Modelling the earliest events of t(8;21) acute myeloid leukaemia in human embryonic stem cell-derived definitive haematopoietic progenitor cells

Monica Nafria i Fedi

ORCID Identifier: 0000-0002-5329-7966

A thesis submitted to the University of Birmingham
and the University of Melbourne for the degree of
DOCTOR OF PHILOSOPHY

Joint PhD student, Universitas 21 Studentship

Institute for Cancer and Genomic Sciences
College of Medicine and Dentistry, University of Birmingham (United Kingdom)

Murdoch Children's Research Institute
Faculty of Medicine and Dentistry, University of Melbourne (Australia)

September 2019

ABSTRACT

The t(8;21) translocation generates the aberrant transcription factor RUNX1-ETO and occurs in approximately 10% of all acute myeloid leukaemias. *RUNX1-ETO* transcripts can be detected *in utero* and in cells of patients in remission, but its sole expression is insufficient to cause overt leukaemia. Given that t(8;21) patient cells present additional mutations, the epigenetic reprogramming directly mediated by RUNX1-ETO remains unclear. To address this question, we generated human Embryonic Stem Cell lines carrying an inducible RUNX1-ETO transgene, which we subsequently differentiated into definitive haematopoietic progenitors. We show that induction of RUNX1-ETO in already formed progenitors (i) blocks differentiation at an immature stage, (ii) induces a cell-type specific and reversible cell cycle arrest, (iii) abrogates the RUNX1-mediated gene expression program by interfering with RUNX1 binding, resulting in downregulation of haematopoietic, cell cycle as well as DNA repair genes, (iv) closes down a large part of the chromatin accessibility pattern present in adult haematopoietic multipotent progenitors and (v) alters the differentiation of a defined sub-population of progenitors. Our data are consistent with the idea that RUNX1-ETO establishes a precondition for leukaemic transformation by maintaining a reservoir of quiescent pre-leukaemic multipotent progenitors with susceptibility to expand upon acquisition of additional oncogenic events.

Dedicated to my beloved grandmother, Rosa.

My angel, my biggest supporter and my greatest inspiration.

DECLARATION

This declaration certifies that:

i) this thesis comprises only my original work towards the degree of Doctor of Philosophy;

ii) due acknowledgement has been made in the text to all other material used; and

iii) this thesis is fewer than the maximum word limit in length, exclusive of tables, figures, bibliographies and appendices.

Monica Nafria i Fedi

September 2019

PREFACE

I have been responsible for the generation and interpretation of the data presented in this thesis, with exception of:

i) The dual reporter SOX17$^{mCHERRY/w}$ RUNX1C$^{GFP/w}$ H9 human Embryonic Stem Cell line, which was previously generated in the laboratory of Andrew Elefanty (University of Melbourne, AUS) by Elizabeth NG (Ng et al., 2016).

ii) The bioinformatic analyses, which were performed by Dr Peter Keane in the laboratory of Prof Constanze Bonifer (University of Birmingham, UK).

A summary of the chapters 'Results' and 'Discussion' have been submitted for publication to Cell Reports on 24th of September 2019. A preprint (non-peer reviewed) of the article submitted for publication (Nafria et al., 2019) has been made accessible by submission to BioRxiv (doi: https://doi.org/10.1101/748921).

I would like to acknowledge all the funding bodies that have made this PhD research project possible, including Universitas 21, Cancer Research UK and Children's Cancer Foundation (Australia). I would like to express my gratitude to the Murdoch Children's Research Institute (MCRI, AUS) for the award of a PhD top-up scholarship. I would also like to thank the generous travelling scholarships awarded by The Henry and Rachael Ackman trust (Department of Paediatrics, The University of Melbourne) and International Society for Experimental Haematology (ISEH), which allowed me to attend an overseas conference.

ACKNOWLEDGEMENTS

Foremost, I would like to express my most sincere gratitude to both my supervisors, Prof Constanze Bonifer and Prof Andrew Elefanty, for their astronomical support, unmeasurable wisdom and outstanding supervision. I want to acknowledge the enormous commitment they have had to my project even on their busiest schedules and the trust they have placed in me from the very first day.

This work could have not been possible without the guidance and expertise of Dr Elizabeth Ng. She has been the best teacher in regard to the haematopoietic differentiation protocol and all the 'secret' tips for the best cell culture (and flamenco dancing!). I want to deeply thank Elizabeth and Andrew for all the emotional support and the invaluable experiences that they have provided me during my stay in Melbourne. I will be forever grateful for their boundless care and for making me feel at home.

I want to express my gratitude to all the past and present members of the Bonifer-Cockerill (UK) and the Elefanty (AUS) laboratories. I am deeply indebted to Dr Peter Keane for his tremendous work and expertise on performing the bioinformatic analyses and for his commitment to my project. I want to thank Prof Peter Cockerill for his input and stimulating scientific discussion and Prof Edouard Stanley for all the troubleshooting on my endless cloning experiments.

I also want to extend my gratitude to Dr Paloma Garcia and to Yvonne Barnett, who have been helpful, friendly and extremely supportive during the last year of my PhD. I also want to thank Freya Bruveris, not only for her expertise and knowledge but for all the fun times we shared together. In particular, there are two people who have

made my stay in Australia memorable and that, without them, nothing would have been the same: Ana Rita Leitoguinho and Jasna Kusur. They made me laugh on a daily basis – even on 14-hour long days in cell culture – and they lifted me up every time I felt frustrated due to unsuccessful experiments. I will always be grateful for the unconditional love, support and the best friendship, which I will keep forever, they have given to me.

My deepest appreciation goes to all my family, chiefly to my parents, who have always believed in me and encouraged me to embark on any adventure, specially this one. Above all, I would like to thank my grandma, my role model and an example of hard work and success, for supporting everything I do and loving me without conditions.

I would also like to thank all my true friends, mainly Carla, Victor and Marta, who kept by my side, even though being miles apart, and supported me during all this journey. Also, I would like to extend my thanks to the numerous international flatmates I have had along this journey, specially to Iraia, Sean, Cameron and Amanda, who had the most enormous patience and with whom I lived unforgettable moments.

Last but not least, I would like to thank my partner Duval. He has been here for me at all times, made smooth all the ups and downs experienced during the last year of my PhD, cared about my happiness and well-being and, importantly, celebrated all my little victories. I thank him for constantly reminding me that I could do anything.

TABLE OF CONTENTS

FIGURES

TABLES

SUPPLEMENTARY FIGURES

SUPPLEMENTARY TABLES AND DATASETS

ABBREVIATIONS

AA2P: Ascorbic acid 2-phosphate
ACT: Activin A
AGM: Aorta-gonad-mesonephros
AML: Acute myeloid leukaemia
AP-1: Activator Protein 1
ATAC: Assay for Transposase-Accessible Chromatin
BMP4: Bone morphogenetic protein 4
BSA: Bovine Serum Albumin
CDK: Cyclin-dependent kinase
ChIP: Chromatin Immunoprecipitation
CLP: Common lymphoid progenitor
CMP: Common myeloid progenitor
Ct: Cycle thresholds
CTCF: CCCTC-binding factor
d: day
DD: Destabilization domain
DHS: DNaseI hypersensitive site
DMEM: Dulbecco's Modified Eagle Medium
DMSO: Dimethyl sulfoxide
DNMT: DNA (cytosine-5)-methyltransferase
Dox: Doxycycline
dpc: Days post-conception
DTT: Dithiothreitol
EB: Embryoid Body
ecDHFR: *Escherichia coli* dihydrofolate reductase
EHT: Endothelial to haematopoietic transition
ELP: Early lymphoid progenitor
EMP: Erythro-myeloid progenitor
EoBP: Eosinophil/basophil progenitor
EPO: Erythropoietin
ES: Embryonic Stem
FACS: Fluorescence-activated cell sorting
FCS: Foetal Calf Serum
FGF2: Fibroblast Growth Factor 2
FLT3: FMS-like tyrosine kinase 3 receptor

FPKM: Fragments per kilobase of transcript per million mapped reads
GM-CSF: Granulocyte-macrophage colony-stimulating factor
GMP: Granulocyte/monocyte progenitor
HAT: Histone acetyltransferase
HDAC: Histone deacetylase
hLDL: Human low-density lipoproteins
hr: hour
hPSC: Human pluripotent stem cell
HRP: Horseradish peroxidase
HSC: Haematopoietic stem cell
IGF2: Insulin-like growth factor 2
IL: Interleukin
IMDM: Iscove's Modified Dulbecco's Media
iPSC: induced pluripotent stem cell
ITS-E: Insulin-Transferrin-Selenium-E
IVL: Involucrin
Kb: Kilobase
KDM: Lysine demethylases
KMT: Lysine methyltransferases
KOSR: KnockOut Serum Replacer
LMPP: Lymphoid-primed multipotent progenitor
LSD1: Lysine-specific histone demethylase 1
MACS: Magnetic-activated cell sorting
MAPK: Mitogen-Activated Protein Kinase
MDS: Myelodysplastic syndrome
ME: Megakaryocyte/erythrocyte progenitor
MeCP: Methyl-CpG-binding protein
MEF: Mouse embryonic fibroblast
miRNA: microRNA
MLP: Multi-lymphoid progenitor
MPL: Myeloproliferative virus oncogene
MPP: Multipotent progenitor
MTG: Monothioglycerol
NEAA: Non-essential amino acids
PBS: Phosphate buffered saline
PCR: Polymerase Chain Reaction
Pen/Strep: Penicillin/Streptomycin
PFHMII: Protein-Free Hybridoma Medium II

PI: Propidium iodide

PIC: Proteinase inhibitor cocktail

Pol II: RNA polymerase II

PVA: Polyvinyl alcohol

RE: Restriction endonuclease

Rh: Recombinant human

RT-qPCR: Reverse Transcription quantitative PCR

rtTA: reverse tetracycline-controlled transactivator

s: second

SCF: Stem cell factor

SDS-PAGE: Sodium dodecyl sulphate polyacrylamide gel

Seq: Sequencing

TAE: Tris-acetate-EDTA

TALENS: Transcription activator-like effector nucleases

TBE: Tris/Borate/EDTA

TF: Transcription factor

TPO: Thrombopoietin

TRE: Tetracycline Responsive Element

TSS: Transcription Start Site

VEGF: Vascular Endothelial Growth Factor

1 INTRODUCTION

1.1 Chromatin structure

Each single cell carries all the information necessary for the development and function of all the distinct cell types that constitute a living organism. This information is coded into 3 billion DNA bases spanning approximately 2 metres of DNA (Bloom and Joglekar, 2010), which is compacted into the cell nucleus. Packaging of the DNA is a tightly regulated and dynamic process, since access to distinct DNA sequences is required during gene transcription and DNA replication. To achieve this, the genome is organized into chromatin, which can be opened up to allow access to the transcription and replication machinery.

The first stage of compaction begins with the assembly of the DNA in a structure termed the 'nucleosome', which consists of ~146 bp of DNA wrapped around a symmetrical protein core complex formed by four pairs of histones: H2A, H2B, H3 and H4 (Figure 1.1) (Kornberg and Lorch, 1999; Luger et al., 1997; Oudet et al., 1975). Nucleosomes are formed every 160-240 bp (Kornberg, 1974), generating a ~10 nm chromatin thread that resembles ‘beads on a string’. Nucleosome complexes are compacted to a higher-degree thanks to the linker histone H1, which protects the linker DNA from nucleases and stabilizes a ‘zig-zag’ nucleosome structure that is the basis of a ~30 nm chromatin fibre (Figure 1.1) (Allan et al., 1980; Oudet et al., 1975). The packaging of the nucleosomes is dynamic and can be modulated resulting into two chromatin states: inactive condensed (heterochromatin), which includes H1, or active (euchromatin), which harbours mobilized nucleosomes facilitating access to transcription factors and co-factors.

The highest order of chromatin structure involves the folding of the chromatin fibre into large genomic regions separated by conserved boundary regions within the three-dimensional nuclear space (Figure 1.1). These organized territories are known as topologically associated domains (TADs) and promote regulatory long-range interactions within the domain (Bickmore and van Steensel, 2013; Dekker and Misteli, 2015; Tang et al., 2015). In other words, TADs bring DNA sequences from distinct genomic sites in close physical proximity, and hence allow the simultaneous coordination of their chromatin status and transcriptional regulation (Dixon et al., 2012). TAD boundaries are mediated and stabilized by structural proteins, such as CCCTC-binding factor (CTCF) and cohesin (Faure et al., 2012; Handoko et al., 2011). Other factors, such as mediator, co-operate with structural proteins to divide TADs into smaller sub-domains that facilitate shorter range interactions between transcriptional regulatory elements (Bonora et al., 2014).

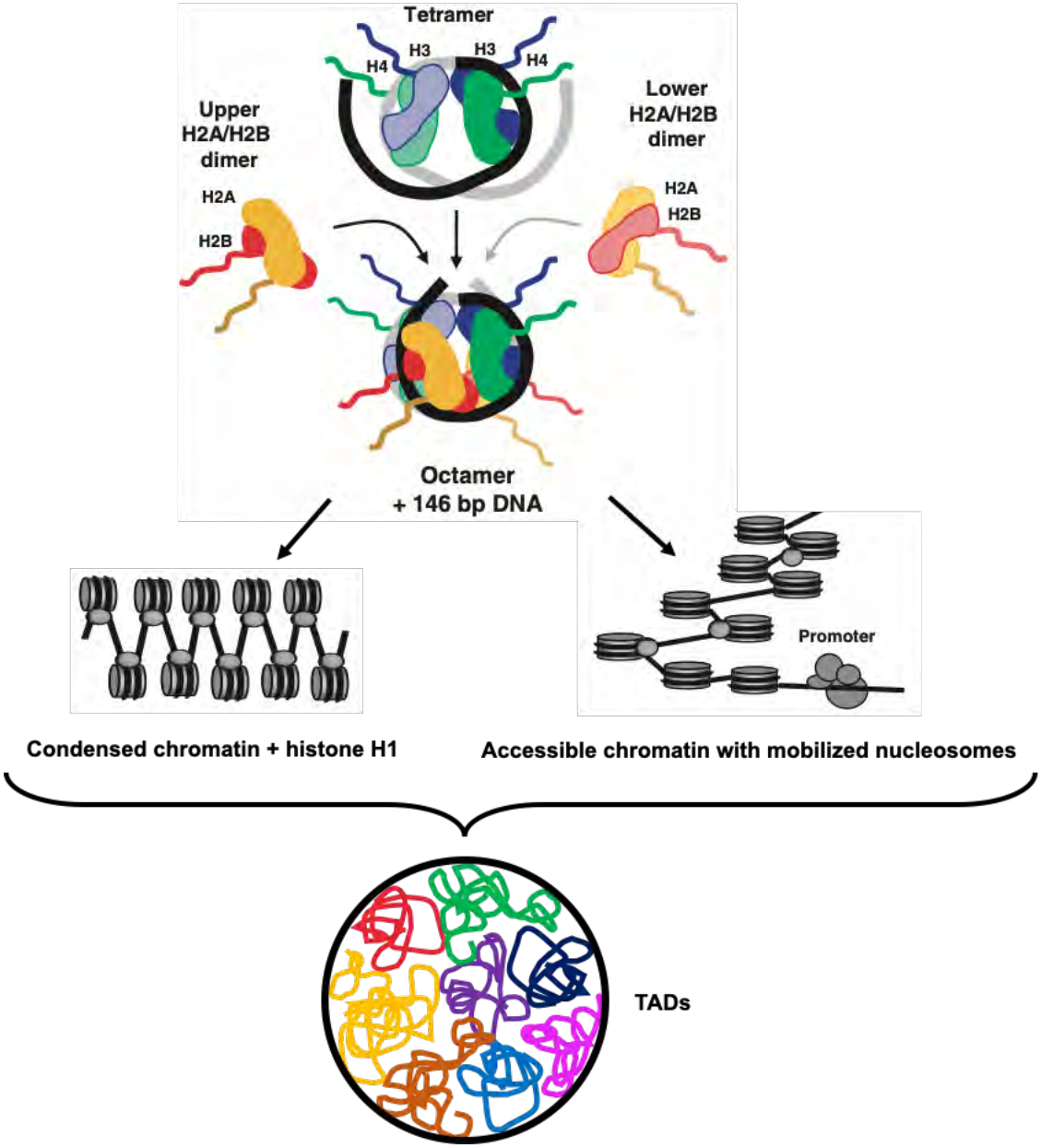


Figure 1.1: Chromatin structure and degrees of compaction

The histone octamer is assembled to DNA to form the first degree of compaction: the nucleosome. Two pairs of H3 and H4 are loaded onto DNA before the incorporation of two H2A/H2B dimers above and below the H3/H4 tetramer, resulting in the wrapping of ~146 bp of DNA. Nucleosomes are assembled into a higher order structure by the incorporation of histone H1, which binds on the edge of the nucleosomes and results in the formation of condensed chromatin fibre. Nucleosome positioning is mobile and can be modulated by different transcription factors and chromatin modifiers, resulting in accessible chromatin state. Within the nucleus, chromatin is organized in three-dimensional domains called topologically associated domains (TADs), which bring into proximity genes with their respective regulatory elements and hence include similarly regulated genes. Figure adapted from (Cockerill, 2011).

1.2 Transcriptional regulation in eukaryotes

By nature, chromatin consists a repressive state, restricting access to transcription factors (TFs) to regulatory elements and ultimately preventing transcriptional activation. Coordinated gene expression is finely regulated by the binding of sequence-specific TFs bringing in chromatin modifiers, leading to stable phenotypic changes on the chromatin state that allow the recruitment of the basal transcriptional machinery. Binding of TFs to regulatory regions results in the recruitment of chromatin remodellers, which create a permissive chromatin environment by actively modifying the chromatin structure. Simultaneously, the maintenance of an accessible chromatin state is mediated by multiple chromatin modifiers through modulation of the DNA methylation, nucleosome occupancy, and post-translational modifications of histones.

1.2.1 Transcription factors

Unique patterns of gene expression in different cell types and during development is dictated by the presence of specific TFs that bind to DNA to open up the chromatin, hence facilitating the recruitment of other transcriptional and chromatin regulators, and to control the rate of transcription of their target genes. TFs are composed of modular structures including DNA binding, transactivation and dimerization/interaction domains. TFs bind to specific DNA sequences, termed 'binding motifs'. Their affinity for a particular motif is determined by the three-dimensional structure of the DNA-binding domain, which can take a number of forms,

such as zinc finger, basic leucine zipper, helix-turn-helix and basic-loop-helix (Lee and Young, 2000).

TFs are able to simultaneously access their target sequences on nucleosomes or in compacted chromatin through cooperative binding with other TFs. However, pioneer TFs are able to actively open and bind chromatin, enabling other TFs to bind, hence being crucial for transcriptional processes requiring sequential TF binding (reviewed in Zaret and Carroll 2011). Some TFs acting downstream to pioneer TFs, such as TAL1/SCL, have a role in 'priming' regulatory sequences through stable binding for their activation at a later time during cell-fate decision (Lichtinger et al., 2012; Org et al., 2015).

TFs can directly recruit and interact with other co-factors to co-operatively bind and enhance their DNA-binding affinity or to act synergistically upon binding in order to activate or repress transcription (Spitz and Furlong, 2012). Recruited co-activators and co-repressors act as 'hub' proteins by integrating TFs and epigenetic modifying enzyme signalling (Watson et al., 2012). Co-factors do not bind DNA directly, but can bind other TFs, and include proteins involved in the initiation of transcription as well as chromatin remodelling and histone modifying enzymes (Fishburn et al., 2005).

TFs activate gene expression mainly by (i) recruiting chromatin remodellers, such as SWI/SNF, to regulatory elements to increase DNA accessibility and allow binding of other co-activators (de la Serna et al., 2005), (ii) facilitating assembly of co-factor complexes at promoters prior to the start of transcription (Heo et al., 2008), (iii) enhancing the activity of the general transcriptional machinery via recruitment of co-activators, chromatin modifiers and factors that mediate post-translational

modifications to Pol II (Jonkers and Lis, 2015), and (iv) mediating physical interactions between regulatory elements such as enhancers and promoters (Chen et al., 2012; Levantini et al., 2011).

TFs can also mediate transcriptional repression, as exemplified by Zinc finger multiprotein TF families harbouring repressive domains, such as the Krüppel-associated box. Upon binding to DNA, these repressive TFs lead to gene silencing through recruitment of co-repressors, such as the heterochromatin protein 1 (HP1) and histone deacetylases and methyltransferases (Sripathy et al., 2006).

1.2.2 Transcriptional regulatory elements

Transcription of a protein-coding gene is modulated by different types of regulatory elements including promoters, located at the transcription start site (TSS), and distal elements such as enhancers, silencers and insulators. Regulatory elements are recognized and bound by TFs in a DNA sequence-specific manner. The co-operative assembly of multi-protein complexes, including TFs and co-factors, replace the existing nucleosomes (reviewed in Merika and Thanos, 2001), resulting in the generation of highly accessible regions that are sensitive to DNase I digestion, denominated DNase I hypersensitive sites (DHS) (Stalder et al., 1980). DHS appear free of nucleosomes but can also be alternatively assembled with unstable nucleosomes containing the H2AZ and H3.3 histone variants (reviewed in Cockerill, 2011).

1.2.2.1 Promoters

Promoter regions constitute the proximal promoter and core promoter region. The proximal promoter region contains binding sites for the binding of TF and activator proteins (Maston et al., 2006), whilst the core promoter contains the TSS of the protein-coding gene and several conserved motifs, which are necessary for the assembly of the transcriptional pre-initiation complex and the basic transcription machinery (Smale and Kadonaga, 2003; Zhang, 1998). The core promoter often includes a TATA box motif and other motifs such as the downstream core, downstream promoter and motif ten elements (DCE, DPE and MTE, respectively), which serve as docking site for the general TFs such as the TFIID subunit TBP or TFIIID (Morris et al., 2004).

The combination and frequency of the distinct sequence elements is variable between the different promoters in the genome. For example, only a subset of promoters is thought to contain a canonical TATA box motif (Gershenzon and Ioshikhes, 2005). Promoters that lack the core promoter motifs, including the TATA box, tend to be enriched with high CG content regions, called CG islands (Blake et al., 1990). Promoters have therefore been classified into three categories based on the presence or absence of TATA box and CG islands within the promoter region (Lenhard et al., 2012): type I promoters have TATA box and low CG content and are associated with inducible and tissue-specific genes, type II promoters lack the TATA motif, have CG sequences over the TSS and are associated with constitutively expressed housekeeping genes, and type III promoters display spread CG rich regions and are associated with developmental genes, which are usually regulated through interactions with distal elements, such as enhancers.

1.2.2.2 Enhancers

The activity of promoters can be temporally and spatially modulated through interactions with multiple distal regulatory elements, denominated enhancers (de Villiers and Schaffner, 1981). Enhancers are constituted by a collection of TF binding motifs and their active state is usually associated with DHS, bound TFs and co-factors as well as H3K4me1 and H3K27ac histone marks (Rada-Iglesias et al., 2011).

A single promoter can be simultaneously regulated by several enhancers, as exemplified by the *Fdf8* promoter (Marinić et al., 2013). Unlike promoters, which are located upstream of the TSS, enhancers may be located either up- or down-stream to the target gene, within introns or even megabases away from the interacting promoter (Mifsud et al., 2015). Therefore, genes regulated by the same subset of enhancers often encode proteins that are tissue and developmental-stage specific (Spitz and Furlong, 2012).

1.2.2.3 Silencers and Insulators

Certain regulatory elements are not involved in transcriptional activation. These negative regulatory elements include silencers and insulators, which interfere with promoter-enhancer interactions.

Silencers are involved in repressing transcription and, similar to enhancers, may be located distal to their target genes. Silencers contain multiple binding sites for TFs, such as CT-rich motifs (Petrykowska et al., 2008), that recruit repressive co-factors (reviewed in Privalsky, 2004). Key examples of corepressors recruited to silencers

include the Silencing Mediator of Retinoic acid and Thyroid hormone receptors (SMRT) and the Nuclear hormone receptor Co-Repressor (N-CoR) (Chen and Evans, 1995; Sande and Privalsky, 1996). TFs and associated co-repressors may exert their functions by directly competing with co-activators for binding DNA (Harris et al., 2005), by recruiting epigenetic modifying enzymes to create a repressive chromatin structure (Srinivasan and Atchison, 2004) or by interfering with the assembly of the basal transcriptional machinery near the TSS, as exemplified by PAX5 (Tagoh et al., 2006).

Insulators, which are located within gene boundaries, create discrete domains of gene regulation by isolating a locus from the regulatory activities occurring in neighbouring domains, such as blocking the spread of repressive DNA methylation or preventing unwanted enhancer-promoter interactions (Recillas-Targa et al., 2002).

1.2.3 Regulation and maintenance of the chromatin state

1.2.3.1 Nucleosome remodelling and histone variants

Nucleosomes prevent transcription by physical obstruction, hence reducing accessibility of TFs to the DNA sequences wrapped around the histone core (Lee et al., 1993; Wasylyk and Chambon, 1979).

Nucleosome occupancy can be altered by ATP-dependent chromatin remodelling complexes, which are able to alter histone-DNA contacts and promote exchange of histone complexes and other factors by using the energy from ATP hydrolysis. The most studied remodelling complexes include SWI/SNF, nucleosome-remodelling

factor (NURF) and BRG1-associated factor (BAF) (reviewed in Chen and Dent 2014; Guan et al. 2013; Skene and Henikoff 2013; Smith and Meissner 2013).

Nucleosome remodelling is further tuned by the exchange of canonical histones for histone variants. Histone variants exhibit minor differences in their amino acid sequences, compared to their canonical counterparts. Incorporation of histone variants alters the interaction strength within the nucleosome core, affecting its stability and hence the degree of compaction of the chromatin fibre (Becker and Workman, 2013). Unlike canonical histones, histone variants are incorporated into chromatin independently of DNA replication, and result in distinct chromatin functions (Albig and Doenecke, 1997). Using H2 variants as examples; H2A.Z is enriched at transcription start sites and correlates with the presence of hypomethylated DNA, whilst phosphorylated H2A.X marks DNA double-strand breaks (reviewed in Henikoff and Smith, 2015).

1.2.3.2 Histone modifications

The nucleosome carries regulatory information, because histones can undergo post-translational modification (PTMs). Such PTMs occur mainly on the amino-terminus (N-terminal) histone tails, which project outside the nucleosome making them accessible to chromatin modifying enzymes (Luger et al., 1997). These modifications primarily include acetylation (ac), methylation (me), phosphorylation, ubiquitylation, sumoylation, ADP ribosylation and deamination (reviewed in Cosgrove et al., 2004; Kouzarides, 2007). PTMs can also occur within the globular domain of the histones (Tropberger and Schneider, 2013). An example is H3K79me (Zhang et al., 2002a),

which correlates with active gene expression and transcript abundance. In active genes, H3K79me3 is found at the transcriptional start site, whilst H3K79me1 extends over the gene body (Steger et al., 2008).

Histone PTMs influence chromatin compaction by affecting interaction between nucleosomes. In addition, combinatorial patterns of histone PTMs indirectly dictate gene transcription and chromatin state by recruiting chromatin regulators, which contain PTM 'reader' domains (Clements et al., 2003; Vettese-Dadey et al., 1996). The first histone-modification-binding module that was identified was the bromodomain from p300/CBP-associated factor (PCAF), which binds to acetylated histones (Dhalluin et al., 1999). Another example, heterochromatin protein 1 (HP1), binds methylated H3K9 (Bannister et al., 2001; Lachner et al., 2001).

Histone modifications are added to, or removed from, chromatin by distinct enzymes – commonly known as 'writers' and 'erasers', respectively – including histone acetyltransferases (HAT) and deacetylases (HDAC), lysine methyltransferases (KMT) and demethylases (KDM), arginine methyltransferases, kinases and phosphatases, or ubiquitylation enzymes (E1, E2, E3) and deubiquitylases. The first writers discovered were HATs (Brownell et al., 1996), which act as transcriptional co-activators, and include the TATA-box binding protein associated factor TFIID subunit 1 (TAF1) (Mizzen et al., 1996), CREB-binding protein and p300 (CBP/p300) (Bannister and Kouzarides, 1996; Ogryzko et al., 1996) and PCAF (Yang et al., 1996). Soon after the discovery of HATs, HDACs were identified as transcriptional co-repressors (Taunton et al., 1996). Later on, several KMTs, containing an evolutionary conserved SET domain, were described, including SUV39H1 (Tschiersch et al., 1994), G9a (Tachibana et al., 2001), Trithorax factors (such as

MLL) (Milne et al., 2002) and EZH2 (Czermin et al., 2002). Interestingly, histone lysine methylation can either be repressive, such as SUV39H1-mediated H3K9me3 (Rea et al., 2000), or activating, such as MLL-mediated H3K4me3 (Milne et al., 2002). Moreover, some KMTs lacking the SET domain are able to methylate the globular domain of histones. This includes DOTL1, which mediates H3K79me (van Leeuwen et al., 2002). Histone lysine methylation can me erased by KDMs, such as lysine-specific histone demethylase 1 (LSD1), which acts as a transcriptional co-repressor (Shi et al., 2004). Methylation can also occur at histone arginines, mediated by the co-activator-associated arginine methyltransferase 1 (CARM1) (Chen et al., 1999) or the protein arginine N-methyltransferase 1 (PRMT1) (Wang et al., 2001).

The combined modulation of histone variants and histone modifications affects nucleosome occupancy, regulating chromatin mobility within the nucleus, and co-operates with TFs and proteins from the DNA replication and repair machinery.

1.2.3.3 DNA methylation

DNA bases can be directly modified by methylation or hydroxymethylation of cytosines, resulting in chromatin condensation (Hotchkiss, 1948). DNA methylation results in gene repression (Razin and Riggs, 1980) and occurs mainly in CG dinucleotides (Bird et al., 1985), with the exception of the CG islands commonly located at promoters, which are generally unmethylated (Larsen et al., 1992). DNA methylation is regulated by DNA (cytosine-5)-methyltransferase (DNMT) enzymes, including DNMT1, DNMT3a and DNMT3b. *De novo* DNA methylation during early

development is established by DNMT3A and (Hsieh, 1999; Okano et al., 1999). DNMT1 prefers to methylate new CG dinucleotides with neighbouring methylated DNA (Pradhan et al., 1999), hence maintaining DNA methylation through cell divisions.

DNA methylation can modulate transcription, mainly via two mechanisms. The first mechanism involves direct interference of the methyl groups with the binding of TFs to their specific motif sequences. This scenario can also lead to transcriptional activation through interference with the binding of transcriptional repressors, as exemplified by the prevention of CTCF binding resulting in activation of Igf2 expression (Bell and Felsenfeld, 2000). The second mechanism involves the binding of methyl-CpG-binding protein (MeCP) complexes to the DNA methyl groups (Meehan et al., 1989).

MeCPs affect chromatin structure and gene expression, as exemplified by MeCP2, which recruits HDACs through binding the transcriptional co-repressor mSIN3A (Nan et al., 1998). Moreover, MeCP2 can also bind hemi-methylated DNA and recruit DNMT1 (Kimura and Shiota, 2003), which maintains DNA methylation upon stabilization by multi-domain factors that form a bridge between H3K9me and hemi-methylated DNA (Liu et al., 2013). In turn, *de novo* DNA methylation can be prevented by KMTs, which are recruited by some TFs that can bind to CG islands with high affinity (Blackledge et al., 2010). Therefore, DNA methylation and histone modifications work in concert to regulate gene expression.

1.2.4 Bivalent domains and reversible chromatin marks

Many chromatin signatures have been described, assigning functional patterns to each histone modification. This is well exemplified by H3K4me3 and H3K27me3, which are associated with active promoter elements and repressed developmentally-controlled gene regions, respectively (Boyer et al., 2006). However, H3K4me3 and H3K27me3 can also co-occur within the same genomic region (Bernstein et al., 2006), and more recent studies suggest that these co-exist asymmetrically on different histone peptides within the same mono-nucleosome (Sen et al., 2016; Voigt et al., 2012). These intermediate signatures harbouring both active and repressive marks have been referred to as bivalent domains.

Unlike genetic alterations, the vast majority of epigenetic modifications have been proved to be reversible. One of the few exceptions is exemplified by *de novo* promoter methylation mediated by Krüppel-associated TFs during early murine embryogenesis, which leads to irreversible gene repression (Wiznerowicz et al., 2007). Reversal of chromatin marks can be mediated by chromatin-modifying enzymes (known as 'erasers') or by chemical inhibition. Erasers of epigenetic marks include HDACs, KDMs and phosphatases, which remove histone modifications, as well as the ten-eleven-translocation-1 (TET1) family of enzymes, which converts 5mC to 5hmC resulting in fully demethylated DNA (He et al., 2011b). The reversal of aberrant chromatin marks by chemical inhibition is important in the context of disease, as dysfunctional chromatin regulators become good drug targets for the development of directed therapies, as has been the case for DNMT and/or HDAC inhibitors (Dawson and Kouzarides, 2012).

1.2.4.1 Chromatin regulation in pluripotent stem cells and differentiated cells

Epigenetic control is crucial for cell-type identity and development. Pluripotent cells have extensive regions of transcriptionally accessible chromatin, consisting of hypomethylated DNA marked by histone acetylation and H3K4me (Figure 1.2) (Azuara et al., 2006; Guenther et al., 2007; Lister et al., 2009; Meissner et al., 2008; Meshorer et al., 2006). On one hand, genes involved in pluripotency remain active by retaining H3K27ac at their enhancers (Creyghton et al., 2010). Conversely, lineage-specific genes are kept silent, and a proportion of genes remain in a 'poised' state. Poised genes harbour H3K27me3 and H3K4me3 bivalent domains in both enhancer and promoter regions and present paused RNA polymerase II in the promoter-proximal region. Bivalent domains therefore allow for rapid initiation of transcription of a lineage-specific program upon removal of the repressive H3K27me3 modification (Azuara et al., 2006; Bernstein et al., 2006; Mikkelsen et al., 2007; Pan et al., 2007; Zhao et al., 2007). In contrast, genes involved in the regulation of other developmental programs are silenced through repressive H3K27me3 marks at promoter regions (Creyghton et al., 2010; Gifford et al., 2013; Rada-iglesias et al., 2011; Wamstad et al., 2012; Xie et al., 2013). Therefore, differentiated cells are characterized by the presence of large chromatin domains harbouring repressive marks, such as H3K27me3, H3K9me2 and H3K9me3, which prevent expression of non-specific lineage programs (Figure 1.2) (Pauler et al., 2009; Wen et al., 2009; Zhu et al., 2013).

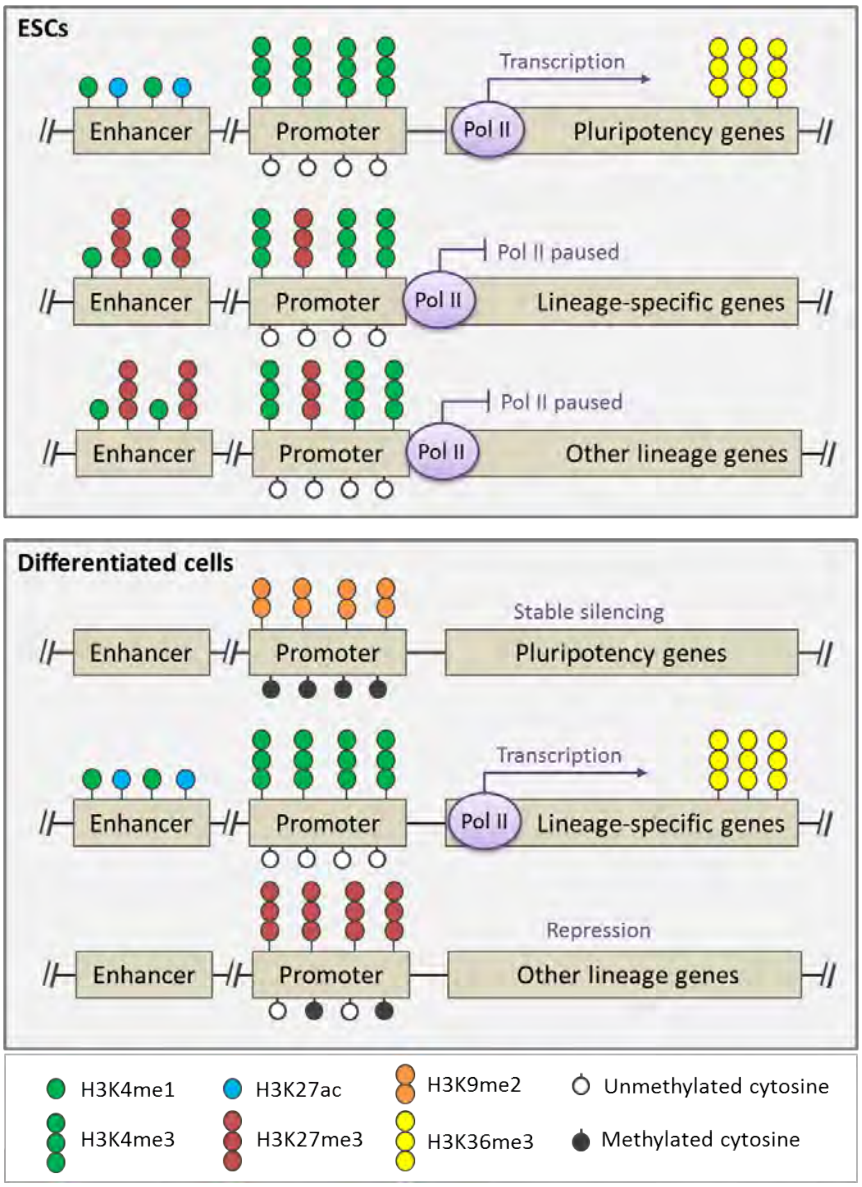


Figure 1.2: Chromatin modifications in pluripotent and differentiated cells

In embryonic stem cells (ESCs), pluripotency genes are kept active by maintaining hypomethylated DNA and active histone modifications at the genomic regulatory regions, including H3K4me1 and H3K27ac at enhancers and H3K4me3 at promoters. Actively transcribed pluripotency genes are enriched with H3K36me3 across their gene body. Conversely, lineage-specific genes contain bivalent domains in both enhancers and promoters, consisting of both repressive H3K27me3 and active H3K4me histone modifications. In differentiated cells, pluripotency genes are stably repressed through DNA methylation, acquisition of and H3K9me2 and loss of H3K4me3 at promoter regions. Lineage-specific genes lose the repressive marks at their bivalent domains and acquire active H3K27ac at enhancers, leading to transcriptional elongation. Genes from other lineages are silenced upon removal of the active marks at the bivalent domains. Pol II: RNA polymerase II. Figure adapted from *Chen and Dent 2014* (Chen and Dent 2014).

1.3 The basal transcriptional machinery

Transcription of protein-coding genes begins with the assembly of Pol II and a large set of general TFs (TFIIB, TFIID, TFIIE, TFIIF and TFIIH) at the promoter regions near the TSS to form a conserved pre-initiation complex (PIC). The classical model of stepwise PIC assembly usually commences with the binding of TFIID, which contains the TATA box-binding protein (TBP) and TBP-associated factors (TAFs) with promoter specific function. Binding of TFIID is mediated by other members of the general TFs (TFIIB and TFIIA) and the resulting upstream promoter complex is subsequently joined by Pol II and TFIIF. Binding of TFIIE and TFIIH result in the formation of the complete PIC, which remains in a closed state. Transcriptional elongation is dependent on phosphorylation of the carboxy-terminal (C-ter) domain of Pol II at Serine 5 and the presence of ATP, which allows Pol II to open the DNA within the PIC. This results in the formation of a transcriptional 'bubble' that permits Pol II to initiate transcription of a few nucleotides of RNA. Subsequently, Pol II pauses and accumulates 30-60 nucleotides downstream of the TSS (Figure 1.3) (reviewed in Grünberg and Hahn, 2013; Sainsbury, Bernecky and Cramer, 2015).

Prior to productive elongation, Pol II is held paused in promoter-proximal regions by negative elongation factor (NELF) and DRB-sensitivity-inducing factor (DSIF) (Muse et al., 2007; Wada et al., 1998; Yamaguchi et al., 1999), as well as by the core promoter elements and the +1 nucleosome (Kwak et al., 2013; Li and Gilmour, 2013; Weber et al., 2014) (Figure 1.3). NELF interacts with the RNA sequences that protrude from Pol II and colocalizes with DSIF to stabilize the paused Pol II (Yamaguchi et al., 2002). Pol II pausing is aided by the presence of the +1 nucleosome, which hinders elongation by generating a physical barrier (Li and

Gilmour, 2013; Weber et al., 2014). Pol II pausing is required to keep the promoter free from nucleosomes and accessible to TFs, since knockdown of NELF not only leads to a reduction of Pol II pausing but to an increase of nucleosome occupancy within the promoter region (Core et al., 2012; Gilchrist et al., 2010). In addition, Pol II pausing facilitates the assembly of transcriptional and co-transcriptional complexes, such as the splicing machinery (Barboric et al., 2009).

In order to progress to productive transcriptional elongation, Pol II must be released from the promoter-proximal region (Figure 1.3). This process begins with the recruitment of the positive transcription elongation factor-b (P-TEFb) kinase complex by several TFs and by co-activators, such as bromodomain-containing protein 4 (BRD4) and super elongation complex (SEC), forming the activating complex (Rahl et al., 2010; Smith et al., 2011). The P-TEFb kinase complex contains the cyclin dependent kinase 9 (CDK9) (Lin et al., 2002), which releases paused Pol II by phosphorylating (i) the C-terminal domain of Pol II at Serine 2, (ii) NELF, dissociating it from Pol II, and (iii) DSIF, turning it into a positive elongation factor (reviewed in Adelman and Lis, 2012; Bacon and D'Orso, 2019; Peterlin and Price, 2006; Zhou et al., 2012). Release of Pol II into elongation is also influenced by SEC, which interacts with other co-activators associated with transcriptional elongation, including polymerase-associated factor 1 (PAF1), Mediator complex and Integrator complex (Gardini et al., 2014; He et al., 2011a; Kim et al., 2010; Takahashi et al., 2011; Wier et al., 2013). SEC may indirectly facilitate the recruitment of P-TEFb through recruitment of Mediator complex, which aids in the formation of promoter-enhancer interactions (Takahashi et al. 2011; reviewed in Allen and Taatjes 2015). After release of Pol II from the core promoter, some components of the PIC remain

assembled acting as a scaffold for subsequent recruitment of Pol II and re-initiation of transcriptional elongation (Hahn, 2004). Elongation rates can be modulated by different factors including elongation factors, such as DSFI and PAF1, chromatin modifiers and remodellers, affecting histone marks and nucleosome displacement, specific DNA sequences and presence of introns and exons (Alexander et al., 2010; Fuchs et al., 2014; Jonkers et al., 2014; Veloso et al., 2014).

Pol II continues productive elongation until it reaches the end of the gene body, where it pauses at poly(A) sequences. In order to allow Pol II release from its paused state, CDK9 phosphorylates (i) a regulatory subunit of the Protein Phosphatase 1 (PP1), which in turn, prevents dephosphorylation of DSIF and (ii) maintains phosphorylation of Pol II at Serine 2, which, together allow Pol II transcription through the poly(A) site. Reactivation of PP1 results in the dephosphorylation of DSIF and transcription termination through recruitment of termination factors, including cleavage and polyadenylation-specific factor (CPSF) and cleavage-stimulatory factor (CSTF) (reviewed in Bacon and D'Orso, 2019). Termination factors cleave and polyadenylate the nascent RNA, resulting in productive termination (Core et al., 2008).

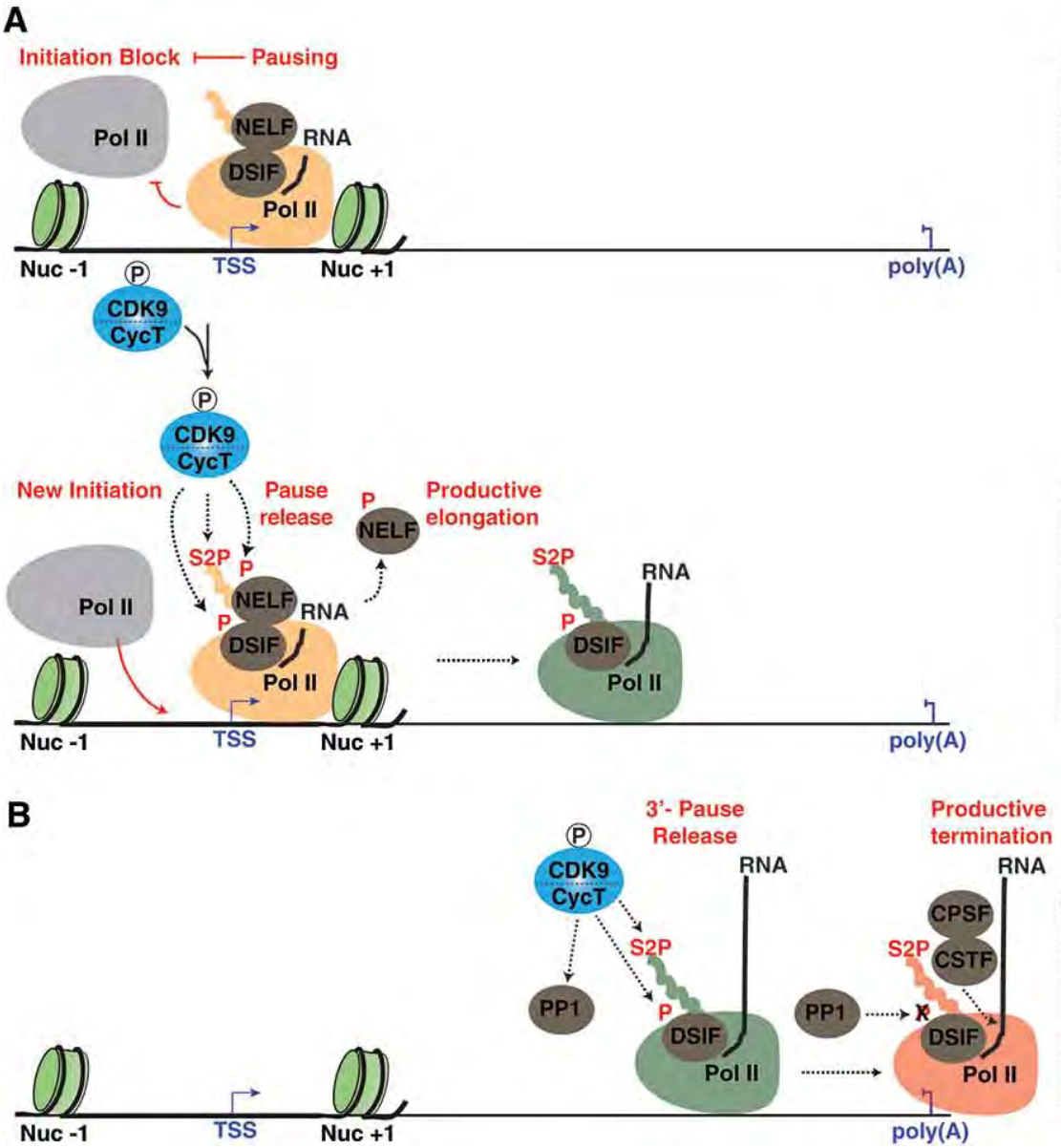


Figure 1.3: RNA polymerase II recruitment, pausing and release

(A) Pol II is recruited by general TFs and, together, forms the pre-initiation complex, which interacts with the core elements within the promoter region. Following a transcription of a short length of DNA nucleotides, the pausing factors NELF and DSIF retain Pol II paused at the promoter-proximal region. Other factors assisting in Pol II pausing include the core promoter elements and the +1 nucleosome (Nuc +1), which creates a physical barrier for Pol II preventing elongation. Pol II is released into elongation upon recruitment of CDK9, which phosphorylates Pol II at Ser2 (S2P) as well as both NELF and DSIF. Phosphorylation causes eviction of NELF from Pol II and conversion of DSIF into a positive elongation factor.

(B) Pol II pauses at poly(A) sites, where it is phosphorylated by CDK9 at Ser2. CDK9 also phosphorylates protein phosphatase 1 (PP1, preventing the dephosphorylation of DSIF. Pol II is then released from its paused state and transcribes through the poly(A) site, resulting in the re-activation of PP1, dephosphorylation of DSIF, and transcription termination through recruitment of termination factors (CPSF and CSTF), which cleave and polyadenylate the nascent RNA. Figure modified from Bacon and D'Orso, 2019.

1.4 Haematopoiesis

1.4.1 Haematopoietic development in the embryo

Our current understanding of vertebrate haematopoietic development comes largely from studies in non-human model organisms (reviewed in Crisan and Dzierzak, 2016; Kauts et al., 2016; Kim et al., 2014; Medvinsky et al., 2011; Robertson et al., 2016). Due to limited availability of early human embryos, the study of human haematopoietic development has been restricted to morphological studies of small numbers of embryos, *in vitro* assays and, *in vivo* functional evaluation of emerging human HSCs in immunodeficient mouse models (Shultz et al., 2012).

Haematopoiesis in mammalian embryos comprises at least three discrete programs: (i) primitive, (ii) definitive erythroid/myeloid progenitor (EMP) and (iii) intraembryonic definitive progenitors. Each wave of progenitor formation takes place in different tissues and/or spans different temporal windows and generates cells with different potential (Figure 1.4). Despite differences in the duration of the developmental time course, mouse and human embryonic haematopoiesis share many similarities.

In order to ensure embryonic survival prior to the establishment of circulation and HSC formation, haematopoietic progenitors are generated in the yolk sac from an extraembryonic mesodermal progenitor, the haemangioblast. A primitive wave of yolk sac progenitors, at 16-18.5 human days post-conception (dpc), mainly includes nucleated erythrocytes and rare primitive macrophages and megakaryocytes (Bloom and Bartelmez, 1940; Luckett, 1978). The first primitive erythrocytes observed within the intra-embryonic cardiac cavity mark the onset of circulation at 21-22 dpc (Tavian et al., 1999). A second wave of yolk sac progenitors includes EMPs, megakaryocytic

and lymphoid progenitors that migrate to transiently colonise the foetal liver (Chen et al., 2011; McGrath et al., 2015; Palis et al., 1999; Tober et al., 2007). In the human embryo, this may occur around 28-35 dpc (Migliaccio et al., 1986).

