A., MANNARI, D., JAMES, S. R., WILLIAMSON, D., DUNNE, J., HOOGENKAMP, M., WU, M., CARE, M., MCNEILL, H., CAUCHY, P., CULLEN, M., TOOZE, R. M., TENEN, D. G., YOUNG, B. D., COCKERILL, P. N., WESTHEAD, D. R., HEIDENREICH, O. & BONIFER, C. 2012. Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia,* 26**,** 1829-1841.

PTASINSKA, A., ASSI, SALAM A., MARTINEZ-SORIA, N., IMPERATO, MARIA R., PIPER, J., CAUCHY, P., PICKIN, A., JAMES, SALLY R., HOOGENKAMP, M., WILLIAMSON, D., WU, M., TENEN, DANIEL G., OTT, S., WESTHEAD, DAVID R., COCKERILL, PETER N., HEIDENREICH, O. &

BONIFER, C. 2014. Identification of a Dynamic Core Transcriptional Network in t(8;21) AML that Regulates Differentiation Block and Self-Renewal. *Cell Reports,* 8**,** 1974-1988.

PTASINSKA, A., PICKIN, A., ASSI, S. A., CHIN, P. S., AMES, L., AVELLINO, R., GRÖSCHEL, S., DELWEL, R., COCKERILL, P. N., OSBORNE, C. S. & BONIFER, C. 2019. RUNX1-ETO Depletion in t(8;21) AML Leads to C/EBPα- and AP-1-Mediated Alterations in Enhancer-Promoter Interaction. *Cell reports,* 28**,** 3022-3031.e7.

PULVERER, B. J., HUGHES, K., FRANKLIN, C. C., KRAFT, A. S., LEEVERS, S. J. & WOODGETT, J. R. 1993. Co-purification of mitogen-activated protein kinases with phorbol ester-induced c-Jun kinase activity in U937 leukaemic cells. *Oncogene,* 8**,** 407-415.

PUNDHIR, S., POIRAZI, P. & GORODKIN, J. 2015. Emerging applications of read profiles towards the functional annotation of the genome. *Frontiers in genetics,* 6**,** 188-188.

QI, L. S., LARSON, M. H., GILBERT, L. A., DOUDNA, J. A., WEISSMAN, J. S., ARKIN, A. P. & LIM, W. A. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell,* 152**,** 1173-83.

QUINLAN, A. R. 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics,* 47**,** 11.12.1-34.

RADA-IGLESIAS, A., BAJPAI, R., SWIGUT, T., BRUGMANN, S. A., FLYNN, R. A. & WYSOCKA, J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature,* 470**,** 279-283.

RAINGEAUD, J., WHITMARSH, A. J., BARRETT, T., DÉRIJARD, B. & DAVIS, R. J. 1996. MKK3- and MKK6-regulated gene expression is mediated by the p38 mitogen-activated protein kinase signal transduction pathway. *Molecular and Cellular Biology,* 16**,** 1247.

RAMIREZ-CARROZZI, V. R., BRAAS, D., BHATT, D. M., CHENG, C. S., HONG, C., DOTY, K. R., BLACK, J. C., HOFFMANN, A., CAREY, M. & SMALE, S. T. 2009. A Unifying Model for the Selective Regulation of Inducible Transcription by CpG Islands and Nucleosome Remodeling. *Cell,* 138**,** 114-128.

RAMÍREZ, F., DÜNDAR, F., DIEHL, S., GRÜNING, B. A. & MANKE, T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res,* 42**,** W187-91.

RAMPAL, R., ALKALIN, A., MADZO, J., VASANTHAKUMAR, A., PRONIER, E., PATEL, J., LI, Y., AHN, J., ABDEL-WAHAB, O., SHIH, A., LU, C., WARD, PATRICK S., TSAI, JENNIFER J., HRICIK, T., TOSELLO, V., TALLMAN, JACOB E., ZHAO, X., DANIELS, D., DAI, Q., CIMINIO, L., AIFANTIS, I., HE, C., FUKS, F., TALLMAN, MARTIN S., FERRANDO, A., NIMER, S., PAIETTA, E., THOMPSON, CRAIG B., LICHT, JONATHAN D., MASON, CHRISTOPHER E., GODLEY, LUCY A., MELNICK, A., FIGUEROA, MARIA E. & LEVINE, ROSS L. 2014. DNA Hydroxymethylation Profiling Reveals that WT1 Mutations Result in Loss of TET2 Function in Acute Myeloid Leukemia. *Cell Reports,* 9**,** 1841-1855.

RANGATIA, J., VANGALA, R. K., TREIBER, N., ZHANG, P., RADOMSKA, H., TENEN, D. G., HIDDEMANN, W. & BEHRE, G. 2002. Downregulation of c-Jun Expression by Transcription Factor C/EBPα Is Critical for Granulocytic Lineage Commitment. *Molecular and Cellular Biology,* 22**,** 8681-8694.

RAUSCHER, F. J., MORRIS, J. F., TOURNAY, O. E., COOK, D. M. & CURRAN, T. 1990. Binding of the Wilms' tumor locus zinc finger protein to the EGR-1 consensus sequence. *Science,* 250**,** 1259.