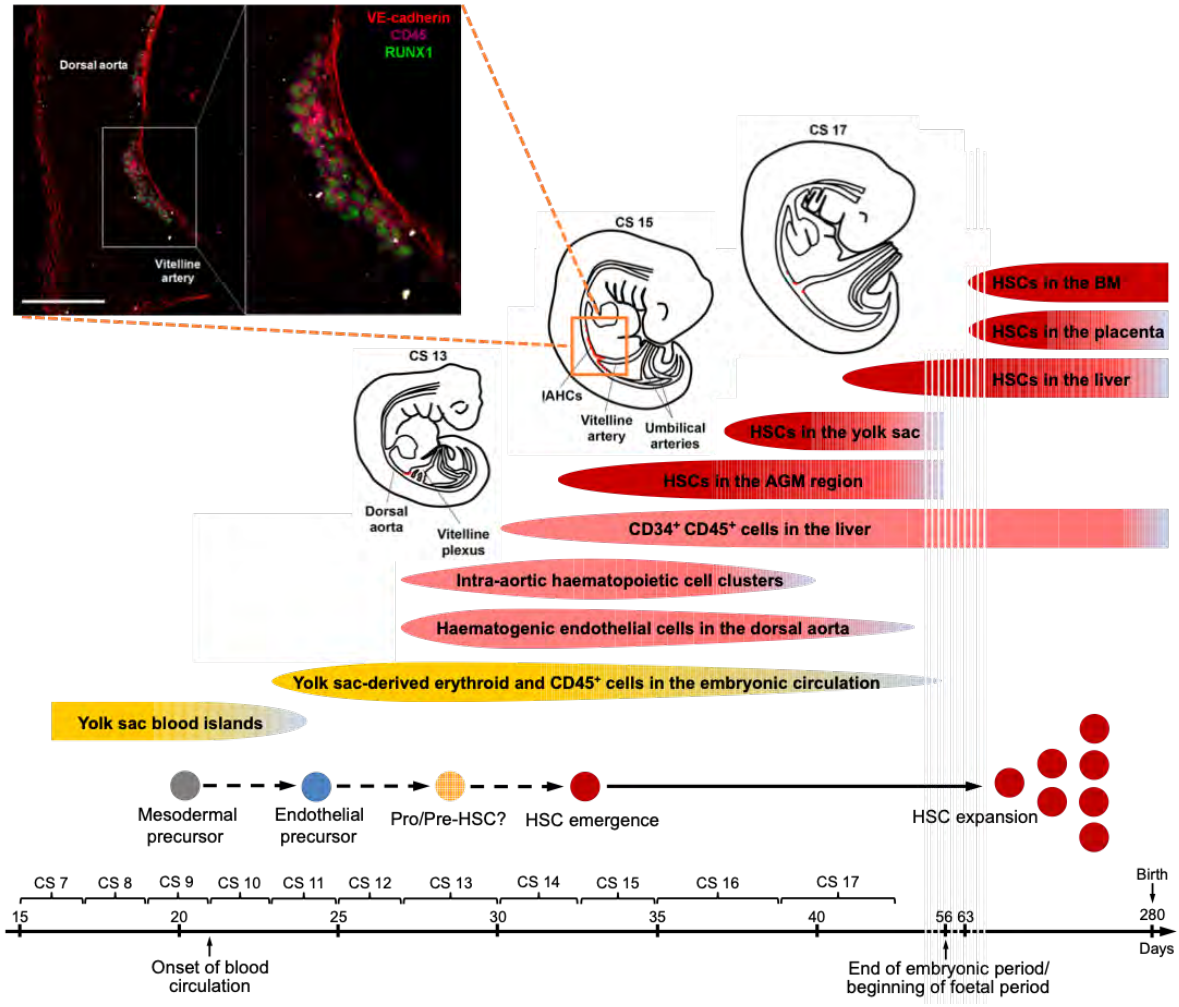


Figure 1.4: Chronology of human embryonic haematopoietic development

The first human haematopoietic cells are generated from an extra-embryonic mesodermal precursor (grey), the haemangioblast, in the blood islands of the extraembryonic yolk sac during 16-18.5 days post-conception (dpc). This primitive wave of progenitors includes nucleated erythrocytes and rare primitive macrophages as well as megakaryocytes, and ensures embryonic survival prior to the establishment of circulation, which occurs at 21-22 dpc. The first CD34+ CD45+ progenitor cells originate from intra-aortic haematopoietic clusters (IAHC) from haemogenic endothelium (HE, blue) located in the ventral side of the dorsal aorta, in a process known as endothelial-haematopoietic transition (EHT) at 27 dpc. Presence of HSCs has been identified from 30-42 dpc in the aorta-gonad-mesonephros (AGM), which is the first region harbouring HSC activity within the embryo. Individual progenitor cells detach from the HE clusters and seed the foetal liver to expand and mature. After 56 dpc, blood progenitors colonize the bone marrow, which will remain the main haematopoietic niche throughout adult life. Carnegie stages (CS) are stages of human development as defined by external morphological features. Yellow: yolk sac haematopoietic differentiation, salmon pink: definitive haematopoietic lineages, red: *bona-fide* HSCs. White lines within the coloured bubbles represent a

change in time scale (omission of weeks). Immunofluorescence image shows VE-cadherin+CD45+RUNX1+ human embryonic IAHCs at CS 16. Right image corresponds to a sagittal confocal section of the boxed area. Scale bar: 0.05 mm. Modified figure, adapted from *Ivanovs et al., 2017* (Ivanovs et al., 2017).

The intraembryonic program generates the first pre haematopoietic stem cells (pre-HSC) and repopulating HSC, with self-renewal and long-term engraftment capacity that give rise to the adult haematopoietic system. The aorta-gonad-mesonephros (AGM) region develops from para-aortic splanchnopleura mesoderm and includes the dorsal aorta, genital ridges and mesonephros. The ventral side of the dorsal aorta contains specialized vascular cells with endothelial potential, known as haemogenic endothelium (HE). The human aortic endothelial lining of the AGM is marked by the expression of CD34 from 19 dpc (Oberlin et al., 2002). The first HSCs originate within the AGM region from HE in a process known as endothelial-haematopoietic transition (EHT). During EHT, intra-aortic clusters of haematopoietic cells emerge from the endothelial cell layer from day 10.5 in the mouse and 27 dpc in the human embryo and will generate the first pre-HSCs, HSCs and progenitor cells (de Bruijn et al., 2000; Chen et al., 2009; Cumano et al., 1996; Eilken et al., 2009; Godin et al., 1995; Lancrin et al., 2009; Medvinsky and Dzierzak, 1996; Medvinsky et al., 2011; Müller et al., 1994; Rybtsov et al., 2014; Taoudi et al., 2008; Zovein et al., 2008). Thereafter, haematopoietic cells detach from the aortic clusters and migrate to the foetal liver where they expand and mature (Bertrand et al., 2010; Boisset et al., 2010; De Bruijn et al., 2002; Jaffredo et al., 1998; Kissa and Herbomel, 2010; Lam et al., 2010; North et al., 2002). This developmental phenomenon has been widely studied in the murine embryo, where the transition from aortic cluster pre-HSCs to foetal liver HSCs occurs within 24 hours (Rybtsov et al., 2016). It is thought that human HSCs undergo a

similar developmental process, however, the seeding of the foetal liver with HSCs takes at least one week after the emergence of aortic pre-HSCs, suggesting that the molecular mechanisms of human pre-HSC emergence and maturation to HSCs may differ (Ivanovs et al., 2011).

The human yolk sac and para-aortic splanchnopleura, two mesodermal-derived structures, exhibit distinct haematopoietic potentials as observed in *ex vivo* cultures of these embryonic tissues isolated prior to the onset of circulation at 21 dpc. The yolk sac generates myeloid and natural killer cells only, whilst the para-aortic splanchnopleura forms a wider range of haematopoietic progenitors including B- and T- lymphoid cells (Tavian et al., 2001). Moreover, cobblestone area-forming cells, an *in vitro* assay predictive of multipotent precursors and HSCs, were present in explants from the dorsal aorta but not from the yolk sac (Oberlin et al., 2002). Transplantation studies of human cells into immunodeficient mice indicate that the first long-term multilineage HSCs are generated in the AGM around 30-32 dpc and remain present predominantly within this region up until 42 dpc, when HSCs are found in the foetal liver. A low number of definitive HSCs with regenerative potential is generated in the human AGM (1-2 per embryo), however, the presence of nascent, pre-HSCs could have been underestimated, as they cannot be yet detected by direct transplantation of cells from the human AGM into mice (Ivanovs et al., 2011). Indeed, this conclusion is reached in more detailed studies of mouse AGM by the same investigators (Rybtsov et al., 2016). The first human AGM-derived HSCs are contained within a population of CD34+VE-Cadherin+CD45+KIT+CD90+endoglin+ RUNX1+CD38-/lo CD45RA- cells (Ivanovs et al., 2014). HSC development is also dependent on cell adhesion and migration, which is regulated through surface

molecules expressed in the intra-aortic endothelium and cell clusters, including ALCAM, VE-Cadherin, CD34, CD43, CD44, CD164, VCAM1 (Drew et al., 2005; Ivanovs et al., 2014).

The first blood cells seen in the foetal liver are primitive yolk-sac erythrocytes and CD45+ monocyte/macrophage cells at 22-23 dpc (Tavian et al., 1999). From 27-30 dpc, an increasing number of CD34+CD45+ progenitors – most likely representing yolk-sac EMPs and pre-HSCs generated in the intra-aortic clusters – seed the liver, which remains a crucial site for HSC expansion and haematopoietic differentiation. Bone marrow formation marks the end of the embryonic period at 56 dpc, with formation in the long bones antedating bone marrow formation in the vertebrae (Chen and Weiss, 1975). The bone marrow remains the main haematopoietic niche throughout adult life providing the adequate environment to sustain proper numbers and function of HSCs (Zovein et al., 2008).

1.4.2 *In vitro* modelling of human haematopoietic development and signalling pathways

Many laboratories have attempted to recapitulate early human blood development through the *in vitro* differentiation of human pluripotent stem cells, using a variety of protocols (reviewed in Ditadi et al., 2016; Wahlster and Daley, 2016). The main approaches that have been used are: (i) monolayer culture (D'Amour et al., 2005; Jiang et al., 2007), (ii) co-culture with a layer of stromal cells, such as mouse bone marrow stromal (S17 or OP9) or mouse yolk-sac endothelial (C166) cell lines, to induce lineage differentiation (Kaufman et al., 2001; Vodyanik et al., 2005), and (iii)

embryoid body (EB) culture (Chadwick et al., 2003; Schuldiner et al., 2000). Haematopoietic differentiation of murine and human pluripotent stem cells mimic the molecular and phenotypic events that occur during early embryogenesis (Keller, 1995; Keller et al., 1993; McGrath et al., 2015; Ng et al., 2016). Differentiating cells sequentially pass from stem cell to primitive streak like stages and then to haematopoietically patterned mesoderm, which leads to the generation of primitive and definitive haematopoiesis. However, the majority of protocols for *in vitro* haematopoietic differentiation mimic extra-embryonic yolk sac-like haematopoiesis, rather than generating the intra-embryonic AGM-like blood cell formation.

In order to reproduce intra-embryonic haematopoiesis and HSC development *in vitro*, the signalling landscape underlying progenitor formation within the AGM niche needs to be mimicked through addition of soluble cell-based growth factors at the adequate amount and in the correct temporal window. Haematopoietic development, including mesodermal specification and patterning as well as the formation of the AGM, is regulated by several growth factors involving distinct signalling pathways. Mesodermal patterning is regulated by FGF, BMP4, WNT and ACTIVIN signals (Kimelman, 2006; Nostro et al., 2008; Wang and Nakayama, 2009; Woll et al., 2008; Yu et al., 2011). Primitive streak genes in early mesoderm can be induced by BMP4 alone, which, in combination with vascular endothelial growth factor (VEGF), also supports the initial generation of haematopoietic cells from in vitro cultures (Pick et al., 2007). Recently, it has been demonstrated that addition of growth factors and inhibitors to modulate these signalling pathways allows the mesoderm to be patterned towards intra-embryonic, AGM-like haematopoiesis, inhibiting extra-embryonic yolk sac-like haematopoiesis. The primitive erythroid programme can be

suppressed by concurrent BMP4 activation and activin suppression (Kennedy et al., 2012) and addition of a WNT agonist (Gertow et al., 2013). Induction of the definitive program can similarly be achieved via activin inhibition and WNT stimulation, using the factors individually (Kennedy et al., 2012; Sturgeon et al., 2014) or in combination (Ng et al., 2016), during mesoderm patterning. These signals upregulate expression of *CDX* genes, which in turn activate selective expression of *HOXA* genes, resulting in a *HOXA*-signature similar to that seen in human foetal liver and cord blood-derived haematopoietic progenitors (Dou et al., 2016; Ng et al., 2016). Although we are not able to distinguish AGM-derived progenitors and pre-HSCs from yolk-sac-type progenitors by known cell surface markers (Ditadi and Sturgeon, 2016), they differentially express *HOXA* genes (Dou et al., 2016; Ng et al., 2016) (Figure 1.5).

In the embryo, development of the arterial and venous vasculature is dependent on NOTCH signalling (Lawson et al., 2001). Polarization of the mouse dorsal aorta and the ventral location of the haemogenic endothelium and the intra-aortic clusters results from asymmetric modulation of BMP signalling within the AGM, which displays a ventral to dorsal gradient across the aorta (Marshall et al., 2000; Souilhol et al., 2016; Wilkinson et al., 2009). BMP4 also plays a role in the formation the HSC niche, since BMP4 induces expression of molecules from the extracellular-matrix, such as tenascin C and fibronectin (Molloy et al., 2008), and is expressed in the ventral side of the dorsal aorta at 28 dpc (Ivanovs et al., 2011). BMP4 is able to upregulate KIT expression, which may facilitate initiation of HSC formation (Marshall et al., 2007). Moreover, *KIT* is expressed in the human intra-aortic clusters and the first emerging HSCs (Ivanovs et al., 2014). Mouse studies suggest that sustained BMP4 signalling is detrimental to HSC development (Souilhol et al., 2016). *BMP4*

expression is therefore downregulated, and *TGFb* expression upregulated, near the intra-aortic clusters. It is hypothesized that TGFβ may reorganize the extracellular matrix within the niche, thus preventing HSC expansion in the AGM (Marshall et al., 2000). FLT3 might also be instrumental in human HSC maturation, since FLT3 and its ligand are expressed in the intra-aortic clusters (Marshall et al., 1999). Recent studies have highlighted the importance of retinoic acid signalling for the formation of embryonic HSCs (Chanda et al., 2013). Therefore, because the presence of these growth factors is crucial for proper embryonic haematopoietic development, it is likely that their inclusion in *in vitro* systems of haematopoietic differentiation is required to simulate the human AGM niche and generation of intra-embryonic haematopoietic progenitors (Ng et al., 2016).

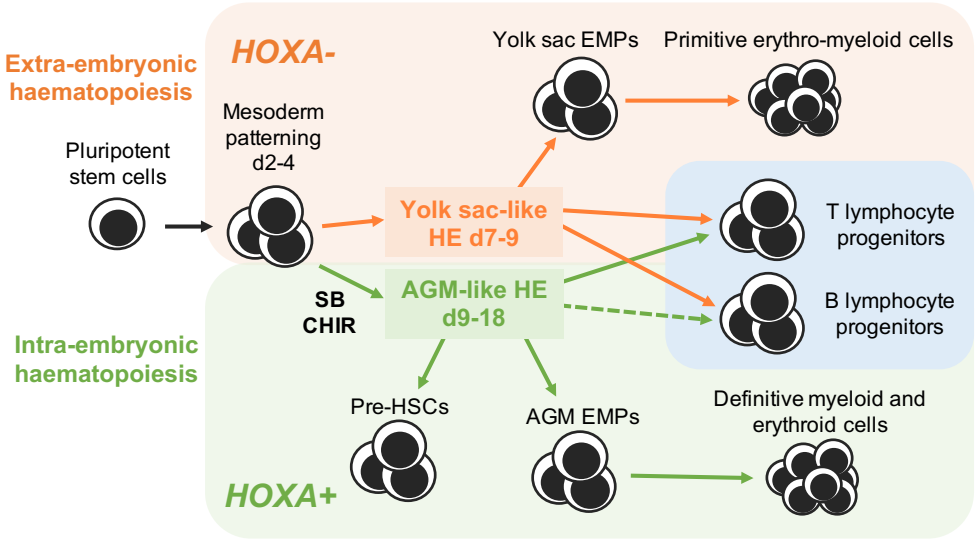


Figure 1.5: *HOXA* expression signature during *in vitro* haematopoietic differentiation

During the time of mesoderm patterning (days 2-4 of differentiation), cells are primed towards activation of the extra-embryonic yolk-sac (orange) or intra-embryonic AGM (green) haematopoietic programs. This switch can be modulated *in vitro* by addition of the activin antagonist SB431542 (SB) and the WNT agonist CHIR99021 (CHIR). Extra- and intra- embryonic progenitors differ in their *HOXA* expression signature, which is upregulated in endothelial and definitive immature haematopoietic cells. Both programmes are able to generate erythro-myeloid progenitors (EMPs) and T lymphocytes from haemogenic endothelium (HE). B cells have been generated via the extra-embryonic programme but B-cell derivation from AGM-like progenitor cells has not been proved yet (dashed arrow). The intra-embryonic programme also generates progenitors with pre-HSC features, but more mature adult-type HSC progeny have not been demonstrated yet.

1.4.2.1 Generation of definitive progenitors from human pluripotent stem cells

Most protocols for haematopoietic differentiation *in vitro* generate progenitors resembling those originating in the embryonic yolk sac during the first two waves of haematopoiesis, comprising mainly primitive erythroid, erythro-myeloid and T-lymphoid progenitors but lacking HSC activity (reviewed in Ackermann et al. 2015). Both pre-HSCs and rare HSCs with long-term haematopoietic repopulation capacity are only generated in the embryonic AGM region (Boisset et al., 2010; Ivanovs et al., 2011; Medvinsky et al., 1993; Rybtsov et al., 2011, 2016; Taoudi et al., 2008; Tavian et al., 1996). Indeed, comparison of transcriptional profiles from human ESC-derived haematopoietic CD34+ cells, generated under a yolk sac-like differentiation protocol, and human CD34+ cord blood repopulating cells, demonstrated a key difference – whilst the cord blood cells expressed members of the *HOXA* cluster of homeobox genes, this gene signature was absent from the ESC-derived CD34+ progenitors (Ng et al., 2016). *HOX* genes are involved primarily in the specification of the anterior to posterior axis during mesoderm patterning (McGinnis and Krumlauf, 1992). However, several *HOXA* genes are expressed in HSCs (Lebert-Ghali et al., 2010; Novershtern et al., 2011) and are important for haematopoietic development in both mouse (Argiropoulos and Humphries, 2007; Di-Poï et al., 2010) and human (Dou et al., 2016; Zhou et al., 2016). For this reason, *Ng et al.* developed a protocol that depends on the generation of *HOXA*-expressing cultures to reproduce human AGM-like haematopoiesis in a dish (Ng et al., 2016). Furthermore, unlike some previous methods, which involved the purification of endothelial progenitors prior to blood formation (Ditadi and Sturgeon, 2016), this approach utilizes whole EBs in an adherent culture, allowing the formation of a supportive stroma, *SOX17*-expressing

vascular structures and subsequent formation of *RUNX1C*-expressing haematopoietic progenitors. Although there is yet no evidence of generation of long-term engrafting cells in these cultures, the RUNX1C+ progenitors harbour high clonogenic activity and ability to home in the bone marrow (Ng et al., 2016).

Upregulated *HOXA* gene expression is mediated through modulation of WNT and ACTIVIN signalling pathways between d2 and d4 of differentiation. In this context it is relevant that (i) commitment to primitive and definitive haematopoiesis occurs during this period (Kennedy et al., 2012; Sturgeon et al., 2014) and (ii) *HOXA* genes are first expressed in the embryonic primitive streak (Deschamps and van Nes, 2005), which is the site for mesoderm induction and patterning. *HOXA* cluster induction *in vitro* is achieved through exposure of cultures to a pulse of a WNT agonist (CHIR99021) and an ACTIVIN receptor-like kinase inhibitor (SB431542) from d2 to d4 of differentiation, the period during which expression of *MIXL1* and *BRACHYURY* primitive streak genes is greatest. This treatment leads to a transient upregulation of *CDX* genes, which in turn regulate *HOX* gene expression (Deschamps and van Nes, 2005). These two small molecules were chosen on the basis of previous evidence showing suppression of primitive haematopoiesis through WNT3A supplementation during mesoderm patterning in human ES cells (Gertow et al., 2013), and production of definitive haematopoietic lineages from human pluripotent stem cells by exposure to ACTIVIN antagonists or WNT agonists (Kennedy et al., 2012; Sturgeon et al., 2014).

The above described differentiation protocol relies on the use of a serum-free medium base that includes recombinant human (rh) protein components – albumin, transferrin and insulin – and is free of purified animal or human products, hence reducing variability due to reagent batch effects and increasing reproducibility of the

differentiation. Other key features of this media formulation are: (i) a reduced use of lipid components, restricted to synthetic cholesterol and the essential linoleic and linolenic fatty acids, and (ii) the use of polyvinyl alcohol (PVA), which enhances initial formation of EBs (Ng et al., 2008). EB cultures are set up using a spin-based method, in which undifferentiated human ES cells are aggregated by centrifugation onto round-bottom 96-well plates, resulting in formation of EBs of uniform size in each well, which guarantees synchronized development. During the first four days, EB plates are cultured in media supplemented with growth factors that induce differentiation to haematopoietic mesoderm (Davis et al., 2008; Ng et al., 2005; Pick et al., 2007). Addition of the Activin-inhibitor SB and WNT-agonist CHIR from day 2 to 4 allows the subsequent upregulation of *HOXA* gene expression and suppression of primitive haematopoiesis (Ng et al., 2016), since during this time frame (i) Activin/Nodal signalling is involved in the formation of the primitive early erythroid progenitors (Kennedy et al., 2012) and (ii) Wnt-signalling is required for the specification of definitive progenitors (Sturgeon et al., 2014). Expression of the arterial vascular *SOX17* marker and stem-like CD34 markers arises from d3 and are maximal at d7. Vascular structures are seen to form after d8 of differentiation. The EHT occurs from d10-14 of differentiation, once the vascular structures have developed. From that time on, RUNX1C+ and RUNX1C- progenitors are generated from SOX17+ haemogenic endothelium, and subsequently detach and proliferate. Progenitor formation continues until about day 28-30, and haematopoietic cells in suspension will progress through more mature lineages depending on the cytokines used. The laboratory of Prof Elefanty has been able to generate cells of the myeloid, erythroid and T-cell lineages using this protocol (Ng et al., 2016).

1.4.3 Blood cell formation in the adult

Blood cell development in the adult originates from long-term self-renewing HSCs, believed to lie at the apex of a differentiation hierarchy (Figure 1.6). HSCs give rise to a family of multipotent progenitors (MPPs), which possess intermediate or short-term self-renewal and transient repopulation potential (Adolfsson et al., 2001; Morrison et al., 1997). The earliest lineage commitment decision occurs downstream of MPPs, in which they lose the self-renewal potential and segregate into myeloid and lymphoid lineages. These two main branches were first inferred from the identification of common myeloid and lymphoid progenitors in the mouse (CMPs and CLPs, respectively) (Akashi et al., 2000; Kondo et al., 1997; Reya et al., 2001). Progenitor intermediates lose their multilineage potential and differentiate into mature cells, by progressively repressing lineage programs until an individual one becomes dominant. For example, MPPs lose their erythroid and megakaryocytic potential to give rise to an intermediate with myelo-lymphoid potential, as demonstrated by the isolation of lymphoid-primed multipotent progenitors (LMPPs) from the mouse (Adolfsson et al., 2005; Kawamoto et al., 1999, 2010). LMPPs gradually become restricted to the lymphoid lineage by losing their myeloid potential and differentiating into early lymphoid progenitors (ELPs), which then become CLPs that finally differentiate to T, B and natural killer lymphocytic cells (Månsson et al., 2007; Welner et al., 2008).

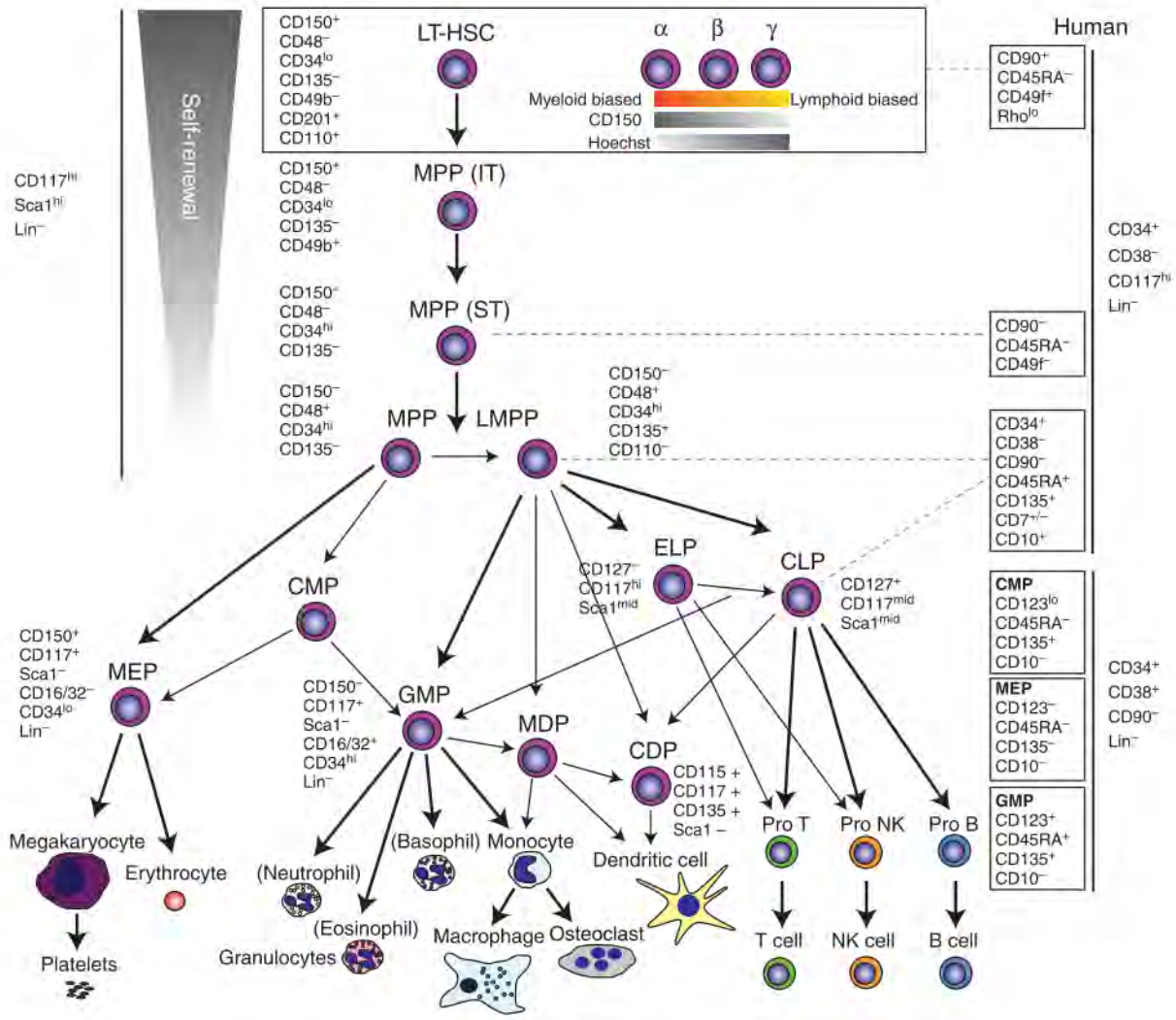


Figure 1.6: Adult haematopoietic differentiation in the mouse and human

Long-term HSCs sit at the apex in this hierarchical model of adult haematopoiesis. They give rise to all mature blood cell lineages through several progenitor cell stages and distinct differentiation routes. The differentiation stages and lineages can be identified by specific surface markers, some of which are distinct in both human and mouse. Mouse surface molecules for the identification of each cell type are indicated and respective human markers are plotted on the right within boxes. HSC: haematopoietic stem cell, MPP: multipotent progenitor, LT: long-term repopulating, IT: intermediate-term repopulating, ST: short-term repopulating, LMPP: lymphoid-primed MPP, ELP: early lymphoid progenitor, CLP: common lymphoid progenitor, CMP: common myeloid precursor, GMP: granulocyte/monocyte progenitor, MEP: megakaryocyte–erythrocyte progenitor, CDP: common dendritic progenitor, MDP: monocyte–dendritic cell progenitor, NK: natural killer cell). Figure reproduced from *Rieger and Schroeder 2012* (Rieger and Schroeder, 2012).

MPPs also give rise to CMPs, which generate granulocyte/monocyte progenitors (GMPs) and to megakaryocyte/erythrocyte progenitors (MEPs). As reported in mouse studies, the majority of MEPs are derived from *Flk2*+ CMPs and not directly from MPPs (Boyer et al., 2011; Buza-Vidas et al., 2011). MEPs are the precursors of erythrocytes and megakaryocytes, which generate red blood cells and platelets. GMPs differentiate into granulocytes (neutrophils, eosinophils and basophils), monocytes (macrophages and osteoclasts) and monocyte/dendritic progenitors that subsequently become common dendritic progenitors and dendritic cells. Mature myeloid cells harbour innate immune roles: neutrophils and macrophages are important for defence against bacteria, eosinophils and basophils provide immunity against parasites and mast cells located in the gut and skin control allergies (Koenderman et al., 2014). Lymphoid differentiation occurs gradually with multiple intermediates, which could be explained by dependence on progressive silencing of myeloid programs. This idea is supported by the observation of widespread methylated myeloid promoters in mouse lymphoid lineages and by the resulting impairment of B and T cell mouse specification upon loss of the DNA methyltransferase Dnmt1 (Bröske et al., 2009; Ji et al., 2010).

In reference to human haematopoiesis, myeloid development seems to be consistent with the classical model stablished in the mouse. However, different cell surface markers are expressed (Figure 1.6). CMPs and GMPs express CD123 and CD135, and the CMP-GMP transition is marked by acquisition of CD45RA (Doulatov et al., 2010; Manz et al., 2002). Lymphoid development differs in mice and human due to the distinct molecular mechanisms of immune responses (Mestas and Hughes, 2004). For this reason, a broader term to identify the earlier lymphoid progenitors,

named multi-lymphoid progenitors (MLPs), was defined. MLPs expressing CD34 and the earlier B cell marker CD10 were first shown to be largely lymphoid restricted, since cells were able to derive B, T and natural killer cells but not myelo-erythroid lineages (Galy et al., 1995). Such cells retained lymphoid potential in both cord blood and bone marrow and are found in circulation through life (Six et al., 2007). Later studies have identified that MLPs also contain myeloid, but not erythro-megakaryocytic, potential. MLPs are restricted to the CD34+CD38-Thyr-CD45RA+ (Thyr-CD45RA+) compartment and are able to generate myeloid, monocytic and B cells upon NSG mice engraftment (Doulatov et al., 2010). These findings have been reproduced using Thyr-CD45RA+ cells from adult bone marrow of AML patients (Goardon et al., 2011). A fraction of cells expressing the above mentioned markers plus the early T cell marker CD7 also harbour multi-lymphoid potential (Hao et al., 2001; Hoebeke et al., 2007), however, they may not sustain lymphopoiesis in the adult since are present in cord blood but decrease after birth (Haddad et al., 2006).

A revised model was proposed in 2013 based on the expression of the primitive surface marker CD133 (Figure 1.7), in which MPPs give rise to CD133+ LMPPs and CD133- EMPs (Görgens et al., 2013a, 2013b). The CD133+ fraction contained all long-term MLPs as well as the GMP progenitors that were restricted to form neutrophils. The model also predicted that the CD133- EMPs were able to form MEPs and eosinophil/basophil progenitors (EoBPs). Subsequently, a human single-cell study suggested that CMPs and MEPs with multilineage potential existed *in utero* but not in adulthood, hence proposing a developmental-stage dependent model of human haematopoiesis in which adult blood development occurs from multipotent cells, such as HSCs and MPPs, that differentiate into committed uni-lineage

progenitors (Notta et al., 2015). A recent study that evaluated the lineage potential of individual human adult bone marrow progenitor cells using single-cell RNA-Seq and functional techniques (Drissen et al., 2019) was in broad agreement with the 2013 revision by *Görgens et al., 2013*. This latest revision proposes an early bifurcation of multipotent MPPS into cells with either (i) neutrophil/monocyte as well as lymphoid potential or (ii) megakaryocyte, erythroid and eosinophil/basophil/mast cell potential. In contrast to the model proposed in 2013, this study observed strict early segregation of neutrophil/monocyte from the basophil/mast cell and eosinophil progenitors (Figure 1.7).

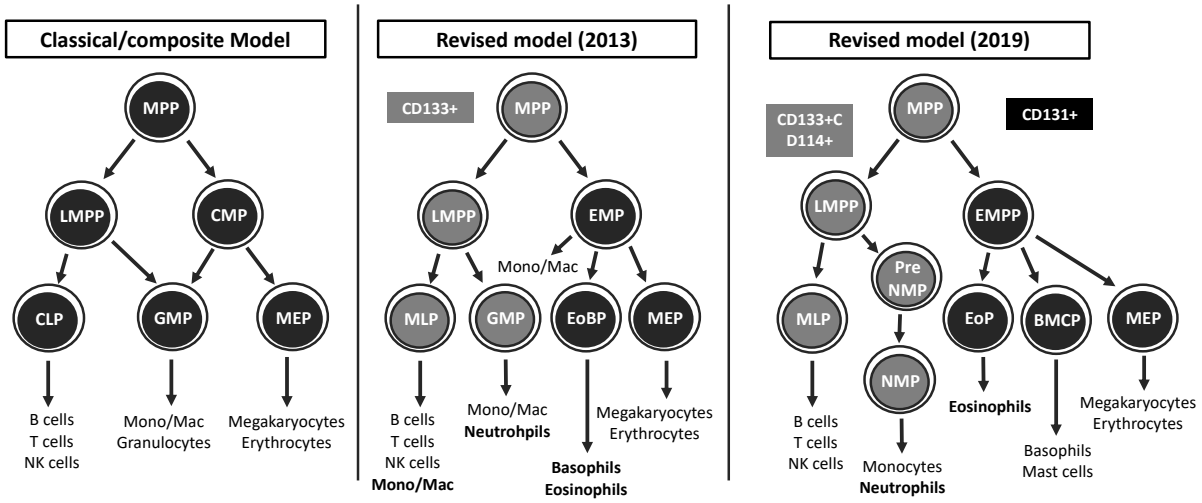


Figure 1.7: Revised models of human haematopoiesis

The composite model (Adolfsson et al. 2005), the 2013 revised model (Görgens, Radtke, Möllmann, et al. 2013) and the 2019 revised model based on single-cell RNAseq and functional data (Drissen et al., 2019). (MPP: multipotent progenitor, LMPP: lymphoid-primed MPP, CLP: common lymphoid progenitor, CMP: common myeloid precursor, GMP: granulocyte/monocyte progenitor, MEP: megakaryocyte–erythrocyte progenitor, MLP: multi-lymphoid progenitor, EMP: erythro-myeloid progenitor, EoBP: eosinophil/basophil progenitor, Mono/Mac: monocytes/macrophages, EMPP: erythroid- megakaryocyte primed multipotent progenitor, BMCP: Basophil-mast cell common progenitor).

1.4.4 Key transcription factors regulating haematopoietic specification

Haematopoietic development is tightly regulated by the complex interplay of cell-stage-specific TFs, which must be expressed in a temporal and dynamic fashion to activate lineage-specific genes at the correct developmental stage and appropriate level (Figure 1.8) (Goode et al., 2016). Understanding the temporal expression and function of each TF has been possible thanks to both *in vitro* and *in vivo* studies.

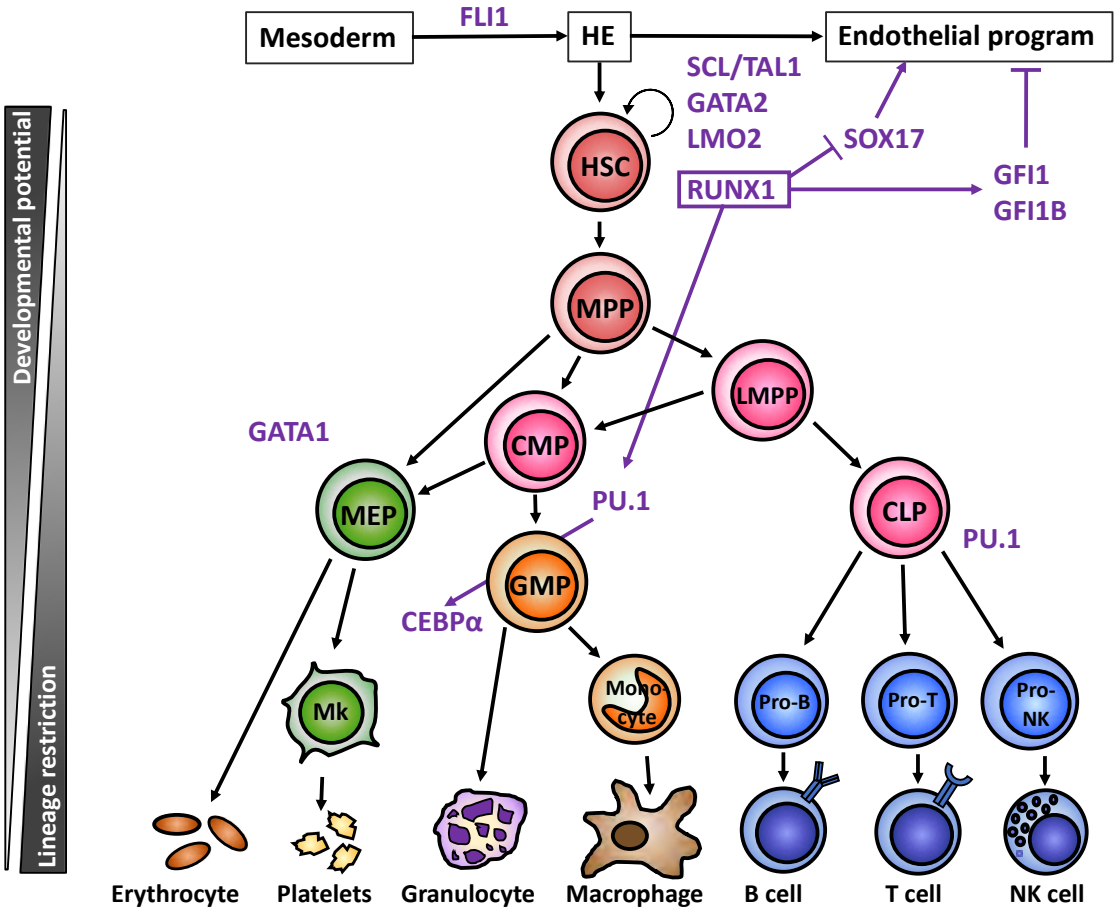


Figure 1.8: Haematopoietic specification is regulated via the interplay of key TFs

Schematic of the embryonic developmental tissues (mesoderm, HE and endothelium) and the haematopoietic progenitors generated after the EHT. Haematopoietic progenitors appear organized from most undifferentiated HSC (at the top of the hierarchy), to lineage restricted cells. Key TFs regulating early haematopoietic specification as well as lineage differentiation are indicated in purple, as well as key interactions. The generation of HSC and haematopoietic progenitors from HE requires not only the activation of the lineage-specific TFs but also the downregulation of the endothelial program, which entails the RUNX1-mediated downregulation of SOX17 and activation of the endothelial repressors GFI1 and GFI1B.

At the top of this hierarchical network, *FLI1* – a member of the E-twenty-six (ETS) TF family – regulates the haematopoietic and endothelial gene expression programs, acting upstream of key haematopoietic genes such as *SCL/TAL1, GATA2* and *LMO2* (Liu et al., 2008; Spyropoulos et al., 2000). Mice lacking *Scl/Tal1* present vascular defects and are unable to generate haematopoietic cells (Robb et al., 1995). Later studies confirmed SCL/TAL1 is required for the *in vitro* formation of endothelial and haematopoietic cells (D'Souza et al., 2005) as well as for HE development (Lancrin et al., 2009). Other genes important for proper haematopoietic development at early stages include LMO2 (Jaffredo et al., 2005; Warren et al., 1994; Yamada et al., 1998) and also SP1 (Gilmour et al., 2014; Krüger et al., 2007).

Expression of RUNX1 is required for the generation of intra-aortic clusters, HSCs and for the EHT, but not thereafter (Cai et al., 2000; Chen et al., 2009; Lancrin et al., 2009; North et al., 2002). FLI1 and SCL/TAL1 together with C/EBPβ prime a large number of haematopoietic transcriptional regulator genes, such as *SPI1* (PU.1), and of genes that control the cellular morphological changes required for EHT, such as *GFI1* and its paralog *GFI1B*. In turn, RUNX1 expression mediates the reorganization of these priming factors, which is a critical process during the EHT(Goode et al., 2016; Lichtinger et al., 2012). RUNX1 establishes a haematopoietic transcriptional network through upregulation of haematopoietic TFs, such as *SPI1* (PU.1), as well as downregulation of the endothelial program. The latter is exemplified by the RUNX1-mediated upregulation of the transcriptional repressors *GFI1* and *GFI1B*, which bind to regulatory regions of cardiovascular and vascular genes and recruit the histone demethylase LSD1, resulting in gene silencing (Thambyrajah et al., 2015). Expression of GFI1 and GFI1B, together with RUNX1, mark the HE stage (Lancrin et

al., 2012) and are required to enable HSC emergence within the AGM (Thambyrajah et al., 2015). Another HSC regulator is SOX17, which contributes to the proliferation and preservation of their undifferentiated state (Clarke et al., 2013; Nobuhisa et al., 2014). However, the generation of haematopoietic progenitors requires the downregulation of *SOX17* by RUNX1 binding, resulting in downregulation of the endothelial program (Lichtinger et al., 2012) (Nakajima-Takagi et al., 2013; Nobuhisa et al., 2014).

RUNX1 also upregulates expression of several TFs regulating myelopoiesis. *SPI1* (*PU.1*) is a direct target of RUNX1 (Hoogenkamp et al., 2007; Huang et al., 2008) and is upregulated only after the onset of *RUNX1* expression (Lancrin et al., 2009; Lichtinger et al., 2012; Swiers et al., 2013). A balanced expression of *PU.1* is required for proper development of myeloid and lymphoid lineages (Leddin et al., 2011; McKercher et al., 1996; Rosenbauer et al., 2006; Scott et al., 1997) as well as HSC maintenance (Staber et al., 2013). RUNX1 regulates *PU.1* expression by coordinating chromatin unfolding at *cis*-regulatory elements and reorganizing the binding of the *PU.1* priming factors SCL/TAL1 and FLI1 (Hoogenkamp et al., 2009; Lichtinger et al., 2012). In turn, *PU.1* expression is then maintained by the C/EBP family of TFs, as they remain bound to regulatory elements forming stable complexes that keep transcription active even in the absence of RUNX1 (Leddin et al., 2011). C/EBPα and C/EBPβ are both indispensable for early embryonic development (Begay et al., 2004) and its expression is also dependent on RUNX1. *CEBPA* is expressed following *RUNX1* and *SPI1* expression (Burda et al., 2009; Lichtinger et al., 2012). However, *CEBPB* appears already expressed in the first stage of HE, before *RUNX1* expression (Goode et al., 2016; Hoogenkamp et al., 2009; Lichtinger

et al., 2012). Another PU.1 target is *CSF1R*, which is absolutely required for macrophage differentiation and whose expression also requires binding of RUNX1 and other PU.1-induced factors (Hoogenkamp et al., 2009).

Definitive haematopoiesis requires the expression of *ERG* (Loughran et al., 2008), which, together with FLI1, regulates HSC homeostasis and normal megakaryopoiesis (Kruse et al., 2009). The generation of HSCs and their survival after the EHT also requires expression of GATA2 (De Pater et al., 2013). Moreover, GATA2 blocks erythroid differentiation in haematopoietic progenitors and promotes megakaryocytic differentiation (Ikonomi et al., 2000). Conversely, GATA1 is involved in the maturation of red blood cell precursors (Fujiwara et al., 1996; Pevny et al., 1991; Simon et al., 1992; Weiss et al., 1994). *NFE2* expression regulates erythropoiesis (Shivdasani, 1996) and is bound and primed by SCL/TAL1, FLI1 and C/EBPβ and upregulated upon RUNX1 binding (Lichtinger et al., 2012; Wang et al., 2010). Definitive erythroid differentiation demands expression of *KLF1* (Pilon et al., 2011; Tallack et al., 2010), which acts synergistically with KLF2 (Alhashem et al., 2011; Pang et al., 2012; Vinjamur et al., 2014)

Later in embryo development, the *C-MYB* proto-oncogene plays an important role in controlling haematopoiesis in the foetal liver (Mucenski et al., 1991). Other factors important for haematopoietic development are MEIS1, which is needed for definitive haematopoiesis as well as for megakaryocytic differentiation (Azcoitia et al., 2005), and C/EBPε, which is essential for terminal granulocytic differentiation (reviewed in Lekstrom-Himes 2001, Yamanaka et al. 1997).

1.5 Runt-related transcription factor 1 (RUNX1)

1.5.1 RUNX1 structure

Runt-related transcription factor 1 (RUNX1) – also known as acute myelogenous leukemia-1 (AML-1), core binding factor alpha 2 (CBFA-2) or polyoma enhancer binding protein 2 alpha B (PEBP-2αB) – is absolutely required for definitive haematopoiesis and one of the most frequently mutated genes in human leukaemia. It belongs to the heterodimeric family of core-binding factors (CBFs), which includes the CBFα subunits RUNX1, RUNX2 and RUNX3 and the CBFβ subunit encoded by *Cbfb*. *Runx1* was the first CBF gene to be isolated, and it was cloned from the t(8;21) chromosomal translocation, which is the most frequent recurrent translocation in acute myeloid leukaemia (Miyoshi et al., 1991).

The RUNX1 protein (Figure 1.9) is characterized by (i) a RUNT homology DNA-binding domain of 128 amino acids in the N-terminal that specifically recognizes the PyGPyGGT consensus sequence, (ii) a nuclear matrix targeting signal or localization site (NLS), (iii) a transactivation domain at the carboxyl-terminus (C-ter) and (iv) a 5-amino acid VWRPY motif at the 3' end of the protein that mediates interactions with the transcriptional co-repressor Groucho/TLE (Aronson et al., 1997; Walrad et al., 2010; Zeng et al., 1998). The RUNT domain is highly homologous to the RUNT protein of *Drosophila melanogaster*, which is involved in segmentation, neurogenesis and sexual determination during embryonic development (Daga et al., 1992; Duffy and Gergen, 1991; Kania et al., 1990). C-terminal of the RUNT domain, RUNX1 contains transactivation domains (Kurokawa et al., 1996; Meyers et al., 1995), two inhibitory domains (Imai et al., 1998; Kanno et al., 1998; Levanon et al., 1998) and a nuclear matrix targeting signal (Zeng et al., 1998). These C-terminal RUNX1

sequences are crucial for haematopoiesis since they are required for RUNX1 sub-cellular localization, biochemical functions and activation of haematopoietic target genes (Dowdy et al., 2010).

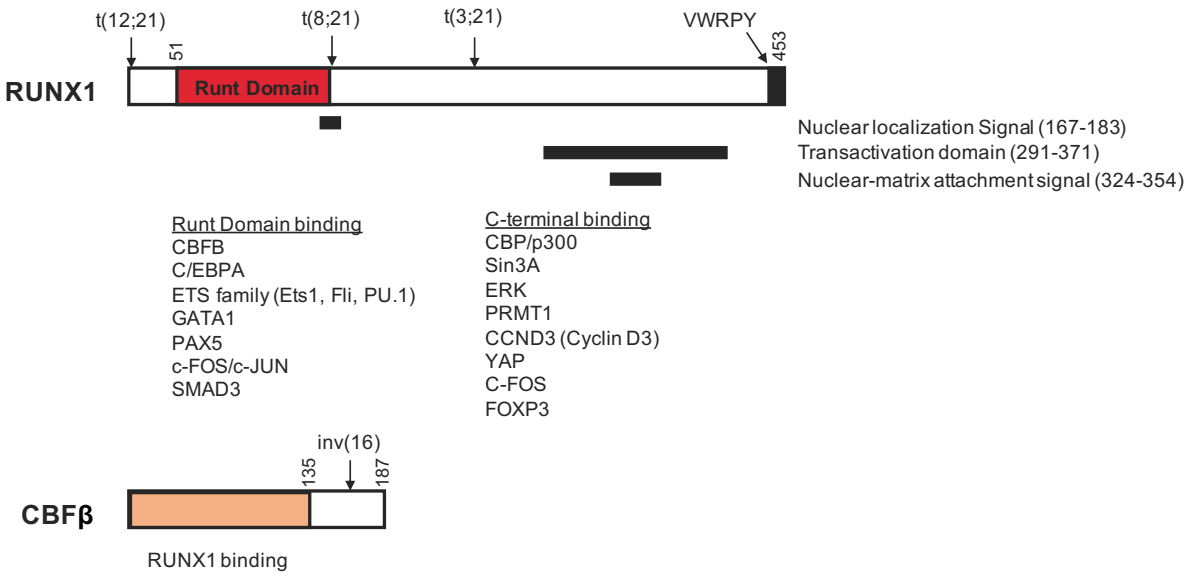


Figure 1.9: Structural domains and interaction partners of the CBF-family proteins RUNX1 and CBFβ

The RUNX1b isoform is shown with all its major domains. It contains an N-terminal Runt domain (red) that is shared between all CBFα subunits and mediates DNA-binding activity, heterodimerization with the core-binding factor-β (CBFβ) and interaction with other proteins. RUNX1 also contains sequences that are carboxyl-terminal to the Runt domain and are responsible for defined biochemical functions, including: a nuclear localization signal, a transactivation domain, attachment to the nuclear-matrix signal, and sequences important for interactions with other DNA-binding and non-DNA-binding proteins, such as CBP/p300 and Sin3A. RUNX1 also contains a 5-amino acid VWRPY motif, located in the C-ter, required for interactions with transcriptional co-regulators. The CBFβ subunit contains a heterodimerization domain for RUNX1 (orange) in the N-terminal. Common translocation breakpoints for both proteins are indicated with an arrow. Amino acids are indicated with numbers. CBF: core-binding factor, C/EBP: CCAAT/enhancer binding protein, YAP: YES- associated protein. Structural domains and protein interactions adapted from *Speck and Gilliland 2002* and *Lam and Zhang 2012* (Lam and Zhang, 2012; Speck and Gilliland, 2002).

1.5.2 The RUNX1 and CBFβ complex

RUNX1 function critically depends on the CBFβ subunit. Although CBFβ does not directly bind DNA by itself, it stimulates RUNX1 DNA-binding activity and protects it from proteolysis by forming a heterodimeric complex through the Runt domain

(Ogawa et al., 1993). The RUNT domain within the full-length RUNX1 isoform contains adjacent sequences that inhibit its DNA binding. Heterodimerization with CBFβ blocks this sequence-mediated auto-inhibition and allows correct binding of RUNX1 to DNA (Kanno et al., 1998). Importantly, both RUNX1 and CBFβ subunits are required for definitive haematopoiesis (North et al., 1999; Wang et al., 1996).

1.5.3 RUNX1 promoters and isoforms

Three main *Runx1* isoforms are transcribed in mice (Fujita et al., 2001) and human (Miyoshi et al., 1995). The *RUNX1* locus is under the control of distal (P1) and proximal (P2) alternative promoters, which generate one longer (RUNX1C) and two shorter (RUNX1A and RUNX1B) protein isoforms, respectively (Figure 1.10). RUNX1A, consisting of 250 amino acids, and RUNX1B, consisting of 453 amino acids, are transcribed from the same promoter (proximal, P2) but undergo different splicing events (Miyoshi et al., 1995). RUNX1A contains the same N-terminal region as RUNX1B, but lacks the regulatory sequences located in the C-terminal domain of the RUNX1B and RUNX1C isoforms (Tanaka et al., 1995). Conversely, RUNX1C is transcribed from the distal promoter (P1) and hence is the longest isoform, consisting of 480 amino acids. RUNX1B and RUNX1C isoforms have the same C-terminal sequences but differ in their 5' untranslated region (UTR) and N-terminal sequences (Miyoshi et al., 1995).

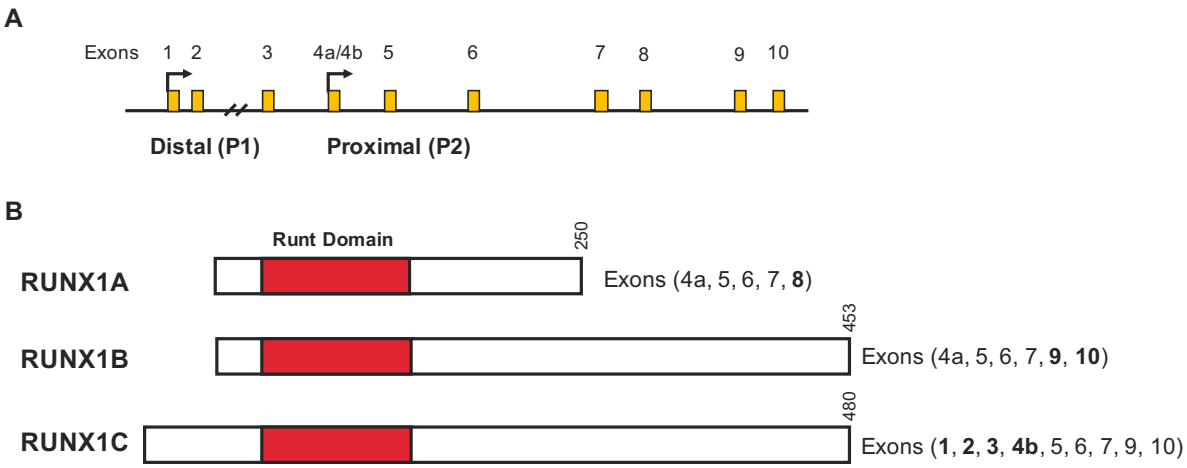


Figure 1.10: Structure of the RUNX1 genomic locus and protein isoforms

(A) Schematic representation of the RUNX1 genomic locus located on chromosome 21. The proximal and distal promoters are indicated with an arrow and exons are represented with yellow boxes. (B) The three main RUNX1 transcriptional isoforms, including RUNX1A, RUNX1B and RUNX1C. Amino acids and exons are indicated for each isoform. The shared Runt homology domain is coloured in red. Figure adapted from *Lam and Zhang 2012* (Lam and Zhang, 2012).

Different isoforms play different roles during embryonic haematopoiesis and in regulating HSC identity. Transcription of the RUNX1A isoform is more common in human cord blood CD34+ progenitors and its overexpression promotes engraftment of mouse progenitors upon transplantation (Tsuzuki et al., 2007). However, RUNX1A may act as a dominant-negative regulator of the other two isoforms, as suggested in studies showing leukaemia development upon RUNX1A overexpression in a mouse transplantation model (Liu et al., 2009). This result could be due to RUNX1A lacking the C-terminal regulatory domains required for normal RUNX1 function during haematopoiesis. The RUNX1B and RUNX1C isoforms mark different stages during haematopoietic development. Haematopoietic progenitors are generated from hemogenic endothelium that expresses the proximal isoform, *RUNX1B,* which has been shown to be required for definitive haematopoiesis in the AGM. The expression of the distal isoform, *RUNX1C*, is upregulated in definitive haematopoietic

progenitors, which coincides with the loss of the endothelial phenotype. Loss of the distal *RUNX1C* isoform does not affect haematopoietic commitment in mouse embryos, whereas mutation of the proximal promoter affects *RUNX1B* transcripts, leading to impaired AGM and foetal liver haematopoiesis, and results in neonatal lethality (Bee et al., 2010; Pozner et al., 2007; Sroczynska et al., 2009).

Promoter regulation in the *RUNX1* locus is dependent on a regulatory element located approximately 23 kb upstream of the distal promoter TSS (Ng et al., 2010; Nottingham et al., 2007). This element is enriched in binding sites for several haematopoietic TFs such as ETS members, GATA2 and LMO2 (Nottingham et al., 2007). In addition, this element has been shown to be active in haemogenic endothelium and in mouse HSCs (Bee et al., 2009; Ng et al., 2010).

1.5.4 RUNX1 post-translational modifications

RUNX1 can also be regulated at the protein level by post-translational modification. Phosphorylated RUNX1 has been detected in CD34+ haematopoietic progenitors (Erickson et al., 1996; Tanaka et al., 1996). Depending on the modified amino acid residue, RUNX1 phosphorylation can have opposing effects on RUNX1 function: it enhances RUNX1 transcriptional activity upon Extracellular Signal-Regulated Kinase (ERK)-dependent phosphorylation at serine 249 and 266 (Tanaka et al., 1996), increases its transactivation potential (Zhang et al., 2008a), and also decreases RUNX1 stability by promoting its degradation via the ubiquitin-proteasome pathway (Biggs et al., 2006; Zhang et al., 2008a).

However, mutation of these serines does not affect haematopoiesis in murine studies (Tachibana et al., 2008). Ubiquitin-mediated degradation of RUNX1 is positively and negatively regulated by CDK phosphorylation and heterodimerization with the CBFβ subunit, respectively (Biggs et al., 2006; Huang et al., 2001).

In addition, RUNX1 C-terminal sequences downstream to the RUNT domain can be methylated by the arginine methyltransferase PRMT1. This methylation inhibits RUNX1 interaction with the co-repressor Sin3A, which enhances RUNX1-dependent upregulation of *CD41* and *PU.1* (Zhao et al., 2008).

1.5.5 The role of RUNX1 during development

Runx1-deficient mice die at E12.5 due failure of definitive haematopoiesis and extensive haemorrhages (Okuda et al., 1996; Wang et al., 1996). RUNX1 is indispensable for the formation of intra-aortic haematopoietic clusters at E11.5 from the HE located on the ventral floor of the dorsal aorta and the adjacent vitelline and umbilical arteries (North et al., 1999). There is subsequently no formation of HSCs in the AGM and foetal liver in the absence of RUNX1 (Bertrand et al., 2010; Boisset et al., 2010; Chen et al., 2009; Eilken et al., 2009; Kissa and Herbomel, 2010; Lancrin et al., 2009; North et al., 2002; Yokomizo et al., 2001; Zovein et al., 2008). However, RUNX1 is dispensable for development after the emergence of blood cells from haemogenic endothelium (Chen et al., 2009). During adult haematopoiesis, RUNX1 is constitutively expressed in all definitive haematopoietic lineages excluding mature erythrocytes (North et al., 2004). The RUNX1 knock-out phenotype is less severe during terminal differentiation of megakaryocytic and lymphocytic lineages than

before the formation of HSCs, suggesting a distinct role for RUNX1 in either establishing or maintaining the different haematopoietic gene expression programs (Growney et al., 2005; Ichikawa et al., 2004).

RUNX1 controls the expression of several genes crucial for blood cell development. At the HE stage, RUNX1 initiates the transcription of key lineage-determining genes, which are already primed by SCL/TAL1, FLI1 and C/EBPβ, by reorganising these TF assemblies, altering histone marks and mediating long distance interactions between regulatory elements (Levantini et al., 2011; Lichtinger et al., 2012). RUNX1 has a role as a pioneer factor, since it has been shown to influence the initiation of the *Spi1*-related gene expression programme by unfolding the chromatin at the *Spi1* locus even before the emergence of active histone marks (Hoogenkamp et al., 2009).

RUNX1 function relies on its interaction partners, which allow for tissue-specific regulation of RUNX1 target genes. RUNX1 binding partners include: p300, mSIN3A, ETS1, FLI1, PU.1, C/EBPα and GATA1 among others. Interaction with epigenetic regulators, such as the acetyltransferase P300, reinforces the transcriptional activation of RUNX1 targets and also regulates RUNX1 activity (Kitabayashi et al., 1998). Moreover, RUNX1 also interacts with the co-repressor histone deacetylase mSIN3A to regulate its own phosphorylation, stability and nuclear localization (Imai et al., 2004). In order to further reinforce its activity, RUNX1 also co-operates with other proteins either by direct interaction with DNA-sequence-specific binding factors, such as FLI1, or indirectly, working with epigenetic regulators to enhance chromatin accessibility (reviewed in Lichtinger et al. 2010).

The correct function of RUNX1 can be perturbed in several ways, leading to leukaemogenesis, including: (i) point mutations, (ii) translocations fusing *RUNX1* to other genes resulting in chimeric proteins - including t(8;21), t(3;21) and t(12;21) in childhood B cell acute lymphoblastic leukaemia (ALL), (iii) and/or rearrangements within chromosome 16 involving the CBFβ subunit, that adversely affects the DNA-binding activity of RUNX1. Haploinsufficiency of *Runx1* as a result of mutations impairing the function of the RUNT DNA binding domain or the transactivation domain causes familial thrombocytopenia, an inherited platelet disorder in which patients also have a predisposition to AML (Song et al., 1999). Patients present with defective platelet formation, which correlates with the phenotype observed in the *Runx1* conditional knockout mouse model (Growney et al., 2005; Ichikawa et al., 2004). However, loss-of-function of RUNX1 alone is not sufficient to cause leukaemogenesis, and a 'second hit' mutation is required.

1.6 Acute myeloid leukaemia (AML)

Acute myeloid leukaemia (AML) is a clonal disorder of proliferating neoplastic myeloid cells with impaired differentiation and is characterized by more than 20% of immature precursors (termed 'blast cells') present in the bone marrow (Kumar, 2011). Incidence increases with age, with an average age of 67 years at diagnosis. AML is the most common leukaemia in adults and accounts for 25% of all adult cases. Younger patients commonly achieve complete remission upon conventional anthracycline- and cytarabine-based treatment, but long-term survival rates are only about 30-40% with the outcome being even poorer for older patients.

AML has been historically classified into different subtypes, using two established systems: the French-American-British (FAB) and the World Health Organization (WHO). The FAB classifies AMLs as 8 subtypes based on cell morphology and histochemistry (Bennett *et al.*, 1985), whilst the WHO classification takes into consideration the cell differentiation state as defined by molecular markers and chromosomal translocations (Vardiman et al., 2009). The WHO system defines a minimum of 17 subclasses of AML classified into 4 subgroups: i) AML with recurrent genetic aberration, ii) AML with multilineage dysplasia, iii) therapy-related AML and myelodysplastic syndrome (MDS), and iv) other AML.

1.6.1 Development and associated mutagenesis

The normal activity of TFs and/or regulatory proteins can be perturbed by three main mechanisms: (i) mutagenesis; including mutations in the coding sequence, gene amplification or translocation of coding regions adjacent to a highly active enhancer, (ii) chromosomal translocations that result in the generation of aberrant fusion

proteins and (iii) aberrant signalling linked to TF activity. The altered function of TFs can impair the tightly regulated gene expression during haematopoiesis, producing a block in blood cell differentiation and resulting in leukaemia (reviewed in Bonifer and Bowen 2010).

AML develops following the accumulation of somatic mutations in haematopoietic progenitors, which alter their proliferation, self-renewal and differentiation properties. The transcriptional network of mutated progenitors is redirected towards activation of a self-renewal program, eventually resulting in the emergence of rapid proliferating leukaemic blast cells. For that to occur, the co-operation of several complementary genetic alterations is required. These can be broadly classified into two groups (Figure 1.11): (i) those that confer increased proliferative capacity, usually associated with genes regulating signalling pathways such as *FLT3* or *RAS*, and (ii) those that impair haematopoietic differentiation, associated with genes encoding TF and other regulators controlling the expression of the haematopoietic gene expression program, for example *CEBPA*, *RUNX1* or mixed lineage leukaemia (*MLL*) (Dash and Gilliland, 2001; Gilliland and Tallman, 2002; Kelly and Gilliland, 2002). Acquisition of mutations encompassing both these groups results in the clonal expansion of the original single mutated, pre-leukaemic clone and the subsequent appearance of sub-clonal populations harbouring both the initial mutation along with secondarily acquired alterations. Exceptions to this scenario are translocations involving the *MLL* gene, which encodes a DNA-binding protein that contains a SET domain with H3K4 methyltransferase activity (Milne et al., 2002). *MLL* translocations result in the fusion of the N-terminus DNA-binding domain of *MLL* to the C-terminus of diverse genes, the most frequent ones being: *AF4*, *AF9*, *AF10*, *ENL* and *ELL* (Meyer et al., 2006).

What makes this translocation unique is their ability to directly transform, not only self-renewing HSCs, but also myeloid progenitor cells into leukaemic stem cells (Cozzio et al., 2003), independently of collaborative mutational events.

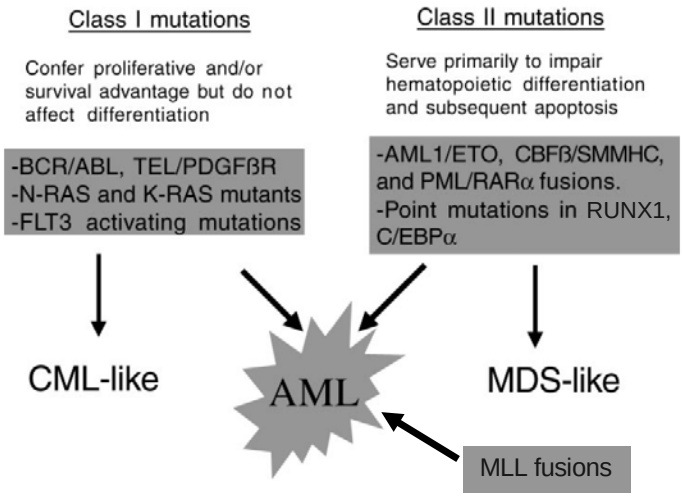


Figure 1.11: Two-hit model of leukaemogenesis

The two-hit model hypothesizes that AML develops in a multistep manner due to the co-operation of two groups of mutations classified depending on its effect. Class I mutations confer a survival or proliferative advantage, whilst class II mutations impair the function of transcription regulators important for blood development. When class I and class II mutations occur alone, they may develop a chronic myelogenous leukaemia (CML)-like or myelodysplastic syndrome (MDS)-like phenotype, respectively. AML develops when both types of mutations are acquired, with exception of *MLL* translocations, with can directly induce leukaemic transformation in both HSCs and myeloid progenitors. Figure adapted from *Kelly and Gilliland 2002* (Kelly and Gilliland, 2002).

AML may also originate secondary to an antecedent MDS or myeloproliferative disorder, *de novo* or consequent to prior radio- or chemo-therapy (Lindsley et al., 2015). Interestingly, dominant clones at the MDS stage may not accurately predict the pattern of future disease, because it may be smaller subclonal stem cell populations that become dominant during progression to AML (Chen et al., 2019). This concept of nonlinear clonal evolution emerging from the premalignant stem cell compartment highlights the importance and difficulty in trying to predict targets relevant to later disease progression in the stem cell populations of MDS patients when assessing pharmacological interventions to treat evolving disease.

AML exhibits a broad array of cytogenetic profiles, ranging from an ostensibly normal karyotype, to balanced translocations or inversions, other chromosome aberrations (deletions, monosomies and trisomies) and complex karyotypes (Byrd et al., 2002; Gaidzik and Döhner, 2008). Moreover, it presents a highly heterogeneous array of acquired secondary molecular alterations within cytogenetically-defined groups, which may also provide relevant diagnostic and prognostic information (Lindsley et al., 2015; Marcucci et al., 2011; The Cancer Genome Atlas Research Network, 2013). Specifically, the genes most frequently harbouring mutations in AML (Figure 1.12) include: *Nucleophosmin 1* (*NPM1*) (Cheng et al., 2010; Falini et al., 2006, 2007), *DNA methyltransferase 3A* (*DNMT3A*) (Marcucci et al., 2012; Shlush et al., 2014), *Fms-like Tyrosine Kinase 3* (*FLT3*) (Kayser et al., 2009; Kelly et al., 2002), *Isocitrate Dehydrogenase* (*IDH*) (Marcucci et al., 2010, 2011; Patel et al., 2012), *Ten–Eleven Translocation 2* (*TET2*) (Chou et al., 2011b), *RUNX1* (Marcucci et al., 2011; Meyers et al., 1993; Tang et al., 2009), *CEBPA* (Döhner et al., 2015; Mrózek et al., 2007), *Additional Sex Comb-Like 1* (*ASXL1*) (Metzeler et al., 2011), *Mixed Lineage Leukaemia* (*MLL*) (Caligiuri et al., 1994, 1998; Meyers et al., 1993), *Tumour Protein p53* (*TP53*) (Haferlach et al., 2008), c-KIT (Paschka et al., 2006), and mutations in both splicing-factor genes (Cazzola et al., 2013; Lindsley et al., 2015) and members of the DNA cohesion complex (Thota et al., 2014). Recent studies suggest a role for methylation-associated C/D box small nucleolar RNAs and ribonucleoproteins in increasing leukaemic self-renewal potential (Zhou et al., 2017). The function of the master haematopoietic regulator RUNX1 is commonly altered by intragenic mutations as well as by chromosomal rearrangements.

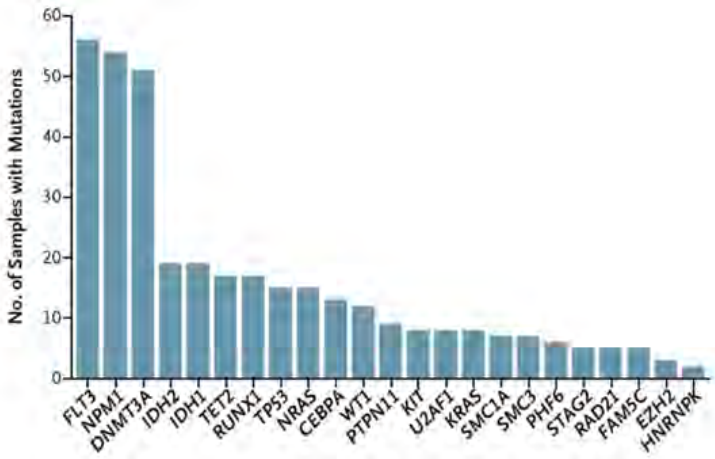


Figure 1.12: Frequently mutated genes in AML

Significantly mutated genes and number of samples carrying each mutation. Data and figure obtained from (The Cancer Genome Atlas Research Network, 2013).

1.6.2 AML with CBF chromosomal rearrangements

Besides mutations of *RUNX1*, a second group termed CBF AML is characterized by the altered function of CBF proteins due to dysfunctional chimeric proteins, as a result of chromosomal rearrangements. These include the t(8;21) and t(3;21) translocations, involving *RUNX1*, and rearrangements within chromosome 16 (inv(16) / t(16;16)), involving the RUNX1 heterodimeric partner *CBFB*. The t(8;21) or inv(16) rearrangements account for approximately 12% of younger adults with *de novo* AML (Grimwade et al., 2010). Conversely, t(3;21) is more commonly found in therapy-related AML, being present in only 1% of newly AML-diagnosed younger adults (Grimwade et al., 2010; Lugthart et al., 2010; Rubin et al., 1990). Unlike t(8;21) and inv(16), which carry a relative favourable prognosis, the t(3;21) confers an adverse prognosis with the poorest survival among the other cytogenetic entities (Grimwade et al., 2010). These rearrangements are able to promote leukaemogenesis, but secondary co-operating mutations are required to develop overt leukaemia.

1.7 The t(8;21) fusion protein: RUNX1-ETO

1.7.1 Pathogenesis of t(8;21) leukaemia

The t(8;21) translocation represents one of the most frequent cytogenetic aberrations in AML, occurring in 7% of adult *de novo* AML and in nearly 40% of M2-subtype AML (AML with maturation) (Grimwade et al., 2010; Peterson et al., 2007; Rowley, 1984). Although t(8;21) AML patients have a relatively favourable prognosis, this is not the case for older patients who are unable to tolerate intensive chemotherapy. Moreover, given the high heterogeneity within this type of leukaemia, up to 40% of t(8;21) patients will relapse after initial remission. There is therefore an unmet medical need to successfully treat all t(8;21) patients.

This AML rearrangement initially promotes the expansion of myeloid progenitor cells with upregulation of self-renewal and proliferation transcriptional programs, leading to expansion of both HSC and GMP populations harbouring leukaemic initiating potential, as demonstrated in conditional mouse models (Cabezas-Wallscheid et al., 2013; Higuchi et al., 2002) and in human HSCs (Mulloy et al., 2002). However, several RUNX1-ETO-expressing mouse models have not been successful in fully reproducing the human t(8;21) disease and its associated molecular mechanisms. Heterozygous RUNX1-ETO knock-in mice (Okuda et al., 1998; Yergeau et al., 1997) showed embryonic lethality due to failure in establishing definitive haematopoiesis. Neither non-lethal transgenic mouse models (Higuchi et al., 2002; Rhoades et al., 2000; Yuan et al., 2001) nor RUNX1-ETO-transduced bone marrow transplantation assays (de Guzman et al., 2002; Higuchi et al., 2002; Mulloy et al., 2003) developed AML. These findings indicate that expression of the t(8;21) fusion protein alone shifts cells to a pre-leukaemic state, but that additional secondary mutations are required

for complete transformation to leukaemia (Kuchenbauer et al., 2005; Rhoades et al., 2000; Yuan et al., 2001).

Fusion transcripts of the t(8;21) translocation have been detected in neonatal blood samples though Guthrie spot analyses, which allows the identification of mutations that had occurred *in utero* (Wiemels et al., 2010). This finding indicates that the chromosomal rearrangement event can occur at an early stage of embryonic development. However, expression of the fusion protein causes a specific phenotype depending on the developmental stage (Regha et al., 2015). Germline expression of RUNX1-ETO results in embryonic lethality with haemorrhage and a severe block of the aortic EHT and foetal liver haematopoiesis (Okuda et al., 1998; Yergeau et al., 1997) – reminiscent of the phenotype observed in *Runx1* knock out mice (Okuda et al., 1996; Okuda et al., 1998; Wang et al., 1996; Yergeau et al., 1997). Conversely, conditional expression after birth gives rise to rapid growing cells with self-renewal properties that become malignant after acquisition of additional mutations (Higuchi et al., 2002; Rhoades et al., 2000).

1.7.2 Structure of RUNX1-ETO

The t(8;21) translocation generates the RUNX1-ETO chimeric protein by fusing the *RUNX1* gene to the Eight-twenty-One (*ETO*, *RUNX1T1* or *MTG8*) gene, located on chromosomes 21 and 8, respectively, with breaks at 8q22 and 21q22.3 (Erickson et al., 1992; Gao et al., 1991; Miyoshi et al., 1991, 1993; Nisson et al., 1992; Rowley, 1973). The translocation breakpoints are located in intron 5 within chromosome 21 (Zhang et al., 2002b) and in introns 1a and 1b within chromosome 8 (Tighe and

Calabi, 1995; Tighe et al., 1993; Xiao et al., 2001; Zhang et al., 2002b). Studies on human CD34+ cord blood cells shed light onto the mechanisms by which this translocation is generated, showing that activated WNT/β-catenin signalling enhances transcription and spatial proximity of both *RUNX1* and *ETO* genes and that *in vitro* long-term treatment with WNT3A is able to induce the translocation event (Ugarte et al., 2015).

The RUNX1-ETO protein consists of the N-terminal 177 amino acids of RUNX1 (exons 1-5), including the Runt homology domain, fused in frame to almost the entire nuclear co-repressor ETO protein (exons 2-11). The ETO protein has a modular structure comprising four Nervy homology regions (NHR), which are evolutionary conserved domains named because of their sequence similarity to the *Nervy* mRNA transcript that localises to the neural precursors cells upon ubiquitous expression of the ultrabithorax (*Ubx*) gene in *Drosophila* (Feinstein et al., 1995). Each NHR mediates a distinct function. The NHR1 stabilizes the formation of complexes of high molecular weight, including interaction with N-CoR (Wei et al., 2007) and the E proteins E2A and HEB (Zhang et al., 2004a). The NHR2 constitutes a core-repressor domain together with its flanking sequences, as it recruits mSIN3A and induces transcriptional repression. Moreover, the NHR2 domain aids the formation of RUNX1-ETO homotetramers through hydrophobic and ionic interactions, due to its amphipathic helix structure. Lastly, the NHR3 and NHR4 co-operate to bind N-CoR and mediate interactions with other proteins via the two zinc finger motifs located in NHR4 (Davis et al., 2003; Hildebrand et al., 2001). The tetramerization and repressor functions from the NH2 domain and the NHR4 zinc-finger structures, respectively, have been shown to be critical for RUNX1-ETO transforming capacity (Chen-

Wichmann et al., 2019; Liu et al., 2006; Minucci et al., 2000). Nuclear localization of RUNX1-ETO is mediated through a critical nuclear localization signal (NLS) located between the NHR1 and NHR2 domains (Barseguian et al., 2002; Odaka et al., 2000). The structure of RUNX1-ETO and the different domains with their interacting proteins is represented in Figure 1.13 (reviewed in Lin et al., 2017; Peterson and Zhang, 2004).

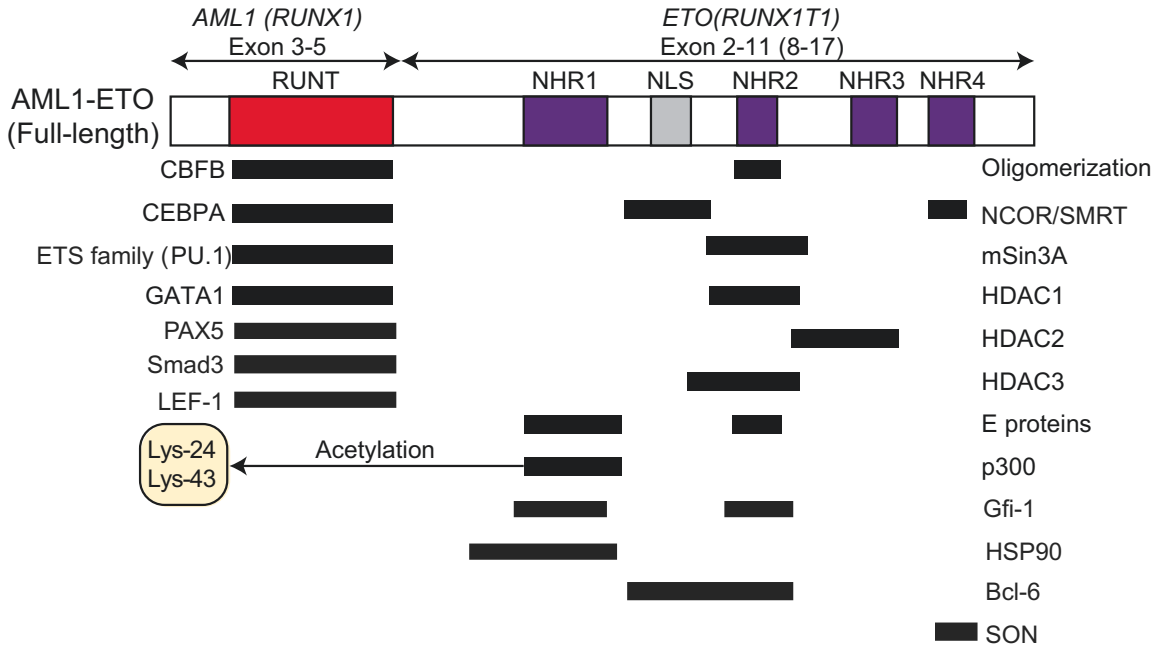


Figure 1.13: Structure and interacting proteins of RUNX1-ETO fusion protein

Diagram of the full-length RUNX1-ETO illustrating its structural domains alongside interactions with other proteins. RUNX1-ETO comprises the N-terminal portion of RUNX1 (exons 3-5), including the DNA-binding RUNT homology domain (RHD), and almost the entire ETO protein (exons 2-11), including four Nervy homology regions (NHR1-4). The oligomerization domain is located on NHR2. The nuclear localization signal (NLS) is located between the NHR1 and NHR2 domains. Figure adapted from Peterson and Zhang 2004 (Lin et al., 2017a; Peterson and Zhang, 2004).

1.7.2.1 Isoforms

Two distinct C-terminal truncated RUNX1-ETO isoforms harbouring reduced transcriptional repressor activity have been identified in t(8;21) patients (Figure 1.14).

RUNX1-ETOtr (or *RUNX1-ETO11a*) lacks the NHR4 domain as a result of aberrant splicing due to a frameshift mutation. Expression of this alternate spliced variant with wild-type ETO recovers its repressor activity and stimulates further oligomerization of fusion proteins, which might contribute to malignant transformation (Kozu et al., 2005).

RUNX1-ETO9a lacks the NHR3-4 domains due to alternative splicing at exon 9 (Yan et al., 2006) and leads to AML development in mouse transplantation models (Yan et al., 2004, 2006). However, its presence or absence does not make a difference with respect to clinical outcome (Agrawal et al., 2019) and its ability to cause AML in the mouse was dependent on supra-physiological expression levels (Link et al., 2016), questioning the clinical relevance of this observation.

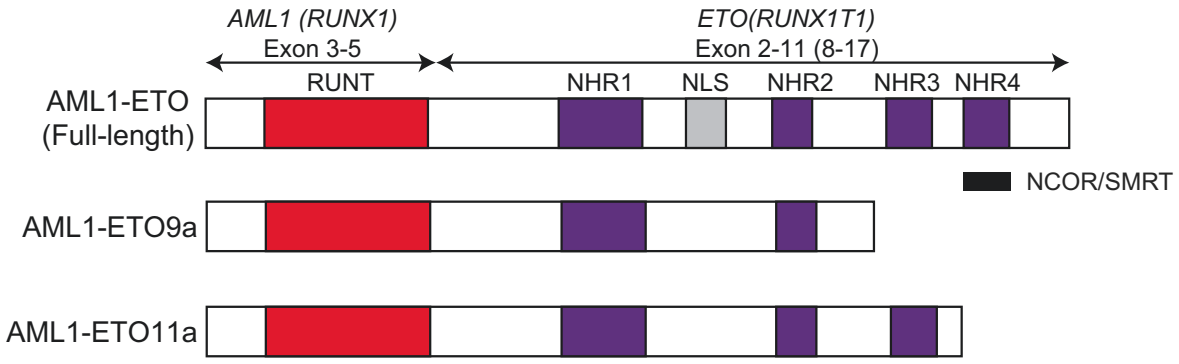


Figure 1.14: Structure of full-length RUNX1-ETO and two shorter isoforms

The truncated RUNX1-ETO isoforms RUNX1-ETO9a and RUNX1-ETO11a lack the NHR3-4 and NHR4, respectively. As a result, both proteins lack one of the interacting domains to NCOR/SMRT.

1.7.2.2 Post-translational modifications

RUNX1-ETO function is modulated by post-translational modifications mediated by epigenetic modifiers that are bound by RUNX1-ETO. Site-specific lysine acetylation at Lys-24 and Lys-43 by p300 is required for RUNX1-ETO-dependent self-renewal in CD34+ cord blood cells and leukaemic transformation in mouse models (Wang et al., 2011a, reviewed in Goyama et al., 2015). RUNX1-ETO is also methylated at Arg-142 by bound PRMT1 (Shia et al., 2012, reviewed in Goyama et al., 2015). RUNX1-ETO stability is regulated through ubiquitination upon interaction with the UbcH8 ubiquitin E2 enzyme and the SIAH1 E3 ligase (Krämer et al., 2008). Although phosphorylation of ETO and RUNX1-ETO on serine and threonine residues has been described in Kasumi-I cells (Erickson et al., 1996), its associated functions during leukaemogenesis remain unclear. Nonetheless, the identification of phosphorylated RUNX1-ETO suggests that its function might be regulated by various kinases and proteins from signalling pathways.

1.7.3 The RUNX1-ETO co-factor complex and interacting proteins

RUNX1-ETO controls leukaemogenesis by forming a stable transcriptional co-factor complex that contains multiple DNA-binding domain proteins and co-localizes genome wide with other TFs to dysregulate gene expression (Sun et al., 2013). Proteins within the RUNX1-ETO co-factor complex include the RUNX1-binding partner CBFβ, the LIM-domain protein LMO2 and its partner LDB1, which operate as bridging factors, the E proteins HEB and E2A, and the E-box-binding TF LYL1 (Figure 1.15).

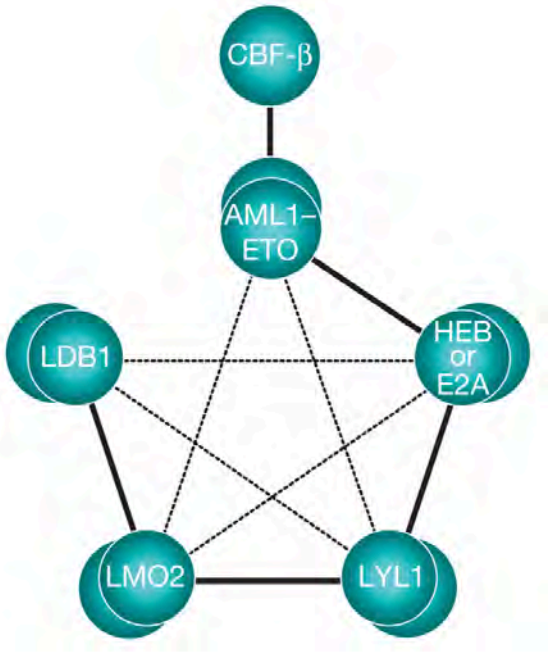


Figure 1.15: The RUNX1-ETO-containing transcription factor complex

Schematic of the components and the interaction network assembling and stabilizing the RUNX1-ETO-containing TF complex. Strong and weak interactions are depicted with thick and thin lines, respectively. Double spheres represent potential homodimerization of those components. Figure reproduced from Sun et al., 2013.

RUNX1-ETO interaction with CBFβ increases its binding to DNA and it is important for leukaemogenesis (Roudaia et al., 2009). However, the critical event for RUNX1-ETO's transforming activity is its oligomerization, mediated by the NHR2 domain (Kwok et al., 2009).

RUNX1-ETO directly represses target genes by interacting with the corepressor proteins N-CoR, SMRT and the HDAC(1-3) complex (Figure 1.16) (Amann et al., 2001; Davis et al., 2003; Gelmetti et al., 1998; Hildebrand et al., 2001; Lausen et al., 2004; Lutterbach et al., 1998a, 1998b; Meyers et al., 1993; Peterson et al., 2007; Trombly et al., 2015; Wang et al., 1998; Zhang et al., 2001). The main interaction partners and the respective binding domain to RUNX1-ETO are represented in Figure 1.13. In addition, RUNX1-ETO has been also shown to interact with the anti-apoptotic protein SON (Ahn et al., 2008). RUNX1-ETO can also operate as an indirect repressor through interference with the activity of other TFs, as exemplified by interaction with the E-box proteins contained within the RUNX1-ETO co-factor

complex (Figure 1.16) (Gardini et al., 2008; Guo et al., 2009; Park et al., 2009; Sun et al., 2013; Zhang et al., 2004b). RUNX1-ETO binds the E proteins HEB and E2A through the NHR1 domain (Gardini et al., 2008; Guo et al., 2009; Park et al., 2009; Zhang et al., 2004b) and through a dimeric-binding interface established upon NH2-mediated oligomerization (Sun et al., 2013). Importantly, the E-box-binding TFs within the co-factor complex mediate the recruitment of RUNX1-ETO to a selection of target genes, enabling RUNX1-ETO binding to DNA independently of its RUNT domain (Sun et al., 2013). Indirect repressor mechanisms important for the block in differentiation also include the sequestration of several haematopoietic TFs, such as CEBPA, GATA1 and PU.1, through the RUNX1-ETO DNA-binding domain (Burel et al., 2001; Choi et al., 2006; Elagib et al., 2003; Vangala et al., 2003; Westendorf et al., 1998).

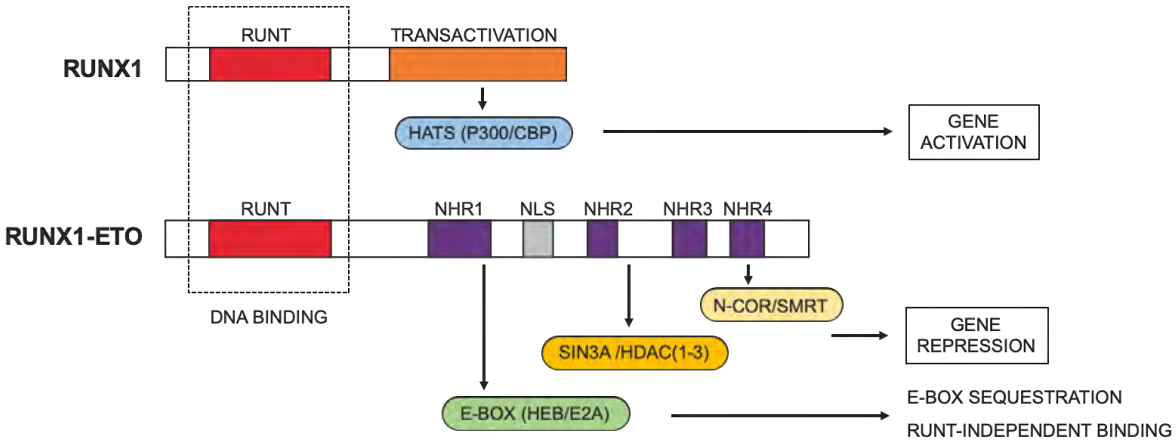


Figure 1.16: RUNX1 and RUNX1-ETO gene regulation by co-factor recruitment

The master haematopoietic regulator RUNX1 binds DNA through its N-terminal domain (Runt) and recruits histone acetyl transferases (HATs), such as P300/CBP, resulting in activation of genes that regulate blood development. The RUNX1-ETO fusion protein also contains the Runt DNA binding domain. However, RUNX1-ETO recruits corepressor proteins N-CoR, SMRT and the Sin3A/HDAC(1-3) complex to RUNX1-binding sites, resulting in direct gene repression leading to a differentiation block of myeloid progenitors. In addition, RUNX1-ETO may exert indirect gene repression through interference with the E proteins HEB and E2A present within the transcription co-factor complex and may bind DNA independently of the Runt domain via E-protein-mediated recruitment to DNA binding sites.

Despite its main role as a repressor, RUNX1-ETO has been reported to activate gene expression through interactions with transcription co-activators, including the lysine acetyltransferase p300/CBP (Wang et al., 2011a) and PRMT1 (Shia et al., 2012). However, using mass-spectrometry the association with p300 directly with RUNX1-ETO could not be confirmed (Sun et al., 2013) and re-ChIP experiments showed a strong preference of RUNX1-ETO for HDACs as compared to p300 (Ptasinska et al., 2014). In contrast to the stable interactions within the RUNX1-ETO transcription factor complex, RUNX1-ETO interactions with both co-activators and co-repressors (NCOR/Sin3A/HDACs) appears to be dynamic and context-dependent. RUNX1-ETO can also lead to gene upregulation through interference with intergenic negative regulatory elements, as it has been previously reported for RUNX1 (Setoguchi et al., 2008).

1.7.4 Molecular pathogenesis of RUNX1-ETO

1.7.4.1 Deregulation of gene expression by RUNX1-ETO

Transcriptional changes mediated by RUNX1-ETO in haematopoietic progenitors have been described in both mouse and human models (Dunne et al., 2006; Lo et al., 2012; Ptasinska et al., 2012; Tonks et al., 2007). Resulting gene expression profiles upon knock-down of both RUNX1 and RUNX1-ETO are inversely correlated, suggesting an opposed regulation of gene expression by both TFs (Ben-Ami et al., 2013). Although there is a 60-80% overlap of RUNX1-ETO and RUNX1 binding sites (Ben-Ami et al., 2013; Gardini et al., 2008; Ptasinska et al., 2012), their genome-wide distributions are dissimilar; RUNX1 binding is enriched at promoter regions, whilst

RUNX1-ETO binding is more abundant at intergenic regions harbouring distal elements (Ben-Ami et al., 2013; Okumura et al., 2008). The preferential binding of both TFs to specific regions could be explained by different affinity for their RUNX1-consensus sequence motif [5'-TGTGGT-3']. RUNX1 prefers longer motifs harbouring two additional thymidines at the 3' site (Okumura et al., 2008). RUNX1-ETO binds strongly to regions harbouring multimerized RUNX1 motifs and is more efficient than RUNX1 in binding to the short consensus motifs (Li et al., 2016; Okumura et al., 2008).

RUNX1-ETO is also able to deregulate expression of genes that are not RUNX1 targets (Shimada et al., 2000). Later studies have shown that binding of the RUNX1-ETO complex to DNA can occur independently of the presence of RUNX1 motifs (Maiques-Diaz et al., 2012; Ptasinska et al., 2012) and stabilization of binding may be facilitated by several interaction partners, such as the ETS-family members ERG and FLI1 (Martens et al., 2012; Ptasinska et al., 2012). ERG and RUNX1-ETO bind together at the same genomic loci, as confirmed by sequential re-ChIP experiments, whilst FLI1 binds prior to the binding of RUNX1-ETO, as shown in inducible RUNX1-ETO cell lines (Martens et al., 2012). This idea is also supported by several ChIP-Seq studies showing enrichment of E-box motifs within RUNX1-ETO binding regions (Gardini et al., 2008; Sun et al., 2013; Zhang et al., 2004b), which highlights the crucial role of E-box TFs within the co-factor complex in regulating RUNX1-ETO-mediated gene regulation.

Other studies have described the effect of RUNX1-ETO by siRNA-mediated knockdown and upon withdrawal of Dox in RUNX1-ETO inducible models. In these studies, reduction of RUNX1-ETO results in a decrease in leukaemic blasts, which

differentiate to mature myeloid cells (Cabezas-Wallscheid et al., 2013; Dunne et al., 2006; Martinez Soria et al., 2009; Ptasinska et al., 2012). It has been previously shown that RUNX1-ETO and RUNX1 form different complexes that dynamically compete for the same genomic sites (Meyers et al., 1995; Mulloy et al., 2002; Ptasinska et al., 2012, 2014). Studies on differentiated progenitors derived from mouse ES cells have shown that a similar set of genes is targeted by RUNX1-ETO before and after the endothelial to haematopoietic transition, but with differing outcomes affecting gene expression. This study showed that, through direct competition with RUNX1 binding, RUNX1-ETO disrupts pre-existing RUNX1 complexes, thus interfering with both activating and repressing RUNX1 activity (Regha et al., 2015).

Despite RUNX1-ETO's dominant negative effect on RUNX1 (Okuda et al., 1998; Yergeau et al., 1997), expression of both proteins must be balanced in order to maintain cell survival and the leukaemic phenotype (Ben-Ami et al., 2013). This study showed that presence of a wild-type RUNX1 allele prevents apoptosis of *RUNX1-ETO*-expressing cells, as RUNX1 knock-down induces apoptosis of t(8;21) Kasumi-1 cells. Requirement of RUNX1 activity for the survival of leukaemic cells has been also demonstrated in a mouse genetic model of MLL fusion AML (Goyama et al., 2013).

RUNX1-ETO-mediated transcriptional repression of its target genes involves the co-operation of several interaction partners. These include NCoR, which co-occupies many of the RUNX1-ETO-bound target genes (Trombly et al., 2015) and the DNA methyltransferase DNMTA1, which also interacts with RUNX1-ETO (Liu et al., 2005). Nevertheless, presence of RUNX1-ETO does not always result in transcriptional

repression. RUNX1-ETO can indirectly mediate gene activation by interfering with the repressive function of RUNX1 (Ptasinska et al., 2012) or by causing the upregulation of Activator Protein 1 (AP-1) family members (Martinez-Soria et al., 2019).

1.7.4.2 Disruption of the chromatin structure by RUNX1-ETO

RUNX1-ETO transcriptional regulation differs depending on the developmental stage in an *in vitro* mouse ES cell differentiation system (Regha et al., 2015), indicating a link between RUNX1-ETO binding and the chromatin status. RUNX1-ETO binding is also linked to accessible chromatin regions presenting the coactivator p300 and intermediate acetylation levels (Maiques-Diaz et al., 2012; Saeed et al., 2012). However, re-ChIP experiments performed in our lab have shown that binding of RUNX1-ETO and p300 in the same genomic loci is mutually exclusive; RUNX1-bound elements have preference for p300, whilst RUNX1-ETO-bound loci preferentially bind HDAC2 (Ptasinska et al., 2014). Intriguingly, removal of RUNX1-ETO results in a global redistribution of TF complexes within pre-existing accessible chromatin sites, a redistribution of RUNX1 binding, and a genome-wide reversal of the altered epigenome, indicating a role for RUNX1-ETO in mediating genome-wide changes of chromatin structure (Ptasinska et al., 2014).

1.7.5 Dysregulated genes and pathways in RUNX1-ETO leukaemia

1.7.5.1 Myeloid transcription factor, tumour suppressor and other target genes

The RUNX1-ETO fusion protein blocks progenitors at an early myeloid stage through repression of genes that promote haematopoietic differentiation (Hatlen et al., 2012).

Expression of haematopoietic genes also appears indirectly dysregulated by RUNX1-ETO through interference with the activity of TFs. CEBPα is a TF that regulates expression genes of the myeloid lineage, and it is required for the formation of mature granulocytes (Zhang et al., 1997). *CEBPA* is a direct RUNX1-ETO target gene, downregulated through (i) direct interaction of RUNX1-ETO with CEBPα itself, resulting in sequestration of CEBPα from its promoter and hence suppressing its autoactivation (Pabst et al., 2001) and (ii) direct binding of RUNX1-ETO to an enhancer element located ~40 kb downstream of the *CEBPA* gene (Ptasinska et al., 2014), which is crucial for specific *CEBPA* expression in myeloid cells (Avellino et al., 2016). A similar mechanism has been shown for *SPI1* (PU.1), where physical binding of RUNX1-ETO to PU.1 results in downregulation of its transactivation capacity in t(8;21) cell lines (Vangala et al., 2003).

RUNX1-ETO has been shown to interfere with the normal expression of tumour suppressor genes. RUNX1-ETO has been shown to repress *RUNX3* (Cheng et al., 2008), which has a tumour suppressor role in solid tumours. In addition, RUNX1-ETO also represses the expression of *CDKN2A* ($p14^{ARF}$) and *NF1* through a dominant negative effect on RUNX1 activity (Linggi et al., 2002; Yang et al., 2005). Strikingly, the cell cycle inhibitor *CDKN1A* ($p21^{WAF1}$) is upregulated in RUNX1-ETO-expressing cells (Berg et al., 2008; Yan et al., 2004). One possible explanation for this could be a role of $p21^{WAF1}$ pathway in preventing exhaustion of leukaemic stem cells (Viale et al., 2009).

Upregulation of several anti-apoptotic genes upon expression of RUNX1-ETO has also been described. *BCL2* is upregulated in leukaemic cell lines (Klampfer et al., 1996) and it is a direct target of RUNX1-ETO co-operatively with ERG and FLI1

(Martens et al., 2012). In addition, RUNX1-ETO has been shown to indirectly upregulate BCL-XL expression through activation of thrombopoietin/ myeloproliferative virus oncogene (TPO/MPL) signalling (Chou et al., 2012). However, the mechanisms underlying RUNX1-ETO-dependent gene activation remain unclear.