RAUSCHER III, F. J., COHEN, D. R., CURRAN, T., BOS, T. J., VOGT, P. K., BOHMANN, D., TJIAN, R. & FRANZA JR, B. R. 1988. Fos-associated protein p39 is the product of the jun proto-oncogene. *Science,* 240**,** 1010-1016.

RAY, D., KWON, S. Y., TAGOH, H., HEIDENREICH, O., PTASINSKA, A. & BONIFER, C. 2013. Lineage-inappropriate PAX5 expression in t(8;21) acute myeloid leukemia requires signaling-mediated abrogation of polycomb repression. *Blood,* 122**,** 759-69.

RICCIARDI, M. R., MIRABILII, S., ALLEGRETTI, M., LICCHETTA, R., CALARCO, A., TORRISI, M. R., FOÀ, R., NICOLAI, R., PELUSO, G. & TAFURI, A. 2015. Targeting the leukemia cell metabolism by the CPT1a inhibition: functional preclinical effects in leukemias. *Blood,* 126**,** 1925-9.

RIDKER, P. M., EVERETT, B. M., THUREN, T., MACFADYEN, J. G., CHANG, W. H., BALLANTYNE, C., FONSECA, F., NICOLAU, J., KOENIG, W., ANKER, S. D., KASTELEIN, J. J. P., CORNEL, J. H., PAIS, P., PELLA, D., GENEST, J., CIFKOVA, R., LORENZATTI, A., FORSTER, T., KOBALAVA, Z., VIDA-SIMITI, L., FLATHER, M., SHIMOKAWA, H., OGAWA, H., DELLBORG, M., ROSSI, P. R. F., TROQUAY, R. P. T., LIBBY, P. & GLYNN, R. J. 2017. Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *New England Journal of Medicine,* 377**,** 1119-1131.

RITCHIE, M. E., PHIPSON, B., WU, D., HU, Y., LAW, C. W., SHI, W. & SMYTH, G. K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research,* 43**,** e47-e47.

ROBERTS, J. 2019. *University of Birmingham Thesis*.

ROBINSON, P. J. J., FAIRALL, L., HUYNH, V. A. T. & RHODES, D. 2006. EM measurements define the dimensions of the "30-nm" chromatin fiber: Evidence for a compact, interdigitated structure. *Proceedings of the National Academy of Sciences,* 103**,** 6506.

RONGVAUX, A., WILLINGER, T., MARTINEK, J., STROWIG, T., GEARTY, S. V., TEICHMANN, L. L., SAITO, Y., MARCHES, F., HALENE, S., PALUCKA, A. K., MANZ, M. G. & FLAVELL, R. A. 2014. Development and function of human innate immune cells in a humanized mouse model. *Nature Biotechnology,* 32**,** 364-372.

ROSARIO, S. R., LONG, M. D., AFFRONTI, H. C., ROWSAM, A. M., ENG, K. H. & SMIRAGLIA, D. J. 2018. Pan-cancer analysis of transcriptional metabolic dysregulation using The Cancer Genome Atlas. *Nature Communications,* 9**,** 5330.

ROSENBLUH, J., XU, H., HARRINGTON, W., GILL, S., WANG, X., VAZQUEZ, F., ROOT, D. E., TSHERNIAK, A. & HAHN, W. C. 2017. Complementary information derived from CRISPR Cas9 mediated gene deletion and suppression. *Nature Communications,* 8**,** 15403.

RUSSLER-GERMAIN, D. A., SPENCER, D. H., YOUNG, M. A., LAMPRECHT, T. L., MILLER, C. A., FULTON, R., MEYER, M. R., ERDMANN-GILMORE, P., TOWNSEND, R. R., WILSON, R. K. & LEY, T. J. 2014. The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer Cell,* 25**,** 442-54.

SABARI, B. R., DALL'AGNESE, A., BOIJA, A., KLEIN, I. A., COFFEY, E. L., SHRINIVAS, K., ABRAHAM, B. J., HANNETT, N. M., ZAMUDIO, A. V., MANTEIGA, J. C., LI, C. H., GUO, Y. E., DAY, D. S., SCHUIJERS, J., VASILE, E., MALIK, S., HNISZ, D., LEE, T. I., CISSE, I. I., ROEDER, R. G., SHARP, P. A., CHAKRABORTY, A. K. & YOUNG, R. A. 2018. Coactivator condensation at super-enhancers links phase separation and gene control. *Science (New York, N.Y.),* 361**,** eaar3958.

SAINSBURY, S., BERNECKY, C. & CRAMER, P. 2015. Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology,* 16**,** 129.

SAKAMOTO, Y., YOSHIDA, M., SEMBA, K. & HUNTER, T. 1997. Inhibition of the DNA-binding and transcriptional repression activity of the Wilms' tumor gene product, WT1, by cAMP-dependent protein kinase-mediated phosphorylation of Ser-365 and Ser-393 in the zinc finger domain. *Oncogene,* 15**,** 2001-2012.

SAKUMA, T., NAKADE, S., SAKANE, Y., SUZUKI, K. T. & YAMAMOTO, T. 2016. MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nat Protoc,* 11**,** 118-33.

SALDANHA, A. J. 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics,* 20**,** 3246-8.

SANADA, F., TANIYAMA, Y., MURATSU, J., OTSU, R., SHIMIZU, H., RAKUGI, H. & MORISHITA, R. 2018. Source of Chronic Inflammation in Aging. *Frontiers in Cardiovascular Medicine,* 5.

SANBORN, A. L., RAO, S. S. P., HUANG, S.-C., DURAND, N. C., HUNTLEY, M. H., JEWETT, A. I., BOCHKOV, I. D., CHINNAPPAN, D., CUTKOSKY, A., LI, J., GEETING, K. P., GNIRKE, A., MELNIKOV, A., MCKENNA, D., STAMENOVA, E. K., LANDER, E. S. & AIDEN, E. L. 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences,* 112**,** E6456.

SANJUAN-PLA, A., MACAULAY, I. C., JENSEN, C. T., WOLL, P. S., LUIS, T. C., MEAD, A., MOORE, S., CARELLA, C., MATSUOKA, S., BOURIEZ JONES, T., CHOWDHURY, O., STENSON, L., LUTTEROPP, M., GREEN, J. C., FACCHINI, R., BOUKARABILA, H., GROVER, A., GAMBARDELLA, A., THONGJUEA, S., CARRELHA, J., TARRANT, P., ATKINSON, D., CLARK, S. A., NERLOV, C. & JACOBSEN, S. E. 2013. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature,* 502**,** 232-6.

SAW, S., WEISS, A., KHOKHA, R. & WATERHOUSE, P. D. 2019. Metalloproteases: On the Watch in the Hematopoietic Niche. *Trends in Immunology,* 40**,** 1053-1070.

SAWADA, S., SCARBOROUGH, J. D., KILLEEN, N. & LITTMAN, D. R. 1994. A lineage-specific transcriptional silencer regulates CD4 gene expression during T lymphocyte development. *Cell,* 77**,** 917-29.

SCHONES, D. E., CUI, K., CUDDAPAH, S., ROH, T. Y., BARSKI, A., WANG, Z., WEI, G. & ZHAO, K. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell,* 132**,** 887-98.

SCHORPP-KISTNER, M., WANG, Z. Q., ANGEL, P. & WAGNER, E. F. 1999. JunB is essential for mammalian placentation. *Embo j,* 18**,** 934-48.

SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research,* 13**,** 2498-2504.

SHERR, C. J., BEACH, D. & SHAPIRO, G. I. 2016. Targeting CDK4 and CDK6: From Discovery to Therapy. *Cancer discovery,* 6**,** 353-367.

SIMEONOV, D. R., GOWEN, B. G., BOONTANRART, M., ROTH, T. L., GAGNON, J. D., MUMBACH, M. R., SATPATHY, A. T., LEE, Y., BRAY, N. L., CHAN, A. Y., LITUIEV, D. S., NGUYEN, M. L., GATE, R. E., SUBRAMANIAM, M., LI, Z., WOO, J. M., MITROS, T., RAY, G. J., CURIE, G. L., NADDAF, N., CHU, J. S., MA, H., BOYER, E., VAN GOOL, F., HUANG, H., LIU, R., TOBIN, V. R., SCHUMANN, K., DALY, M. J., FARH, K. K., ANSEL, K. M., YE, C. J., GREENLEAF, W. J., ANDERSON, M. S., BLUESTONE, J. A., CHANG, H. Y., CORN, J. E. & MARSON, A. 2017. Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature,* 549**,** 111-115.

SINGH, R. & VALCÁRCEL, J. 2005. Building specificity with nonspecific RNA-binding proteins. *Nat Struct Mol Biol,* 12**,** 645-53.

SOMERVILLE, T. D., WISEMAN, D. H., SPENCER, G. J., HUANG, X., LYNCH, J. T., LEONG, H. S., WILLIAMS, E. L., CHEESMAN, E. & SOMERVAILLE, T. C. 2015. Frequent Derepression of the Mesenchymal Transcription Factor Gene FOXC1 in Acute Myeloid Leukemia. *Cancer Cell,* 28**,** 329-42.

SONG, J., TEPLOVA, M., ISHIBE-MURAKAMI, S. & PATEL, D. J. 2012. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science,* 335**,** 709-712.

SONG, W. J., SULLIVAN, M. G., LEGARE, R. D., HUTCHINGS, S., TAN, X., KUFRIN, D., RATAJCZAK, J., RESENDE, I. C., HAWORTH, C., HOCK, R., LOH, M., FELIX, C., ROY, D. C., BUSQUE, L., KURNIT, D., WILLMAN, C., GEWIRTZ, A. M., SPECK, N. A., BUSHWELLER, J. H., LI, F. P., GARDINER, K., PONCZ, M., MARIS, J. M. & GILLILAND, D. G. 1999. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet,* 23**,** 166-75.

SPECK, N. A. & BALTIMORE, D. 1987. Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. *Molecular and Cellular Biology,* 7**,** 1101-1110.

STARCK, J., COHET, N., GONNET, C., SARRAZIN, S., DOUBEIKOVSKAIA, Z., DOUBEIKOVSKI, A., VERGER, A., DUTERQUE-COQUILLAUD, M. & MORLE, F. 2003. Functional cross-antagonism between transcription factors FLI-1 and EKLF. *Mol Cell Biol,* 23**,** 1390-402.