Other targets of RUNX1-ETO include signal transducers, cell-cycle regulators and DNA-associated proteins. The B-lymphocyte marker CD48 is downregulated by RUNX1-ETO in a HDAC-dependent manner and its repression may have a role in evasion of the immune response (Elias et al., 2014). The DNA-ATPase/helicase-encoding gene *RUVBL1* has been described as a RUNX1-ETO-upregulated target in *Drosophila* (Breig et al., 2014). A role for the G1-S-specific cell cycle regulator CCND2 in self-renewal and propagation of RUNX1-ETO leukaemic cells has recently been shown. RUNX1-ETO promotes proliferation and cell-cycle progression through maintenance of *CCND2* expression (i) directly by binding to an upstream element and (ii) indirectly by promoting AP-1 binding to the *CCND2* promoter (Martinez-Soria et al., 2019). However, the mechanisms behind upregulation of AP-1 members upon RUNX1-ETO expression remain unclear. Other downregulated genes through direct binding of RUNX1-ETO include the membrane-protein encoding genes *MS4A3* and *NKG7* as well as the ribonuclease *RNASE2* (Ptasinska et al., 2014).

Two recent studies have shed new light onto the connections within the gene regulatory networks resulting from RUNX1-ETO expression (Assi et al., 2019; Ptasinska et al., 2019). RUNX1-ETO plays a role in promoter-enhancer element interactions, which are mediated by cooperative binding of RUNX1-ETO-regulated TFs, such as AP-1.

1.7.5.2 MicroRNAs

RUNX1-ETO has been shown to cause both up-and downregulation of microRNA (miRNA) genes to promote leukaemogenesis. Downregulated miRNAs include the tumour suppressors miR-9, miR-223 and miR193a. Forced expression of miR-9 in RUNX1-ETO cell lines induced differentiation and reduced leukaemic growth (Emmrich et al., 2014). miR-223 is involved in myelopoiesis and is repressed by RUNX1-ETO through a HDAC- and DNA methylation-dependent mechanism (Fazi et al., 2007). miR-193a is downregulated through RUNX1-ETO-dependent recruitment of chromatin modifiers, including HDACs and DNMTs, to RUNX1-binding sites. Interestingly, miR-193a has a negative effect on the expression of factors within the RUNX1-ETO complex, including RUNX1-ETO itself (Li et al., 2013). Presence of RUNX1-ETO upregulates several miRNAs, such as miR-24 and miR-126, that regulate signalling pathways to promote leukaemogenesis. miR-24 stimulates myeloid proliferation through downregulation of the mitogen activated protein kinase (MAPK) phosphatase-1, a MAPK-signalling inhibitor (Zaidi et al., 2009). miR-126 regulates haematopoietic and leukaemic stem cells through repression of the PI3K/AKT pathway (Lechman et al., 2012, 2016). Strikingly, both overexpression and knockout of miR-126 promote RUNX1-ETO-induced leukaemogenesis (Li et al., 2015). However, lack of miR-126 sensitises leukaemic cells to standard chemotherapy, suggesting a role for this miRNA in contributing to drug resistance (Li et al., 2015).

1.7.5.3 Signalling pathways

RUNX1-ETO affects the normal activity of many signalling pathways through transcriptional dysregulation of the genes encoding their modulators. Enhanced TPO/MPL signalling increases survival and leukaemogenesis in t(8;21) AML and results in the upregulation of the anti-apoptotic protein BCL-XL, which in turn activates the JAK/STAT and PI3K/AKT pathways (Chou et al., 2012; Pulikkan et al., 2012). Moreover, downregulation of the protein tyrosine phosphatase CD45 in RUNX1-ETO-expressing leukaemic cells also results in an activated JAK/STAT pathway (Lo et al., 2012). The PI3K/AKT pathway enhances the survival and proliferation capacity of RUNX1-ETO-expressing cells through an MPL-dependent antiapoptotic mechanism (Pulikkan et al., 2012). Targeting components of the AKT signalling pathway might be a potential therapeutic strategy, since inhibition of Inhibitor of DNA-binding 1 (ID1) – a RUNX1-ETO-upregulated protein – blocks initiation and progression of t(8;21) leukaemia in mice by abrogating the activation of protein kinase B1 (AKT1) and results in apoptosis of AML cells, but not wild-type haematopoietic progenitors (Wang et al., 2015). The levels of activity of the PIK3/AKT pathway might be differently regulated in LSCs or during leukaemic progression, given that miR-126 has been shown to downregulate multiple components of this pathway to enhance quiescence, self-renewal and therapeutic resistance (Lechman et al., 2012, 2016).

RUNX1-ETO expression leads to an activation of the WNT signalling pathway through downregulation of the WNT-antagonist *SFRP1* (Cheng et al., 2011) and through upregulation of γ-Catenin (JUP) (Müller-Tidow et al., 2004; Zheng et al., 2004), the Groucho-related amino-terminal enhancer of split (AES) (Steffen et al.,

2011) and the *COX2* gene (Yeh et al., 2009; Zhang et al., 2013b), which is important for tumour initiation, growth and self-renewal.

Several other pathways are also been reported to be upregulated in RUNX1-ETO expressing cells. RUNX1-ETO expression upregulates the nerve growth factor receptor TRKA (NTRK1), which contributes to the expansion of RUNX1-ETO-expressing human blood cells (Mulloy et al., 2005). Enhanced reactive oxygen species (ROS) promoted maintenance of leukaemic cells in a RUNX1-ETO *Drosophila* model (Sinenko et al., 2010). NF-KB signalling is also upregulated in RUNX1-ETO-expressing cells through a failure of inhibition of this pathway by native RUNX1 (Nakagawa et al., 2011). The UBASH3B/CBL pathway is indirectly activated by RUNX1-ETO through miR-9-dependent upregulation of UBASH3B. This pathway promotes the expansion of RUNX1-ETO cells by activating the TPO/MPL signalling pathway (Goyama et al., 2016).

1.7.6 RUNX1-ETO and DNA damage

There is evidence that the first oncogenic events initiating leukaemia, such as RUNX1-ETO, may promote mutagenesis and hence facilitate the acquisition of co-operating mutations (Araten et al., 2013; Krejci et al., 2008). Ectopic RUNX1-ETO expression results in the downregulation of genes involved in DNA repair pathways, including *BRCA2*, *OGG1*, *POLE* and *ATM*, and increases the level of phosphorylated TP53 and of the DNA-double-strand-break marker γH2AX (Alcalay et al., 2003; Krejci et al., 2008). These findings suggest increased DNA damage and a mutator phenotype conferred by RUNX1-ETO. A more recent study confirmed a RUNX1-

ETO-conferred predisposition to acquire mutations, both spontaneously and after treatment with genotoxic agents (Forster et al., 2016). This study also described RUNX1-ETO-dependent repression of *OGG1*, a DNA glycosylase involved in base excision and repair, hence hindering the repair of the RUNX1-ETO -associated mutagenesis. In line with this, RUNX1-ETO-expressing cells display increased levels of DNA damage that can be explained by a reduced expression of genes involved in the homologous recombination (HR) repair pathway, including *ATM*, *BRCA1*, *BRCA2* and *RAD51* (Esposito et al., 2015). Overall, these studies suggest that the acquisition of cooperating secondary mutations in RUNX1-ETO-expressing cells might be facilitated by the repression of DNA repair pathways mediated by RUNX1-ETO.

1.7.7 Collaborative genetic aberrations in t(8;21) AML

Transcripts of the *RUNX1-ETO* translocation have been detected in non-leukaemic cells of patients in long term remission (Miyamoto et al., 2000; Shima et al., 2014). Several studies have shown that RUNX1-ETO is able to shift cells to a pre-leukaemic state in both humans and mice, but that additional secondary mutations are required in order to cause overt leukaemia (Higuchi et al., 2002; Kuchenbauer et al., 2005; Rhoades et al., 2000; Shima et al., 2014; Yuan et al., 2001). Indeed, in transgenic and conditional knock-in mouse models (Higuchi et al., 2002; Yuan et al., 2001) and in retrovirally transduced human CD34+ cells (Mulloy et al., 2002, 2003), RUNX1-ETO alone was not able to induce AML. Collaborative secondary mutations commonly involve: (i) additional chromosomal aberrations, (ii) the activation of signalling pathways, (iii) mutations in genes encoding epigenetic regulators.

Additional cytogenetic abnormalities are detected in 70% of t(8;21) leukaemias, commonly involving loss of sex chromosomes, 9q deletion and trisomy 8 (Krauth et al., 2014; Kuchenbauer et al., 2006). Loss of sex chromosomes involves both X and Y chromosomes and hence candidate genes whose loss may enhance RUNX1-ETO leukaemogenesis and are therefore likely to be located on the common pseudo-autosomal regions (PARs) within the sex chromosomes. The GM-CSF receptor alpha-subunit (CSF2RA) is located in PAR1 and has been identified as a tumour suppressor in a murine transplantation model of RUNX1-ETO leukaemia, since lack of GM-CSF signalling promotes the development of RUNX1-ETO-induced AML (Matsuura et al., 2012).

Mutations affecting signalling pathways in t(8;21) AML mostly occur in growth factor receptors, including mutations in tyrosine kinase receptors such as *KIT* and *FLT3*, and other RAS pathway-activating mutations including *CBL, NRAS*, *KRAS* and *JAK2* (Kuchenbauer et al., 2006; Wang et al., 2005, 2011b). The co-existence of RUNX1-ETO fusion protein and *KIT* mutations – affecting either the activation loop, the tyrosine kinase or the receptor dimerization domains – is enough to develop AML in murine transduction and transplantation models (Nick et al., 2012; Wang et al., 2011b) and in a human CD34+ *ex vivo* progenitor cell expansion assay (Wichmann et al., 2015).

Epigenetic regulators, such as *ASXL1, ASXL2* and *IDH*2, are also frequently mutated in RUNX1-ETO leukaemia (Krauth et al., 2014; Madan et al., 2018; Micol et al., 2014). Mutated *ASXL1* and *ASXL2* may co-operate with RUNX1-ETO in a similar manner to induce leukaemogenicity, as mutations in these genes are mutually exclusive (Krauth *et al.*, 2014).

Expression of *RUNX1-ETO* together with a mutant *KIT*, *KRAS*, *NRAS* or *FLT3* in mouse progenitors was sufficient to develop leukaemia in mice (Nick et al., 2012; Schessl et al., 2005; Wang et al., 2011b; Zhao et al., 2014). However, a single second hit was insufficient in human cord blood cells (Chou et al., 2011a; Goyama et al., 2016; Wichmann et al., 2015). This indicates that several hits are required to induce overt human RUNX1-ETO leukaemia, as human cells might be more resistant to transformation. The expression of RUNX1-ETO already predisposes to the acquisition of these cooperating mutations in a protein-level-dependent manner, both spontaneously and after treatment with DNA damage-inducing agents (Forster et al., 2016; Wang et al., 2005). Increased mutagenesis upon treatment with genotoxic agents can be explained by propensity for faulty DNA repair, as exemplified by downregulation of *OGG1* (Forster et al., 2016).

1.8 Aims and objectives

Many studies have suggested an HSC as the cell of origin establishing the pre-leukaemic clone. Clinical studies have identified the t(8;21) translocation in non-leukaemic HSCs, indicating that RUNX1-ETO might represent the first oncogenic event. The Bonifer group and others have extensively studied the effect of RUNX1-ETO in many scenarios, including *in vitro* mouse and human ES cell differentiated progenitors, mouse models, t(8;21) cell lines, oncogene-expressing human CD34+ blood cells and patient samples. The majority of these studies agree that expression of RUNX1-ETO blocks myeloid differentiation, enhances self-renewal and is required to maintain the leukemic phenotype. However, different models present dissimilarities on their findings, which can be explained by variable dosages of RUNX1-ETO expression, the cellular context including the progenitor type, adaptations of cell lines to culture conditions and the presence of co-existing mutations. These caveats have hindered the understanding of the earliest epigenetic reprogramming events occurring after RUNX1-ETO induction. In addition, there are several questions, such as the direct RUNX1-ETO mechanisms leading to gene dysregulation, that need to be answered.

Therefore, the overall aims of this project include:

i) The generation of an *in vitro* human model of t(8;21) AML that represents the cell of origin acquiring the translocation event. We plan to generate inducible RUNX1-ETO human ES cell lines and use them in an *in vitro* protocol that allows the generation of definitive haematopoietic progenitors from *HOXA*+ endothelium, resembling progenitor cell populations generated within the intra-embryonic human AGM.

ii) The characterization of the phenotypic effects resulting from RUNX1-ETO induction at different stages of the *in vitro* definitive haematopoiesis, including progenitor cell formation, clonogenic capacity and cell growth.

iii) Comparison of the effects resulting from RUNX1-ETO induction on the distinct cell populations generated in our *in vitro* cultures.

iv) The identification of the immediate effects of RUNX1-ETO induction on reprogramming the epigenome of the *in vitro* generated progenitors, in order to understand how the first genetic alteration alters the regulatory network.

Establishing a *bona fide* inducible system of RUNX1-ETO expression in human definitive progenitors will also allow us to investigate the outcomes of introducing collaborative mutations during the generation of fully transformed cells and will allow future screens for therapeutically active agents against pre-leukaemic cells.

2 MATERIALS AND METHODS

2.1 Generation of RUNX1-ETO, RUNX1-ETO K-RAS(G12D) and RUNX1-ETO KIT(N822K) knock-in vectors

AAVS1-targeting vectors containing a RUNX1-ETO cDNA transgene were generated as follows. The pSIEW-RUNX1-ETO vector, containing the RUNX1-ETO cDNA, was a gift from Olaf Heidenreich (Newcastle, UK) (Bomken et al., 2013). The pTREG-CAGGS-Tet3G plasmid for gene targeting into the AAVS1 locus, used as a backbone to insert the transgene, was a present from Su-Chun Zhang (Wisconsin, US) (Qian et al., 2014). Final vectors comprised: (i) an HA-tagged RUNX1-ETO transgene under the regulation of a tetracycline-inducible promoter (TRE3G-CMV), (ii) a Puromycin N-acetyltransferase resistance cassette with gene expression to be driven by endogenous promoter after insertion and (iii) a cytomegalovirus early enhancer/chicken beta actin (CAG) promoter driving expression of a modified Tet3G reverse tetracycline-controlled transactivator (rtTA), between two homology arms of 804 bp (5') and 837 bp (3') to the AAVS1 genomic locus (Figure 2.1). Amplification of the RUNX1-ETO cDNA insert from the pSIEW-RUNX1-ETO donor plasmid was conducted by Polymerase Chain Reaction (PCR) using Phusion High-fidelity DNA Polymerase (New England Biolabs cat# M0530S). Forward and reverse primers included the restriction endonuclease (RE) sites SalI (5') and MluI (3'), respectively, for subsequent cloning into the multiple cloning site of the pTREG-CAGGS-Tet3G knock-in plasmid (Table 2.1). Forward primers were designed with a human influenza hemagglutinin (HA) tag sequence following the Kozak sequence containing the start codon. Reactions of 50-µl volume were prepared containing 250 ng pSIEW-RUNX1-ETO template DNA 1X Phusion HF Buffer, 200 µM dNTPs, 0.5 µM forward primer,

0.5 µM reverse primer, 3% DMSO and 1-unit Phusion DNA Polymerase. Reactions were transferred to a Thermocycler with the following conditions: initial denaturation at 98ºC for 30 seconds, 5 cycles (denaturation at 98ºC for 10 seconds, annealing at 63ºC for 15 seconds, elongation at 72ºC for 150 seconds), 30 cycles (denaturation at 98ºC for 10 seconds, annealing and elongation at 72ºC for 150 seconds) and final extension at 72ºC for 5 minutes. PCR products were purified using gel electrophoresis with 0.7-1 % agarose Tris-acetate-EDTA (TAE) (40 mM Tris, 1mM EDTA and acetic acid added to pH 8.0) gels containing 0.5 µg/ml ethidium bromide. Agarose gels were run in 1X TAE buffer supplemented with 0.5 µg/ml ethidium bromide. DNA products were visualised using a Bio-Rad Gel Doc™ XR+ System and excised using a UV trans-illuminator (Bio-Rad). Excised products were purified from the gel using the QIAquick® Gel Extraction Kit (Qiagen cat# 28704) as described in the manufacturer's protocol and eluted in 30 µl Elution Buffer. DNA concentration was quantified using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific). The knock-in pTREG-CAGGS-Tet3G plasmid and 500 - 1000 ng insert DNA were digested with 5 units of SalI and MluI RE for 3 hours at 37ºC and digested products were purified using agarose gel electrophoresis. Inserts were ligated into digested knock-in plasmids in molar ratio of 1:3 vectors to insert using a T4 DNA ligase (New England Biolabs cat# M0202S) following the manufacturer's protocol. The ligated reaction was transformed into One Shot™ TOP10 Chemically Competent *E. coli* (Thermo Fisher Scientific cat# C404010) by heat shock at 42 ºC. Bacteria carrying the ligated product were selected on 100 µg/ml Ampicillin agar plates and plasmid DNA was purified from selected Ampicillin-resistant individual colonies on silica-based columns using QIAprep® Spin Miniprep kit (Qiagen cat# 27104) protocol

as described by the manufacturer. Successful insert recombination into vectors was confirmed by RE digestion and lack of point mutations in plasmid inserts was confirmed by Sanger sequencing, conducted in the sequencing service of the Biosciences department at the University of Birmingham (UK).

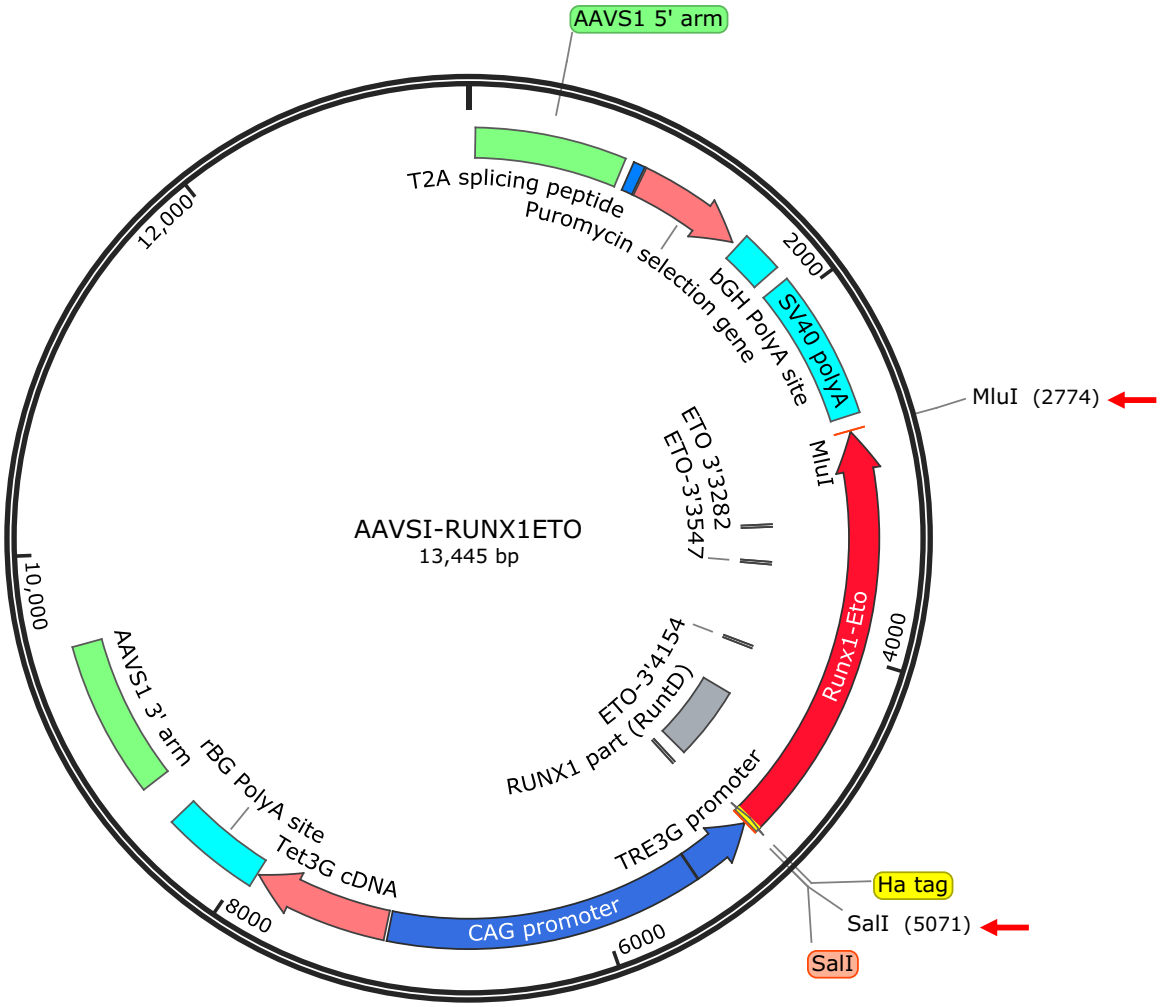


Figure 2.1: Map of the inducible RUNX1-ETO plasmid for knock-in to the AAVS1 locus

Schematic of the maps corresponding to the inducible RUNX1-ETO-containing plasmid for transgene knock-in into the human AAVS1 locus, including: 5' and 3' homology arms to the AAVS1 human locus, a T2A splicing peptide followed by a puromycin resistance cassette (expression under the endogenous promoter upon knock-in), a Tet3G reverse tetracycline-controlled activator (rtTA, Tet3G cDNA) under expression of a constitutive cytomegalovirus early enhancer/chicken beta actin (CAG) promoter and an HA-tagged *RUNX1-ETO* cDNA under control of a Tetracycline-inducible promoter (TRE3G-CMV). The *RUNX1-ETO* transgene was cloned substituting a GFP gene, using the SalI and MluI restriction endonuclease (RE) sites (indicated with red arrows). Primers used for sequencing the vector after cloning the cDNA are indicated in grey boxes.

GAPDH-targeting vectors containing either K-RAS(G12D) or KIT(N822K) cDNA were generated as follows. The pGAPTrap-T2A-ecDHFR-Ires-Meo plasmid for gene targeting into the GAPDH locus, used as a backbone to insert the transgene, was previously generated in the laboratory of Ed Stanley (Melbourne, AUS) (Kao et al., 2016). Final vectors comprised: a 3,3 kb 5' GAPDH homology arm, a T2A peptide sequence fused in frame with the GAPDH coding sequence, the corresponding transgene (K-RAS(G12D) or KIT(N822K)) fused to an *E. coli* dihydrofolate reductase (ecDHFR) domain, an internal ribosomal entry site (IRES), a G418 neomycin selectable marker, and a 4.2 kb 3' GAPDH homology arm (Figure 2.2). For the KIT(N822K) vector only, sequences encoding a FLAG tag were cloned into BamHI RE site downstream of the ecDHFR domain of the pGAPTrap-T2A-ecDHFR-Ires-Meo (mouse neomycin) vector. KIT(N822K) and K-RAS(G12D) transgenes were inserted upstream and downstream, respectively, of the ecDHFR domain into the GAPTrap targeting vector by InFusion cloning (Takara Bio cat# 121416) following the manufacturers protocol. Briefly, GAPTrap targeting vectors were linearized using AscI or SalI RE sites for the K-RAS or KIT cloning strategy, respectively, and de-phosphorilated with Antartic phosphatase (New England Biolabs cat# M0289S) to prevent plasmid re-ligation. Transgenes were amplified from donor plasmids by PCR using a high-fidelity Platinum™ Taq DNA Polymerase (Thermo Fisher Scientific cat# 10966026), as described in the manufacturer's protocol, by using a 3-minute extension and overlapping primers to the boundaries of the RE used to linearize the GAPtrap targeting vector. Both linearized GAPTrap plasmids and insert PCR products were purified by agarose gel electrophoresis, as described above, and ligated by In-Fusion cloning with a 2:1 insert to vector molar ratio. Reactions were

transformed into Stellar Competent Cells (Takara Bio cat# 636766) and ligated products were purified from Ampicillin-resistant selected individual colonies as above. Successful insert recombination into vectors was confirmed by RE digestion and the lack of point mutations in plasmid inserts was confirmed by Sanger sequencing, conducted by Australian Genome Research Facility at Walter and Eliza Hall Institute (Melbourne, Australia). Generation of the KIT vector required the insertion of a 93-bp intron - the human Granulocyte-macrophage colony-stimulating factor (GM-CSF) intron 1 - containing a STOP codon between an exon-exon junction within the KIT cDNA to prevent bacteria from expressing the KIT and selecting clones with acquired STOP point mutation. Primers used and corresponding descriptions are listed in (Table 2.1).

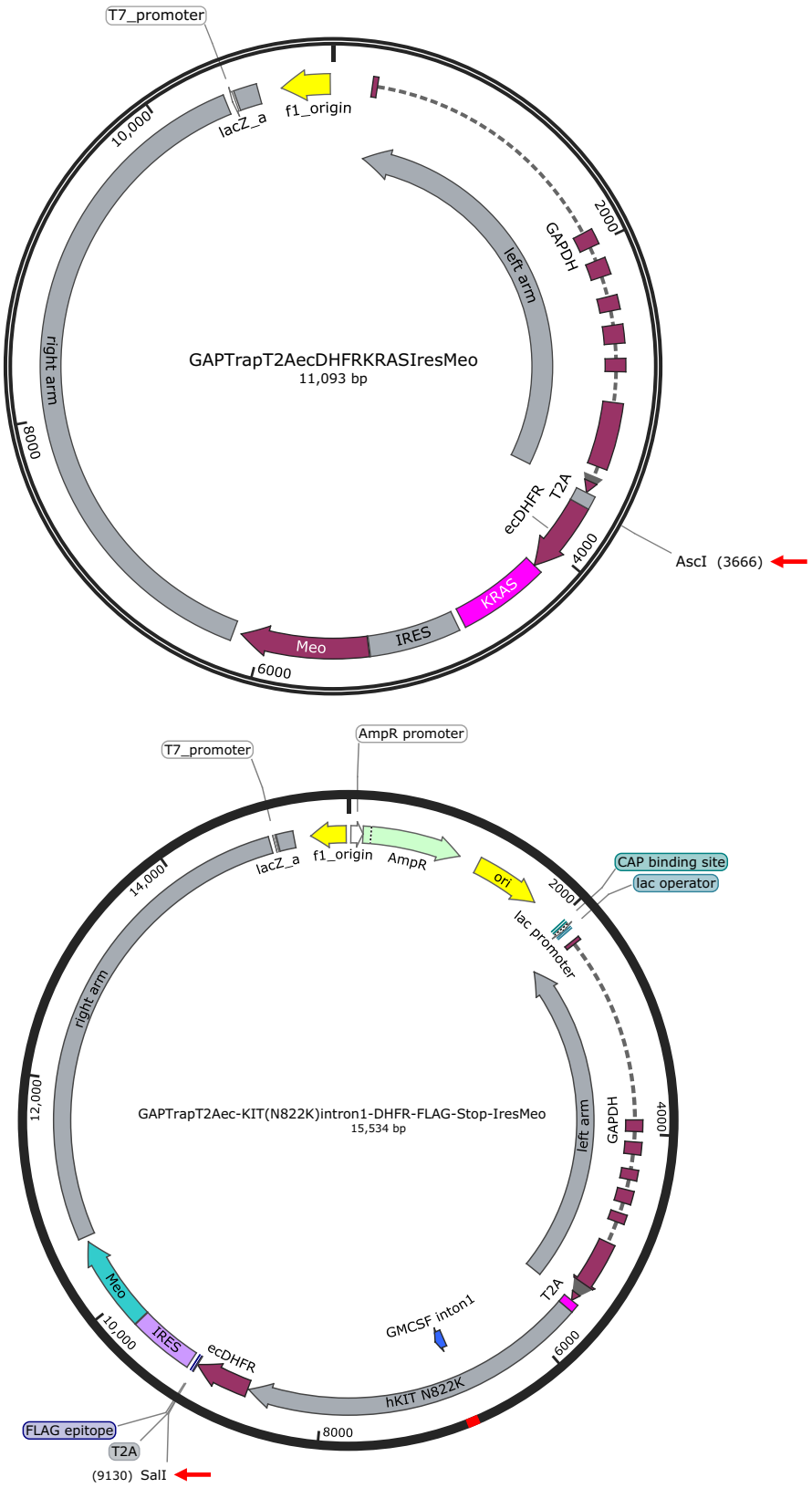


Figure 2.2: Maps of the K-RAS(G12D) and KIT(N822K) GAPTrap plasmids

Schematic of the maps corresponding to the GAPTrap plasmids containing a K-RAS(G12D) or a KIT(N822K) transgene, respectively. Vectors included: homology arms for gene targeting into the

human GAPDH locus, a T2A sequence following the GAPDH coding sequence (to allow expression of the transgene under the endogenous GAPDH promoter), cDNA sequence of the corresponding transgene fused to an *E. coli* dihydrofolate reductase (ecDHFR) domain, an internal ribosomal entry site (IRES) and a G418 neomycin selection cassette. Sequences encoding a FLAG tag were cloned into a BamHI RE site downstream of the ecDHFR domain of the KIT(N822K) vector only. The K-RAS(G12D) transgene was inserted downstream of the ecDHFR domain by using an AscI RE site. The KIT(N822K) transgene was inserted upstream of the ecDHFR domain by using a SalI RE site, and included a 93-bp intron (GM-CSF intron 1) containing a STOP codon between two KIT exons codon to prevent bacteria from expressing the KIT and selecting clones with point mutations. The AscI and SalI Re sites used to linearize the vectors and clone the transgenes are indicated with red arrows.

2.2 Growth and maintenance of human Pluripotent Stem Cells (hPSCs)

The dual reporter SOX17$^{mCHERRY/w}$RUNX1C$^{GFP/w}$ H9 human Embryonic Stem (ES) Cell line previously generated in the laboratory of Andrew Elefanty (Melbourne, AUS) was used (Ng et al., 2016). Human ES cells were routinely co-cultured with mitotically inactivated primary mouse embryonic fibroblasts (MEFs) in a defined serum-free media at 37°C in a humidified incubator with 5% CO2 and low (5%) O_2 conditions. The media was filter sterilized and consisted of DMEM Nutrient Mixture F-12 1x + L-glutamine Na bicarbonate (DMEM/F12, cat# 11320-033) supplemented with 20% KnockOut Serum Replacer (KOSR, cat# 10828028), 100x non-essential amino acids (NEAA, cat# 11140-050), 200 mM GlutaMAX^{TM}I 100x (cat# 35050-061), 55 mM 2-Mercaptoethanol (cat# 21985-023) (all Thermo Fisher Scientific) and 10 ng/ml basic Fibroblast Growth Factor (FGF2, PeproTech cat# 100-18B). All cell centrifugations were done at 300 x g for 3 minutes at 4ºC. Human ES cell cultures were passaged with TrypLE™ Select Enzyme (Thermo Fisher Scientific cat# 12563011). Cells were cryopreserved in 10% Dimethyl sulfoxide Hybri-Max™ (DMSO, Merck cat# D2650) and CJ2 solution consisting of 20x Choline Chloride (382 mg/ml in dH_2O), 0.01 mM $CaCl_2.2H_2O$, 2.68 mM KCl, 1.47 mM KH_2PO_4, 6.54 mM $K_2HPO_4.3H_2O$, 0.5 mM $MgCl_2.6H_2O$ and 5.5 mM D-glucose in dH_2O.

2.3 Growth of feeder cells

Irradiated MEFs were cultured in feeder media, consisting of DMEM high glucose (4.5 g/l) without glutamine and Na pyruvate (Thermo Fisher Scientific cat# 11960044) supplemented with 10% FCS, 200 mM GlutaMAXI and 1x Pen/Strep. T75 flasks were pre-coated with media for a minimum of 8 h and MEFs were plated at concentration of 1-1.5 x 10^5 cells/flask to support human ES cell cultures.

2.4 Establishment and validation of human ES cell lines expressing RUNX1-ETO, RUNX1-ETO K-RAS(G12D) or RUNX1-ETO KIT(N822K)

RUNX1-ETO cell lines were generated by Transcription activator-like effector nuclease (TALENs)-mediated transgene knock-in into the AAVS1 locus of the dual reporter SOX17$^{mCHERRY/w}$RUNX1C$^{GFP/w}$ human ES cell H9 line. K-RAS(G12D) and KIT(N822K) were introduced by TALENs-mediated transgene knock-in into the GAPDH locus of the RUNX1-ETO SOX17$^{mCHERRY/w}$RUNX1C$^{GFP/w}$ human ES cell H9 lines generated previously.

Corresponding parental cell lines were harvested, and 1 x 10^6 cells were resuspended with 5 µg transgene-containing plasmid DNA and 1 µg of each TALEN pair in a total volume of 100 µl Resuspension Buffer R from the Neon™ Transfection System Kit (Thermo Fisher Scientific cat# MPK1096). Human ES cells were electroporated using Neon tips in an electroporation tube containing 3 ml buffer E2 with the following conditions: 2 pulses of 1050 v and 30 ms. Cells were plated on 60 mm dishes pre-seeded with MEFs and positive clones were selected by addition of the respective antibiotic; RUNX1-ETO-targeted cell lines were selected with 1 µg/ml Puromycin Dihydrochloride (Thermo Fisher Scientific cat# A1113802) for two days, KIT and KRAS-targeted RUNX1-ETO cell lines were selected with 50 µg/ml

Geneticin (G418, Thermo Fisher Scientific cat# 10131035) for 7 to 10 days. Non electroporated cells were also subjected to antibiotic treatment as a negative control. Following antibiotic selection, media was refreshed every 2 days until visible colonies were formed. Cells were harvested and subsequently subjected to single cell sorting (section 2.7) on 96-well plates pre-seeded with MEFs and incubated in low O_2 conditions for 7 to 10 days or until a visible colony was formed. Individual clones were expanded and tested for transgene insertion after genomic DNA extraction (sections 2.4.1 and 2.4.2). For all lines, genomic integrity was confirmed using the Illumina HumanCytoSNP-12 v2.1 array by the Victorian Clinical Genetics Services (Murdoch Children's Research Institute, The Royal Children's Hospital, Melbourne, Australia).

2.4.1 Genomic DNA extraction

Genomic DNA extraction from lysed cells was performed using a vacuum manifold. Adherent cells on a 48-well plate were washed with PBS and lysed by overnight incubation at 37°C with 100 µl lysis buffer (0.5% SDS, 10 mM EDTA, 10 mM Tris, 70 mM NaCl) containing 0.2 mg/ml Proteinase K (Sigma cat# P2308) added fresh. Cell lysates were mixed by pipetting with 500 µl binding buffer PB (Qiagen cat# 19066) and transferred to a Whatman UNIFILTER 96-Well Microplate (GE Healthcare Life Sciences cat# 7700-7201). The plate was placed on the vacuum manifold and lysate suspension was passed through the filter by suction, not exceeding -10 Hg of pressure. Filters were washed by 2-minute incubation with 400 µl of buffer PE (Qiagen cat# 19065), which was subsequently discarded using the vacuum manifold. Excess of liquid was removed by tapping off the plate onto tissue paper and filters were dried using the suction of the vacuum manifold for 2 minutes. DNA was eluted

after 1-minute incubation with 70 µl TE buffer (10 mM Tris-HCl containing 1 mM EDTA) pre-warmed at 70°C into a clean collection 96-well plate.

2.4.2 Genomic Polymerase Chain Reaction

Correct transgene integration was screened by genomic PCR amplification using the GoTaq Green Master Mix (Promega #M712) or the Platinum™ Taq DNA Polymerase (Thermo Fisher Scientific cat# 10966026). Forward and reverse primers (Table 2.2) were designed to amplify the boundaries of the genomic insertion, covering a sequence of the locus and a sequence of the insert, respectively. Homozygous or heterozygous targeting of the AAVS1 locus was identified with a pair of primers designed to cover both locus sides, which only yielded fragment amplification in the presence of a wild type allele.

For the AAVS1 locus screen, reactions of 12 µl of volume were prepared containing 1X Gotaq mastermix, 0.5 µM forward primer, 0.5 µM reverse primer, <250ng genomic DNA and nuclease-free H_2O to the final volume. A negative control reaction without DNA and a positive control reaction with DNA of a cell line targeted to our locus of interest were prepared. Reactions were transferred to a Thermocycler with the following conditions: initial denaturation at 95ºC for 2 minutes, 33 cycles (denaturation at 95ºC for 30 seconds, annealing at 60ºC for 30 seconds, elongation at 72ºC for 2 minutes) and a final extension at 72ºC for 6 minutes.

For the GAPDH locus screen, reactions of 20 µl volume were prepared containing 2 units Platinum Taq DNA Polymerase, 0.2 µM forward primer, 0.2 µM reverse primer, <250ng genomic DNA, 1X PCR buffer –Mg, 1.5 mM $MgSO_4$, 0.8 mM dNTP mix, 5% DMSO and nuclease-free H_2O to the final volume. Reactions were

transferred to a Thermocycler with the following conditions: 95ºC for 2 minutes, 35 cycles (95ºC for 30 seconds, 65ºC for 30 seconds, 68ºC for 3 minutes 30 seconds) and 68ºC for 10 minutes.

PCR products were purified using agarose gel electrophoresis with 1% agarose (Bioline cat# BIO-41027) in 1x Tris/Borate/EDTA (TBE) buffer solution containing 5 µl/ml RedSafe Nucleic Acid Staining Solution (Intron Biotechnology cat# 21141). Agarose gels were run at 100 V in 1x TBE buffer and DNA products were visualised using a Bio-Rad Gel Doc™ XR+ System.

2.5 Haematopoietic differentiation

Haematopoietic differentiation of hPSCs was performed following the spin Embryo Body (EB) method in STAPEL medium (Ng et al., 2008, 2016). A confluent (95-98%) T75 flask of hPSCs was used to set up 10 Spin-EB plates of 60 EBs/plate. Low attachment 96-well plates (Sterile, Costar cat# COR3788) were prepared by adding 70 µl sterile ddH_2O to the outer 30 wells of each plate. On the day of differentiation (day 0), hPSCs were passaged 1:1 onto flasks of the same surface area pre-coated with feeder media only. Cells were incubated for 2 hours in low O_2 conditions to allow the cells to re-attach. Alternatively, cells may be passaged onto low-density feeders, such that they will be 90-98% confluent the next day, the day prior to EB set-up. Adherent cells were rinsed with PBS, gently dissociated by 3-minute incubation at 37ºC with Accutase™ solution (Merck cat# A6964) and harvested in Iscove's Modified Dulbecco's Media (IMDM) with no phenol red (Themo Fisher Scientific cat# 21056-023) with 0.2% recombinant human albumin Albucult (Albumedix). Cells were pelleted by centrifugation and mixed into the main corresponding volume of STAPEL

media supplemented with 0.5CH BVSAF cytokines (herein referred as STAGE1): 20 ng/ml bone morphogenetic protein 4 (BMP4, R&D Systems cat# 314-BP), 25 ng/ml vascular endothelial growth factor (VEGF, PeproTech cat# 100-20), 25 ng/ml stem cell factor (SCF, PeproTech cat# 300-07), 7.5 ng/ml ACTIVIN A (ACT, R&D Systems cat# 338-AC), 10 ng/ml FGF2 and 0.5 μM CHIR99021 (Tocris Biosciences cat# 4423). All human cytokines used for haematopoietic differentiation were recombinant. STAPEL media consisted of a mixture of 50% IMDM, 50% Ham's F12 (Thermo Fisher Scientific cat# 11765-062) and 0.05% Polyvinyl alcohol (PVA, Merck cat# P8136) supplemented with: 0.5% Albumin, 1x Gibco® Protein Free Hybridoma Medium II (PFHMII, Thermo Fisher Scientific cat# 12040077), 0.03% α-Monothioglycerol (MTG, Merck cat# M6145), 2.2 μg/ml SyntheChol (Merck cat# S5442), 100 ng/ml Linolenic and Linoleic acids (Merck cat# L1376 and L2376), 0.005 mg/ml Soybean Oil (lecithin) (Merck cat# S7381), 2 mM GlutaMAXI, 50 μg/ml Ascorbic acid 2-phosphate (AA2P, Merck cat# A8960), 50 μg/ml L-Ascorbic acid (Merck cat# A4403) and 1x Insulin-Transferrin-Selenium-E (ITS-E, In Vitria cat# 777ITS092). Albumin could also be substituted for a 1:1 mixture of albumin from rice endosperm (ScienCell Research Labs cat# OsSA) and Bovostar acid stripped Bovine Serum Albumin (BSA, Bovogen cat# BSAS 0.1).

Following cell resuspension in STAGE1 media, volumes of 80 μl were distributed into the 60 inner wells of each low attachment 96-well plate using a multichannel pipette. Therefore, approximately 5 ml differentiation medium per plate was required. Cells were aggregated into EBs to the bottom of the wells by centrifugation and plates were incubated at 37°C with 5% CO_2 and high (air levels) O_2 conditions. On day 1 of the differentiation, EB formation was evaluated under the microscope: the forming

EBs presented a smooth defined edge. Approximately 4-6 hours before the 48-hour time point from setup was reached (day 1.6-1.7), 20 µl of STAPEL supplemented with 3.5 µM SB431542 (Sapphire Bioscience cat# 13031) and 3 µM CHIR99021 were added to each well to promote adequate mesoderm patterning, giving a final volume of 100 µl per well. At day 4 of differentiation, 70-80 µl of STAGE 1 media were carefully removed from each well using a multichannel pipette and 100 µl of STAPEL supplemented with BVSIF cytokines (herein referred as STAGE2) were added to each well. STAGE2 cytokines consisted of 5 ng/ml BMP4, 50 ng/ml VEGF, 50 ng/ml SCF, 10 ng/ml Insulin-like growth factor 2 (IGF2, PeproTech cat# 100-12) and 10 ng/ml FGF2.

At day 7 of differentiation, 20 EBs were harvested to evaluate mesoderm patterning and differentiation kinetics by flow cytometric analysis (section 2.6 and 2.9). At day 8, EBs were transferred onto adherent plates, which had been pre-coated for at least 60 minutes with Matrigel solution (IMDM with 1x Pen/Strep and 1:200 Corning® Matrigel® Growth Factor Reduced phenol red-free (In Vitro Technologies cat# FAL356231). Approximately 15-25 EBs or 70-85 EBs were used for each well of a 6-well plate or 10-cm plate, respectively. EBs may be plated at a higher density for early harvesting and analysis (short time after the adherent culture start date). Plates were incubated overnight to allow for proper attachment of the EBs to the surface and next day media was gently replaced with STAPEL supplemented with STAGE3 cytokines: 20 ng/ml BMP4, 100 ng/ml SCF, 100 ng/ml FMS-like tyrosine kinase 3 receptor (FLT3) ligand (PeproTech cat# 300-19), 50 ng/ml VEGF, 50 ng/ml Thrombopoietin (TPO, PeproTech cat# 300-18), 25 ng/ml Interleukin (IL) 3 (PeproTech cat# 200-03), 25 ng/ml IL6 (PeproTech cat# 200-06), 20 ng/ml IGF2, 10

ng/ml FGF2 and 1x Pen/Strep. Plates were toped up with media every 3 days and half-media changes – with care not to aspirate the floating progenitors, which lie close to the plate surface – were performed when the media capacity of the plate was reached. STAGE3 cytokines and concentrations may vary depending on specific hPSC lines and experiments, with inclusion of BMP4 and IGF2 being optional and concentrations lowered down to half when media was regularly changed. Upon Dox treatment, adherent cultures were supplemented with a 5-FACTOR cytokine mix including 100 ng/ml SCF, 100 ng/ml FLT3 ligand, 50 ng/ml TPO, 25 ng/ml IL3 and 25 ng/ml IL-6.

2.6 Harvesting of haematopoietic cultures

For analysis of non-adherent EBs at day 7, EBs were collected using a multichannel pipette and transferred to a sterile collection tube. EBs were pelleted by 1-minute centrifugation and dissociated into single cells using a 21 or 23 G needle after a 30-minute incubation with TrypLE™ select enzyme. Cell suspensions were filtered through a 35 μm nylon mesh strainer into 5-ml round-bottom polypropylene sterile test tubes for subsequent flow cytometric analysis (section 2.9).

For analysis of adherent EB cultures, supernatants containing floating blood progenitors were collected and remaining adherent cells (including EBs and stroma) were washed with 1 ml PBS and incubated with TrypLE™ select at 37°C for 15 minutes. The stromal layer was collected using a cell scraper and EBs were detached with a pipette and transferred to a sterile tube containing 1 ml of 2 mg/ml Collagenase Type 1 or Type 4 (Worthington, CLS-1 or CLS-4). EBs were incubated for 15-30 minutes (when using CLS-1) or 30-45 minutes (when using CLS-4) in a

37°C water bath and dislodged using a 23 or 25 G needle. Dissociated EBs were collected together with the other cell fractions, which were filtered through a 40µm large filter cap and pelleted for subsequent single cell sort or flow cytometric analysis (sections 2.7 and 2.9, respectively).

2.7 Single cell sorting

Cells were single sorted using a Fluorescence-activated cell sorting (FACS) FACS Aria cell sorter. Following filtering of single-cell suspensions, pellets were resuspended in 200 µl FACS wash buffer (PBS with 2% FCS). Isotypes and single stainings were performed using 5 µl cells and the remaining cell suspension was stained with the correspondent antibody combination. Antibodies against CD9 and EpCam were used for sorting undifferentiated hPSCs and CD34, CD45 and CD90 for sorting haematopoietic progenitors (Table 2.3). Cell suspensions were incubated on ice and in the dark for 15 minutes, washed two times with 2 ml FACS wash and resuspended in 300 µl FACS wash with 1:1000 Propidium iodide (PI) prior to cell sort.

2.8 Magnetic-activated cell sorting of haematopoietic progenitors

The $CD34^{high}$ haematopoietic cell population was separated from the floating fraction of the cultures by magnetic-activated cell sorting (MACS) using CD34 MicroBeads Kit UltraPure human (Miltenyi Biotec cat# 130-100-453) and a Mini (MS) & Midi (ML) MACS™ Starting Kit (Miltenyi Biotec cat# 130-091-632). MS columns were used for 2 x 10^8 total cells with a maximum capacity of 10^7 CD34+ recovered cells per column. The floating cell fraction from the haematopoietic in vitro cultures was filtered through a 30 µm strainer in FACS wash buffer and cells were counted and kept on

ice. For the magnetic labelling, cells were pelleted by centrifugation and up to 10^8 total cells were resuspended in 300 µl ice-chilled MACS buffer (PBS, 2% FCS and 2 mM EDTA). Cell suspensions were mixed with 100 µl FcR blocking reagent, to block unwanted binding of antibodies and increase the specificity of antibody labelling, and 100 µl CD34 microbeads Ultrapure and incubated for 30 minutes at 4°C. Unbound antibody-beads were washed by addition of 10 ml MACS buffer and cells were pelleted by centrifugation at 300 x g for 10 minutes. Meanwhile, MS columns were placed on the magnet (Mini MACS separator) and equilibrated with 500 µl of MACS buffer. A 23 or 25 G needle was placed at the end of the column to slow down the flow rate, aiming to increase the binding of the cell-microbead complexes to the column. Cell pellets were resuspended in 500 µl MACS buffer and loaded into the equilibrated column. Unlabelled CD34- cells were cleared off the column with 3x washes using 500 µl MACS buffer. The column was then removed from the MACS separator and placed into a sterile 15 ml collection tube. CD34+ cell fraction was collected in 1 ml MACS buffer, which was loaded on top of the column and flushed out using the column plunger. Cells were pelleted to remove the MACS buffer and used for downstream analysis immediately.

2.9 Flow cytometry

Flow cytometric analysis was performed using a BD Fortessa analyzer. Following filtering of single-cell suspensions, cell pellets were resuspended in 200 µl FACS wash buffer (PBS with 2% FCS) and incubated with corresponding antibodies (Table 2.4) for 15 minutes on ice and in the dark. Cell suspensions were then washed twice with FACS wash and resuspended in 300 µl FACS wash with 1:1000 PI for subsequent flow cytometric analysis.

For intracellular flow cytometry, cell pellets were fixed by resuspension in three times the pellet volume of Fixation/Permeabilization solution (BD Pharmingen cat# 554722) and subsequent incubation on ice for 30 minutes. Cell suspensions were pelleted by centrifugation and washed twice using 1x Perm/Wash™ buffer (BD Pharmingen cat# 554723) for subsequent 30-minute incubation on ice and in the dark with primary conjugated antibody (Anti-HA tag [16B12] DyLight® 650-conjugated, Abcam cat# ab117515) diluted 1:100 in 1x Perm/Wash™ buffer. Cells were washed again twice with 1x Perm/Wash™ buffer and resuspended in 200 µl of FACS wash prior to flow cytometric analysis.

2.10 Intracellular immunostaining

Adherent undifferentiated ES cells on a 48-well plate were washed with PBS and fixed and permeabilized by 15-minute incubation at room temperature with 300 µl of 4% Paraformaldehyde and 0.5% Triton solution. Cells were washed three times with 1x Perm/Wash™ buffer and non-specific binding of proteins to the antibody was blocked by 15-minute incubation with 300 µl Perm/Wash™ buffer containing 5% 2 mg/ml goat IgG or 10% FCS. Cells were washed three times with 1x Perm/Wash™ buffer and incubated for 1 hour with Anti-HA tag Daylight 650-conjugated antibody and DAPI, diluted 1:100 and 1:100, respectively, in 1x Perm/Wash™ buffer. Cells were washed three times with 1x Perm/Wash™ buffer and 300 µl of PBS were added per well for subsequent epifluorescence imaging.

2.11 Imaging

Epifluorescence images of the *in vitro* haematopoietic cultures and immunostainings were taken using the 10x and 20x objectives of a Zeiss AxioObserver Z1 microscope

and a Zeiss AxioCam monochrome camera and were processed with the Zen Blue software.

Confocal images of the *in vitro* haematopoietic cultures were taken with a Zeiss LSM780 microscope using a 10x objective and processed with Zen Black software.

Epifluorescence and confocal images were exported as separate layers in JPEG format and assembled in Adobe Photoshop when required. Brightness and contrast adjustments were applied equally to all images.