STEIDL, U., ROSENBAUER, F., VERHAAK, R. G., GU, X., EBRALIDZE, A., OTU, H. H., KLIPPEL, S., STEIDL, C., BRUNS, I., COSTA, D. B., WAGNER, K., AIVADO, M., KOBBE, G., VALK, P. J., PASSEGUE, E., LIBERMANN, T. A., DELWEL, R. & TENEN, D. G. 2006a. Essential role of Jun family transcription factors in PU.1 knockdown-induced leukemic stem cells. *Nat Genet,* 38**,** 1269-77.

STEIDL, U., ROSENBAUER, F., VERHAAK, R. G. W., GU, X., EBRALIDZE, A., OTU, H. H., KLIPPEL, S., STEIDL, C., BRUNS, I., COSTA, D. B., WAGNER, K., AIVADO, M., KOBBE, G., VALK, P. J. M., PASSEGUÉ, E., LIBERMANN, T. A., DELWEL, R. & TENEN, D. G. 2006b. Essential role of Jun family transcription factors in PU.1 knockdown–induced leukemic stem cells. *Nature Genetics,* 38**,** 1269-1277.

STOLL, R., LEE, B. M., DEBLER, E. W., LAITY, J. H., WILSON, I. A., DYSON, H. J. & WRIGHT, P. E. 2007. Structure of the Wilms Tumor Suppressor Protein Zinc Finger Domain Bound to DNA. *Journal of Molecular Biology,* 372**,** 1227-1245.

STOPKA, T., AMANATULLAH, D. F., PAPETTI, M. & SKOULTCHI, A. I. 2005. PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. *The EMBO journal,* 24**,** 3712-3723.

STRUHL, K. 1987. The DNA-binding domains of the jun oncoprotein and the yeast GCN4 transcriptional activator protein are functionally homologous. *Cell,* 50**,** 841-846.

STUNNENBERG, H. G. & VERMEULEN, M. 2011. Towards cracking the epigenetic code using a combination of high-throughput epigenomics and quantitative mass spectrometry-based proteomics. *Bioessays,* 33**,** 547-51.

SUN, X.-J., WANG, Z., WANG, L., JIANG, Y., KOST, N., SOONG, T. D., CHEN, W.-Y., TANG, Z., NAKADAI, T., ELEMENTO, O., FISCHLE, W., MELNICK, A., PATEL, D. J., NIMER, S. D. & ROEDER, R. G. 2013. A stable transcription factor complex nucleated by oligomeric AML1–ETO controls leukaemogenesis. *Nature,* 500**,** 93.

TAHIROV, T. H., INOUE-BUNGO, T., MORII, H., FUJIKAWA, A., SASAKI, M., KIMURA, K., SHIINA, M., SATO, K., KUMASAKA, T., YAMAMOTO, M., ISHII, S. & OGATA, K. 2001. Structural Analyses of DNA Recognition by the AML1/Runx-1 Runt Domain and Its Allosteric Control by CBFβ. *Cell,* 104**,** 755-767.

TAKAHASHI, K. & YAMANAKA, S. 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell,* 126**,** 663-676.

TALBERT, P. B. & HENIKOFF, S. 2017. Histone variants on the move: substrates for chromatin dynamics. *Nat Rev Mol Cell Biol,* 18**,** 115-126.

TATE, J. G., BAMFORD, S., JUBB, H. C., SONDKA, Z., BEARE, D. M., BINDAL, N., BOUTSELAKIS, H., COLE, C. G., CREATORE, C., DAWSON, E., FISH, P., HARSHA, B., HATHAWAY, C., JUPE, S. C., KOK, C. Y., NOBLE, K., PONTING, L., RAMSHAW, C. C., RYE, C. E., SPEEDY, H. E., STEFANCSIK, R., THOMPSON, S. L., WANG, S., WARD, S., CAMPBELL, P. J. & FORBES, S. A. 2019. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res,* 47**,** D941-d947.

TCGA 2013. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *New England Journal of Medicine,* 368**,** 2059-2074.

THOMA, F., KOLLER, T. & KLUG, A. 1979. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol,* 83**,** 403-27.

327

TIBES, R., KEATING, M. J., FERRAJOLI, A., WIERDA, W., RAVANDI, F., GARCIA-MANERO, G., O'BRIEN, S., CORTES, J., VERSTOVSEK, S., BROWNING, M. L. & FADERL, S. 2006. Activity of alemtuzumab in patients with CD52-positive acute leukemia. *Cancer,* 106**,** 2645-2651.

TILL, J. E. & MCCULLOCH, E. A. 1961. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiation research,* 14**,** 213-222.

TOLHUIS, B., PALSTRA, R. J., SPLINTER, E., GROSVELD, F. & DE LAAT, W. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell,* 10**,** 1453-65.

TOWERS, T. L., LUISI, B. F., ASIANOV, A. & FREEDMAN, L. P. 1993. DNA target selectivity by the vitamin D3 receptor: mechanism of dimer binding to an asymmetric repeat element. *Proceedings of the National Academy of Sciences,* 90**,** 6310.

TSAI, F. Y., KELLER, G., KUO, F. C., WEISS, M., CHEN, J., ROSENBLATT, M., ALT, F. W. & ORKIN, S. H. 1994. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature,* 371**,** 221-6.

TUNA, M., CHAVEZ-REYES, A. & TARI, A. M. 2005. HER2/neu increases the expression of Wilms' Tumor 1 (WT1) protein to stimulate S-phase proliferation and inhibit apoptosis in breast cancer cells. *Oncogene,* 24**,** 1648.