2.12 Colony-forming-unit assays

Colony-forming-unit (CFU) assays were performed as reported (Ng et al., 2016). Briefly, cells from the floating fractions and EBs of day 18-24 cultures were dissociated and 3-5 x 10^3 cells were cultured in 1% methylcellulose supplemented with: 5-FACTOR cytokine mix (as used in the haematopoietic differentiation), 10 µg/ml human low density lipoproteins (hLDL, Stem Cell Technologies cat# 02698) and 5 U/ml erythropoietin (EPO, PeproTech cat# 100-64). For the preparation of 1% methylcellulose, 40 ml serum-free 2.6% MethoCult™ H4100 (Stem Cell Technologies #01400) was mixed with an equal volume of 2x STAPEL-P medium (STAPEL medium made with IMDM containing 2x supplements and without PFHMII) plus 20 ml of 1x STAPEL medium to give a final volume of 100 ml. The bottle was shaken to mix the various layers and let sit at room temperature allowing the bubbles to dissipate. Cells were cultured either with or without Dox and each condition was set up in triplicate in ultra-low attachment 24-well plates (cat# NUN144530). Plates were scored for haematopoietic CFUs after 7 to 10 days.

2.13 Replating assays

Replating assays were conducted on floating haematopoietic progenitors plated at a known concentration on wells of a 6-well plate that had been pre-coated with Matrigel solution. Cells were harvested, counted and replated weekly. Alive and dead cells were determined using a FL Countess-II Automated Cell Counter (Thermo Fisher Scientific) after Trypan-Blue staining.

2.14 Apoptosis cell staining

Cell death was analyzed by flow cytometry detection of the apoptosis marker AnnexinV and of a DNA-intercalator, using the eBioscience AnnexinV Apoptosis Detection Kit APC (Thermo Fisher Scientific cat# 88-8007-74). Cells were resuspended in 1X binding buffer (10X solution: 0.2 µm sterile filtered 0.1M Hepes (pH 7.4), 1.4M NaCl and 25 mM CaCl2) at a concentration of $1x10^6$ cells/ml. For each staining, including a double staining of AnnexinV and the DNA intercalator and single stainings, $1x10^5$ cells suspended in a volume of 100 µl were used. AnnexinV-APC staining was conducted using 1:20 dilution of the antibody followed by 15-minute incubation at room temperature and in the dark. A volume of 400 µl of binding buffer was added to each tube prior to addition of the DNA intercalator. Vybrant DyeCycle Violet Stain (Thermo Fisher Scientific cat# V35003) was used as a DNA intercalator at 1 µM concentration and incubated for 30 minutes at 37°C in the dark before proceeding to flow cytometric analysis.

2.15 Cell cycle analysis

Cell cycle was analyzed by flow cytometry using incorporation of the thymidine analog BrdU and of a fluorescent cell-membrane permeable DNA intercalator. First,

25 µM BrdU (Sigma cat# B5002) was added to the cells and incubated for 3 hours. Thereafter, around 3-4 million floating blood progenitor cells were collected, washed with PBS and resuspended in 250 µl PBS. Cells were fixed on a vortex shaker at low speed (1000 rpm) by drop-wise addition of 750 µl ice-cold pure ethanol to the cells to achieve a final concentration 75% ethanol. Cells could be stored at 4°C up to a week before proceeding to BrdU staining. Suspensions of fixed cells were transferred to a microcentrifuge tube and cells were pelleted by centrifugation at 500 x g for 5 minutes and 4°C. Cells were washed with PBS by flicking the tube and cell pellets were hydrated by a 20-minute incubation in PBS. Cells were then resuspended in 200 µl 2N HCl and incubated for 20 minutes to denature the double-stranded DNA and allow binding of the antibody to the BrdU nucleoside. Cells were subsequently washed twice with PBS and then washed twice with blocking buffer (5% FBS, 0.1% NaN_3, 0.1% TritonC100 in PBS), which is required to avoid denaturation of the antibodies. After this step, samples were divided into two tubes for the IgG control and the BrdU staining and subjected to RNAseA treatment (100 µg/ml) in PBS for 30 minutes at 37°C. Cell pellets were resuspended in 40 µl of undiluted IgG-FITC or BrdU-FITC conjugated antibodies (BD Pharmingen cat# 556028), respectively, and incubated at room temperature for 50 minutes. Cells were washed twice with PBS, resuspended in 400 µl PBS containing 1 µM Vybrant DyeCycle Violet Stain and incubated for 30 minutes at 37°C in the dark before proceeding to flow cytometric analysis.

2.16 RNA isolation

2.16.1 Chaotropic salt RNA Lysis buffer-based protocol

Total RNA was isolated from cell cultures by using the ISOLATE II RNA Mini Kit (Bioline cat# BIO-52073). Up to 5 x10^6 cells were collected and lysed with 350 µl RNA Lysis (RLY) Buffer and 3.5 µl ß-mercaptoethanol. Homogenized samples were stored at -80 °C. RNA was isolated following the manufacturer's protocol and final elution was performed in 30 µl or 10 µl RNase-free water, for the mini kit or the Micro kit, respectively. The eluate was passed twice through the column to increase the elution efficiency. RNA concentration was quantified using a Nanodrop 2000c spectrophotometer. Total RNA was directly used for cDNA generation using random hexamer priming (section 2.18.2).

2.16.2 TRIzol-based protocol

Total RNA was isolated from cell cultures using TRIzol Reagent (Thermo Fisher Scientific cat# 15596026). Cell pellets were lysed and homogenized with 1 ml of TRIzol and incubated 5 minutes at room temperature. For each 1 ml of TRIzol, 0.2 ml of chloroform were added and hand-mixed for 15 seconds and reactions were incubated for 3 minutes at room temperature. Phases were separated by centrifugation at 12,000 x g for 15 minutes at 4°C and the RNA-containing upper-aqueous phase was pipetted out, avoiding pipetting any of the interphase or phenol-chloroform organic layers. RNA was subsequently isolated from the aqueous phase by addition of 5 µg of UltraPure glycogen (Thermo Fisher Scientific cat# 10814010) and 0.5 ml of 100% isopropanol per 1 ml TRIzol used. Samples were incubated at room temperature for 10 minutes and centrifuged at 12,000 x g for 10 minutes at 4°C.

RNA pellets were washed with 1 ml of 75% ethanol, vortexed and then centrifuged at 16,000 x g for 10 minutes at 4°C. RNA pellets were air-dried for a maximum of 10 minutes and resuspended in 17 µl DNase/RNase-free water (Thermo Fisher Scientific cat# 10977049). Following incubation at 55-60°C for 10 minutes, 1 µl RNase-free DNase I (Thermo Fisher Scientific cat# EN0521) and 2 µl DNase I Buffer were added to each sample and reactions were incubated at 37°C for 30 minutes. RNA was isolated following the manufacturer's "Clean-up of RNA from Reaction Mixtures" protocol from the Nucleospin RNA kit (Macherey-Nagel cat# 740955.50). RNA was eluted in 30 µl RNase-free water passed twice through the column to increase the elution efficiency. RNA concentration was quantified using a Nanodrop 2000c spectrophotometer. Total RNA was directly used for cDNA generation using oligo $(dT)_{18}$ priming (section 2.18.1).

2.17 RNA library preparation and sequencing

RNA-sequencing (seq) libraries were prepared using a TruSeq® Stranded mRNA Library Prep kit (Illumina cat# 20020594) following the Low Sample (LS) workflow according to manufacturer's instructions. Briefly, polyA-containing mRNA was purified with 50 µl oligo-dT-attached magnetic (RNA Purification) beads and subjected to two rounds of purification. Thereafter, the RNA was fragmented and primed with random hexamers for cDNA synthesis by addition of 19.5 µl Fragment-Prime-Finish buffer and incubation at 94°C for 8 minutes. The cleaved RNA was then reverse transcribed into first strand cDNA using 1 µl Superscript II reverse transcriptase (Invitrogen) and 7 µl First Strand Synthesis Mix, where false DNA-dependent synthesis was prevented by Actinomycin D included in the Synthesis mix, and samples were subsequently incubated (25°C 10 minutes, 42°C 15 minutes, 70°C

15 minutes). Second strand cDNA was synthesized by incubation at 16°C for 1 hour using a Second Strand marking mix containing dUTP instead of dTTP, which allowed quenching of the second strand during the library amplification step. RNAseH was included during the second strand synthesis to remove the RNA template. The resulting blunt-ended cDNA was purified from the reaction mix by addition of 90 µl of AMPure XP (Beckman Coulter cat# A63881) magnetic beads to each sample and 15-minute incubation at room temperature, followed by subsequent removal of the supernatant, two washes with 80% ethanol and elution in 17.5 µl of resuspension buffer. In order to prevent simultaneous ligation of the blunt-ended cDNA during the adapter ligation step – and therefore concatenate template formation – 3' ends of the fragments were adenylated to provide a complementary overhang to the corresponding single 'T' nucleotide on the 3' end of the adaptors, by addition of 12.5 µl A-Tailing mix and 2.5 µl resuspension buffer prior to incubation (37°C for 30 minutes and 70°C for 5 minutes). Single-index Illumina adaptors were ligated to the ends of the double-stranded adenylated cDNA fragments for later hybridization onto a flow cell, allowing for multiplexed sequencing. This was performed by addition of 2.5 µl RNA Adaptor Index (diluted 1:4), 2.5 µl Ligation Mix and 2.5 µl resuspension buffer following a 10-minute incubation at 30°C and addition of 5 µl Stop Ligation Buffer. Ligated fragments were subjected to two rounds of purification with 42 µl and 50 µl, respectively, of AMPure magnetic beads and DNA fragments containing adaptor sequences on both ends were eluted in 22.5 µl and selectively enriched by 15 cycle PCR amplification. Reactions were prepared containing 20 µl of eluted DNA fragments, 5 µl PCR Primer Cocktail and 25 µl PCR Master Mix and were incubated as follows: 98°C for 30 seconds, 15 cycles (98°C for 10 seconds, 60°C for 30

seconds, 72°C for 30 seconds) and final extension 72°C for 5 minutes. On this amplification step, the second strand is effectively quenched during amplification, given that the Polymerase used does not incorporate past dUTP.

Libraries were subjected to a quality control by determining the average fragment size using a High Sensitivity DNA chip on an Agilent Technologies 2100 Bioanalyser™ instrument and were quantified using the real-time quantitative PCR (RT-qPCR)-based method KAPA Library Quantification Kit for Illumina Sequencing Platforms (Roche cat# KR0405), following the manufacturer's protocol, on an Applied Biosystems StepOne Plus RT-qPCR system. Libraries were run in a pool of twelve indexed libraries in a NextSeq (Illumina) machine using sequencing by synthesis chemistry and a NextSeq® 500/550 High Output 150 cycle sequencing kit v2 (cat# FC-404-2002), obtaining 75 bp paired-end reads. Sequencing Service was performed at the Genomics Birmingham sequencing facilities (Institute of Cancer & Genomic Sciences, University of Birmingham, United Kingdom).

For some other samples, RNA was sequenced at 20M reads per sample (2x75 bp) after TruSeq stranded mRNA library preparation. RNA Quality Controls, library prep, library Quality Controls, library pooling, pool QCs, and sequencing were performed by the Sequencing Service and Development Platform (Translational Genomics Unit, Victorian Clinical Genetics Services and Murdoch Children's Research Institute, The Royal Children's Hospital, Melbourne, Australia).

2.18 cDNA synthesis from total RNA

2.18.1 Oligo $(dT)_{18}$ priming protocol

Reverse-transcription of RNA to cDNA with Oligo $(dT)_{18}$ priming was conducted using 1000 ng and the SuperScript™ II Reverse Transcriptase (Thermo Fisher Scientific cat# 18064014) following the manufacturer's protocol. Transgene expression was quantified by RT-qPCR using a DNA-intercalating fluorescent dye-based method (section 2.19.1).

2.18.2 Random hexamer priming protocol

Reverse-transcription of RNA to cDNA with random hexamer priming was performed using 200-500 ng purified mRNA and the Tetro cDNA Synthesis Kit (Bioline cat# BIO-65043) according to the manufacturers' instructions. Samples were stored at -20 °C or directly used for subsequent RT-qPCR with either DNA binding dyes or hydrolysis probes (section 2.19).

2.19 Real-Time Quantitative PCR

2.19.1 DNA binding dye-based methods

Gene expression levels were evaluated by RT-qPCR analysis using the DNA intercalating dyes GoTaq qPCR Master Mix (Promega cat# A6001) and SYBR Green qPCR master mix protocol (Thermo Fisher Scientific cat# 4309155) with primers listed in Table 2.5. Reactions were prepared in MicroAmp™ Fast Optical 96-Well Reaction Plates with Barcode 0.1 ml (Thermo Fisher Scientific cat# 4346906). Reactions using the GoTaq qPCR Master Mix contained 5 µl cDNA dilution (1:15) and 10 µl of a master mix including 0.5 µl 10 µM primer mix, 7.5 µl GoTaq® qPCR

Master Mix and 2 µl DNase-free water per well. Reactions using the SYBR Green qPCR Master Mix contained 2.5 µl cDNA dilution (1:10) and 7.5 µl of a master mix including 0.3 µl 10 µM primer mix, 5 µl SYBR Green master mix and 2.2 µl DNase-free water per well. Standards consisted of a mixed pool of 3 µl from each cDNA sample diluted to 1, 0.2, 0.04 and 0.008 ng/µl (1:5 serial dilutions). A negative control without cDNA sample was performed and reactions were prepared as technical duplicates. Plates were sealed with MicroAmp™ Optical Adhesive Film (Thermo Fisher Scientific cat# 4360954) and centrifuged at 300 x g for 1 minute to collect reactions at the bottom of the wells. Gene expression was quantified in a AB StepOnePlus Real-Time PCR System (Thermo Fisher Scientific cat# 4376600) or a QuantStudio 5 Real-Time PCR System (Thermo Fisher Scientific). SDS software was used to validate the primer specificity, by analysis of the melting curve plot, and to determine the cycle thresholds (Ct). Gene expression was calculated relative to the standard curve and normalized to the expression of the housekeeping gene *GAPDH*, which was used as an internal control.

2.19.2 Hydrolysis probe-based method

Gene expression levels were manually evaluated by RT-qPCR analysis using TaqMan FAST Advanced Master Mix Applied Biosystems (Thermo Fisher Scientific cat# 4444556) and hydrolysis probes listed in Table 2.6. Reactions were prepared in MicroAmp™ Fast Optical 96-Well Reaction Plates with Barcode 0.1 ml by addition of 9 µl of a master mix per well containing 0.5 µl Taqman assay probe, 5 µl 2X TaqMan FAST Advanced Master Mix and 3.5 µl DNase-free water for subsequent resuspension of 1 µl cDNA dilution (1:15). Plates were sealed with MicroAmp™ Optical Adhesive Film and reactions were pelleted by centrifugation. Gene

expression was quantified in a QuantStudio 5 Real-Time PCR System. Primer specificity was validated by analysis of the melting curve plot and Ct were determined. Gene expression was calculated relative to the standard curve and *GAPDH* was used as the reference gene to normalize data.

2.20 Assay for Transposase-Accessible Chromatin using sequencing

Chromatin accessibility was evaluated using the Assay for Transposase-Accessible Chromatin (ATAC)-seq employing a modified protocol to as reported (de Boer et al., 2018). Following single-cell sorting, 50.000 haematopoietic progenitors were pelleted by centrifugation at 300 x g for 10 minutes and 4°C and resuspended in 5 µl sucrose freezing buffer, consisting of 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 10 mM Tris pH 7.4 and 1.5 M sucrose. Cells in sucrose freezing buffer were snap frozen in liquid nitrogen and stored for subsequent ATAC digestions. Tubes were thawed at room temperature and 45 µl of ATAC reaction buffer, consisting of 25 µl Tagmentation DNA Buffer (Illumina cat# FC-121-1030, Nextera DNA Library Prep Kit), 2.5 µl Tn5 Transposase enzyme (Illumina cat# FC-121-1030, Nextera DNA Library Prep Kit), 1 µl of 0.5% Digitonin (Promega cat# G9441) and 16.5 µl water, were added. Reactions were gently pipetted to resuspend the nuclei in the transposition mix and incubated for 30 minutes at 37°C with gentle shaking. Reactions were purified using a MinElute Reaction Cleanup Kit (Qiagen cat# 28204) and transposed DNA was eluted in 10 µl Elution Buffer (10mM Tris buffer, pH 8). Purified DNA was stored at -20°C or directly amplified by PCR. To amplify transposed DNA fragments, reactions of 50 µl were prepared containing: 10 µl transposed DNA, 10 µl Nuclease-Free H_2O, 2.5 µl 25 µM Customized Nextera PCR Primer Adaptor 1, 2.5 µl 25 µM Customized Nextera PCR Primer Adaptor 2 (barcoded) and 25 µl NEBNext High-Fidelity 2x PCR

Master Mix (New England Biolabs cat# M0541). Adaptor 1 was the same for all reactions, whilst Adaptor 2 changed depending on the barcode required for each flow cell. Reactions were transferred to a Thermocycler with the starting following conditions: incubation at 72°C for 5 minutes, initial denaturation at 98ºC for 30 seconds, 5 cycles (denaturation at 98ºC for 10 seconds, annealing at 63ºC for 30 seconds and elongation at 72ºC for 1 minute) and reactions were held at 4°C.

Before proceeding with library amplification, 5 µl of DNA amplification was analysed by RT-qPCR reaction to monitor the PCR cycles, which allowed to determine the optimal number of cycles to amplify the transposed DNA, stopping amplification prior to saturation, in order to reduce GC and size bias. The remaining 45 µl of PCR-amplified DNA was kept on ice. Reactions for RT-qPCR were prepared as follows: 5 µl of 5-cycle PCR-amplified DNA or water control, 4.44 µl Nuclease Free H2O, 0.25 µl 25 µM Customized Nextera PCR Primer Adaptor 1, 0.25 µl 25 µM Customized Nextera PCR Primer Adaptor 2, 0.06 µl 100x SYBR Green I (diluted from 10,000x stock in 10 mM Tris buffer pH 8), 5 µl NEBNext High-Fidelity 2x PCR Master Mix. Reactions were transferred to an ABI StepOne Real-time PCR machine with the following conditions: initial denaturation at 98°C for 30 seconds and 25 cycles (denaturation at 98ºC for 10 seconds, annealing at 63ºC for 30 seconds and elongation at 72ºC for 1 minute). The additional number of cycles needed for the remaining 45 µl 5-cycle PCR-amplified DNA were calculated from the blue fluorescence raw values: the baseline value of the water reaction was subtracted from the plateau value of each sample reaction, then the resulting values were divided by 3 and finally the baseline value was added on. Resulting values were plotted in a linear graph together with the blue fluorescence raw data and the number

of cycles required were determined by the intersection of the calculated values with the fluorescent values. The smaller cycle threshold value was taken when the intersection lied in between two cycle numbers.

The remaining 45 µl of amplified DNA were re-run on the starting PCR conditions with the correspondent calculated number of additional cycles. Adaptor dimers were removed from the libraries by adding 1.2x the sample volume of AMPure magnetic beads and pipetting 10x up and down. Reactions were incubated for 15 minutes at room temperature and the supernatant was removed by using a magnet to pellet the DNA-bead conjugates. Reactions were subjected to 2 washes with 200 µl freshly prepared 80% EtOH and beads were air-dried for 5-10 minutes. DNA was eluted from the beads by 5-minute incubation in water and beads were cleared out using the magnet. Libraries were evaluated using a High Sensitivity DNA chip on an Agilent Technologies 2100 Bioanalyser™ instrument and concentration was measured using a RT-qPCR-based method (KAPA Library Quantification Kit for Illumina Sequencing Platforms). Libraries were also validated by RT-qPCR evaluation of the ratio of open (TBP promoter) to closed regions of DNA (chromosome 18) and active gene body (*β-actin*) (Table 2.7). Libraries were run in a pool of twelve indexed libraries in a NextSeq (Illumina) machine using sequencing by synthesis chemistry and a NextSeq® 500/550 High Output 75 cycle sequencing kit v2 (cat# FC-404-2005), obtaining 75 bp single-end reads.

2.21 Chromatin Immunoprecipitation

Haematopoietic progenitor cells in the floating fraction were harvested and a maximum of 10 x 10^6 cells were pelleted per 15-ml collection tube. On some

occasions, the CD34+ fraction was enriched by MACS. Pellets were washed three times with PBS and single crosslinked by addition of 1 ml per 10^6 cells of IMDM with 10% FCS containing 1% methanol-free Formaldehyde (Thermo Fisher Scientific cat# 28906). Reactions were incubated for 10 minutes at room temperature on a tube roller mixer and stopped by addition of $1/10^{th}$ of the reaction volume of 2M glycine solution. Cells were pelleted by centrifugation at 400 x g for 5 minutes at 4ºC, washed twice with ice-cold PBS and snap frozen for storage at -80ºC.

The Chromatin Immunoprecipitation (ChIP) protocol includes pre-conjugation of the antibody to the magnetic beads, chromatin preparation and sonication, immunoprecipitation (IP) of the chromatin-bound protein complexes, washes of the unbound antibodies, elution, reverse-crosslink of the immunoprecipitated chromatin and DNA isolation.

Antibodies were used against the following antigens (manufacturer, cat #): HA tag (Sigma, H6908), RUNX1 (Abcam, ab23980), H3K27ac (Abcam, ab4729), H3K4me3 (Merk, 07-473), H3K79me2 (Abcam, ab3594). Antibodies were first pre-conjugated to the magnetic beads: 15 µl per each IP of Dynabeads-Protein G (Thermo Fisher Scientific cat# 10003D) were washed with 500 µl 0.1 M citrate phosphate pH 5.0 buffer (4.7 g Citric Acid MW=192, 9.2 g Dibasic Sodium Phosphate (Na_2HPO_4) dehydrate MW=141.96 for 1L). For histone IP, 10 µl of Dynabeads-Protein G were used per reaction. Tubes were placed on a magnetic rack until supernatant was clear and the supernatant was discarded. Beads were resuspended in 15 µl per IP of 0.1 M citrate phosphate pH 5.0 buffer and the appropriate amount of primary antibody and 0.5% BSA were added per reaction. For most antibodies, 2 µg /IP was used,

whilst 1 µg was used for histone modification IPs. Antibody-bead reactions were incubated on a rotating wheel at 4°C for 2 hours.

Meanwhile, cell lysis and chromatin preparation were conducted: cell pellets were resuspended to 1 x 10^7 cells/ml in ice-cold lysis buffer A (10 mM HEPES pH 8.0, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100) with freshly-added proteinase inhibitor cocktail (PIC, Merck cat# P8340). Suspensions were incubated on a rotating wheel for 12 min at 4°C and nuclei was pelleted by centrifugation at 500 x g for 7 min at 4°C. The supernatant was discarded, and nuclei were resuspended to 1 x 10^7 cells/ml in ice-cold lysis buffer B (10 mM HEPES pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.01% Triton X-100) with freshly-added PIC. Tubes were again incubated on a rotating wheel for 12 min at 4°C and chromatin was pelleted by centrifugation at 500 x g for 7 min at 4°C. The supernatant was discarded, and chromatin was resuspended in 300 µl per 2 x 10^6 starting cells of ice-cold IP buffer 1 (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.5% SDS) with freshly-added PIC. Sonication of the chromatin was performed using a Bioruptor® Pico sonication device (Diagenode). For that, 300 µl of each sample were transferred into 1.5 ml Bioruptor® Pico Microtubes with Caps (Diagenode cat# C30010016) and subjected to 6 cycles (30 sec ON, 30 sec OFF) for transcription factor IP or 10 cycles (30 sec ON, 30 sec OFF) for histone IP to produce fragments between 200-500 bp. Insoluble chromatin was pelleted by centrifugation at 16,000 x g for 10 min at 4°C and 300 µl of supernatant was transferred to a clean tube. Two volumes (600 µl) of IP buffer 2 (25 mM Tris-HCl pH 8.0,150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 7.5% glycerol) were added per tube and a 5% of reaction volume was aliquoted into a clean tube for input control preparation. For samples that were

to be immunoprecipitated using the anti-H3K79me2 ab, a concentration of at least 0.2% SDS in the chromatin (after addition of IP buffer 2) was used. This opens up the chromatin within the deep cleft of H3 where K79 sits, allowing the antibody to bind. Samples could be snap frozen and stored at -80°C at this step if needed.

To continue with the IP step, the antibody-bead reactions were washed with 500 µl citrate-phosphate buffer after the 2-hour incubation step and supernatant was removed using a magnetic stand. Antibody-conjugated beads were resuspended in 15 µl per IP of citrate phosphate buffer 0.5% BSA and then 15 µl were added into each sonicated chromatin solution of 2 x 10^6 cells (suspended in approximately 900 µl of IP buffers). Chromatin-antibody reactions were incubated on a rotating wheel at 4°C overnight. Next day, 1 µl glycogen (20 mg/ml) was added to each chromatin-antibody-beads mixture and reactions were subjected to washing steps on a magnetic rack. For that, supernatant was removed and beads were washed with 500 µl per wash, swapping the tube twice and leaving the buffer during 3 min, with the following buffers: once with Wash Buffer 1 (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA pH 8.0, 1% Triton X-100 and 0.1% SDS), twice with Wash Buffer 2 (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, and 0.1% SDS), once with LiCl Buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% NP-40 and 0.5% Na-deoxycholate) and twice with TE/NaCl Buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl and 1 mM EDTA pH 8.0). Alternatively, a different washing protocol (twice with Wash Buffer 1, once with Wash Buffer 2, once with LiCl Buffer and twice with TE/NaCl Buffer) was used to reduce the stringency of the washes of samples with lower amount of immunoprecipitated chromatin. Chromatin was then eluted from the beads in a final volume of 100 µl by doing two consecutive incubations with 50 µl

Elution buffer (100 mM $NaHCO_3$ and 1% SDS) while shaking at 1,200 rpm at room temperature for 15 min. Meanwhile, input controls were incubated with 1 µl RNaseA (10 mg/ml, Roche cat# 10109142001) for 30 min at 37°C and then samples were diluted to a final volume of 100 µl in elution buffer. To proceed with reverse crosslink, reactions were prepared by adding 4 µl NaCl 5M and 0.5 µl of Proteinase K (50 mg/ml, Roche cat# 03115879001) to each IP sample and 1 µl Proteinase K (50 mg/ml) and 1/10 volume (10 µl) 10% SDS to each input control. Chromatin crosslinking was reversed by overnight incubation in a 65 °C water bath. Next day, DNA was isolated by 15-minute incubation at room temperature with 1.8x volume of AMPure beads, two washes with freshly-prepared 80% ethanol and elution with two incubations with 50 µl 0.1x TE (pH 8.0) while shaking 15 minutes at room temperature. Samples were stored at -80°C or directly validated by RT-qPCR.

Prior to library preparation, the enrichment of specific antibody-bound chromatin regions was assessed by RT-qPCR using a SYBR Green master mix and primers designed to amplify human genomic regions (Table 2.8) in a AB StepOnePlus Real-Time PCR System. Standards consisted of genomic DNA from Kasumi-1 cell line diluted to 5, 1, 0.2 and 0.04 ng/µl (1:5 serial dilutions). A negative control without DNA sample was performed and reactions were prepared as technical duplicates. SDS software was used to validate the primer specificity, by analysis of the melting curve plot, and to determine the Ct. Amplification was calculated relative to the standard curve and normalized to the values obtained in the input control. Enrichment of genomic-bound regions was calculated comparing positive control regions (using primers amplifying known binding sites by the immunoprecipitated protein) to an unbound control region (Involucrin (IVL) gene, located in chromosome

1). RT-qPCR validations of RUNX1 and RUNX1-ETO ChIPs at both uninduced (0 Dox) and RUNX1-ETO-induced (5 Dox) conditions are shown in Figure 2.3.

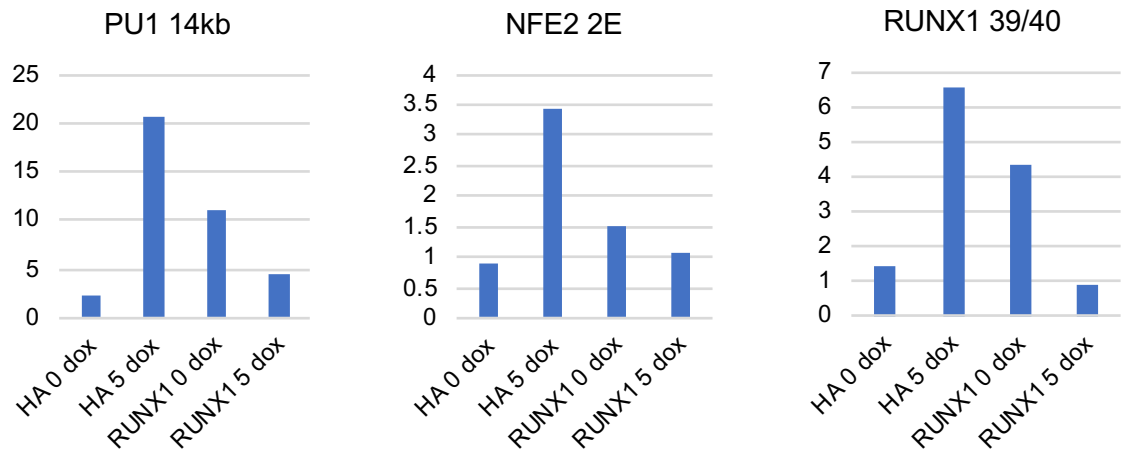


Figure 2.3: Enrichment of RUNX1 and RUNX1-ETO (0 and 5 Dox) ChIP at known binding regions

Manual ChIP qPCR showing enrichment of RUNX1 and RUNX1-ETO binding (at both 0 Dox and 5 Dox conditions) at known binding regions (PU 14 Kb, NFE2, and intronic RUNX1 enhancers). Enrichment is shown as normalized to input and to a control unbound region (IVL).

2.22 ChIP-sequencing (CHIP-seq) library preparation and sequencing

ChIP libraries for Illumina sequencing were prepared using the KAPA Hyper Prep Kit (Roche, KR0961), using the whole ChIP product suspended in a volume of 50 µl. End repair and A-tailing was conducted by 30-minute incubation at 20ºC of the immunoprecipitated DNA fragments with 7 µl End Repair and A-tailing buffer and 3 µl of enzyme mix, followed by 30-minute incubation at 65ºC. Adapters were ligated by 15-minute incubation at 20°C of the reactions after addition of 5 µl 1:10 dilution of adaptor stocks (to make a 200:1 adapter:insert ratio), 5 µl PCR-grade water, 30 µl ligation buffer and 10 µl DNA ligase. A bead-based post-ligation cleanup was performed using 0.8X AMPure XP beads and DNA fragments were eluted in 20 µl to proceed with library amplification. Reactions of 50 µl were prepared containing the

eluted adapter-ligated library, 25 µl KAPA HiFi HotStart ReadyMix (2X) and 5 µl KAPA Library Amplification Primer mix (10X). Libraries were amplified using the following thermocycling protocol: 98ºC for 45 seconds, 16 cycles (98ºC for 15 seconds, 60ºC for 30 seconds and 72ºC for 30 seconds) and 72ºC for 60 seconds. Non-ligated adapters were eliminated by performing a post-amplification cleanup with 1.2X AMPure XP beads and libraries were subjected to size selection in a 2% agarose TAE gel containing 0.5 µg/ml ethidium bromide. Fragments of 200-500 bp in size were excised from the gel and purified using the QIAquick® Gel Extraction Kit as described in the manufacturer's protocol with the following modifications: gels were diluted in QG buffer whilst rotating and libraries were eluted in 12 µl Buffer EB.

Quality control of the libraries was performed using a High Sensitivity DNA chip on an Agilent Technologies 2100 Bioanalyser™ instrument and libraries were quantified using the RT-qPCR-based method KAPA Library Quantification Kit for Illumina Sequencing Platforms, following the manufacturer's protocol. Libraries were run in a pool of twelve indexed libraries in a NextSeq (Illumina) machine using sequencing by synthesis chemistry and a NextSeq® 500/550 High Output 75 cycle sequencing kit v2 (cat# FC-404-2005), obtaining 75 bp single-end reads.

2.23 Single Cell RNA-Seq (scRNA-Seq)

Non-adherent progenitors at day 22 of differentiation (untreated and 24-hour Dox treatment) were sorted for CD45+ CD34+ and RUNX1C+ in a MoFlo cell sorter. Cells were re-suspended in 80 µl at a concentration of 100-1200 cells/µl for evaluation of cell viability prior to loading of 4000 single cells on a Chromium Single Cell Instrument (10X Genomics). Library generation for scRNA-seq was performed by the

Genomics Birmingham Sequencing Facility using the Chromium Single Cell 3′ Library and Gel Bead Kit v2 (10X Genomics). Libraries were paired-end sequenced on an Illumina NextSeq machine using the cycle parameters recommended by 10X Genomics.

2.24 Immunoblotting

Immunoblotting was performed as reported (Lamandé et al., 2011) with some modifications. Briefly, snap-frozen dry cell pellets were defrosted in cell lysis buffer consisting of 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM EDTA, 1% (vol/vol) Triton-X100, and 1:100 Protease Inhibitor Cocktail (2 M N-Ethylmaleimide and 100 mM 4-benzenesulfonyl fluoride hydrochloride). Nuclei and insoluble material were pelleted by centrifugation at 16,000 x g for 10 minutes at 4°C and the supernatant was transferred to a clean tube. Protein lysates were diluted in 4X Bolt™ lithium dodecyl sulphate Sample Buffer (Thermo Fisher Scientific cat# B0007) and 10 mM Dithiothreitol (DTT) and were denatured by incubation at 65°C for 10 minutes. Proteins were separated under denaturing conditions in Bolt 4-12 % Bis-Tris Plus precasted polyacrylamide gels (Thermo Fisher Scientific #NW04120BOX) with Bolt™ MOPS SDS Running Buffer (Thermo Fisher Scientific cat# B000102) at 80 V for the first 20 minutes and 150 V until the end. Precision Plus Protein Standard (#161-0373, BioRad) was used as a molecular weight marker. Proteins were transferred onto 0.2 µm Amersham™ Hybond® LFP PVDF Western blotting membranes (Merck cat# GE10600022) at 20 V for 1 hour using the Mini Blot Module (Thermo Fisher Scientific cat# B1000) as described by the manufacturer. Non-specific protein binding to the membrane was blocked by 1-hour incubation in 5% BSA (diluted in PBS) on a tube roller mixer. Membranes were rinsed with PBST (PBS + 0.1% Tween20) and

incubated in the dark with corresponding primary antibodies (Table 2.9) in PBST with 0.5% BSA. Membranes were rinsed twice and washed twice for 5 minutes with PBST and, when required, incubated in the dark with corresponding secondary antibodies (Table 2.10). Membranes were rinsed twice and washed three times for 5 minutes with PBST, rinsed three times with PBS and air-dried between two filter papers in a 37°C incubator for 30min. Fluorescent detection of antibodies was performed using an Amersham Typhoon NIR Plus Biomolecular Imager (GE Healthcare).

2.25 Statistical analysis

Experiments were analysed using GraphPad Prism versions 5–7 (GraphPad Software Inc.) and Microsoft Excel (Microsoft corporation).

2.26 Bioinformatic data processing and analysis

Bioinformatic analyses were performed by Dr Peter Keane (Prof Bonifer's lab).

2.26.1 Bulk RNA-Seq data analysis

Sequencing adaptors and low quality bases were trimmed from the raw RNA-Seq reads using Trimmomatic v0.32 (Bolger et al., 2014). The processed reads were then aligned to the human genome (version hg38) using Hisat2 v2.1.0 (Kim et al., 2015) with default settings. 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. Gene models from the RefSeq database (O'Leary et al., 2016) were used 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 normalized using the Limma package v3.26.9 (Ritchie et al., 2015) in R v3.5.1. The normalized 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 encyclopedia 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 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.

To carry out gene expression co-variance analysis, gene expression values were first transformed to Z-scores using the scale function in R. These were then hierarchically clustered using complete linkage of the Euclidean distances. Clusters corresponding to sets of genes with similar patterns of expression were then extracted from the dendrogram using the dynamicTreeCut package v1.63 (Langfelder et al., 2008) in R using the hybrid method with a minimum cluster size of 25 genes.

To compare the gene expression profile of the RUNX1-ETO induced cells to that of AML patients with the t(8;21) translocation, RNA-Seq data from t(8;21) patients and

from healthy peripheral blood stem cells (PBSCs) from *Assi et al*., 2019 was downloaded from GEO using the accession GSE108316. These data were aligned and processed as described above. The sets of genes that were up and down regulated in the RUNX1-ETO induced cells was then compared to the gene expression profiles of the t(8;21) AML cells and PBSCs using gene set enrichment analysis (GSEA) using the GSEA software (Subramanian et al., 2005).

2.26.2 ATAC-Seq data analysis

Single-end reads from ATAC-Seq experiments were processed to remove low-quality bases and sequencing adaptors using Trimmomatic. Reads were then aligned to the human genome (version hg38) using Bowtie2 v2.2.6 (Langmead and Salzberg, 2012) with the parameter --very-sensitive-local. Reads that aligned to the mitochondrial genome were removed from further analysis. Potential PCR duplicated reads were identified and removed from the alignments using Picard v2.10.5 (http://broadinstitute.github.io/picard). Open chromatin regions (peaks) were identified using MACS2 v2.1.1 (Zhang et al., 2008b) using the settings --nomodel --nolambda -B --trackline. 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, and then further annotated as either a promoter or distal element using the annoatePeaks.pl function in the Homer software package v4.9.1 (Heinz et al., 2010). A peak was annotated as being within a gene promoter if it was within 1.5 kb of a TSS and as a distal element otherwise.

ATAC 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 normalized 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 normalized 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 2 kb window centred on the peak summits was then calculated 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).

ATAC-Seq data from hematopoietic cell type in various stages of differentiation were obtained from *Corces et al.*, 2016 via GEO using the accession GSE74912. These data were aligned and processed as described above.

2.26.3 ChIP-Seq data analysis

Reads from ChIP-Seq experiments were processed, aligned to the human genome and de-duplicated in the same way described above for the ATAC-Seq data. Peaks from ChIP-Seq experiments targeting the transcription factors RUNX1 and RUNX1-ETO were identified using MACS2 with default settings. These peaks were then compared to the ATAC-Seq data, with only peaks that occurred within open chromatin regions being retained for further analysis. To identify differential binding of RUNX1 between the 0 and 5 Dox datasets, a union of RUNX1 peaks was first 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 normalized as counts per million in R. Peaks that had a fold-difference of at least 2 were considered to be differentially bound between experiments. RUNX1 and RUNX1-ETO target genes were identified by annotating each peak to its closest TSS using the annotatePeaks.pl function in Homer.

Peaks corresponding to the histone modifications H3K27ac and H3K4me3 were called using MACS2 with default settings. These peaks were then filtered against the hg38 blacklist and simple repeat tracks from the UCSC table browser to remove any potential artefacts.

2.26.4 Construction of average profiles

Average profiles for ATAC and ChIP-Seq data were constructed by first normalizing each of the alignment tracks as counts per million (CPM) using the bamCoverage function in deepTools v3.2.0 (Ramírez et al., 2016). These were then plotted using the plotProfile function in deepTools.

2.26.5 Single cell RNA-Seq data analysis

Illumina base call (BCL) files that were generated using the Chromium platform from 10x genomics were de-multiplexed and converted to the fastq format using the mkfastq function in CellRanger v2.1.1. These were then aligned to the human genome (version hg38) using the count function in CellRanger. Gene models from the RefSeq database were used as the reference transcriptome. Unique molecular identifier (UMI) counts were processed and normalized using the Seurat v2.3.4 package (Butler et al., 2018) in R. Cells with less than 1500 detectable transcripts, or that had more than 10% of UMIs aligned to mitochondrial genes were removed from further analysis. Additionally, transcripts that were detected in less than 20 cells were also excluded from analysis. The cell cycle stage for each cell was inferred using the CellCycleScoring function in Seurat. The possible effects of cell cycle stage, as well as sequencing depth (as measured by the total number of UMIs) per cell were removed from the analysis by linear regression using the ScaleData function in Seurat.

Clustering of cells was performed by first combining the datasets from the 0 and 5 dox treated cells into a single dataset using canonical correlation analysis (CCA). This combined dataset was then clustered using the t-distributed stochastic neighbour embedding (t-SNE) method. Cell clusters were identified using the FindClusters function in Seurat, using a resolution value of 0.4. Cell marker genes, corresponding to genes that are enriched on one cluster relative to others, were identified using the FindMarkers function. A gene was considered as a marker gene if it had a log fold-change value greater than 0.5 and could be detected in at least 50% of cells in that cluster. Differential gene expression analysis was also carried out for

each cluster using the FindMarkers function, with genes with a log-fold-change greater than 0.25 and an FDR < 0.05 being considered to be differentially expressed.

Cell trajectory (pseudo-time) analysis was carried out using Monocle v2.10.1 (Qiu et al., 2017; Trapnell et al., 2014). Normalized UMI counts from Seurat were first imported into Monocle using the importCDS function. Cells were then ordered along a pseudo-time trajectory using the discriminative dimensionality reduction with trees (DDRTree) method using the complete set of cell marker genes identified by Seurat to order the cells.

2.27 Tables of primers and antibodies

Table 2.1: Primers used for cloning

Oligonucleotide	Sequence	Orientation	Modifications/Notes	Description
SalI_HA-RUNX1	TACCGTCGACCCGCCATGTACCCATACGACGTCCCAGACTACGCTCGTATCCCCGTAGATGCCAGCACGA	Forward	NA	RUNX1-ETO cloning into AAVS1 plasmid
ETO_MluI	CGCAACGCGTCTACTAGCGAGGGGTTGTCTCTA	Reverse	NA	RUNX1-ETO cloning into AAVS1 plasmid
KRAS(Nter) inFusion into SalI	GCTTTGAGATTCTGGAGCGGCGAATGACTGAATATAAACTTGTGGT	Forward	NA	KRAS InFusion cloning downstream DHFR domain
KRAS(Cter) inFusion into SalI	GAGAGAGGGGCGGATCCGTCTTACATAATTACACACTTTG	Reverse	NA	KRAS InFusion cloning downstream DHFR domain
BamHI-FLAG-STOP	GATCCGACTACAAGGACGACGATGACTAA	Forward	5' Phosphorylation	FLAG tag + STOP insertion into BamHI site downstream DHFR domain
BamHI-STOP-FLAG	GATCTTAGTCATCGTCGTCCTTGTAGTCG	Reverse	5' Phosphorylation	FLAG tag + STOP insertion into BamHI site downstream DHFR domain
KIT(Nter)_inFusion into AscI	TGGAGGAGAATCCTGGCCCGATGAGAGGCGCTCGCGGCGC	Forward	HPLC purified	KIT InFusion cloning upstream DHFR domain
KIT(Cter)_inFusion into AscI	GCAATCAGACTGATCATTTGGCCGCCGACATCGTCGTGCACAAGCA	Reverse	HPLC purified	KIT InFusion cloning upstream DHFR domain
hGMCSF intron 1_KIT (Nter)	GAGAATGAAAGTAATATCAGGTAAGTGAGAGAATGTGGGC	Forward	HPLC purified	InFusion cloning of a STOP-containing INTRON between two exons of KIT
hGMCSF intron 1_KIT (Cter)	GTTAGATGAAGTTCACTTACGTATCTGTAGAAAAGGAAAATGTC	Reverse	HPLC purified	InFusion cloning of a STOP-containing INTRON between two exons of KIT

Table 2.2: Primers for genomic DNA used in transgene screening assays

Oligo name (binding)	Orientation	Sequence
AAVS1 5' screen (AAVS1 5')	Forward	GGACCACTTTGAGCTCTACT
AAVS1 5' screen (T2A)	Reverse	TCCACGTCACCGCATGTTAG
AAVS1 3' screen (TET3G)	Forward	TGCCTGCTGACGCTCTTGACGATT
AAVS1 3' screen (AAVS1 3')	Reverse	GAAGGATGCAGGACGAGAAA
Wild-Type screen (AAVS1 5')	Forward	CCCCTATGTCCACTTCAGGA
Wild-Type screen (AAVS1 3')	Reverse	CAGCTCAGGTTCTGGGAGAG
T2AScreen 3.1	Forward	CCGCATGTTAGAAGACTTCCTCTG
GapScreen_5.2	Reverse	CCACTAGGCGCTCACTGTTCTCTC

Table 2.3: Conjugated antibodies used for single cell sorting

Antibody	Fluorochrome	Manufacturer	Catalogue #	Stock	Dilution
CD9	PE	BD Pharmingen	555372	100 tests	1:100
CD34	PeCy7	BioLegend	343516	100 µg/ml	1:100
CD45	BV-421	BioLegend	304032	25 µg/ml	1:50
CD90	APC	BD Pharmingen	559869	0.2 mg/ml	1:100
CD326 (EpCam)	BV-421	BioLegend	324220	100 tests	1:30

Table 2.4: Conjugated antibodies used for flow cytometry

Antibody	Fluorochrome	Manufacturer	Catalogue #	Stock	Dilution
CD16	PeCy7	BioLegend	302015/302016	200 µg/ml	1:40
CD31	APC	BioLegend	303115	80 µg/ml	1:50
CD34	BV-421	BioLegend	343609	25 tests	1:50
CD34	PeCy7	BioLegend	343516	100 µg/ml	1:100
CD38	APC	BD Pharmingen	555462	100 tests	1:50
CD43	BV-421	BD HORIZON	562916	100 tests	1:50
CD45	BV-421	BioLegend	304032	25 µg/ml	1:50
CD90	APC	BD Pharmingen	559869	0.2 mg/ml	1:100
CD90	BV-421	BioLegend	328121	200 µg/ml	1:100
CD144 (Ve-Cadherin)	APC	BioLegend	348508	100 tests	1:10

Table 2.5: Primers used for RT-pPCR gene expression analysis from total RNA

Primers for cDNA	Forward	Reverse
GAPDH	CCTGGCCAAGGTCATCCAT	AGGGGCCATCCACAGTCTT
RUNX1 (Cter)	CCCTCAGCCTCAGAGTCAGAT	AGGCAATGGATCCCAGGTAT
RUNX1 (Runt Domain)	AACAAGACCCTGCCCATCGCTTTC	CATCACAGTGACCAGAGTGCCAT
RUNX1-ETO junction	TCAAAATCACAGTGGATGGGC	CAGCCTAGATTGCGTCTTCACA
DHFR-KRAS	AATTCCACGATGCTGATGCG	CAAGGCACTCTTGCCTACGC
KIT-DHFR	TCAATTCTGTCGGCAGCACC	CATGGCGTTTTCCATGCCGA
RUNX1-EVI1 junction	CCACAGAGCCATCAAAATCA	TCTGGCATTTCTTCCAAAGG
EVI1 exon 7	AAACCTTTGCCGTCATAAGCG	CGTAGTGCTGAACATTTGTCCACAG

Table 2.6: Taqman probes used for gene expression analysis from total RNA

Name	Probe
GAPDH	Hs99999905_m1
GATA1	HS00231112_m1
GFI1	Hs00382207_m1
GFI1B	Hs01062469_m1
PU.1	HS00231368_m1
RUNX1 COMMON	HS00231079_m1
RUNX1C	Hs01021967_m1
RUNX1T1	Hs00231702_m1

Table 2.7: Primers for validating ATAC

Target	Forward	Reverse
TBP promoter	CTGGCGGAAGTGACATTATCAA	CCCGACCTCACTGAACCC
Chromosome 18	AGGTCCCAGGACATATCCATT	GTTCAAATTGTGTTTTGTGGTTA
β-actin	GCAATGATCTGAGGAGGGAAGGG	AGCTGTCACATCCAGGGTCCTCA

Table 2.8: Primers for human genomic DNA used for ChIP-qPCR enrichment

Primers for gDNA	Forward	Reverse
IVL	GCCGTGCTTTGGAGTTCTTA	CCTCTGCTGCTGCCACTT
PU.1 -14kb enhancer	AACAGGAAGCGCCCAGTCA	TGTGCGGTGCCTGTGGTAAT
CSF1R FIRE enhancer	GCCTGACGCCAACAATGTG	GGCAAAGGAGGGAAGTGAGAG
NFE2 2Enhancer	AATAGCGAGGCCCCTCTTAG	ACCCAAACTGGAACACAA GG
RUNX1 39/40	GATACCGGAAAGGCCTGTGA	AGTGCCTGGAAATGAACGT

Table 2.9: Primary antibodies used for Immunoblotting

Antibody	Fluorochrome	Manufacturer	Cat #	Immunoglobulin	[mg/ml]	Dilution
FLAG tag	APC	BioLegend	637308	Rat IgG2a	0.2	1:1000
KRAS	unconjugated	Abcam	ab206969	Rabbit monoclonal	1.175	1:1000
GAPDH	unconjugated	Novus Biologicals	NB300-221	Mouse monoclonal	1	1:1000
CD117 (c-KIT)	APC	BioLegend	313206	Mouse IgG1	100 tests	1:100

Table 2.10: Secondary antibodies used for Immunoblotting

Fluorochrome	Manufacturer	Catalogue #	Immunoglobulin	[mg/ml]	Dilution
Alexa Fluor 488 anti-mouse	Thermofisher Scientific	A-21202	Donkey anti-Mouse IgG	2	1:1000
Alexa Fluor 488 anti-rabbit	Thermofisher Scientific	A-11034	Goat anti-Rabbit IgG	2	1:1000
Alexa Fluor 594 anti-rabbit	Thermofisher Scientific	A-11012	Goat anti-Rabbit IgG	2	1:1000

3 RESULTS

3.1 Generation of inducible RUNX1-ETO human ES cell lines

Human pluripotent stem cell models with conditional expression of RUNX1-ETO have been previously generated (Mandoli et al., 2016). However, the *in vitro* haematopoietic differentiation systems used remained skewed towards the differentiation of blood progenitors resembling those generated in the embryonic yolk sac. For this reason, we generated an inducible RUNX1-ETO human ES cell line and used a novel method to obtain progenitors resembling those generated in the embryonic AGM, as previously described (Ng et al., 2016).

In order to efficiently track the haematopoietic lineage differentiation, a dual reporter SOX17$^{mCHERRY/w}$ RUNX1C$^{GFP/w}$ human H9 ES cell line – wherein "w" refers to a wild-type allele – was previously generated in Prof Elefanty's lab (Ng et al., 2016) This tracer cell line carries a *mCHERRY* gene in the exon 1 of *SOX17* locus and a *GFP* gene in the exon 1 of *RUNX1* locus, resulting in expression of GFP from the distal promoter (*RUNX1C*) (Figure 3.1). *SOX17* expression marks the haemogenic endothelium and it is required to generate HSCs from the embryonic AGM in the mouse (Clarke et al., 2013; Kim et al., 2007). Despite its critical role in development of the haemogenic endothelium, SOX17 downregulation is required during the EHT to allow haematopoietic differentiation (Clarke et al., 2013; Nakajima-Takagi et al., 2013; Nobuhisa et al., 2014). *RUNX1C* expression is upregulated compared to the RUNX1B isoform in definitive haematopoietic progenitors, which coincides with the loss of the endothelial phenotype, making RUNX1C the dominant RUNX1 isoform in foetal liver blood progenitors (Sroczynska et al., 2009). In contrast to *RUNX1B*,

RUNX1C expression is restricted to haematopoietic cells and defines a subset of CD34+ cells with clonogenic and bone marrow homing activity (Ng et al., 2016). It has been shown that one functional *RUNX1C* isoform allele is sufficient for functional definitive progenitors (Ng et al., 2016; Sroczynska et al., 2009) and that disturbance of both alleles only modestly affects haematopoiesis (Sroczynska et al., 2009). Likewise, heterozygous deletion of Sox17 does not yield any adverse phenotype (Kim et al., 2007). Therefore, the SOX17$^{mCHERRY/w}$ RUNX1C$^{GFP/w}$ human H9 ES cell line allows to easily and safely track the endothelial (*SOX17*, mCHERRY) to haematopoietic (*RUNX1C*, GFP) transition and hence was selected as parental cell line for our RUNX1-ETO-targeting approach.

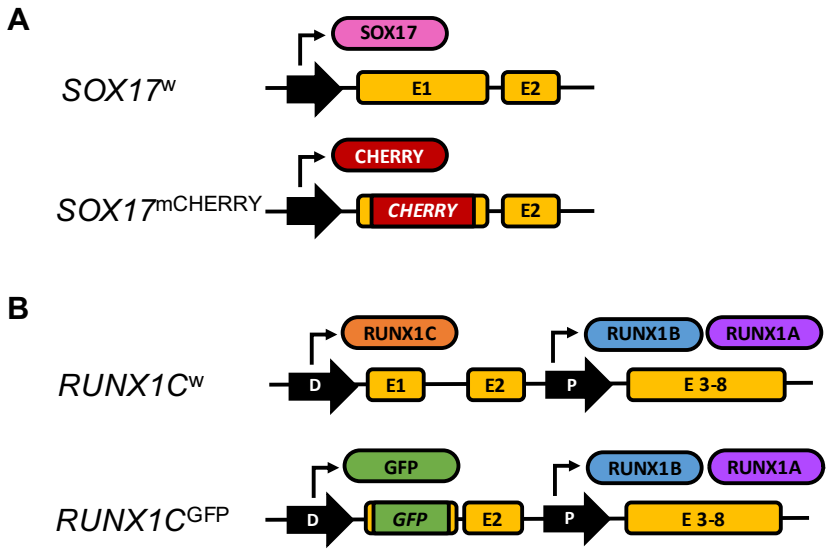


Figure 3.1: Schematic representation of the targeted alleles in the human H9 ES cell dual reporter line

(A) Wild-type and targeted alleles in the *SOX17* locus, with mCHERRY sequence inserted into exon 1. (B) Wild-type and targeted alleles in the *RUNX1* locus. GFP sequence was inserted into exon 1, resulting in expression of GFP from the distal (D) promoter and RUNX1B and RUNX1A isoforms from the proximal (P) promoter within the same allele. Promoters and exons are shown with black arrows and yellow boxes, respectively. Protein products generated from each allele is represented with ovals.

The AAVS1 locus displays an open chromatin conformation and contains insulator elements, thus allowing robust and unperturbed transgene expression (Lamartina et al., 2000; Lombardo et al., 2011; Ogata et al., 2003). For this reason, *RUNX1-ETO* was targeted to the AAVS1 locus of the dual reporter cell line. To achieve inducible activation of the oncogene, RUNX1-ETO was cloned into vectors containing a doxycycline (Dox) regulated promoter. For ease of detection of *RUNX1-ETO* in downstream assays, an HA tag was fused to its N-terminal domain. Vectors also included a puromycin cassette to select successfully targeted clones and homology arms to the human AAVS1 locus, allowing recombination of the cassette after linearization. *RUNX1-ETO*-containing vectors were co-transfected with a pair of previously validated transcription activator-like effector nucleases (TALENs) (Qian et al., 2014) into the dual reporter parental cell line (Figure 3.2). Targeted clones were selected with puromycin and subsequently subjected to single-cell sorting. In order to screen for transgene targeting into the AAVS1 locus, primer pairs were designed amplifying the 5' AAVS1 integration site, as shown in Figure 3.3A. Primer pairs were also designed spanning the AAVS1 integration site to distinguish double from single allele targeting, yielding a PCR amplifiable product only in presence of a residual wild-type allele. TALEN-mediated integration into the AAVS1 locus showed highly efficient double allele targeting of the *RUNX1-ETO* transgene (Figure 3.3B). Selected clones were re-screened for complete integration of the transgene by PCR amplification of the 3' vector-AAVS1 integration junction. In order to avoid the selection of clones harbouring aneuploidies, the molecular karyotype was evaluated by single nucleotide polymorphism (SNP) chromosomal microarray (Illumina Infinium

CoreExome-24 v1.1, 0.50Mb resolution), performed by Victorian Clinical Genetics Services Ltd. No aneuploidies were detected in any of the clones.

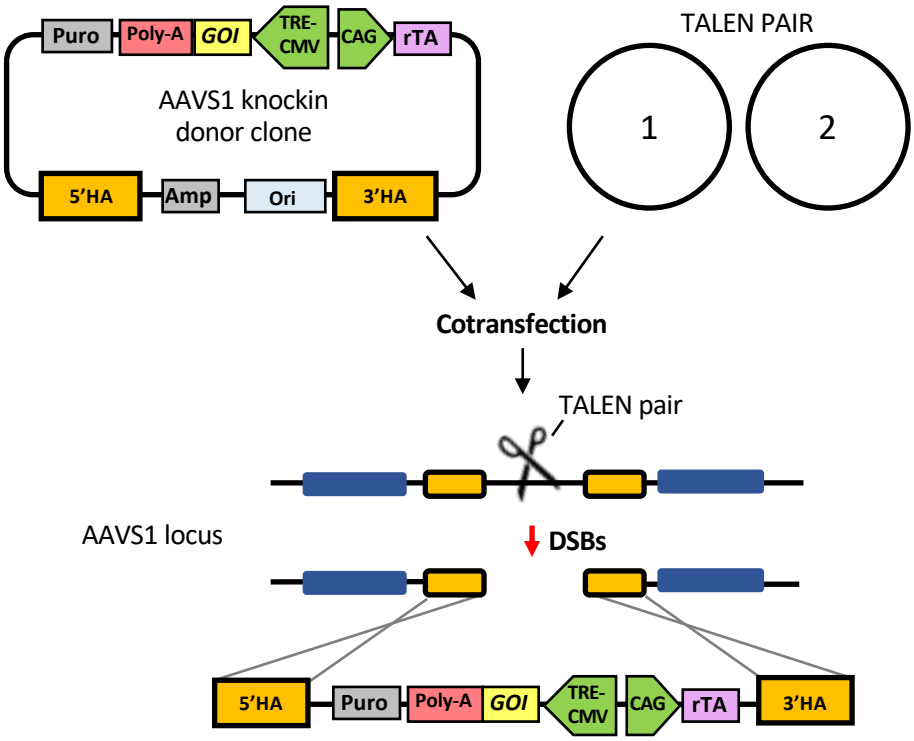


Figure 3.2: Schematic representation of the gene targeting approach

Efficient knock-in of the selected oncogenes into the AAVS1 locus of the SOX17$^{mCHERRY/w}$RUNX1C$^{GFP/w}$ human ESC line H9 was performed as shown above. Genes of interest (GOI) were previously cloned into knock-in plasmids containing a puromycin resistance gene within two AAVS1 homology arms (HA). Knock-in donor plasmids were co-transfected with the TALEN pair and were integrated into the genome via homology recombination with the AAVS1 locus after a TALENs-mediated double strand break (DBS).

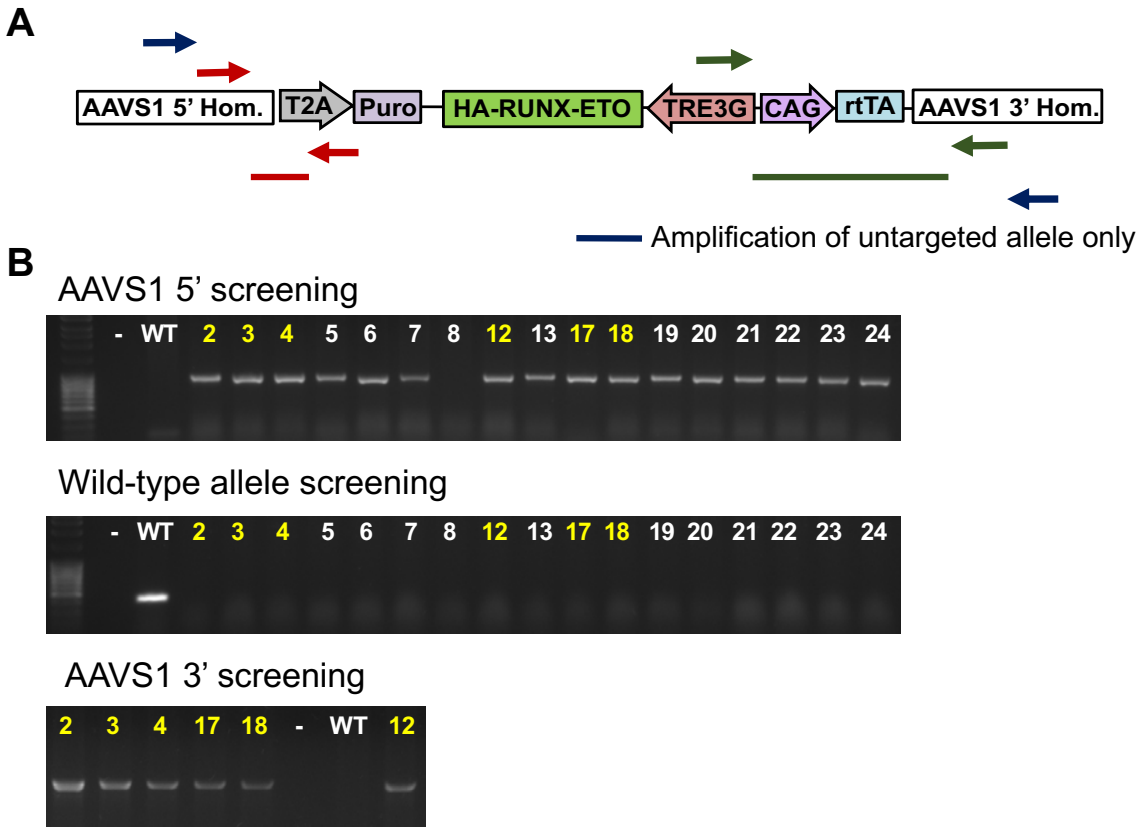


Figure 3.3: Efficient targeting RUNX1-ETO into the AAVS1 human locus

PCR screening to identify targeted clones. (A) Schematic representation of the transgene knock-in strategy into the AAVS1 safe harbour locus, showing the binding of AAVS1 5' (red) and AAVS1 3' (green) primer pairs, which amplify correctly targeted alleles, and of wild-type (blue) primer pair, which amplifies residual untargeted alleles. (B) Agarose gels showing PCR products corresponding to the amplification of the AAVS1 5' breakpoint (956 bp), wild-type allele (500 bp) and AAVS1 3' breakpoint (2,067 bp). Water was used as a negative control (-) and an AAVS1-untargeted cell line was used as a positive control (WT) for the wild-type allele amplification. Selected clones are coloured in yellow.

In order to prove efficient induction of the transgene expression and translation, targeted ESC clones were treated with 1 μg/ml Dox for two days and subjected to single-cell protein detection assays. Intracellular flow cytometry using a monoclonal anti-HA-tag DyLight® 650-conjugated antibody showed expression of RUNX1-ETO after Dox treatment in a pool of targeted cells (Figure 3.4A). RUNX1-ETO expression was also detected upon induction in all the selected single-sorted clones by

immunofluorescence assays, confirming homogeneous expression of the transgene within our cell populations (Figure 3.4B).

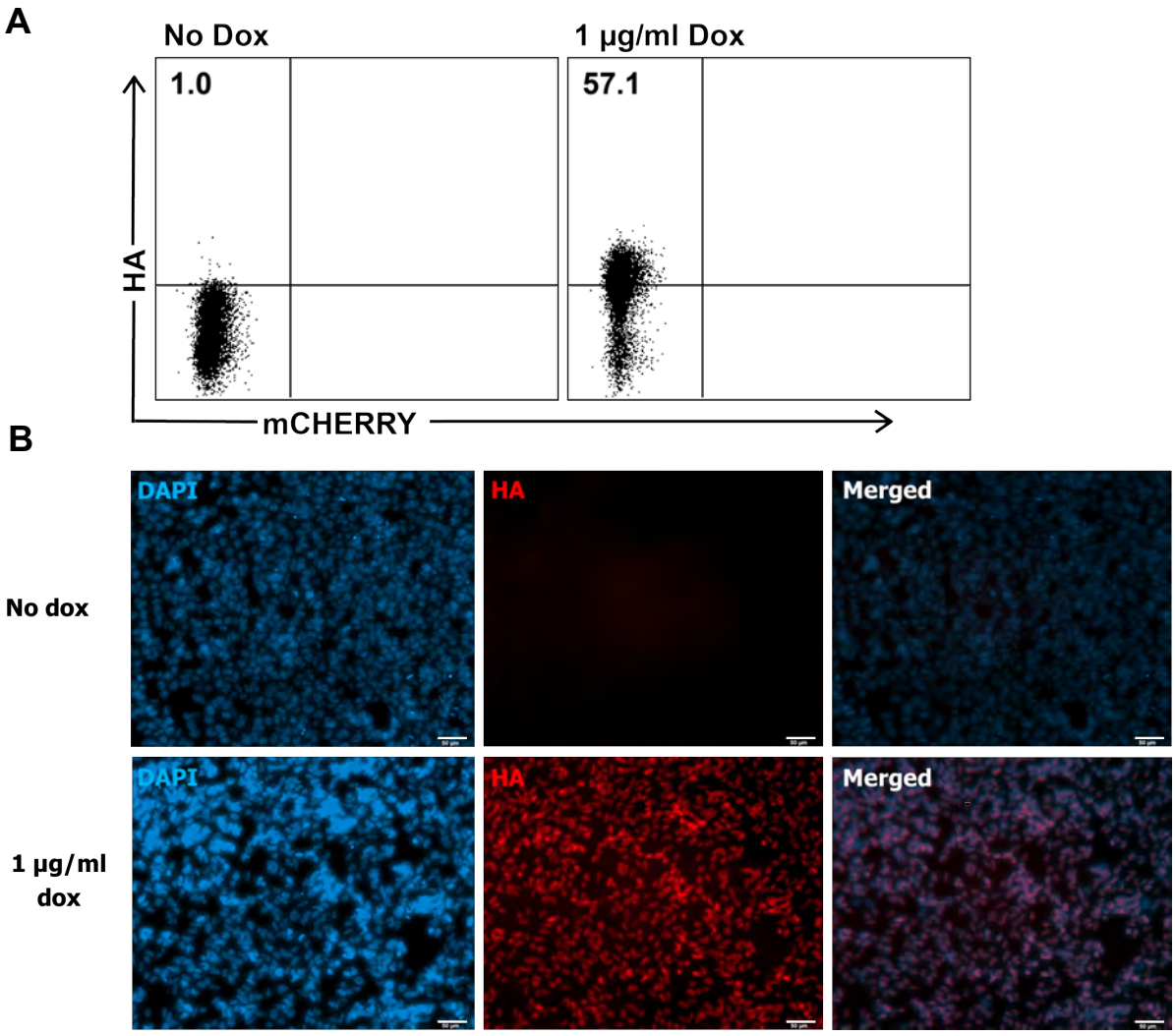


Figure 3.4: Targeted clones showed Dox-inducible *RUNX1-ETO* expression from the AAVS1 locus

Detection of RUNX1-ETO using an anti-HA tag DyLight® 650-conjugated antibody upon addition of 1 µg/ml Dox for two days evaluated by (A) flow cytometry analysis in a pooled population of puromycin-resistant cells after transfection and (B) immunofluorescence assay of clone #18. Immunofluorescence images are representative of six single-sorted clones. Cell nuclei are stained with DAPI (blue) and cells expressing HA-RUNX1-ETO are detected through the anti-HA antibody (red). Fluorescence channels are merged on right panels. Scale bar: 50 µm.

3.2 RUNX1C+ blood progenitors are generated from SOX17+ haemogenic endothelium

Recently, *Ng et al.* have developed a protocol that generates definitive progenitors from hPSCs, which rely in the formation of *HOXA*-expressing cultures resembling human AGM haematopoiesis in a dish (Ng et al., 2016). This protocol uses whole EBs in an adherent culture, allowing the formation of SOX17+ vascular structures and subsequent generation of RUNX1C+ haematopoietic progenitors. Aiming to better reproduce the first events leading to a t(8;21) pre-leukaemia in the embryo, we differentiated our RUNX1-ETO cell line following the method developed in Prof. Elefanty's lab (Ng et al., 2016). The experimental time course, culture conditions including non-adherent spin culture (or the formation of EBs) and adherent culture (for the generation of blood progenitors), equivalent embryonic developmental stages and cytokines used in culture are summarized in Figure 3.5. Importantly, mesoderm patterning to a definitive haematopoietic programme occurs upon addition of the Activin-inhibitor SB and the Wnt-agonist CHIR from day 2 to 4 (Ng et al., 2016). At this stage (d4), EBs appear round with a very defined edge (Figure 3.5, d6 image).

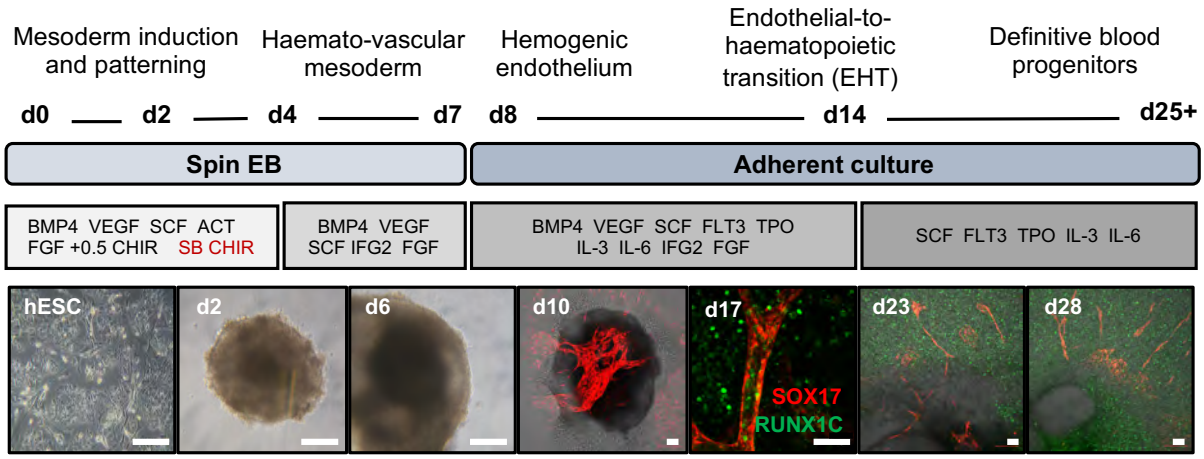


Figure 3.5: Time course of the *in vitro* human definitive haematopoietic differentiation

Diagram showing the time course of the *in vitro* human definitive haematopoietic differentiation as spin culture (for the formation of EBs) and adherent culture (for the generation of blood progenitors). Developmental stages as well as cytokines and factors are indicated. Epifluorescence (hESC, d2 and d6) and confocal images (d10-d28) of the cultures as representatives of each stage are shown. Human Embryonic Stem cells (hESC) appear at 50% confluency on a feeder layer. EBs appear as opaque round structures, surrounded by adherent cells (including stroma, endothelium and the first blood progenitors) from d10. Fluorescence and brightfield channels are merged in images corresponding to d10, d13 and d28. Scale bar: 100 µm. SOX17 (mCHERRY, red) expression marks vascular structures and RUNX1C (GFP, green) marks haematopoietic progenitors.

We were able to visually monitor the formation of SOX17+ (CHERRY) vascular structures and RUNX1C+ (GFP) haematopoietic progenitors thanks to the fluorescent markers of our dual reporter cell line. A selection of epifluorescence and confocal images representing the main phenotypic features of the time course during the *in vitro* human definitive haematopoietic cultures is represented in Supplementary Figure 1. By d8, EBs started expressing the vascular SOX17 (CHERRY) and the stem-like CD34 marker but lacked surface expression of the CD45 haematopoietic marker. During adherent cultures, EBs generate a stromal cell layer of vasculature, containing arterial cells, and haemogenic endothelium, as shown in Figure 3.6A. After the EHT (around d14), RUNX1C+ progenitors started to emerge from the EBs and SOX17+ haemogenic endothelium – including the inner side of the vasculature,

resembling progenitor formation within the AGM – forming cell clusters (Figure 3.6B). In our experiments, cytokines in the media were then changed to a 5-factor cocktail including SCF, FLT3, TPO, IL6 and IL3, given that this combination supports the growth of RUNX1-ETO-expressing CD34+ cord blood cells (Mulloy et al., 2002). Subsequently, haematopoietic progenitors detached from the endothelium and started populating the culture within the following days (Figure 3.7). Generation of haematopoietic progenitors continued until about day 28 (Figure 3.5, d28 image), which further differentiated to more mature stages.

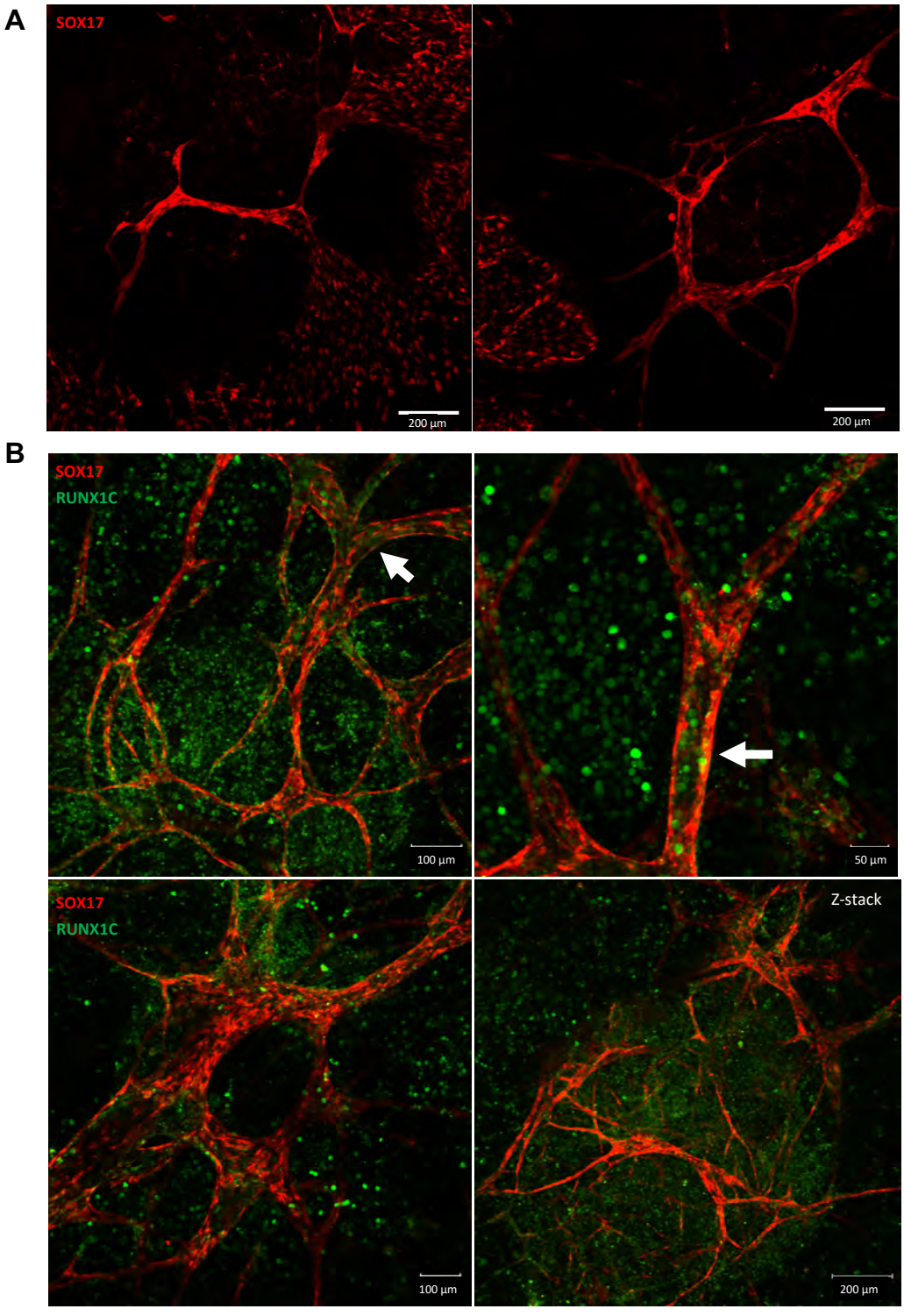


Figure 3.6: RUNX1C+ haematopoietic progenitors emerge from cell clusters located within vascular structures of SOX17+ hemogenic endothelium after the EHT

Confocal images at d12 (A) and d17 (B) of the adherent EB cultures. (A) SOX17+ hemogenic endothelium and vascular structures formed before the EHT. (B) Generation of RUNX1C+ blood progenitors from SOX17+ haemogenic endothelium. Bottom right panel shows an image as combined layers of Z-stack capture. Arrows point intra-vascular haematopoietic cell clusters mimicking the structures observed in the embryonic AGM. Scale bar: 200 µm, 100 µm or 50 µm as indicated. SOX17 (mCHERRY, red) and RUNX1C (GFP, green).

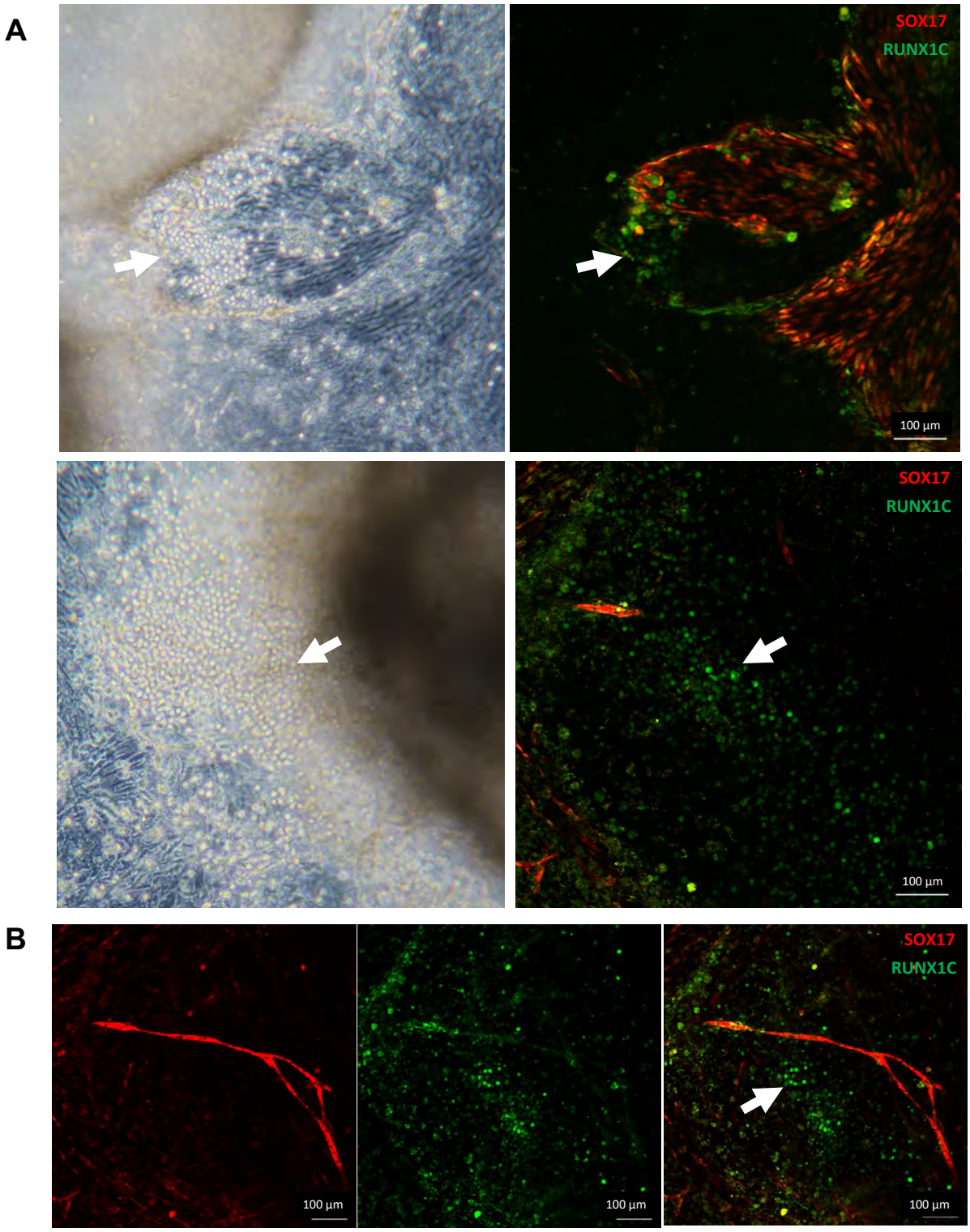


Figure 3.7: RUNX1C+ haematopoietic progenitors detach from the hemogenic endothelium and populate the culture

(A) Brightfield (left) and confocal (right) microscopy images. (B) Confocal single-layer (right and middle panels) and combined layers (left panel) of Z-stack images. Arrows point cell clusters of RUNX1C+ haematopoietic progenitors. Scale bar: 100 µm. SOX17 (mCHERRY, red) and RUNX1C (GFP, green).

3.3 High dosage of RUNX1-ETO reorganizes the vascular structures in a dose-dependent manner and blocks blood formation

In order to evaluate the effect of RUNX1-ETO during embryonic development, a preliminary assay was performed where EB cultures were treated with 0, 100 and 500 ng/ml Dox at d15 of differentiation and analysed at multiple time points thereafter. Half of the media was exchanged to add fresh Dox every three days. Cultures were carefully tracked during haematopoietic differentiation using fluorescence microscopy to visualize the fluorescent reporters. Wild-type EB cultures at d14, before the induction of RUNX1-ETO, presented an extensive SOX17+ (CHERRY+) endothelium, organised into a combination of vascular structures and endothelial cell monolayers, and few forming blood progenitors around the edges of the EBs (Figure 3.8A). Induction of RUNX1-ETO from d15 caused a disorganization of the vascular structures in a dose-dependent manner (Figure 3.8B); induction using 100 ng/ml Dox caused a reduction in SOX17+ (CHERRY+) cell monolayers and a gain in vessel thickness and higher Dox concentrations (500 ng/ml) resulted in aberrant aggregates of cell monolayers and further disorganization of the vessels. In the untreated control, RUNX1C+ blood progenitors appeared from around d15 and were abundant by d24. In contrast, high RUNX1-ETO expression from d15 completely abrogated blood formation (Figure 3.8C). Overall, high RUNX1-ETO dosage disrupts the organization of SOX17+ endothelial structures and abrogates the formation of haematopoietic progenitor cells.

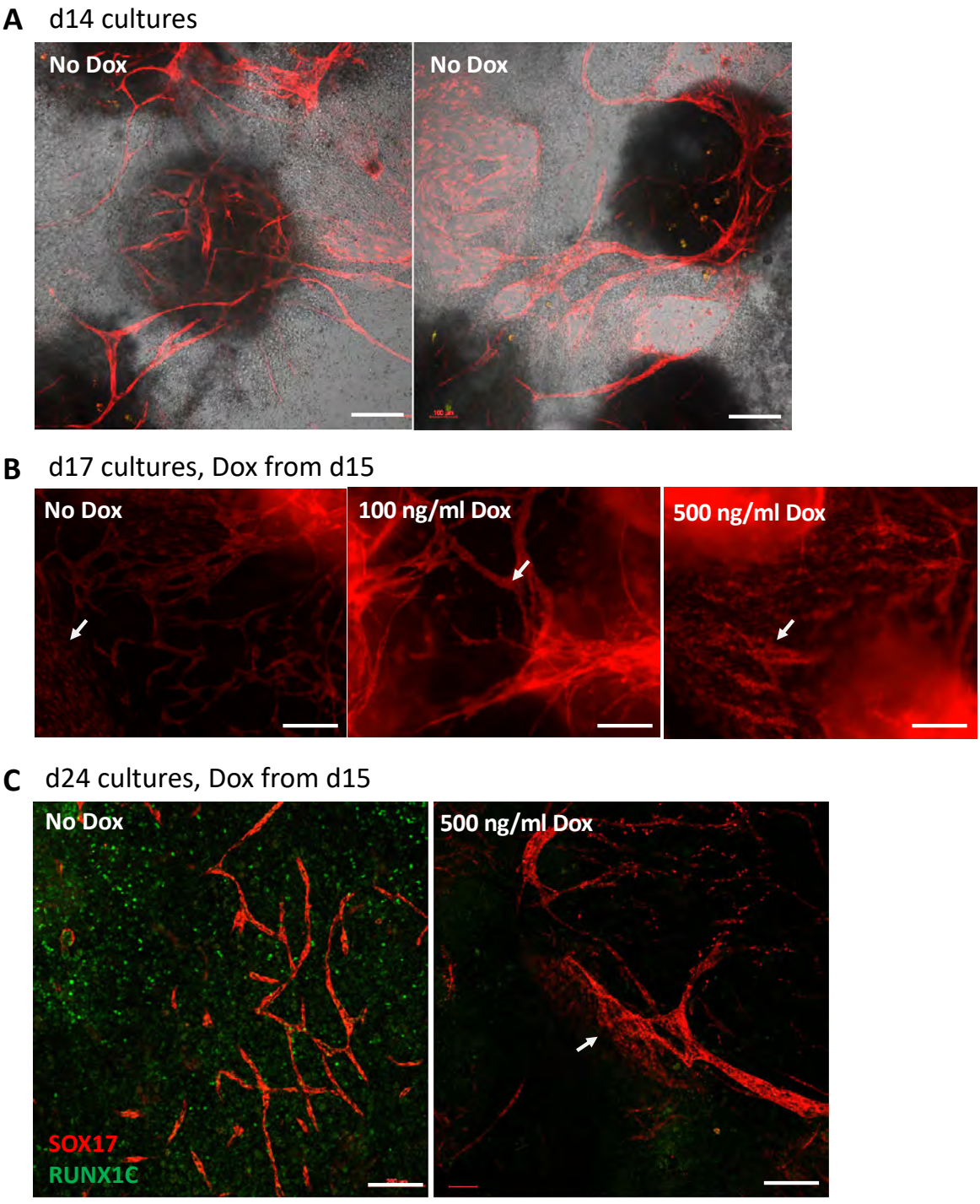


Figure 3.8: High dosage of RUNX1-ETO reorganizes the vascular structures in a dose-dependent manner and blocks blood formation

(A) Confocal images of wild-type EB cultures at d14 of haematopoietic differentiation. EBs appear as opaque round structures surrounded by a stromal layer containing SOX17+ vessels and monolayer endothelial structures. Fluorescence and brightfield channels are merged. (B) Epifluorescence microscopy images of d17 EB cultures upon 2-day treatment with 100 or 500 ng/ml Dox. Arrows point changes in SOX17+ structures upon induction of RUNX1-ETO. (C) Confocal images of d24 EB cultures wild-type and treated with 500 ng/ml Dox from d15 of differentiation. Scale bar: 200 µm. SOX17 (mCHERRY, red) and RUNX1C (GFP, green).

Small numbers of blood progenitors with multilineage capacity and abnormal morphology are present in the foetal liver of RUNX1-ETO knock-in mice (Okuda et al., 1998; Yergeau et al., 1997). Therefore, the lack of blood formation upon *RUNX1-ETO* expression suggested that induction at those levels is toxic for minimal blood cell development in general, and hence for the formation of a pre-leukaemic clone. Previous studies have proven a requirement for a delicate balance between the levels of *RUNX1-ETO* and endogenous *RUNX1* expression for the maintenance of the leukaemic phenotype and cell growth and survival (Ben-Ami et al., 2013; Loke et al., 2017; Ptasinska et al., 2014). For this reason, Dox concentrations ranging from 0 to 100 ng/ml were titrated on d25 mixed haematopoietic progenitors and *RUNX1-ETO* and *RUNX1* mRNA expression was evaluated. *RUNX1-ETO* transcripts could not be detected by qPCR using concentrations below 1 ng/ml Dox. However, *RUNX1-ETO* expression exceeded that of the endogenous *RUNX1* by using concentrations over 10 ng/ml (Figure 3.9A). Addition of 1 μg/ml Dox to untargeted human H9 ES cell line had been previously tested by another PhD student in the lab and proved safe for cell viability and differentiation (data not shown). Thus, a negative control using the parental cell line was not performed in our assays, given the lower working concentrations of Dox used in our experiments as well as the high cost of the differentiation protocol.

Further evaluation of the induction levels of RUNX1-ETO was performed in three different CD34+ haematopoietic progenitors sorted on SOX17 and RUNX1C (S+R-, S-R- and S+R-). Upon sorting, populations of CD34+ progenitors were co-cultured on an AKT-expressing human umbilical vein endothelial stromal layer, to support their growth, and RUNX1-ETO was induced with 3, 5 or 10 ng/ml Dox thereafter. These

studies further demonstrated that both 3 and 5 ng/ml Dox yielded levels of *RUNX1-ETO* transcripts comparable to those of endogenous *RUNX1*. In line with the previous experiment, 10 ng/ml Dox resulted in higher levels of *RUNX1-ETO* expression (Figure 3.9B). Therefore, 5 ng/ml was selected as the optimal Dox concentration and used thereafter. The 3 and 10 ng/ml Dox concentrations were kept as indicators of the effect of a lower and higher RUNX1-ETO dosage.

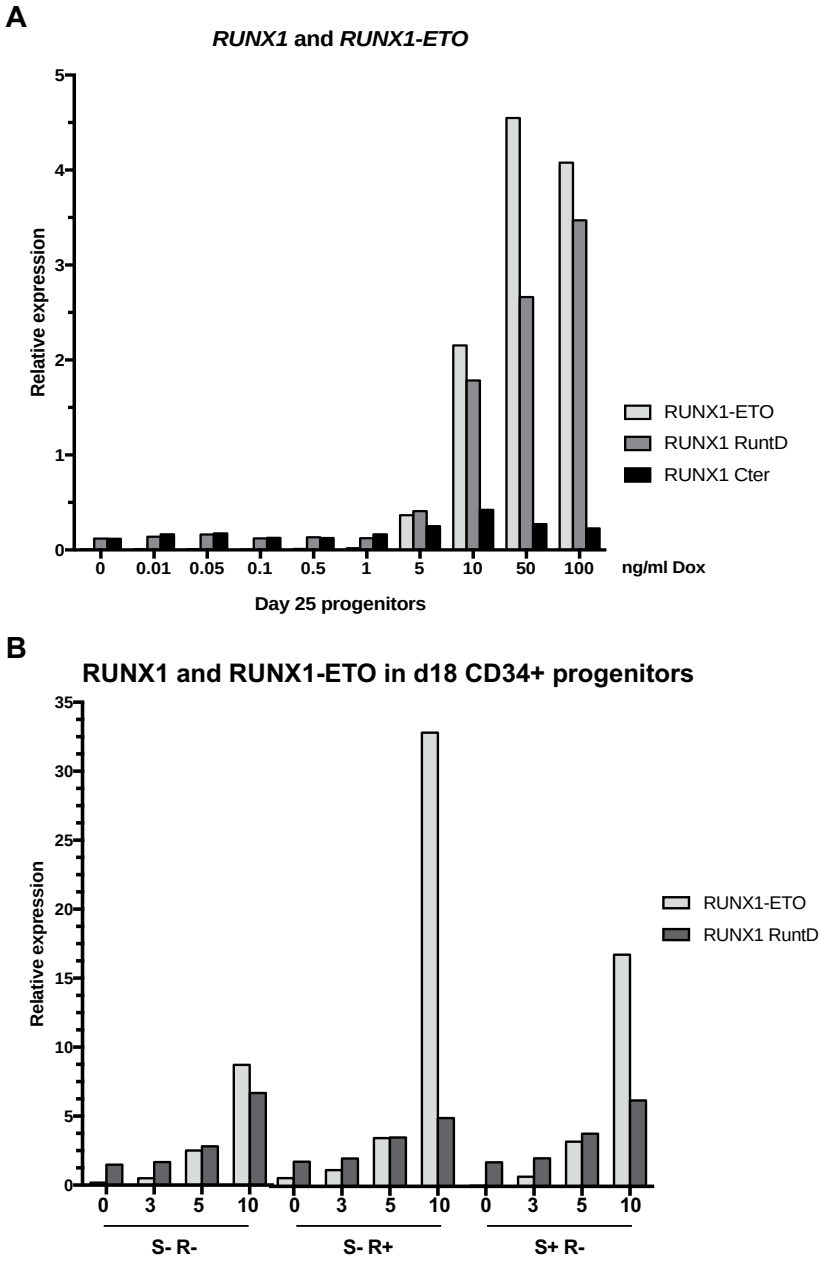


Figure 3.9: Balanced *RUNX1-ETO* levels are achieved using a low Dox concentration

(A) *RUNX1-ETO* and *RUNX1* gene expression in response to Dox titration on d25 haematopoietic progenitors. Primers were designed to amplify: the translocation breakpoint (RUNX1-ETO), the DNA-binding domain present in both gene products (RUNX1 RuntD) and the Carboxy-terminal present only in endogenous RUNX1 (RUNX1 C-ter). Gene expression was normalized to that of GAPDH. N=1.

(B) *RUNX1-ETO* gene expression in response to Dox titration on SOX17 and RUNX1C sorted populations of CD34+ haematopoietic progenitors. Three CD34+ populations were sorted at d18 depending on SOX17 (marked as S) and RUNX1C (marked as R) expression and cultured on AKT-expressing human umbilical vein endothelial stromal layer. Primers were designed to amplify: the translocation breakpoint (RUNX1-ETO), the DNA-binding domain present in both gene products (RUNX1 RuntD) and the Carboxy-terminal present only in endogenous RUNX1 (RUNX1 C-ter). Gene expression was normalized to that of endogenous RUNX1 instead of GAPDH in order to exclude the stromal cells from the analysis.

3.4 Expression of *RUNX1-ETO* at balanced levels before the EHT disrupts vasculature and prevents blood formation

Studies on *in vitro* differentiation of mouse ES cells have described a developmental-stage-specific effect on blood cell development upon RUNX1-ETO induction, highlighting the importance for the oncogenic hit to appear in the right target cell (Regha et al., 2015). As described in the previous section, balanced levels of RUNX1-ETO to those of endogenous RUNX1 in our *in vitro* system can be achieved by using 5 ng/ml Dox. Based on these observations, our experimental strategy was designed to evaluate: (i) to which extent blood differentiation was affected by inducing RUNX1-ETO at balanced levels and (ii) whether the phenotype was affected by the time of Dox treatment. This was achieved using a series of experiments that involved a 7-day Dox treatment starting at different time points of the adherent EB cultures, as outlined in Figure 3.10.

Figure 3.10: Diagram of the experimental strategy of RUNX1-ETO induction for 7 days

Experimental outline used to evaluate the effects of different levels of RUNX1-ETO induction at several time points during the *in vitro* haematopoietic differentiation. Dox (0, 3, 5 or 10 ng/ml) was added at different time points after the adherent EB culture was set up. The floating haematopoietic cell fraction was subsequently harvested 7 days after Dox addition, and used for several assays including flow cytometry analysis, colony forming unit (CFU) assays in methylcellulose, replating assays and mRNA sample collection for gene expression analysis.

Maintenance of the desired RUNX1-ETO induction upon long-term Dox treatment at multiple time points was evaluated by analysis of gene expression levels from total RNA. As expected, *RUNX1-ETO* expression levels were strictly dependent on Dox dosage regardless of the induction time point during differentiation (Supplementary

Figure 2). Importantly, *RUNX1-ETO* expression was undetectable in the untreated samples as compared to an untargeted control cell line, hence indicating that transgene expression was not leaky in the absence of the inducer.

Prior studies using a mouse ES cell *in vitro* differentiation system have described a block of the EHT upon RUNX1-ETO induction (Regha et al., 2015). We therefore wanted to evaluate to which extent this holds true in our human ES cell system of definitive haematopoiesis. To this end, we induced RUNX1-ETO before the EHT (at d10) using 3, 5 or 10 ng/m Dox and evaluated phenotypic changes by confocal microscopy 4- or 6-days post-induction. We observed a Dox-dependent overall reduction of the number of blood progenitors, which also appeared phenotypically abnormal as compared to the wild-type conditions (Figure 3.11, Supplementary Figure 3). Wild-type cultures (no Dox) presented RUNX1C+ progenitors emerging from the EB as individual blood cells by d14 (Figure 3.11A) and/or forming cell clusters from hemogenic endothelium by d16 (Figure 3.11B). However, cultures treated with the low Dox concentration (3 ng/ml) generated reduced number of progenitor cells, which also lacked *RUNX1C* expression (Figure 3.11A). By using 10 ng/ml Dox, progenitor-resembling cells were attached to the endothelium forming big clusters and showing *SOX17* expression by d14 (Figure 3.11A), which should be downregulated at the progenitor stage, and co-expression of both *SOX17* and *RUNX1C* (appearing yellow) by d16 (Figure 3.11B). Moreover, induction of RUNX1-ETO not only affected the generation of blood progenitors but also yielded the appearance of aberrant structures in the cultures including disorganization of the vasculature (Figure 3.11A column 3, Supplementary Figure 3 (top panel), Supplementary Figure 4).

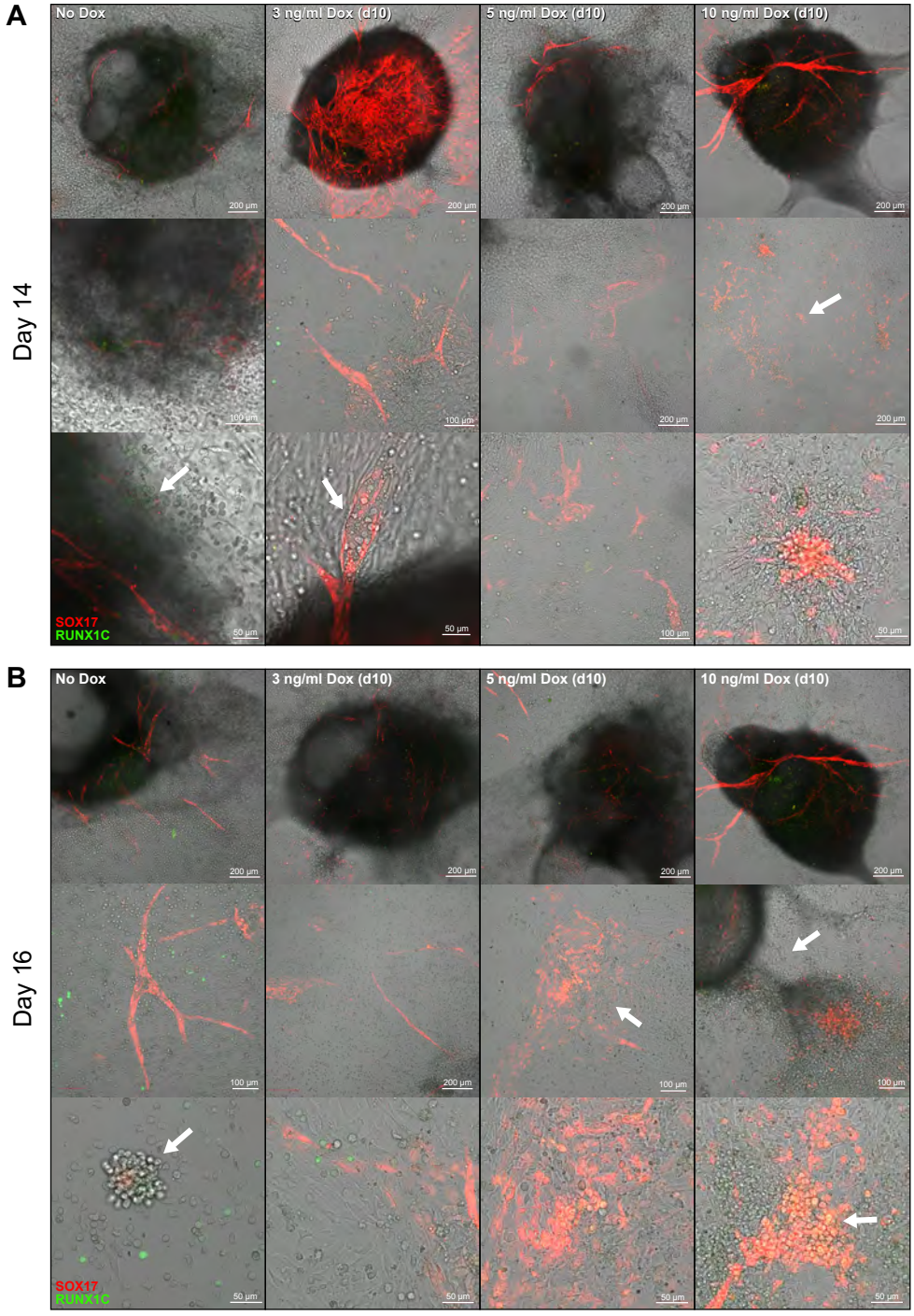


Figure 3.11: RUNX1-ETO induction at balanced levels before the EHT transition disrupts the vascular organization and blocks blood formation

Confocal images of combined Z-stack layers from d14 (A) and d16 (B) of haematopoietic differentiation cultures with RUNX1-ETO induced from d10 (before the EHT) using 3, 5 or 10 ng/m Dox. Each column shows different fields and magnification within the same Dox-treated culture. White arrows in the 0 Dox condition point emerging RUNX1C+ blood progenitors. Arrows in the Dox-treated samples point aberrant structures including disorganized vasculature, RUNX1C- emerging progenitors and *RUNX1C+ SOX17+* co-expressing progenitors. Brightfield and fluorescence field channels are merged. Scale bars: 50, 100 or 200 µm, as indicated. SOX17 (mCHERRY, red) and RUNX1C (GFP, green).