TZELEPIS, K., KOIKE-YUSA, H., DE BRAEKELEER, E., LI, Y., METZAKOPIAN, E., DOVEY, O. M., MUPO, A., GRINKEVICH, V., LI, M., MAZAN, M., GOZDECKA, M., OHNISHI, S., COOPER, J., PATEL, M., MCKERRELL, T., CHEN, B., DOMINGUES, A. F., GALLIPOLI, P., TEICHMANN, S., PONSTINGL, H., MCDERMOTT, U., SAEZ-RODRIGUEZ, J., HUNTLY, B. J. P., IORIO, F., PINA, C., VASSILIOU, G. S. & YUSA, K. 2016. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep,* 17**,** 1193-1205.

ULLMARK, T., JÄRVSTRÅT, L., SANDÉN, C., MONTANO, G., JERNMARK-NILSSON, H., LILLJEBJÖRN, H., LENNARTSSON, A., FIORETOS, T., DROTT, K., VIDOVIC, K., NILSSON, B. & GULLBERG, U. 2017. Distinct global binding patterns of the Wilms tumor gene 1 (WT1) −KTS and +KTS isoforms in leukemic cells. *Haematologica,* 102**,** 336.

VAN GALEN, P., HOVESTADT, V., WADSWORTH II, M. H., HUGHES, T. K., GRIFFIN, G. K., BATTAGLIA, S., VERGA, J. A., STEPHANSKY, J., PASTIKA, T. J., LOMBARDI STORY, J., PINKUS, G. S., POZDNYAKOVA, O., GALINSKY, I., STONE, R. M., GRAUBERT, T. A., SHALEK, A. K., ASTER, J. C., LANE, A. A. & BERNSTEIN, B. E. 2019. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell,* 176**,** 1265-1281.e24.

VAQUERIZAS, J. M., KUMMERFELD, S. K., TEICHMANN, S. A. & LUSCOMBE, N. M. 2009. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet,* 10**,** 252-63.

VERMEULEN, M., MULDER, K. W., DENISSOV, S., PIJNAPPEL, W. W. M. P., VAN SCHAIK, F. M. A., VARIER, R. A., BALTISSEN, M. P. A., STUNNENBERG, H. G., MANN, M. & TIMMERS, H. T. M. 2007. Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell,* 131**,** 58-69.

VIDAL, A. & KOFF, A. 2000. Cell-cycle inhibitors: three families united by a common cause. *Gene,* 247**,** 1-15.

VOLPE, G., WALTON, D. S., DEL POZZO, W., GARCIA, P., DASSÉ, E., O'NEILL, L. P., GRIFFITHS, M., FRAMPTON, J. & DUMON, S. 2013. C/EBPα and MYB regulate FLT3 expression in AML. *Leukemia,* 27**,** 1487-1496.

WADDINGTON, C. 1957. The Strategy of the Genes Routlege Library Editions: 20th Centure Science. George Allen & Unwin Ltd., London.

WADMAN, I. A., OSADA, H., GRÜTZ, G. G., AGULNICK, A. D., WESTPHAL, H., FORSTER, A. & RABBITTS, T. H. 1997. The LIM-only protein Lmo2 is a bridging molecule assembling an

erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *The EMBO Journal,* 16**,** 3145.

WALKER, C. J., OAKES, C. C., GENUTIS, L. K., GIACOPELLI, B., LIYANARACHCHI, S., NICOLET, D., EISFELD, A. K., SCHOLZ, M., BROCK, P., KOHLSCHMIDT, J., MROZEK, K., BILL, M., CARROLL, A. J., KOLITZ, J. E., POWELL, B. L., WANG, E. S., NIEDERWIESER, D. W., STONE, R. M., BYRD, J. C., SCHWIND, S., DE LA CHAPELLE, A. & BLOOMFIELD, C. D. 2019. Genome-wide association study identifies an acute myeloid leukemia susceptibility locus near BICRA. *Leukemia,* 33**,** 771-775.

WALTER, K., COCKERILL, P. N., BARLOW, R., CLARKE, D., HOOGENKAMP, M., FOLLOWS, G. A., RICHARDS, S. J., CULLEN, M. J., BONIFER, C. & TAGOH, H. 2010. Aberrant expression of CD19 in AML with t(8;21) involves a poised chromatin structure and PAX5. *Oncogene,* 29**,** 2927-37.

WANG, E., LU, S. X., PASTORE, A., CHEN, X., IMIG, J., CHUN-WEI LEE, S., HOCKEMEYER, K., GHEBRECHRISTOS, Y. E., YOSHIMI, A., INOUE, D., KI, M., CHO, H., BITNER, L., KLOETGEN, A., LIN, K.-T., UEHARA, T., OWA, T., TIBES, R., KRAINER, A. R., ABDEL-WAHAB, O. & AIFANTIS, I. 2019. Targeting an RNA-Binding Protein Network in Acute Myeloid Leukemia. *Cancer Cell,* 35**,** 369-384.e7.

WANG, W., LEE, S. B., PALMER, R., ELLISEN, L. W. & HABER, D. A. 2001. A functional interaction with CBP contributes to transcriptional activation by the Wilms tumor suppressor WT1. *J Biol Chem,* 276**,** 16810-6.

WANG, Y., XIAO, M., CHEN, X., CHEN, L., XU, Y., LV, L., WANG, P., YANG, H., MA, S., LIN, H., JIAO, B., REN, R., YE, D., GUAN, K.-L. & XIONG, Y. 2015. WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. *Molecular cell,* 57**,** 662-673.

WANG, Z. Q., GRIGORIADIS, A. E., MÖHLE-STEINLEIN, U. & WAGNER, E. F. 1991. A novel target cell for c-fos-induced oncogenesis: development of chondrogenic tumours in embryonic stem cell chimeras. *The EMBO Journal,* 10**,** 2437-2450.

WANG, Z. Q., OVITT, C., GRIGORIADIS, A. E., MOHLE-STEINLEIN, U., RUTHER, U. & WAGNER, E. F. 1992. Bone and haematopoietic defects in mice lacking c-fos. *Nature,* 360**,** 741-5.

WANG, Z. Y., QIU, Q. Q. & DEUEL, T. F. 1993. The Wilms' tumor gene product WT1 activates or suppresses transcription through separate functional domains. *Journal of Biological Chemistry,* 268**,** 9172-9175.

WARBURG, O., WIND, F. & NEGELEIN, E. 1927. THE METABOLISM OF TUMORS IN THE BODY. *The Journal of general physiology,* 8**,** 519-530.

WARD, P. S. & THOMPSON, C. B. 2012. Signaling in control of cell growth and metabolism. *Cold Spring Harbor perspectives in biology,* 4**,** a006783-a006783.

WEINTRAUB, A. S., LI, C. H., ZAMUDIO, A. V., SIGOVA, A. A., HANNETT, N. M., DAY, D. S., ABRAHAM, B. J., COHEN, M. A., NABET, B., BUCKLEY, D. L., GUO, Y. E., HNISZ, D., JAENISCH, R., BRADNER, J. E., GRAY, N. S. & YOUNG, R. A. 2017. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell,* 171**,** 1573-1588.e28.

WELCH, J. S., LEY, T. J., LINK, D. C., MILLER, C. A., LARSON, D. E., KOBOLDT, D. C., WARTMAN, L. D., LAMPRECHT, T. L., LIU, F., XIA, J., KANDOTH, C., FULTON, R. S., MCLELLAN, M. D., DOOLING, D. J., WALLIS, J. W., CHEN, K., HARRIS, C. C., SCHMIDT, H. K., KALICKI-VEIZER, J. M., LU, C., ZHANG, Q., LIN, L., O'LAUGHLIN, M. D., MCMICHAEL, J. F., DELEHAUNTY, K. D., FULTON, L. A., MAGRINI, V. J., MCGRATH, S. D., DEMETER, R. T., VICKERY, T. L., HUNDAL, J., COOK, L. L., SWIFT, G. W., REED, J. P., ALLDREDGE, P. A., WYLIE, T. N., WALKER, J. R., WATSON, M. A., HEATH, S. E., SHANNON, W. D., VARGHESE, N., NAGARAJAN, R., PAYTON, J. E., BATY, J. D., KULKARNI, S., KLCO, J. M., TOMASSON, M. H., WESTERVELT, P., WALTER, M. J., GRAUBERT, T. A., DIPERSIO, J. F., DING, L., MARDIS, E. R. & WILSON, R. K. 2012. The origin and evolution of mutations in acute myeloid leukemia. *Cell,* 150**,** 264-78.

WESTENDORF, J. J., YAMAMOTO, C. M., LENNY, N., DOWNING, J. R., SELSTED, M. E. & HIEBERT, S. W. 1998. The t(8;21) Fusion Product, AML-1–ETO, Associates with C/EBP-α, Inhibits C/EBP-α-Dependent Transcription, and Blocks Granulocytic Differentiation. *Molecular and Cellular Biology,* 18**,** 322-333.

WICHMANN, C., CHEN, L., HEINRICH, M., BAUS, D., PFITZNER, E., ZÖRNIG, M., OTTMANN, O. G. & GREZ, M. 2007. Targeting the Oligomerization Domain of ETO Interferes with RUNX1/ETO Oncogenic Activity in t(8;21)-Positive Leukemic Cells. *Cancer Research,* 67**,** 2280.

WICHMANN, C., QUAGLIANO-LO COCO, I., YILDIZ, Ö., CHEN-WICHMANN, L., WEBER, H., SYZONENKO, T., DÖRING, C., BRENDEL, C., PONNUSAMY, K., KINNER, A., BRANDTS, C., HENSCHLER, R. & GREZ, M. 2014. Activating c-KIT mutations confer oncogenic cooperativity and rescue RUNX1/ETO-induced DNA damage and apoptosis in human primary CD34+ hematopoietic progenitors. *Leukemia,* 29**,** 279.

WIEMELS, J. L., XIAO, Z., BUFFLER, P. A., MAIA, A. T., MA, X., DICKS, B. M., SMITH, M. T., ZHANG, L., FEUSNER, J., WIENCKE, J., PRITCHARD-JONES, K., KEMPSKI, H. & GREAVES, M. 2002. In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *Blood,* 99**,** 3801.