3.5 Balanced *RUNX1-ETO* expression after the EHT allows vasculogenesis and emergence of blood progenitors

Previous studies have described a partial block in myeloid differentiation occurring in *RUNX1-ETO*-expressing myeloid progenitor cells (Regha et al., 2015; Rhoades et al., 2000). For this reason, we evaluated the effect on differentiation upon induction of the oncogene after the EHT, when blood progenitors were already present in culture. Induction during 7 days with 3 and 5 ng/ml Dox at several time points after d15, allowed apparently normal formation of $SOX17^{CHERRY}$ vessel structures and continuous production of $RUNXC^{GFP}$ blood progenitors. However, phenotypically abnormal vasculature and reduced formation of blood progenitors occurred upon 10 ng/ml Dox treatment (Figure 3.12).

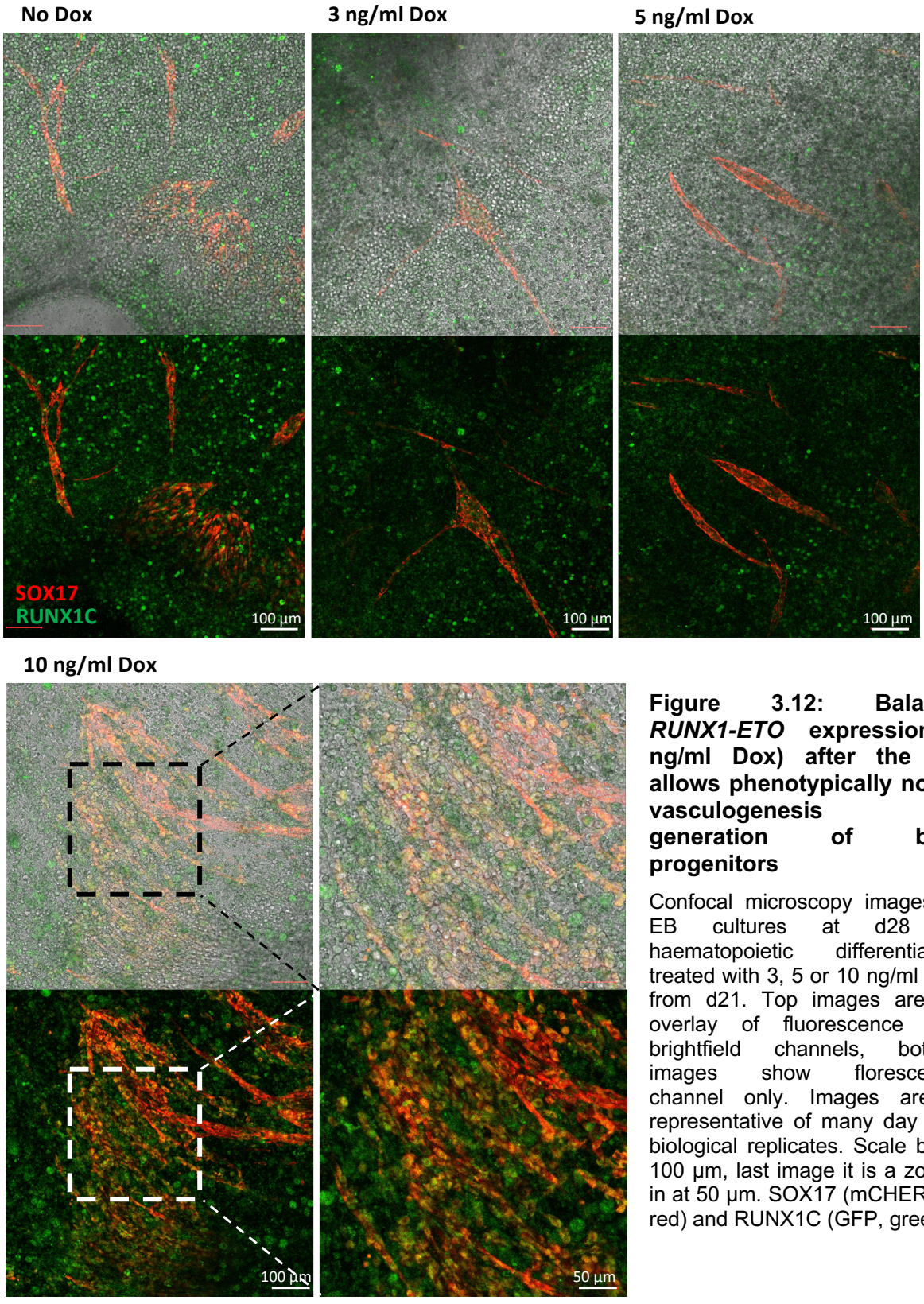


Figure 3.12: Balanced *RUNX1-ETO* expression (5 ng/ml Dox) after the EHT allows phenotypically normal vasculogenesis and generation of blood progenitors

Confocal microscopy images of EB cultures at d28 of haematopoietic differentiation treated with 3, 5 or 10 ng/ml Dox from d21. Top images are an overlay of fluorescence and brightfield channels, bottom images show florescence channel only. Images are a representative of many day and biological replicates. Scale bars: 100 µm, last image it is a zoom-in at 50 µm. SOX17 (mCHERRY, red) and RUNX1C (GFP, green).

3.6 Cells expressing *RUNX1-ETO* retain markers of immature progenitors

Balanced expression of *RUNX1-ETO* permits continuous formation of blood progenitors. We therefore evaluated whether the nature of the cell populations in culture was affected upon expression of the oncogene for 7 days. Blood cells in suspension were characterized by flow cytometric analysis based on expression of haematopoietic markers. Expression of *RUNX1-ETO* resulted in a dose-dependent decrease of RUNX1C+ and CD16+ granulocytic myeloid cells, the latter being barely present in the 10 ng/ml Dox-induced culture. Interestingly, we observed an appearance of a CD34+ CD38- CD90+ population of immature blood progenitors (Figure 3.13). Comparable results were observed following Dox induction of RUNX1-ETO at multiple time points during blood differentiation (Supplementary Figure 5). Moreover, cell populations also expressed additional immature markers, such as CD31 and VECad, upon *RUNX1-ETO* expression (Supplementary Figure 5). Overall, these results are in line with the previously reported studies in the mouse RUNX1-ETO-dependent block of haematopoietic differentiation at an early stage (Regha et al., 2015; Rhoades et al., 2000).

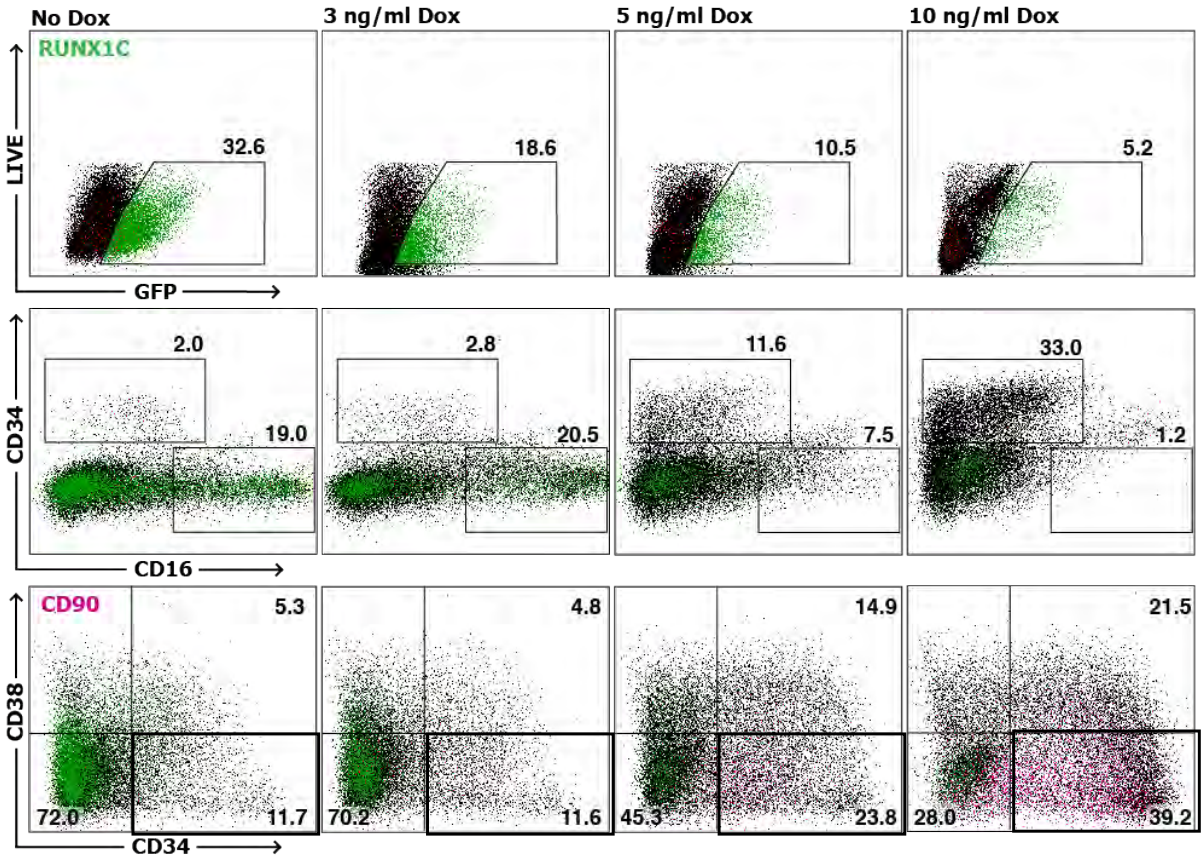


Figure 3.13: *RUNX1-ETO*-expressing cultures retain markers of immature myeloid progenitors

Flow cytometry analysis of the floating fraction of d34 haematopoietic progenitors upon RUNX1-ETO induction from d27. Results are a representative of three biological replicates with comparable results in all the selected stages during haematopoietic differentiation. Average data of flow cytometry analysis is shown in Supplementary Figure 5.

3.7 *RUNX1-ETO* expression results in a reversible decrease in colony forming capacity

In order to further characterise the nature of *RUNX1-ETO* expressing cells in long-term-induced cultures, methylcellulose assays were conducted to assess their clonogenic potential. Cultures were induced for 7 days with a range of Dox concentrations (3, 5 and 10 ng/ml) and progenitors in suspension were harvested and plated into colony-forming unit (CFU) methylcellulose assays. In the methylcellulose wells, either Dox treatment was continued at the same concentration or the inducer was removed (Figure 3.14A). Continued expression of *RUNX1-ETO* in CFU assays resulted in a dose-dependent inhibition of colony formation. However, withdrawal of Dox at the start of methylcellulose cultures, even after prolonged Dox exposure in culture, reversed the reduction of CFU activity, resulting in a dose-dependent increase, suggesting reversibility of the proliferation or differentiation block mediated by RUNX1-ETO (Figure 3.14B). This was found reproducible and highly significant across different biological replicates and clones. Comparable results were observed with Dox treatment at different days during differentiation (Supplementary Figure 6A). In addition, the presence or absence of Dox did not impact on colony shape as compared to the un-induced conditions (Supplementary Figure 6B). Similar results were observed in RUNX1-ETO-induced yolk-sac-like haematopoietic progenitors (Supplementary Figure 6A), which were obtained by omitting the SB/CHIR-depending *HOXA* patterning step during d2-d4 of the differentiation. The reversible decrease in colony-forming capacity together with the accumulation of immature cells in culture, as observed in prior flow cytometric analysis, suggests a role for RUNX1-ETO in retaining highly clonogenic cells in a

quiescent state. A comparable phenotypic effect was observed regardless of the induction time in the progenitor cultures and in both definitive- and yolk-sac-resembling differentiation protocols, hence suggesting that the RUNX1-ETO-specific block of differentiation occurs regardless of the origin and nature of the progenitor cell acquiring the translocation event.

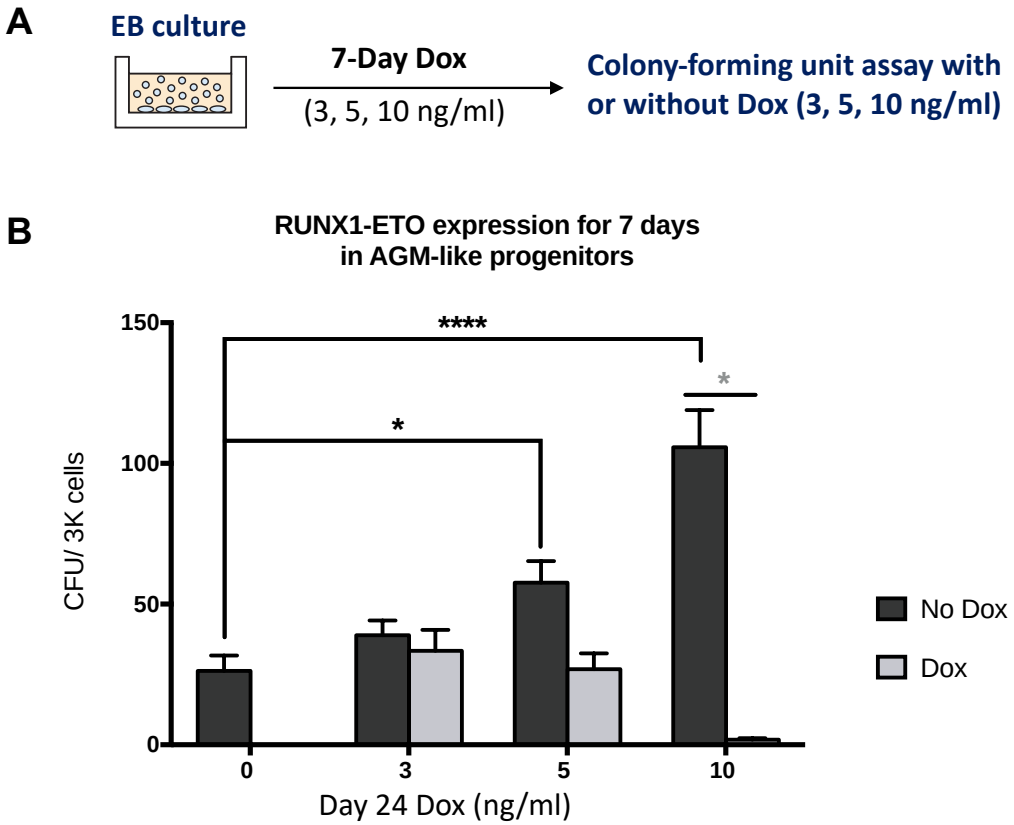


Figure 3.14: *RUNX1-ETO* expression results in a reversible decrease of colony forming capacity

Colony- forming unit (CFU) assays of progenitors from the floating fraction of EB cultures with continued or removed Dox treatment. (A) Diagram depicting the experimental strategy: EB cultures were treated with 3,5, or 10 ng/ml Dox at day 24 for 7 days and plated in methylcellulose (CFU assay) either in presence (light grey) or absence (dark grey) of Dox. (B) CFU assay representing data from three independent biologic replicates using two different clones (clone #2 and #4) conducted in triplicates with 3,000 cells plated per well. Continued Dox treatment in the CFU assay is light grey-coloured, absence of Dox appears as dark grey bars. Comparable results were observed with Dox treatment at different days during differentiation and at yolk-sac-like differentiated progenitors (Supplementary Figure 6). Error bars represent Standard Error of the Mean (SEM). Grey-coloured (*): Multiple T-test, Statistical significance determined using the Holm-Sidak method, with alpha = 0.05. Each row was analysed individually, without assuming a consistent SD. Black-coloured (*): Two-way ANOVA, Statistical significance determined using the Dunnett's multiple comparisons test.

3.8 *RUNX1-ETO* expressed at low levels confers survival but not proliferation to definitive blood progenitors *in vitro*

Previous studies have reported that constitutive *RUNX1-ETO* expression enhanced *in vitro* proliferation of human CD34+ cord blood haematopoietic cells whilst maintaining their self-renewal and differentiation potential (Mulloy et al., 2002, 2003). We therefore evaluated whether balanced levels of *RUNX1-ETO* promoted expansion of human blood progenitors generated in our *in vitro* hPSC differentiation system. To this end, EB cultures were treated with Dox (3, 5,10 ng/ml) for 7 days and the floating haematopoietic cells were subjected to serial replating assays (Figure 3.15A). Cells were replated at 2 $x10^5$ cells/well in fresh medium including the corresponding Dox concentration with which they were originally induced. Cell counts and serial replating of non-adherent cells was performed at 1-2 week intervals. Non-induced haematopoietic progenitors survived poorly, and most cells had died by the second passage. In contrast, all Dox-treated cells underwent an initial transient expansion/growth phase, but the long-term survival of the cultures was inversely related to the concentration of dox used. Whilst 10 ng/ml Dox treatment yielded rapid cell death after the initial growth phase; 5 ng/ml Dox treatment produced a less toxic phenotype. Treatment with 3 ng/ml Dox promoted the prolonged survival of a small subpopulation of cells (Figure 3.15B, Supplementary Figure 7). Cells from EB cultures initially treated with 3 ng/ml Dox from either day 20, 22 or 24, but not untreated cells, could be maintained in replating assays for over 100 days with continued Dox treatment (3ng/ml) (Figure 3.15C, Supplementary Figure 7B). Importantly, none of the different levels of *RUNX1-ETO* expression yielded cell expansion in our experimental system.

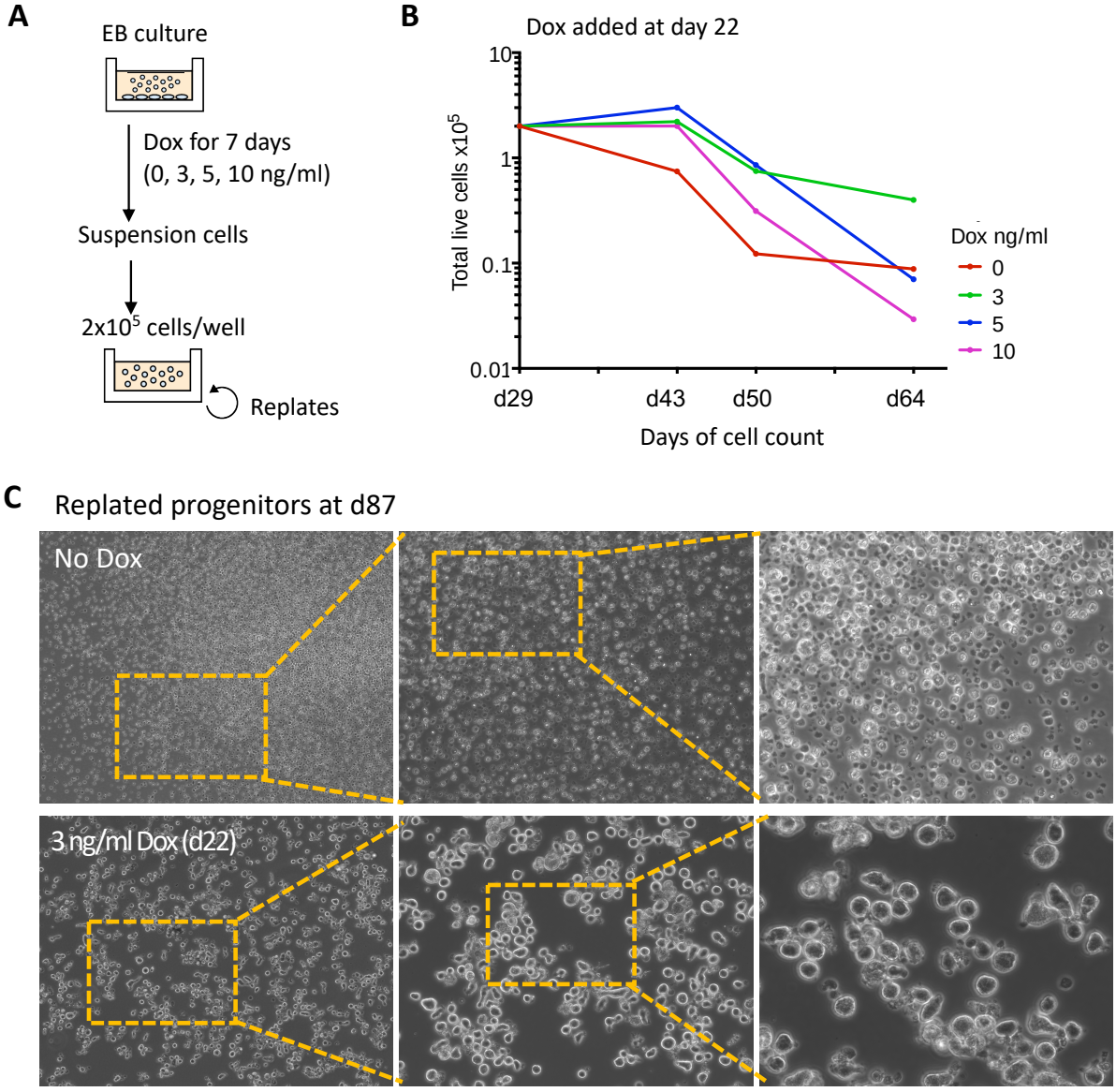


Figure 3.15: *RUNX1-ETO* expressed at low levels confers survival but not proliferation to a subset of *in vitro* definitive blood progenitors

(A) Schematic diagram of the replating assays: EB cultures were treated at different stages of haematopoietic differentiation with 0, 3, 5 or 10 ng/ml Dox for 7 days. Floating progenitors were plated at 2 x10^5 cells/well in the correspondent Dox concentration and were serially replated. (B) Cell count of haematopoietic progenitors from cultures treated at d22 with Dox showing one representative of three biological replicates. Cell growth was measured 3 times as indicated. (C) Brightfield images of haematopoietic progenitors from replating assays at d87 of differentiation uninduced (top row) and treated from d22 onwards with 3 ng/ml Dox (bottom row).

3.9 RUNX1-ETO reduces cell proliferation through a G1-block

To examine the reason for the lack of cell expansion upon RUNX1-ETO expression in our in vitro system, we characterized the changes in cell cycle dynamics of our cultures upon induction of the oncogene. For this reason, we performed a cell proliferation assay by monitoring cell division through pulse labelling with Bromodeoxyuridine (5-bromo-2'-deoxyuridine, BrdU); a synthetic thymidine analogue that gets incorporated into newly synthesized DNA of replicating cells (Gratzner and Leif, 1981). RUNX1-ETO was induced for 4 days in our cultures using 5 ng/ml Dox and 25 μM BrdU was added to cell media for 3.5h. Non-adherent progenitor cells were collected and BrdU incorporation was detected using fluorescent-conjugated antibodies and analysed by flow cytometry. The DNA-intercalator Vybrant-DyeCycle was used as a DNA dye in order to evaluate the distribution of cells in the different cell cycle phases (G0/1, S or G2/M). RUNX1-ETO-induced cultures contained half the amount of non-adherent progenitor cells entering and progressing through the S-phase compared to untreated cells from cultures (Figure 3.16), suggesting a reduction of proliferation due to a block in G0/1.

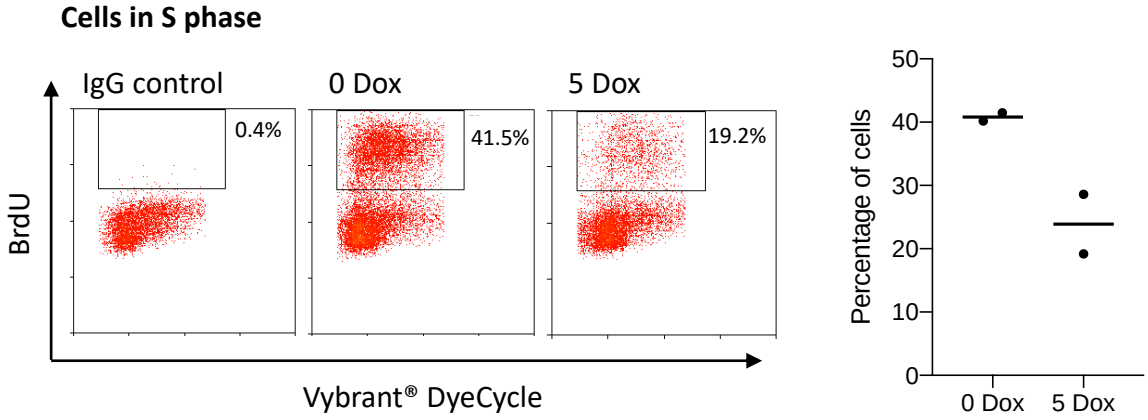


Figure 3.16: RUNX1-ETO induction produces a cell cycle arrest in the G1 phase

Histogram showing cell cycle kinetics of wild-type and RUNX1-ETO induced progenitor cells. EB cultures were induced at d21 with 5 ng/ml Dox for 4 days and then were pulse-labelled with BrdU for 3.5h. Non-adherent cell progenitors were fixed and stained with FITC-conjugated anti-BrdU antibody and Vybrant-DyeCycle. DNA content and cell cycle distribution were analysed by flow cytometric analysis. FITC IgG control is shown. Diagram on the right shows the percentage of cells in S-phase corresponding to two biological replicates (represented as dots) and the median values (represented with a line).

3.10 RUNX1-ETO does not cause cell death via apoptosis

Given the lack of enhanced proliferation and the block at the G1 phase of the cell cycle in RUNX1-ETO induced progenitors, we sought to determine whether the RUNX1-ETO-dependent growth arrest would result in cell death via apoptosis. We therefore induced haematopoietic cultures for four days using 5 ng/ml Dox and non-adherent progenitors were harvested for subsequent stain with AnexinV, which is a marker of apoptosis. Balanced levels of RUNX1-ETO did not lead to cell death via the apoptotic pathway in our differentiation system (Figure 3.17). Whilst this experiment was conducted only once, we did not observe an overall increased cell death when performing routine trypan blue cell counts and flow cytometric analysis in our cultures (data not shown).

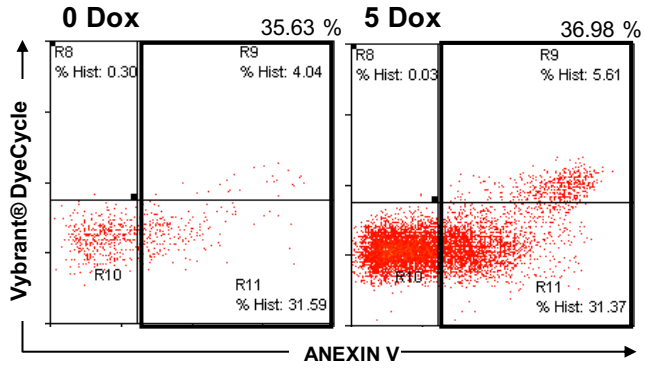


Figure 3.17: RUNX1-ETO does not cause cell death via apoptosis

Histograms showing AnexinV staining of wild-type and RUNX1-ETO induced progenitor cells. EB cultures were induced at d21 with 5 ng/ml Dox for 4 days and then non-adherent cell progenitors were harvested and stained with APC-conjugated anti-AnexinV antibody and Vybrant-DyeCycle. DNA content and apoptosis were analysed by flow cytometric analysis.

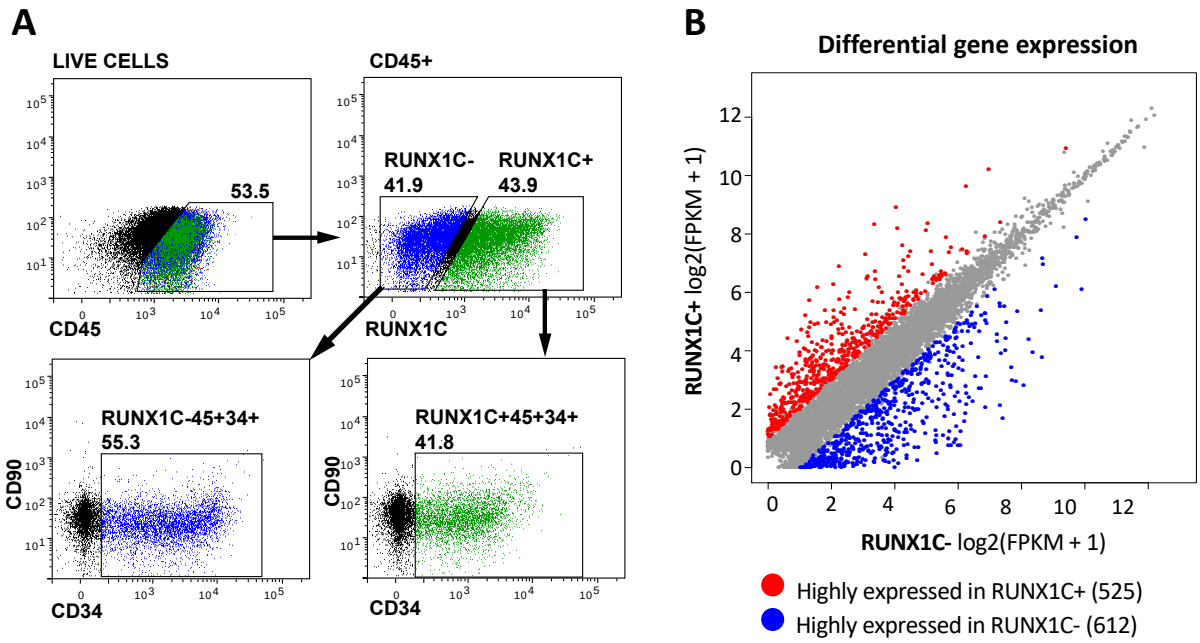
3.11 Expression of *RUNX1C* marks distinct haematopoietic lineages in uninduced cultures

Induction of RUNX1-ETO in progenitors derived from an *in vitro* mouse ES cell differentiation model result in distinct transcriptional outcomes that are dependent on the nature of the induced target population (Regha et al., 2015). We therefore examined the response to RUNX1-ETO induction in different cell populations.

In uninduced conditions, our *in vitro* system generates two main populations of CD45+ CD34+ non-adherent progenitor cells: RUNX1C- and RUNX1C+. RUNX1C- progenitors may still express the RUNX1B isoform from the proximal promoter, as it has been shown that RUNX1B is important for definitive haematopoiesis in the AGM (Sroczynska et al., 2009). *RUNX1C* expression is more prevalent in murine HSC and developing T-cells (Telfer and Rothenberg, 2001). However, transcriptional control, including expression patterns and dosage levels, of the distinct RUNX1 isoforms requires further characterization in the human system. Unpublished data from Elefanty's lab (data not shown) has identified *RUNX1C* expression appearing in CD43+ yolk-sac-like *in vitro* differentiated progenitors. Moreover, *RUNX1C* expression is retained in myeloid progenitors and in mature myeloid cells, such as monocytes, even after loss of colony forming ability. *RUNX1C* was also downregulated during erythroid differentiation and very early during commitment to the lymphoid lineage in the *in vitro* differentiation system. However, molecular analyses of chromatin structure and gene expression patterns to confirm such observations are warranted.

In order to better understand the nature of both cell populations and hence identify the target cells resembling those in which the RUNX1-ETO translocation occurs *in*

utero, prior to the establishment of the pre-leukaemic clone, we performed a deeper molecular characterization of both populations in uninduced conditions. To this end, we took advantage of our lineage tracing system to isolate two cell populations of interest from our wild-type cultures: SOX17-/CD45+/CD34+/ RUNX1C+$^{(GFP+)}$ and SOX17-/CD45+/CD34+/RUNX1C- cells, and performed bulk mRNA gene expression analysis (Figure 3.18A).



C **Up-regulated in RxC+**

Term ID	Term	N. genes	p-value	FDR
KEGG:04015	Rap1 signaling pathway	27	4.08E-09	3.43E-07
KEGG:04510	Focal adhesion	23	4.42E-07	1.86E-05
KEGG:04640	Hematopoietic cell lineage	14	6.65E-06	0.0001
KEGG:04611	Platelet activation	16	5.82E-06	0.0002
KEGG:04512	ECM-receptor interaction	12	2.66E-05	0.0004
KEGG:04151	PI3K-Akt signaling pathway	28	2.47E-05	0.0004
KEGG:04810	Regulation of actin cytoskeleton	20	5.39E-05	0.0006
KEGG:04270	Vascular smooth muscle contraction	13	0.0003	0.0027
KEGG:05200	Pathways in cancer	28	0.0003	0.0028
KEGG:05205	Proteoglycans in cancer	17	0.0008	0.0061
KEGG:04014	Ras signaling pathway	18	0.0011	0.0074
KEGG:04062	Chemokine signaling pathway	15	0.0019	0.0113
KEGG:04022	cGMP-PKG signaling pathway	14	0.0018	0.0116
KEGG:05410	Hypertrophic cardiomyopathy (HCM)	9	0.0025	0.0130
KEGG:05222	Small cell lung cancer	9	0.0027	0.0133
KEGG:05219	Bladder cancer	6	0.0029	0.0136
KEGG:04060	Cytokine-cytokine receptor interaction	19	0.0031	0.0138
KEGG:04540	Gap junction	9	0.0037	0.0155
KEGG:04371	Apelin signaling pathway	11	0.0094	0.0343
KEGG:05202	Transcriptional misregulation in cancer	13	0.0110	0.0386
KEGG:04310	Wnt signaling pathway	11	0.0121	0.0407
KEGG:05414	Dilated cardiomyopathy	8	0.0138	0.0446

Down-regulated in RxC+

Term ID	Term	N. genes	p-value	FDR
KEGG:04640	Hematopoietic cell lineage	23	2.02E-12	4.71E-11
KEGG:04514	Cell adhesion molecules (CAMs)	22	4.90E-08	2.02E-07
KEGG:04658	Th1 and Th2 cell differentiation	15	2.80E-06	1.03E-05
KEGG:04060	Cytokine-cytokine receptor interaction	27	7.98E-06	2.66E-05
KEGG:04620	Toll-like receptor signaling pathway	15	1.33E-05	4.04E-05
KEGG:04064	NF-kappa B signaling pathway	14	2.01E-05	5.63E-05
KEGG:04659	Th17 cell differentiation	14	7.79E-05	0.0002
KEGG:04621	NOD-like receptor signaling pathway	18	0.0001	0.0003
KEGG:04380	Osteoclast differentiation	14	0.0006	0.0014
KEGG:04062	Chemokine signaling pathway	17	0.0009	0.0020
KEGG:04610	Complement and coagulation cascades	9	0.0039	0.0083
KEGG:05202	Transcriptional misregulation in cancer	15	0.0052	0.0104
KEGG:04670	Leukocyte transendothelial migration	11	0.0055	0.0106
KEGG:04066	HIF-1 signaling pathway	10	0.0066	0.0122
KEGG:04668	TNF signaling pathway	10	0.0105	0.0179
KEGG:04662	B cell receptor signaling pathway	7	0.0221	0.0344

Figure 3.18: The wild-type CD45+CD34+RUNX1C+ cell population presents upregulation of genes active in multipotent progenitors and signalling pathways.

(A) Flow cytometry strategy for sorting of d21 cultures based on CD45 (BV), RUNX1C (GFP/FITC) and CD34 (Pe-Cy7) expression.

(B) Clustering of gene expression RNA-Seq data by log2 fold FPKM +1 (fragments per kilobase of transcripts per million mapped reads) values of genes differentially expressed (two-fold change) between RUNX1C+ and RUNX1C- cell populations. Adjusted P value <0.05.

(C) Gene Ontology terms for up- (left) and down- (right) regulated genes in the RUNX1C+ (RxC+) compared to the RUNX1C- cell population.

Alignment statistics, including the number of total and aligned reads as well as the percentage of reads aligned, for each corresponding RNA-Seq library are shown in Table 3.1.

RNA-Seq alignment statistics			
Sample ID	**Total reads**	**Aligned reads**	**% aligned**
RUNX1c_neg_0Dox_Rep1	19642478	19294229	98.23
RUNX1c_neg_0Dox_Rep2	20021355	19668635	98.24
RUNX1c_neg_0Dox_Rep3	19424798	18986780	97.75
RUNX1c_neg_3Dox_Rep1	21378674	20994904	98.2
RUNX1c_neg_3Dox_Rep2	21936702	21061264	96.01
RUNX1c_neg_3Dox_Rep3	19067867	18719083	98.17
RUNX1c_neg_5Dox_Rep1	27420165	27032260	98.59
RUNX1c_neg_5Dox_Rep2	29934399	29599972	98.88
RUNX1c_neg_5Dox_Rep3	28992414	28650589	98.82
RUNX1c_neg_10Dox_Rep1	22885523	22441877	98.06
RUNX1c_neg_10Dox_Rep2	22194071	21765037	98.07
RUNX1c_neg_10Dox_Rep3	17562889	17165571	97.74
RUNX1c_pos_0Dox_Rep1	18344612	18001582	98.13
RUNX1c_pos_0Dox_Rep2	19973654	19613159	98.2
RUNX1c_pos_0Dox_Rep3	19651144	19322123	98.33
RUNX1c_pos_3Dox_Rep1	22091133	21709205	98.27
RUNX1c_pos_3Dox_Rep2	20602306	20258859	98.33
RUNX1c_pos_3Dox_Rep3	16444423	16095980	97.88
RUNX1c_pos_5Dox_Rep1	24377176	24051699	98.66
RUNX1c_pos_5Dox_Rep2	32186546	31812118	98.84
RUNX1c_pos_5Dox_Rep3	25008160	24653300	98.58
RUNX1c_pos_10Dox_Rep1	21341572	20894702	97.91
RUNX1c_pos_10Dox_Rep2	19445370	19051145	97.97
RUNX1c_pos_10Dox_Rep3	13259339	12924741	97.48

Table 3.1: Read alignment statistics of RNA-Seq libraries

RNA-Seq statistics in the RUNX1C+ (RUNX1x_pos) and RUNX1C- (RUNX1c_neg) samples treated with 3, 5 and 10 ng/ml Dox and untreated, showing three sequenced independent biological replicates.

Bioinformatic analyses were performed by Dr Peter Keane (Prof Bonifer's lab) on data from three biological replicates of sorted progenitors. Comparison of differential gene expression in RUNX1C+ versus RUNX1- cell populations showed over 500 and 600 genes up- and down- regulated, respectively, demonstrating the different nature

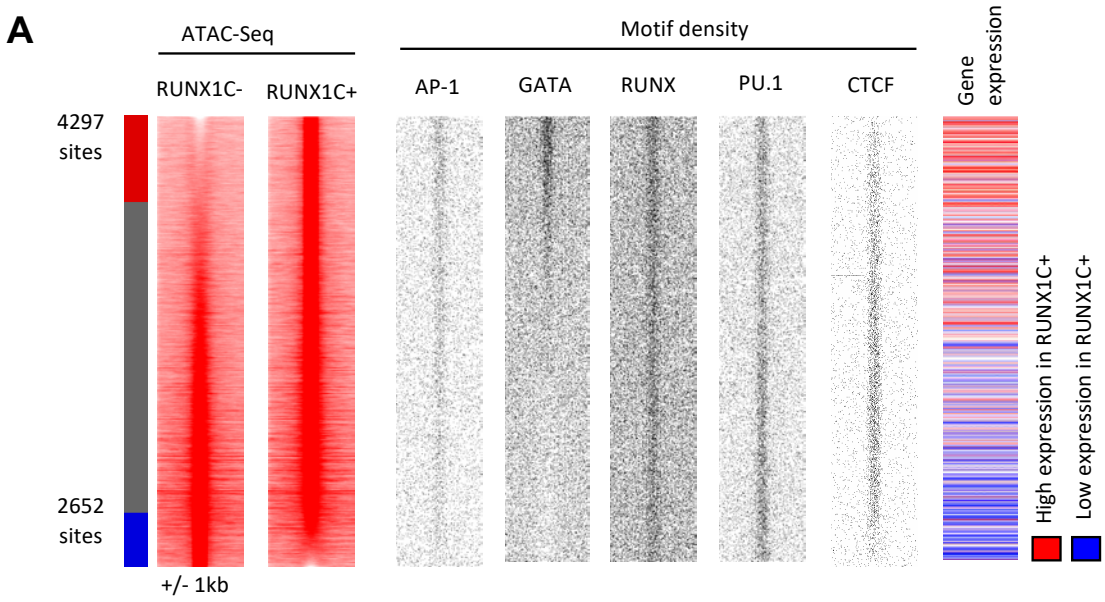
of the two cell populations (Figure 3.18B, Supplementary Table 1). The RUNX1C+ population presented higher expression of the HSC-specific genes *GATA2*, *LYL1*, *MYB*, *NFE2* and *SOX4*, as well as the homeobox genes *MEIS1* and *MEIS2*. In addition, it also expressed the erythroid genes *GATA1*, *GFI1* and *KLF1*. Conversely, the RUNX1C- population has elevated expression of genes regulating monocyte lineage differentiation, such as *CD14, CSF1R, FCGR3A* (CD16a), *KLF4, IRF8, MS4A7, VCAN, FGL2* and, as well as macrophage development, such as *CEBPB, IL10* and *IL13RA1*. Moreover, it also displays high expression of lymphoid progenitor genes, such as *NKG7* and *IL7R*, expressed in Natural killer (NK) cells and T-cell progenitors, respectively.

Understanding the pathways and processes that these genes regulate might aid in the characterization of the two cell populations, for this reason we performed a pathway analysis on the differential expressed genes upon comparison to the KEGG (Kyoto Encyclopaedia of Genes and Genomes) PATHWAY database. Upregulated genes in the RUNX1C+ population are involved in megakaryocyte differentiation, cardiac development and Wnt signalling, which is important for cell fate (Gertow et al., 2013; Sturgeon et al., 2014). Upregulated genes in the *RUNX1C*- population (referred to as downregulated in RUNX1C+ in the differential gene expression analysis) are involved in Th cell differentiation as well as TNF and NK-kappa B signalling pathways, which are activated in both lymphoid lineages and macrophages (Figure 3.18C).

In order to evaluate whether differential gene expression was related to differences in chromatin structure between both populations, we examined genome-wide chromatin accessibility through an Assay for Transposase-Accessible Chromatin followed by

next-generation sequencing (ATAC-seq). ATAC-Seq relies in the mapping of accessible DNA regions by hyperactive Tn5 transposase upon insertion of sequencing adapters into open genomic regions (Buenrostro et al., 2015). For the analyses of the sequenced open chromatin fragments, peaks that displayed a 2-fold differential enrichment in tag counts between samples were considered sample-specific. We observed 4297 and 2652 open chromatin sites specific for RUNX1C+ and RUNX1C-, respectively, that correlate with gene expression (Figure 3.19A, ATAC-Seq). Motif analysis in the RUNX1C+ open chromatin regions showed an enrichment of GATA, RUNX1, PU.1 and AP-1 binding motifs (Figure 3.19B), with GATA sites being exclusively present within the RUNX1C+ specific peaks (Figure 3.19A, Motif density). The RUNX1C- specific peaks were enriched in motifs that were also present in the shared and RUNX1C+ specific peaks, such as PU.1, but were lacking GATA and AP-1 motifs (Figure 3.19A and B). In line with this, RNA-Seq data from our experiments shows that both GATA1 and GATA2 transcription factors are expressed in the RUNX1C+ population, suggesting their binding onto the GATA specific sites.

Overall, analyses of differential gene expression and enrichment of accessible transcription factor motifs suggest that the RUNX1C+ population represents early multipotent and stem-like progenitors as well as erythroid precursors, whilst the RUNX1C- population includes common lymphoid progenitors as well as maturing myeloid cells, being mostly monocytes and primitive macrophages.



B

RUNX1C+ specific sites – 4297 sites

Motif	Match	% sites	p-value
	GATA	35.61	1e-362
	RUNX	30.44	1e-113
	PU.1(ETS)	19.25	1e-223
	AP-1	12.10	1e-108
	CTCF	7.73	1e-162

RUNX1C- specific sites – 2652 sites

Motif	Match	% sites	p-value
	RUNX	28.81	1e-57
	PU.1(ETS)	17.87	1e-213
	CTCF	4.45	1e-72

Figure 3.19: Open chromatin sites specific for the RUNX1C+ cell population are enriched in GATA and AP-1 motifs and correlate with upregulated gene expression

(A) Heatmaps depicting accessible chromatin regions ranked by fold difference between the RUNX1C- and RUNX1C+ CD45+ CD34+ sorted samples. ATAC-Seq peaks were considered sample-specific when displaying a greater than 2-fold enrichment compared to the other sample. Sample-specific sites and number of peaks are indicated alongside being: red the RxC+ specific, blue the RxC- specific and grey the shared peaks. Motif density plots and gene expression at these sites are ranked along the same coordinates as the ATAC-Seq peaks.

(B) Motif enrichment analysis in RxC- and RxC+ specific peaks.

3.12 RUNX1-ETO induction leads to cell-type and dose-dependent changes in gene expression

In order to understand the molecular basis underlying the cell stage-dependent RUNX1-ETO-driven phenotype we sought to examine the gene expression pattern of induced progenitors in the RUNX1C+ and the RUNX1C- population. We therefore used RNA-Seq to examine whether induction of RUNX1-ETO affected distinct sets of genes by performing a comparative gene expression analysis on data from three biological replicates of CD45+CD34+RUNX1C- and CD45+CD34+RUNX1C+ progenitors upon 24-hour Dox exposure (0, 3, 5 or 10 ng/ml) (Figure 3.20).

Figure 3.20: Experimental strategy for the comparison of the RUNX1-ETO-driven effect in RUNX1C- and RUNX1C+ progenitor populations

Outline of the experimental strategy used to evaluate the immediate molecular effects of different levels of RUNX1-ETO induction on distinct progenitors of the *in vitro* haematopoietic differentiation. Dox (0, 3, 5 or 10 ng/ml) was added during 24h in cultures at days d20-d21. Subsequently, CD45+ CD34+ haematopoietic progenitors were sorted for RUNX1C- or RUNX1C+ (GFP+) expression and subjected to gene expression (RNA) and chromatin accessibility (ATAC) experiments followed to high throughput sequencing.

Prior RNA-Seq analyses, transcript levels from a collection of haematopoietic genes were manually quantified by RT-qPCR on both RUNX1C+ and RUNX1C- populations upon 24-hour induction from three biological replicates (Supplementary Figure 8). *RUNX1-ETO* and *RUNX1C* expression were used as a positive control for both the oncogene induction and the RUNX1C+ cell sort, respectively. RUNX1-ETO did not affect overall endogenous *RUNX1* expression – being mostly *RUNX1B* – which appeared sustained regardless of Dox treatment. However, RUNX1-ETO induction

resulted in a slight downregulation of RUNX1C transcript levels in RUNX1C+ populations. Moreover, cells showed downregulation of RUNX1 targets, such as GFI1B, PU.1 and GATA1, in both RUNX1C-/+ populations upon expression of *RUNX1-ETO*, in line with the phenotypic observation of the haematopoietic differentiation block.

As a control for oncogene induction, we first evaluated from our RNA-Seq data the *ETO* (*RUNX1T1*) transcript levels within each replicate. RUNX1-ETO induction levels were consistent across the biological replicates and each Dox concentration (Figure 3.21A). Gene expression analyses showed up- and down-regulation of similar number of genes in both RUNX1C- and RUNX1C+ populations with the same level of RUNX1-ETO induction, with the number of differentially expressed genes correlating to RUNX1-ETO dosage (Figure 3.21B, Supplementary Table 2). An exception was RUNX1-ETO induction using 10 ng/ml Dox in the RUNX1C- population, which yielded less upregulated genes than in the 5-Dox-induced RUNX1C- counterparts (Figure 3.21B, bottom right panel).

Although a similar number of genes was dysregulated in both populations, not all genes responded to RUNX1-ETO induction in the same manner, as some were upregulated, some were downregulated and the expression of others remained unchanged (Figure 3.22). Most of the genes that show a distinct response to RUNX1-ETO correspond to those that were already differentially expressed in both uninduced RUNX1C- and RUNX1C+ populations (Figure 3.22B, orange squares). A comparative analysis showed over 300 differentially expressed genes in RUNX1C+, as compared to the RUNX1C- population, upon induction with 3 and 10 ng/ml Dox (Supplementary Figure 9A), being 236 upregulated and 167 downregulated in the

RUNX1C+ population. Altogether, these analyses demonstrate that RUNX1-ETO dysregulates a different subset of genes depending on the type of progenitor cell.

We observed that a common subset of genes was dysregulated regardless of the RUNX1-ETO expression level, such as *CYP1A1* and *SPINK4*, with 10 ng/ml including the vast majority of the genes responding to 3 ng/ml Dox treatment (Supplementary Figure 9B, Supplementary Table 2). Differentially expressed genes upon treatment were clustered in 8 different groups depending on whether they were up- or down- regulated in each cell population or in both. Induction with 3 ng/ml Dox increased the proportion of upregulated genes (groups 1 (red), 2 (brown) and 3 (green)), whilst 10 ng/ml dox resulted in a larger proportion of downregulated genes overall (groups 6 (purple), 7 (yellow) and 8 (blue)) (Supplementary Figure 9C).