WILLIAMS, D. E., DE VRIES, P., NAMEN, A. E., WIDMER, M. B. & LYMAN, S. D. 1992. The Steel factor. *Developmental Biology,* 151**,** 368-376.

WILSON, A., LAURENTI, E., OSER, G., VAN DER WATH, R. C., BLANCO-BOSE, W., JAWORSKI, M., OFFNER, S., DUNANT, C. F., ESHKIND, L., BOCKAMP, E., LIO, P., MACDONALD, H. R. & TRUMPP, A. 2008. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell,* 135**,** 1118-29.

WILSON, N. K., FOSTER, S. D., WANG, X., KNEZEVIC, K., SCHÜTTE, J., KAIMAKIS, P., CHILARSKA, P. M., KINSTON, S., OUWEHAND, W. H., DZIERZAK, E., PIMANDA, J. E., DE BRUIJN, M. F. T. R. & GÖTTGENS, B. 2010. Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators. *Cell Stem Cell,* 7**,** 532-544.

WINTER, G. E., BUCKLEY, D. L., PAULK, J., ROBERTS, J. M., SOUZA, A., DHE-PAGANON, S. & BRADNER, J. E. 2015. Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science,* 348**,** 1376.

WOOD, A., KROGAN, N. J., DOVER, J., SCHNEIDER, J., HEIDT, J., BOATENG, M. A., DEAN, K., GOLSHANI, A., ZHANG, Y., GREENBLATT, J. F., JOHNSTON, M. & SHILATIFARD, A. 2003. Bre1, an E3 Ubiquitin Ligase Required for Recruitment and Substrate Selection of Rad6 at a Promoter. *Molecular Cell,* 11**,** 267-274.

WOOD, W. I. & FELSENFELD, G. 1982. Chromatin structure of the chicken beta-globin gene region. Sensitivity to DNase I, micrococcal nuclease, and DNase II. *Journal of Biological Chemistry,* 257**,** 7730-6.

WU, S., ZHANG, W., SHEN, D., LU, J. & ZHAO, L. 2019. PLCB4 upregulation is associated with unfavorable prognosis in pediatric acute myeloid leukemia. *Oncol Lett,* 18**,** 6057-6065.

WYSOCKA, J., SWIGUT, T., XIAO, H., MILNE, T. A., KWON, S. Y., LANDRY, J., KAUER, M., TACKETT, A. J., CHAIT, B. T., BADENHORST, P., WU, C. & ALLIS, C. D. 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature,* 442**,** 86-90.

XIE, H., YE, M., FENG, R. & GRAF, T. 2004. Stepwise Reprogramming of B Cells into Macrophages. *Cell,* 117**,** 663-676.

YAMAGUCHI, Y., SHIBATA, H. & HANDA, H. 2013. Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. *Biochim Biophys Acta,* 1829**,** 98-104.

YAMANAKA, R., KIM, G. D., RADOMSKA, H. S., LEKSTROM-HIMES, J., SMITH, L. T., ANTONSON, P., TENEN, D. G. & XANTHOPOULOS, K. G. 1997. CCAAT/enhancer binding protein

epsilon is preferentially up-regulated during granulocytic differentiation and its functional versatility is determined by alternative use of promoters and differential splicing. *Proc Natl Acad Sci U S A,* 94**,** 6462-7.

YAN, M., BUREL, S. A., PETERSON, L. F., KANBE, E., IWASAKI, H., BOYAPATI, A., HINES, R., AKASHI, K. & ZHANG, D.-E. 2004. Deletion of an AML1-ETO C-terminal NcoR/SMRT-interacting region strongly induces leukemia development. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 17186.

YAN, M., KANBE, E., PETERSON, L. F., BOYAPATI, A., MIAO, Y., WANG, Y., CHEN, I. M., CHEN, Z., ROWLEY, J. D., WILLMAN, C. L. & ZHANG, D.-E. 2006. A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. *Nature Medicine,* 12**,** 945.

YANG, L., BRYDER, D., ADOLFSSON, J., NYGREN, J., MÅNSSON, R., SIGVARDSSON, M. & JACOBSEN, S. E. W. 2005. Identification of Lin–Sca1+ kit+ CD34+ Flt3–short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients. *Blood,* 105**,** 2717-2723.

YENGO, R. K., NURMEMMEDOV, E. & THUNNISSEN, M. M. 2018. Structure of WT1 zinc fingers bound to its cognate DNA: Implications of the KTS insert. *bioRxiv***,** 271577.

YOSHIDA, K., SANADA, M., SHIRAISHI, Y., NOWAK, D., NAGATA, Y., YAMAMOTO, R., SATO, Y., SATO-OTSUBO, A., KON, A., NAGASAKI, M., CHALKIDIS, G., SUZUKI, Y., SHIOSAKA, M., KAWAHATA, R., YAMAGUCHI, T., OTSU, M., OBARA, N., SAKATA-YANAGIMOTO, M., ISHIYAMA, K., MORI, H., NOLTE, F., HOFMANN, W.-K., MIYAWAKI, S., SUGANO, S., HAFERLACH, C., KOEFFLER, H. P., SHIH, L.-Y., HAFERLACH, T., CHIBA, S., NAKAUCHI, H., MIYANO, S. & OGAWA, S. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature,* 478**,** 64-69.

YOSHIMI, A., LIN, K.-T., WISEMAN, D. H., RAHMAN, M. A., PASTORE, A., WANG, B., LEE, S. C.-W., MICOL, J.-B., ZHANG, X. J., DE BOTTON, S., PENARD-LACRONIQUE, V., STEIN, E. M., CHO, H., MILES, R. E., INOUE, D., ALBRECHT, T. R., SOMERVAILLE, T. C. P., BATTA, K., AMARAL, F., SIMEONI, F., WILKS, D. P., CARGO, C., INTLEKOFER, A. M., LEVINE, R. L., DVINGE, H., BRADLEY, R. K., WAGNER, E. J., KRAINER, A. R. & ABDEL-WAHAB, O. 2019. Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. *Nature,* 574**,** 273-277.

YOUNG, M. R., LI, J.-J., RINCÓN, M., FLAVELL, R. A., SATHYANARAYANA, B. K., HUNZIKER, R. & COLBURN, N. 1999. Transgenic mice demonstrate AP-1 (activator protein-1) transactivation is required for tumor promotion. *Proceedings of the National Academy of Sciences,* 96**,** 9827.

ZAMBETTI, N. A., PING, Z., CHEN, S., KENSWIL, K. J. G., MYLONA, M. A., SANDERS, M. A., HOOGENBOEZEM, R. M., BINDELS, E. M. J., ADISTY, M. N., VAN STRIEN, P. M. H., VAN DER LEIJE, C. S., WESTERS, T. M., CREMERS, E. M. P., MILANESE, C., MASTROBERARDINO, P. G., VAN LEEUWEN, J., VAN DER EERDEN, B. C. J., TOUW, I. P., KUIJPERS, T. W., KANAAR, R., VAN DE LOOSDRECHT, A. A., VOGL, T. & RAAIJMAKERS, M. 2016. Mesenchymal Inflammation Drives Genotoxic Stress in Hematopoietic Stem Cells and Predicts Disease Evolution in Human Pre-leukemia. *Cell Stem Cell,* 19**,** 613-627.

ZARET, K. S. & CARROLL, J. S. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes & Development,* 25**,** 2227-2241.

ZENZ, R., EFERL, R., KENNER, L., FLORIN, L., HUMMERICH, L., MEHIC, D., SCHEUCH, H., ANGEL, P., TSCHACHLER, E. & WAGNER, E. F. 2005. Psoriasis-like skin disease and arthritis caused by inducible epidermal deletion of Jun proteins. *Nature,* 437**,** 369-75.

ZETSCHE, B., GOOTENBERG, J. S., ABUDAYYEH, O. O., SLAYMAKER, I. M., MAKAROVA, K. S., ESSLETZBICHLER, P., VOLZ, S. E., JOUNG, J., VAN DER OOST, J., REGEV, A., KOONIN, E. V.

& ZHANG, F. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell,* 163**,** 759-71.

ZHANG, D.-E., ZHANG, P., WANG, N.-D., HETHERINGTON, C. J., DARLINGTON, G. J. & TENEN, D. G. 1997a. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein α-deficient mice. *Proceedings of the National Academy of Sciences,* 94**,** 569.

ZHANG, H., KOLB, F. A., JASKIEWICZ, L., WESTHOF, E. & FILIPOWICZ, W. 2004a. Single processing center models for human Dicer and bacterial RNase III. *Cell,* 118**,** 57-68.

ZHANG, J., KALKUM, M., YAMAMURA, S., CHAIT, B. T. & ROEDER, R. G. 2004b. E Protein Silencing by the Leukemogenic AML1-ETO Fusion Protein. *Science,* 305**,** 1286.

ZHANG, P., BEHRE, G., PAN, J., IWAMA, A., WARA-ASWAPATI, N., RADOMSKA, H. S., AURON, P. E., TENEN, D. G. & SUN, Z. 1999. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. *Proceedings of the National Academy of Sciences,* 96**,** 8705.

ZHANG, W. & LIU, H. T. 2002. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Research,* 12**,** 9-18.

ZHANG, X., GUO, C., CHEN, Y., SHULHA, H. P., SCHNETZ, M. P., LAFRAMBOISE, T., BARTELS, C. F., MARKOWITZ, S., WENG, Z., SCACHERI, P. C. & WANG, Z. 2008a. Epitope tagging of endogenous proteins for genome-wide ChIP-chip studies. *Nature methods,* 5**,** 163-165.

ZHANG, X., XING, G., FRAIZER, G. C. & SAUNDERS, G. F. 1997b. Transactivation of an Intronic Hematopoietic-specific Enhancer of the Human Wilms' Tumor 1 Gene by GATA-1 and c-Myb. *Journal of Biological Chemistry,* 272**,** 29272-29280.

ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008b. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology,* 9**,** R137.

ZHAO, J. L., MA, C., O'CONNELL, R. M., MEHTA, A., DILORETO, R., HEATH, J. R. & BALTIMORE, D. 2014. Conversion of danger signals into cytokine signals by hematopoietic stem and progenitor cells for regulation of stress-induced hematopoiesis. *Cell Stem Cell,* 14**,** 445-459.

ZHOU, H.-S., CARTER, B. Z. & ANDREEFF, M. 2016. Bone marrow niche-mediated survival of leukemia stem cells in acute myeloid leukemia: Yin and Yang. *Cancer biology & medicine,* 13**,** 248-259.