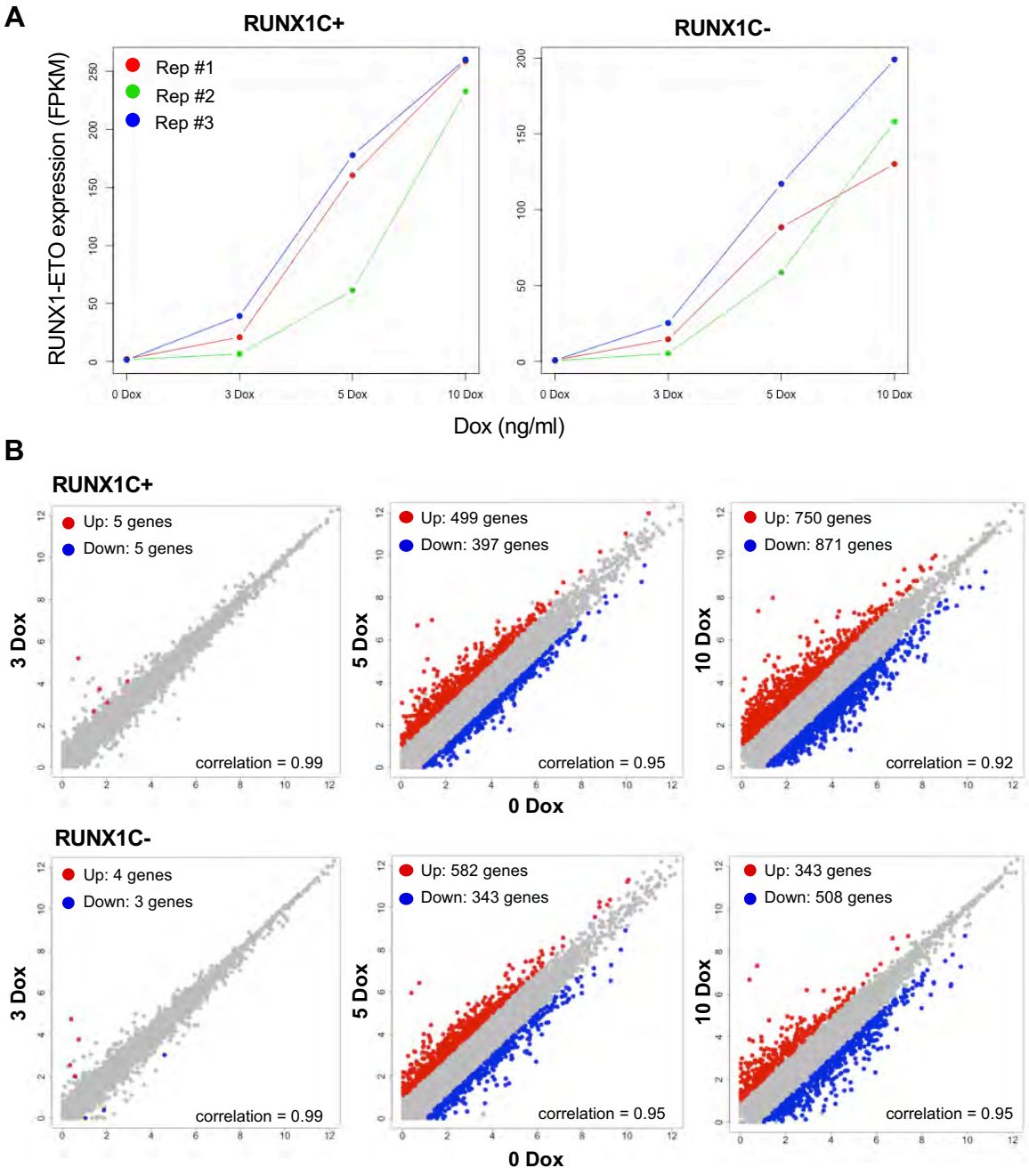


Figure 3.21: RUNX1-ETO induction leads to dose-dependent quantitative changes in gene expression

A) *ETO* (*RUNX1T1*) mRNA transcript levels in both CD45+CD34+RUNX1C+ and CD45+CD34+RUNX1C- progenitors in response to treatment with 3, 5 10 ng/ml Dox for 24h. N=3. Each colour represents a distinct biological replicate.

B) Clustering of gene expression RNA-Seq data by log2 fold FPKM +1 (fragments per kilobase of transcripts per million mapped reads) values of genes differentially expressed (two-fold change) after RUNX1-ETO induction using 3, 5 or 10 ng/ml Dox in both RUNX1C+ and RUNX1C- (CD45+ CD34+) populations. Adjusted P value <0.05.

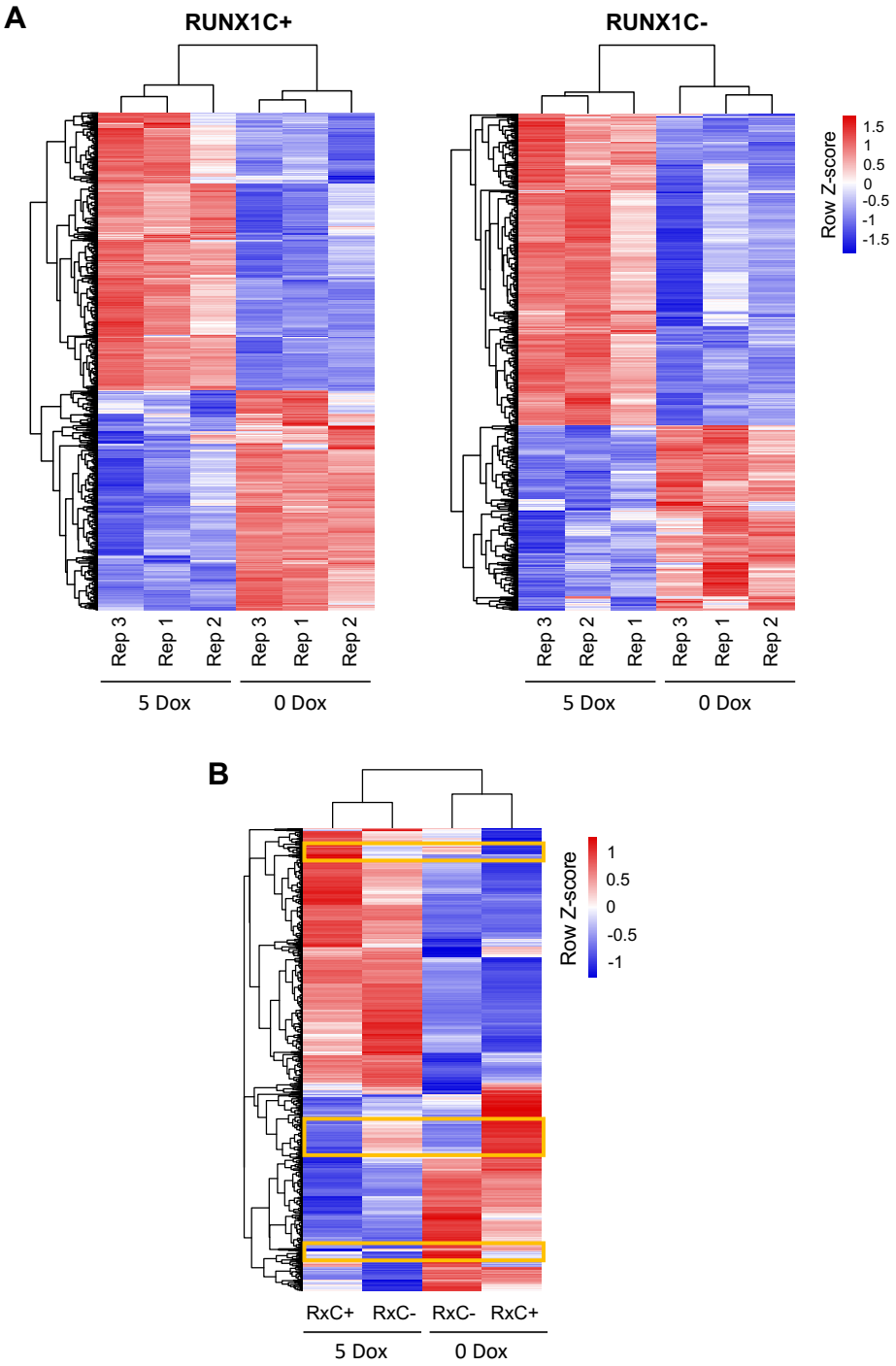


Figure 3.22: Distinct subsets of genes show different response to RUNX1-ETO induction

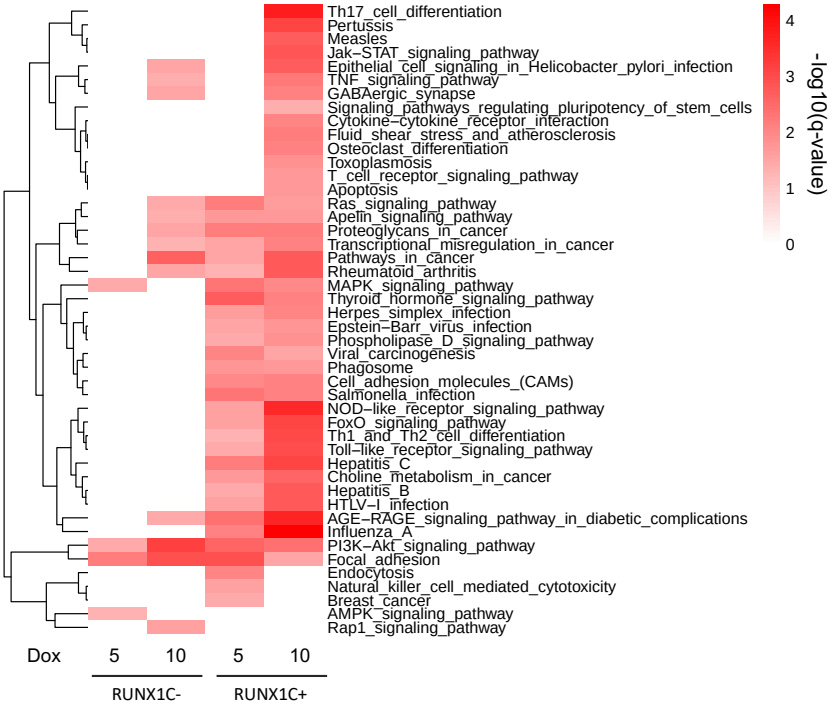
(A) Clustering of gene expression data for the three replicates representing RUNX1C+ (left panel) and RUNX1C- (right panel) CD45+ CD34+ sorted cells both wild-type and after 24-hour RUNX1-ETO induction using 5 ng/ml dox. The figure includes all genes which showed up/down regulation after RUNX1-ETO induction in either the RUNX1C+ or RUNX1C- cell populations.

(B) Clustering of gene expression in RUNX1C- and RUNX1C+ cells upon 24-hour RUNX1-ETO induction using 5 ng/ml dox using the merged data from each replicate in (A).

To examine whether RUNX1-ETO dysregulated distinct associated cellular functions in each population, we assigned differentially expressed genes to their associated KEGG pathways (Figure 3.23, Supplementary Dataset 1). Dysregulated cellular functions associated to gene sets responding to 5 or 10 ng/ml Dox induction differed as well. Upon RUNX1-ETO induction, the RUNX1C+ population upregulated genes, such as *MAPK13*, *MYC*, *NOTCH1*, *PIK3IP1*, *STAT1* and *TP53INP2*, involved in cell signalling pathways including the RAS, Mitogen-Activated Protein Kinase (MAPK), Jak-STAT and PI3K/AKT/mTOR pathways. In contrast, RUNX1-ETO induction in RUNX1C- cells led to upregulation of only a subset of the pathways enriched in the induced RUNX1C+ population, such as the RAS and PI3K/AKT/mTOR pathways. Downregulated pathways in the RUNX1C+ populations included cell cycle (such as *BUB1*), DNA replication and repair (such as *FANCB*), p53 signalling (such as *CHEK2*) and haematopoietic lineages (such as *CD38*). The RUNX1C- population mainly downregulated genes related to the haematopoietic cell lineage (such as *CSF1R* and *IL10*) and genes from other haematopoietic functions such as the complement and coagulation cascade, natural killer cytotoxicity, leukocyte migration and Th cell differentiation. Interestingly, DNA replication and cell cycle pathways were only downregulated in the RUNX1C+ population. KEGG pathway enrichment analysis on the differentially expressed genes in RUNX1C+ compared to the RUNX1C- population further illustrates the heterogeneity of functional pathways depending on cell type and dosage of RUNX1-ETO. For example; Wnt signalling, protein digestion and absorption and phagosome pathways appeared upregulated in RUNX1C+ populations upon 3 ng/ml Dox only, whilst cell cycle genes appeared to be downregulated after treatment with 10 ng/ml Dox only (Supplementary Figure 9D).

Overall, these data show that RUNX1-ETO induction yields a distinct gene expression response in a cell-type and dosage-dependent fashion and therefore evidences the importance of performing adequate RUNX1-ETO induction within the appropriate cell population.

Up-regulated KEGG pathways



Down-regulated KEGG pathways

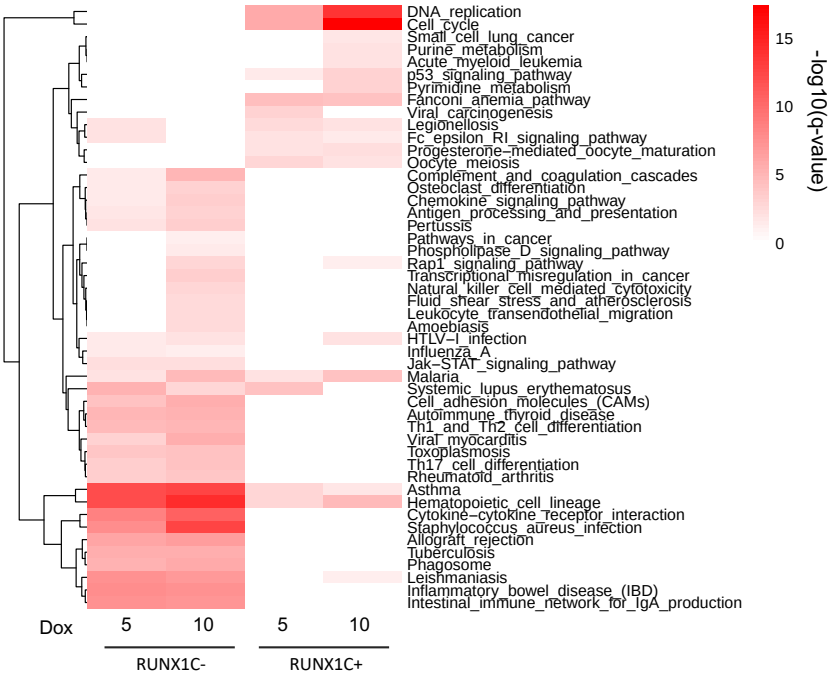


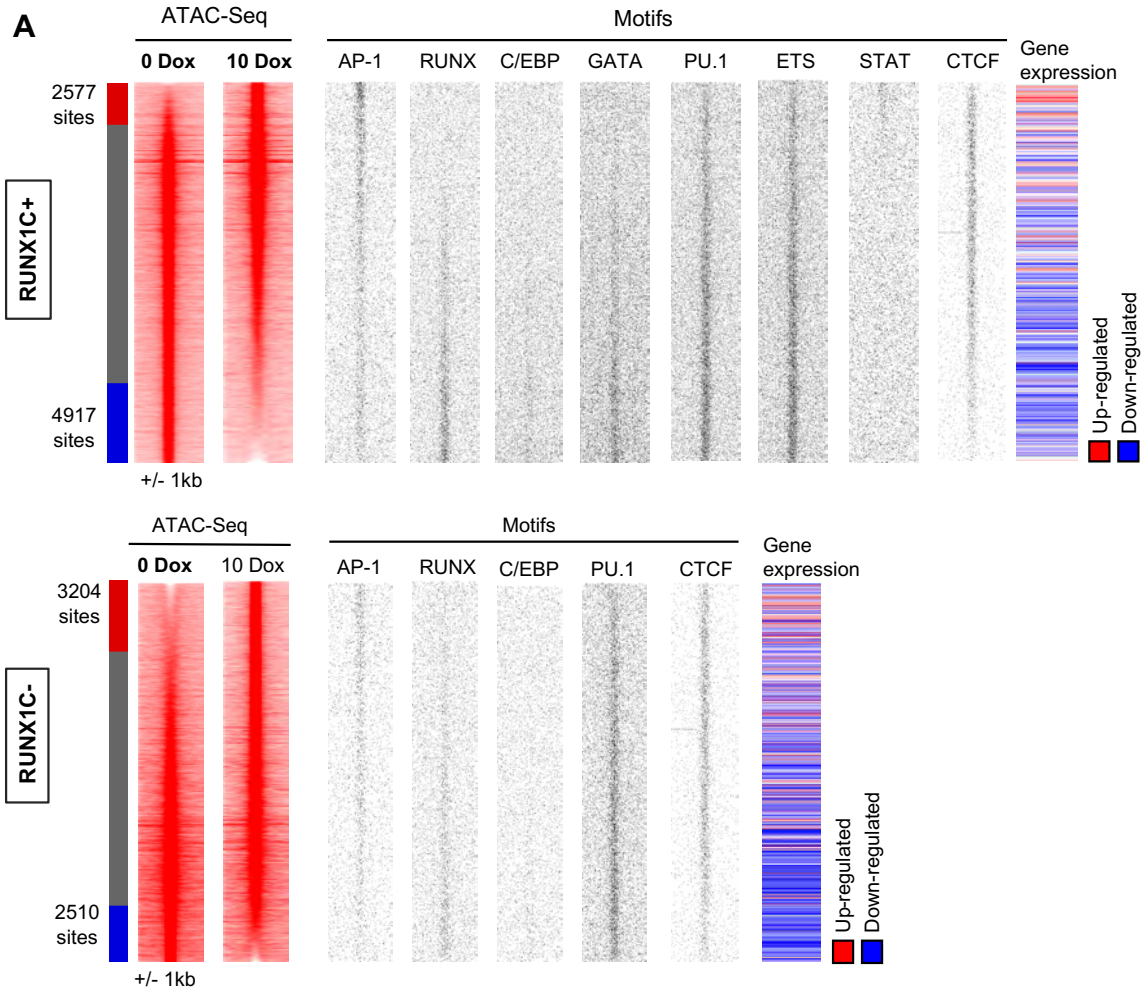
Figure 3.23: Induction of RUNX1-ETO affects distinct pathways and associated cellular functions in each population

Heatmaps clustering KEGG pathways from the differentially expressed genes (up and down regulated) upon RUNX1-ETO induction (5 and 10 ng/ml Dox) in both RUNX1C- and RUNX1C+ sorted populations. Red colour intensity reflects the enrichment significance of the terms in -log10 (q value).

3.13 RUNX1-ETO induction in RUNX1C- and RUNX1C+ populations results in distinct patterns of accessible chromatin that are enriched in similar transcription factor binding motifs

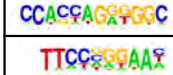
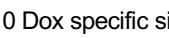
We next wanted to evaluate whether the differential gene expression response after RUNX1-ETO induction was related to a distinct pattern of open chromatin and RUNX1-ETO binding in RUNX1C- and RUNX1C+ populations. In order to examine changes in chromatin accessibility, we performed ATAC-Seq on RUNX1C- and RUNX1C+ cells with or without 10 ng/ml Dox induction (Figure 3.24A). Sample specific peaks were defined by a 2-fold differential enrichment between samples. Induction of RUNX1-ETO with 10 ng/ml Dox resulted in 2577 and 3204 open chromatin sites in the RUNX1C+ and RUNX1C- populations, respectively. The majority of these open chromatin sites in the RUNX1C+ population, but not in the RUNX1C-, correlate with up-regulated gene expression. This result is consistent with the differential gene expression observed in both populations upon induction of the oncogene. Conversely, RUNX1-ETO induction resulted in closure of 4917 sites in the RUNX1C+ samples but only in 2510 inaccessible sites in the RUNX1C-. Shared peaks between both induced cell populations correlated with downregulated genes. Despite the distinct accessible chromatin profile between both induced populations, motif analysis in Dox specific regions showed an enrichment of similar transcription factor binding motifs, including AP-1, ETS and CTCF motifs (Figure 3.24A and B). In addition, 10-Dox specific accessible sites in RUNX1C+ treated cells were also enriched in STAT binding motifs. Both treated populations presented loss of chromatin accessibility in sites harbouring RUNX1, PU.1 and C/EBP. However, RUNX1-ETO induction in RUNX1C+ also resulted in closure of sites enriched in

GATA motifs. This finding is consistent with the presence of accessible GATA sites in uninduced RUNX1C+, but not in RUNX1C-, cells, as demonstrated above (Figure 3.19A). Altogether, this assay suggests that RUNX1-ETO affects chromatin accessibility in a cell-specific manner but resulting accessible sites are enriched in similar transcription factor motifs.

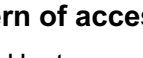


B **RUNX1C+**

10 Dox specific sites – 2577

Motif	Match	% sites	p-value
	ETS	24.06	1e-72
	AP-1	16.14	1e-232
	CTCF	11.02	1e-161
	STAT	8.54	1e-38

0 Dox specific sites – 4917

Motif	Match	% sites	p-value
	GATA	24.08	1e-168
	PU.1(ETS)	20.99	1e-422
	RUNX	20.70	1e-255
	C/EBP	12.24	1e-55
	AP-1	9.52	1e-63
	CTCF	7.81	1e-151

RUNX1C-

10 Dox specific sites – 3204

Motif	Match	% sites	p-value
	CTCF	14.24	1e-289
	PU.1(ETS)	11.24	1e-115
	AP-1	8.77	1e-89

0 Dox specific sites – 2510

Motif	Match	% sites	p-value
	PU.1(ETS)	24.58	1e-159
	RUNX	11.47	1e-65
	C/EBP	6.18	1e-29
	CTCF	5.70	1e-91

Figure 3.24: *RUNX1-ETO*-expressing RUNX1C- and RUNX1C+ cells present a different pattern of accessible chromatin sites that are enriched in similar motifs

(A) Heatmaps depicting accessible chromatin regions ranked by fold difference between the 0 and 10 ng/ml Dox treated samples in RUNX1C+ (above) and RUNX1C- (below) cell populations.

ATAC-Seq peaks were considered sample-specific when displaying a greater than 2-fold enrichment compared to the other sample. Sample-specific sites and number of peaks are indicated alongside, being: red the 10-Dox specific, blue the 0-Dox specific and grey the shared peaks. Motif density plots and gene expression at these sites are ranked along the same coordinates as the ATAC-Seq peaks.

(B) Motif enrichment analysis in 0 and 10 Dox-specific peaks from both RUNX1C+ (left) and RUNX1C- (right) samples.

3.14 RUNX1-ETO dysregulates a similar subset of RUNX1-ETO target genes in both *in vitro* RUNX1C+ and *RUNX1-ETO*-transduced CD34+ cord blood progenitors

In order to further evaluate the significance of the RUNX1-ETO-mediated gene expression response in our progenitors, we sought to identify the shared RUNX1-ETO target genes to previously published data. Given that leukaemic blasts express RUNX1-ETO at higher levels than HSCs (Shima et al., 2014), we compared differentially expressed genes upon 10 ng/ml Dox induction, merged from both RUNX1C- and RUNX1C+ CD34+CD45+ cells, to known RUNX1-ETO-target genes (Figure 3.25A). Known RUNX1-ETO targets were obtained from data of RUNX1-ETO-transduced CD34+ cord blood cells (Lin et al., 2017b) and from t(8;21) patient sample databases (Ptasinska et al., 2012). We observed that a third of the upregulated genes and half of the downregulated genes from our progenitor samples corresponded to known RUNX1-ETO direct targets (Figure 3.25A).

We next wanted to identify the target cell population that yielded an analogous RUNX1-ETO-responding gene expression profile to other *RUNX1-ETO*-expressing systems and patient samples. With this aim, we compared, by hierarchical clustering, RUNX1C- and RUNX1C+ populations with or without 3 or 10 ng/ml Dox induction to gene expression data from wild-type and *RUNX1-ETO*-transduced CD34+ cord blood

cells (Lin et al., 2017b) as well as from peripheral blood stem cells (PBSCs) and t(8;21) patient samples (Ptasinska et al., 2012) (Figure 3.25B). Samples of the same cell nature clustered together, with our early *in vitro* progenitors clustering away from the more mature CD34+ cord blood progenitors and patient blasts, indicating that RUNX1-ETO does not totally override cell intrinsic gene expression programs. However, comparison of gene expression using only known RUNX1-ETO targets highlighted similarities between our induced RUNX1C+ populations and the *RUNX1-ETO*-transduced CD34+ cord blood cells, which represented the most undifferentiated cells within the comparative samples (Figure 3.25C). This highlights the relevance of our model, representing a more undifferentiated progenitor, compared to adult haematopoietic cells, that shares expressed RUNX1-ETO targets with other *RUNX1-ETO*-expressing systems.

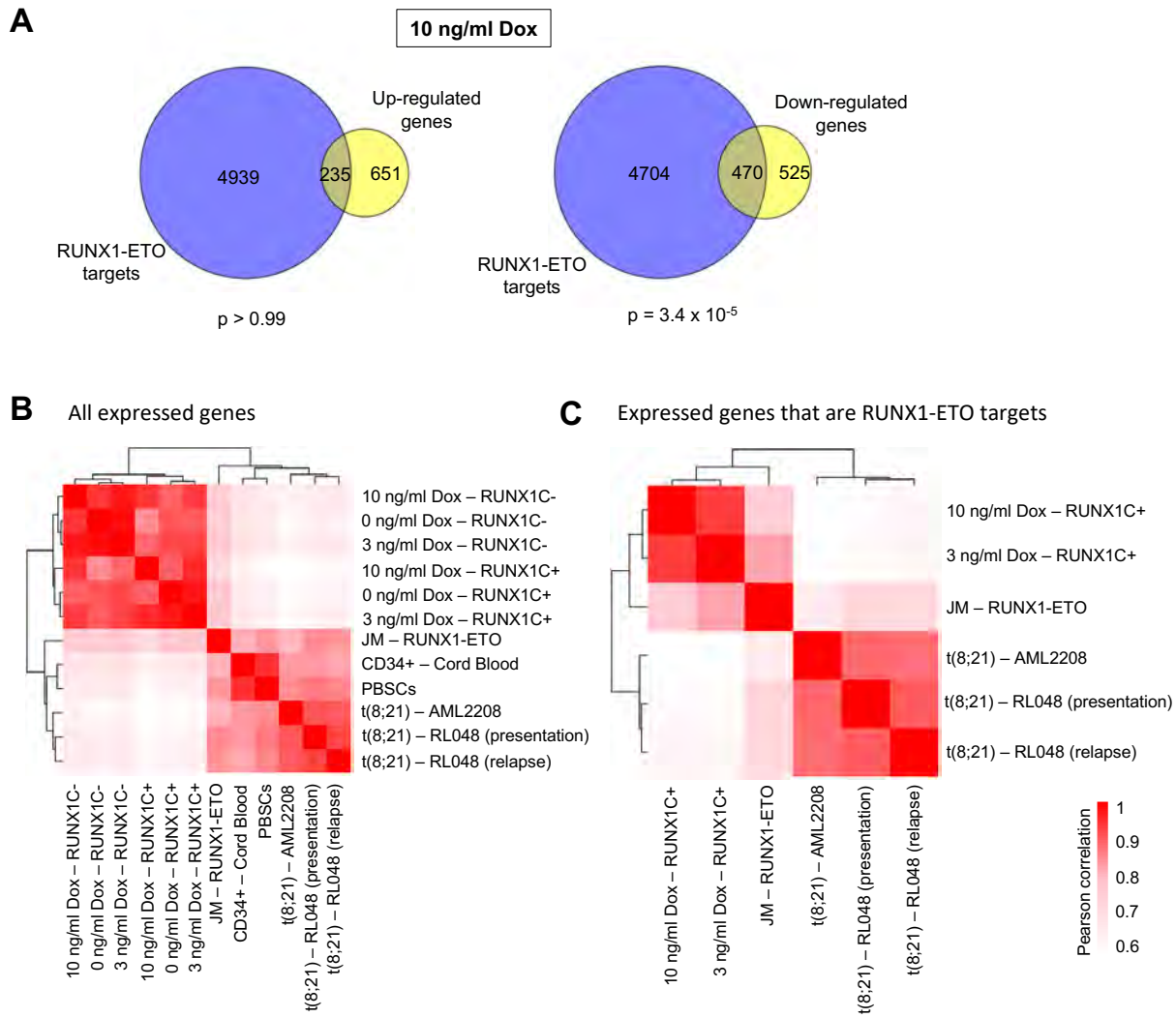


Figure 3.25: RUNX1-ETO dysregulates a similar subset of RUNX1-ETO target genes in both in vitro RUNX1C+ and CD34+ cord blood progenitors

(A) Venn Diagram showing the overlap of upregulated (left) or downregulated (right) genes in CD45+CD34+ progenitors treated with 10 ng/ml Dox for 24 hours to known RUNX1-ETO target genes (from *RUNX1-ETO*-transduced CD34+ cord blood and t(8;21) patient samples). The p value is the result of a hypergeometric test.

(B) Hierarchical clustering based on the strength of correlation of gene expression between samples of RNA-Seq data from: our RUNX1-ETO human ES cell haematopoietic differentiated cells (treated with 0, 3 or 10 ng/ml Dox for 24h and sorted for CD45+CD34+ and RUNX1C- or RUNX1C+), *RUNX1-ETO*-transduced CD34+ cord blood cells (JM-RUNX1-ETO) (Lin et al., 2017b), CD34+ cord blood cells, peripheral blood stem cells (PBSCs) and two patients with t(8;21) (AML2208 and RL048 both at presentation and at relapse) (Ptasinska et al., 2012).

(C) Correlation clustering of RNA-Seq data (as in B) considering only the expressed genes that are known RUNX1-ETO targets, between: our RUNX1-ETO human ES cell haematopoietic differentiated cells (treated with 3 or 10 ng/ml Dox for 24h and sorted for CD45+CD34+RUNX1C+), *RUNX1-ETO*-transduced CD34+ cord blood cells (JM-RUNX1-ETO) (Lin et al., 2017b) and two patients with t(8;21) (AML2208 and RL048 both at presentation and at relapse) (Ptasinska et al., 2012).

3.15 RUNX1-ETO induction in RUNX1C+ cells results in dose-dependent heterogeneous changes in gene expression

RUNX1-ETO-expressing RUNX1C+ progenitors deregulated the same RUNX1-ETO targets as in other *RUNX1-ETO*-expressing systems. For this reason, the RUNX1C+ population was selected to perform further molecular analyses upon induction of RUNX1-ETO.

We next characterized the immediate gene expression response depending on RUNX1-ETO induction levels in RUNX1C+ progenitors and evaluated whether this would yield quantitative and/or qualitative changes. For that, EB cultures were induced around day 20 of differentiation using either 0, 3, 5 or 10 ng/ml Dox for 24 hours and mRNA from CD45+ CD34+ RUNX1C+ sorted haematopoietic progenitors was isolated for subsequent gene expression analyses. We observed a highly dose-dependent gene expression response to RUNX1-ETO induction in RUNX1C+ cells, with a higher number of dysregulated genes with higher Dox concentrations, which was highly reproducible across biological replicates (Figure 3.26A and B). The only exception was the 5 ng/ml Dox induced samples on the replicate number 2, which can be explained due to lower RUNX1-ETO transcript levels as compared to the other replicates (Figure 3.21A). In order to evaluate whether distinct levels of RUNX1-ETO induction would regulate individual genes in a different manner, we performed a covariance analysis on the differentially expressed genes (Figure 3.26C). Interestingly, RUNX1-ETO induction yielded highly heterogeneous changes in gene expression, as distinct subsets of genes displayed a differential response depending on the oncogene dosage. The covariance analysis generated 12 clusters,

grouping together genes showing a similar response to the *RUNX1-ETO* expression levels.

Individual examples of downregulated genes include the cell cycle gene *BUB1,* the TP53 pathway members *CHEK2,* and *CCNB1*, the growth factor receptor gene *KIT* and the *WT1* gene (cluster 2), the stem cell regulator *GATA2* (cluster 3) (Figure 3.27). In contrast, upregulated genes included those involved in signalling pathways, such as *MAPK3* (cluster 7) and immediate early response genes, such as *JUN* and *FOS* (clusters 1 and 11, respectively) (Figure 3.27). The full list of genes within each cluster are listed in Supplementary Table 3.

Overall, these results suggest that (i) RUNX1-ETO produces a different effect depending on the nature of cell population and that (ii) different levels of *RUNX1-ETO* expression dysregulate the same subset of genes but in a different manner.

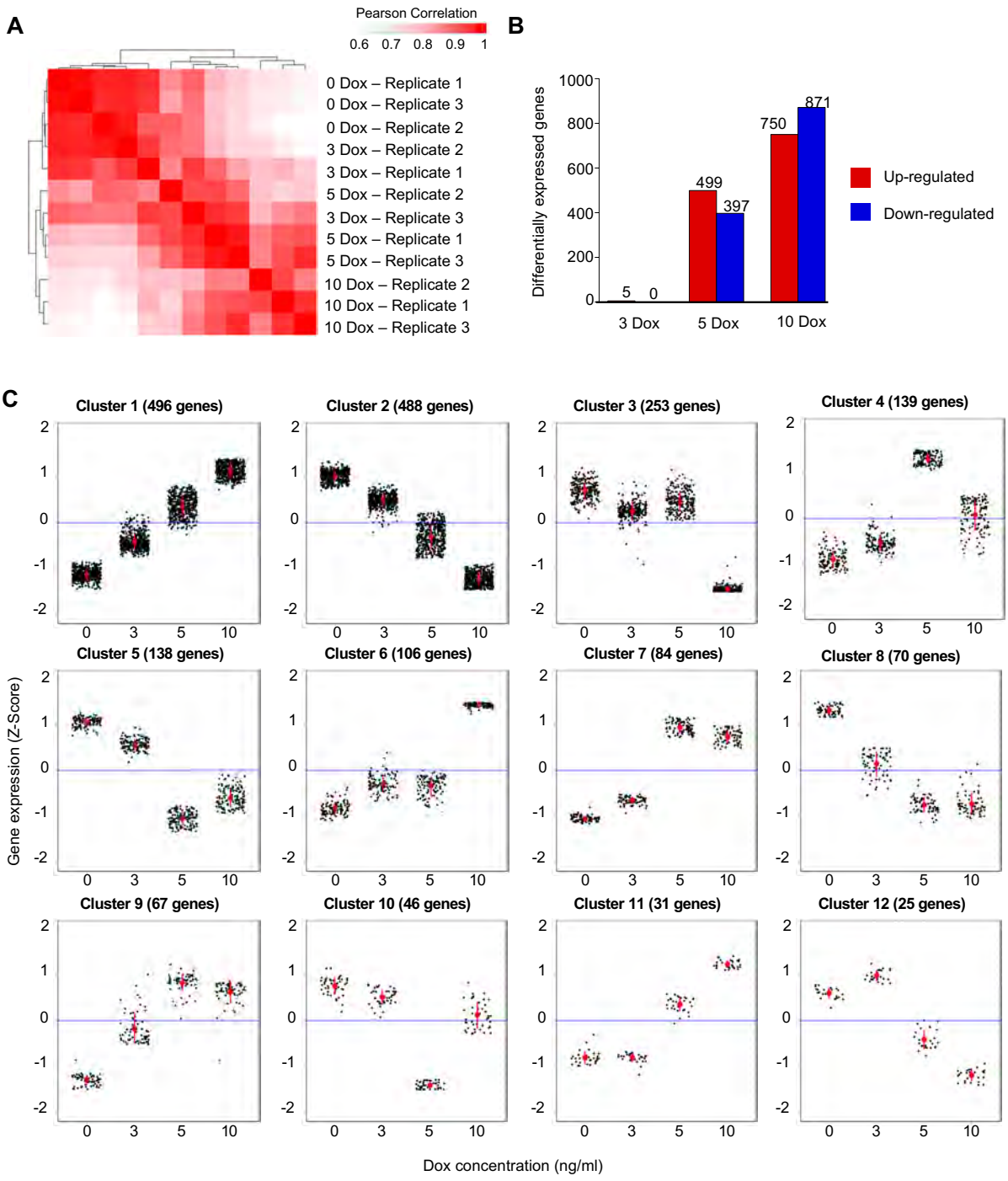


Figure 3.26: RUNX1-ETO induces highly heterogeneous changes in gene expression in a dose-dependent manner

(A) Hierarchical clustering using Pearson correlation coefficients of gene expression in CD45+ CD34+ RUNX1C+ progenitors upon 24-hour Dox exposure (0, 3, 5 or 10 ng/ml) from three biological replicates.

(B) Number of differentially expressed genes between wild-type and Dox-treated (3, 5 and 10 ng/ml) CD45+ CD34+ RUNX1C+ sorted progenitors.

(C) Covariance analysis of gene expression RNA-Seq data by Z-score from CD45+ CD34+ RUNX1C+ sorted progenitor cells upon RUNX1-ETO induction with 3, 5 or 10 ng/ml Dox for 24h, showing 12 clusters/groups of genes with differential expression response to the level of RUNX1-ETO induction. Number of genes comprised on each cluster are indicated. Black dots represent transcript levels for each individual gene. Red dots and bars represent the mean and standard deviation.

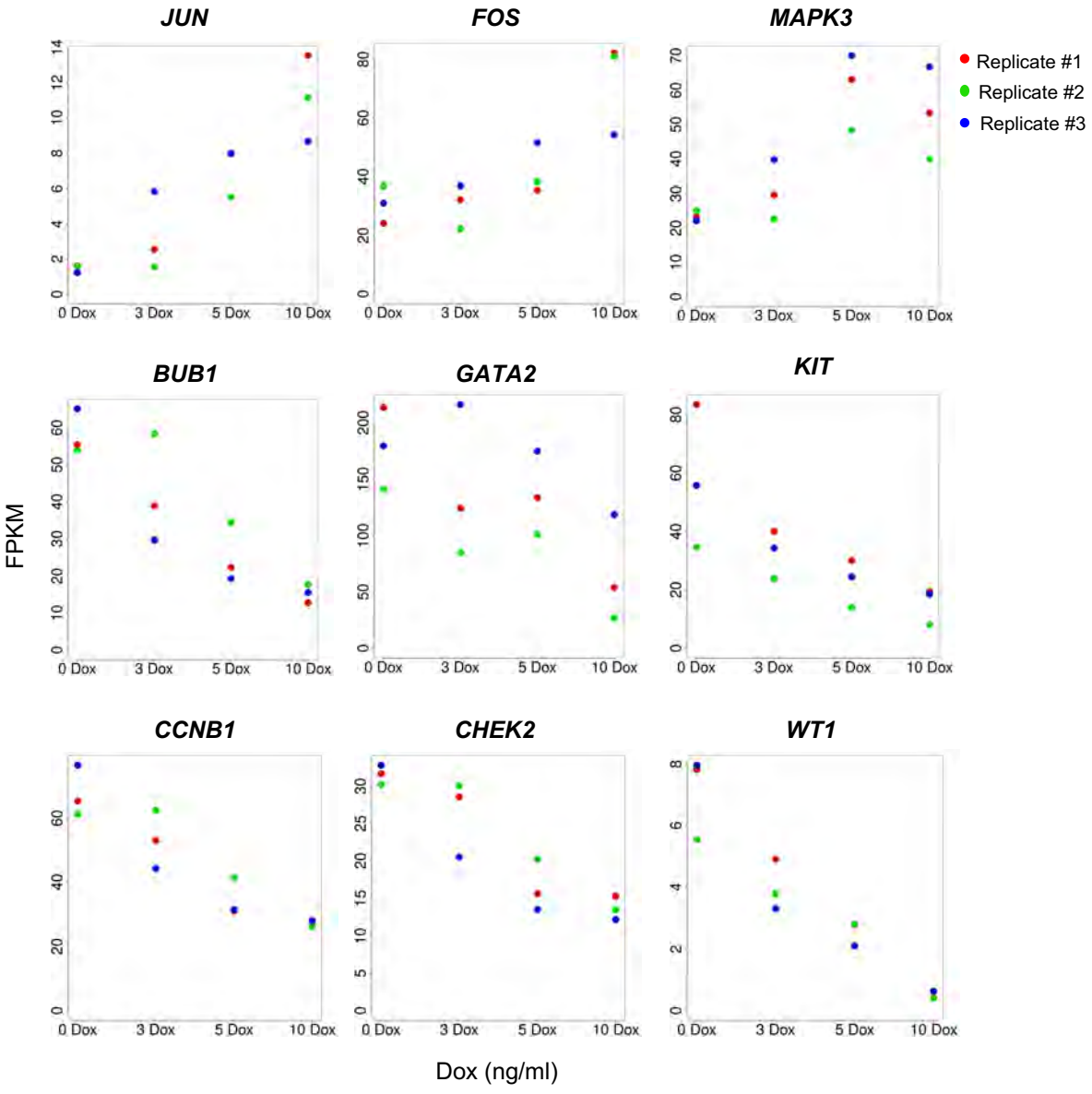


Figure 3.27: Individual genes show distinct responses to RUNX1-ETO dosage

Individual gene examples displaying upregulated (*FOS*, *JUN* and *MAPK3)* and downregulated (*BUB1*, *GATA2, KIT, CCNB1, CHEK2* and *WT1*) mRNA transcripts in CD34+ RUNX1C+ progenitors in response to RUNX1-ETO induction (3, 5 10 ng/ml Dox) for 24h. N=3. Each colour represents a distinct biological replicate.

In order to evaluate whether genes showing a similar expression response to each level of RUNX1-ETO induction were involved in regulating similar cellular functions, we compared associated KEGG pathways to the genes included within each cluster (Figure 3.28). Intriguingly, differentially expressed genes responding to RUNX1-ETO dosage in the same fashion were involved in the regulation of similar cellular activities. For example, downregulated genes belonging to clusters 2 and 3 were involved in regulation of many DNA repair and replication pathways as well as haematopoietic cell lineage. Upregulated genes, belonging to cluster 1, have roles in regulating cell signalling.

In order to further dissect dysregulated cellular functions upon induction of RUNX1-ETO (as shown in Figure 3.23 and Figure 3.28) in the RUNX1C+ population, we generated a KEGG pathways network diagram of associated genes that were at least 2-fold differentially expressed. Induction of RUNX1-ETO using 3 ng/ml Dox did not dysregulate enough genes involved in the same pathway to yield any significant enrichment. However, induction with 5 ng/ml Dox caused upregulation of a large number of signalling genes (such as *MAPK3*, *RRAS*, and *JUN*) and downregulated genes involved in cell cycle (such as *BUB1B* and *CDK1*), DNA replication and repair (such as *BRCA1*, *FANCI*, and *RNASEH2B*) as well as genes related to hematopoietic lineages (such as *CD38*, *IL4* and *KIT*) (Figure 3.29). Similarly, KEGG pathway analysis of the differential expressed genes upon RUNX1-ETO induction with 10 ng/ml Dox showed enrichment of the same major pathways (Supplementary Figure 10): whilst distinct and additional genes were dysregulated by using 10 ng/ml dox, we could observe downregulation of genes involved in myelopoiesis (such as *CEBPA*, *KIT*, *CSF1R* and *MYC*), cell cycle (such as *BUB1*) and DNA replication and

repair genes, including homologous recombination and Fanconi anaemia pathways (such as *BRCA1*, *RAD51*, *RNASEH2A* and *FANCA*) (Supplementary Figure 10). Moreover, genes related to the MAPK and Vascular Endothelial Growth Factor (VEGF) pathways were upregulated as well, as observed upon induction with 5 ng/ml Dox.

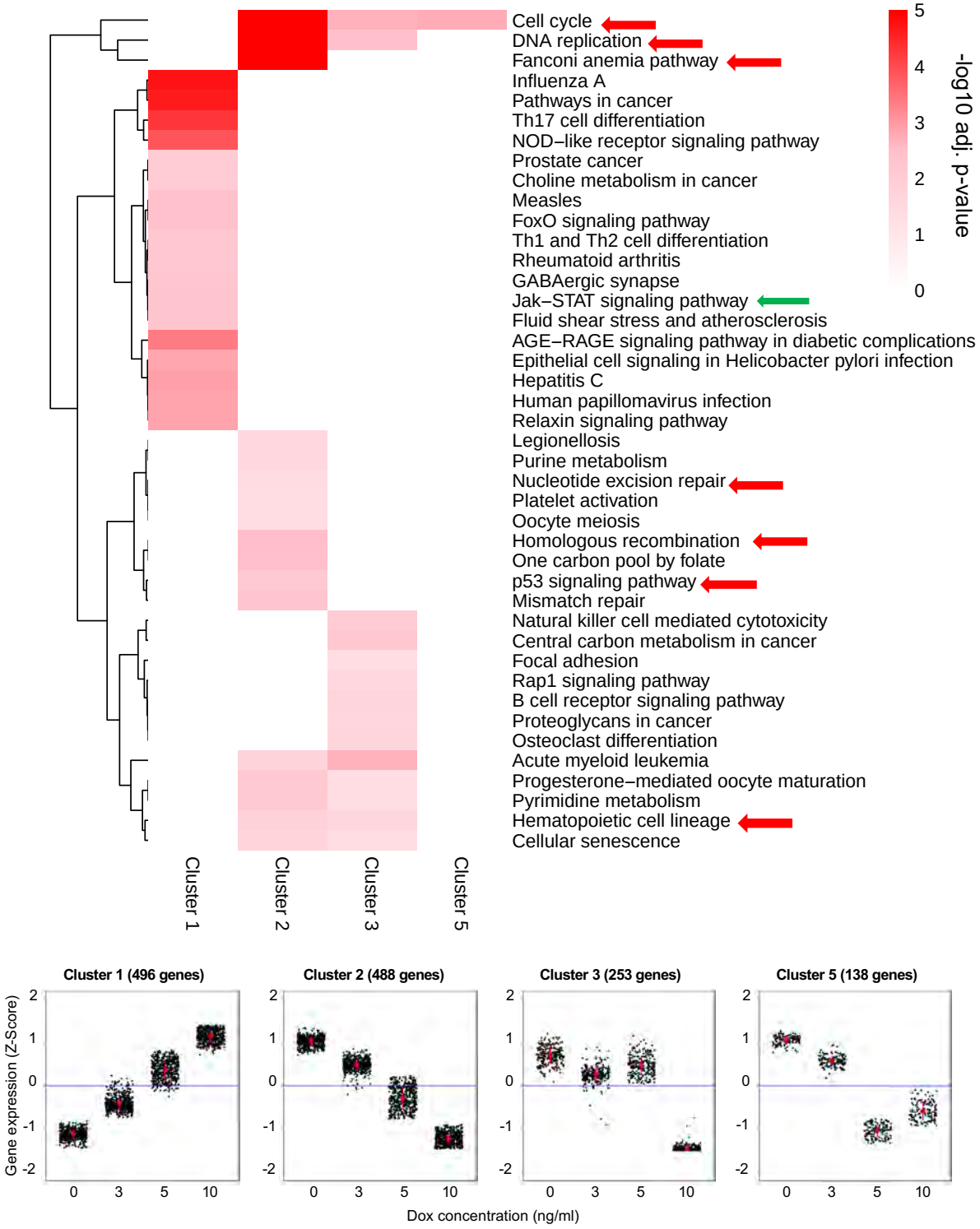
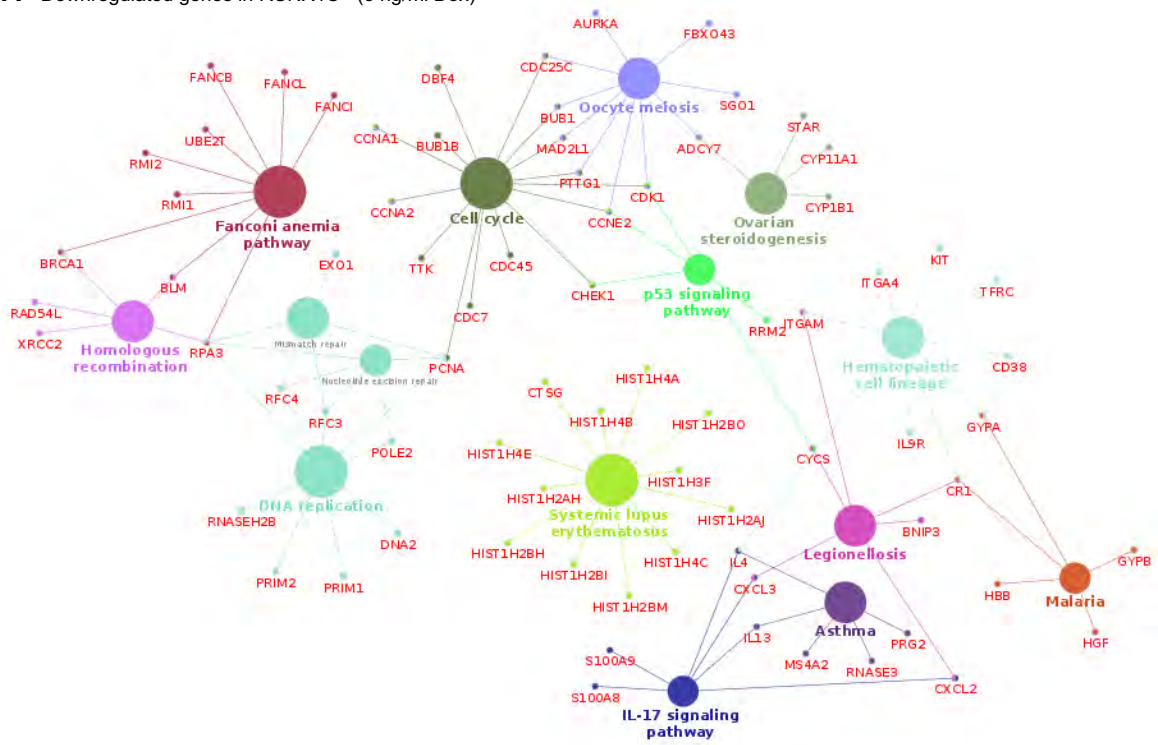


Figure 3.28: Genes responding to RUNX1-ETO in the same fashion are involved in similar cellular activities

Heatmaps clustering enrichment of KEGG pathways from genes included within clusters 1, 2, 3 and 5 of genes with differential expression response to the level of RUNX1-ETO induction. Red colour intensity reflects the enrichment significance of the terms in -log10 (q value). Green and red arrows show upregulated and downregulated relevant pathways, respectively.

A Downregulated genes in RUNX1C+ (5 ng/ml Dox)



B Upregulated genes in RUNX1C+ (5 ng/ml Dox)

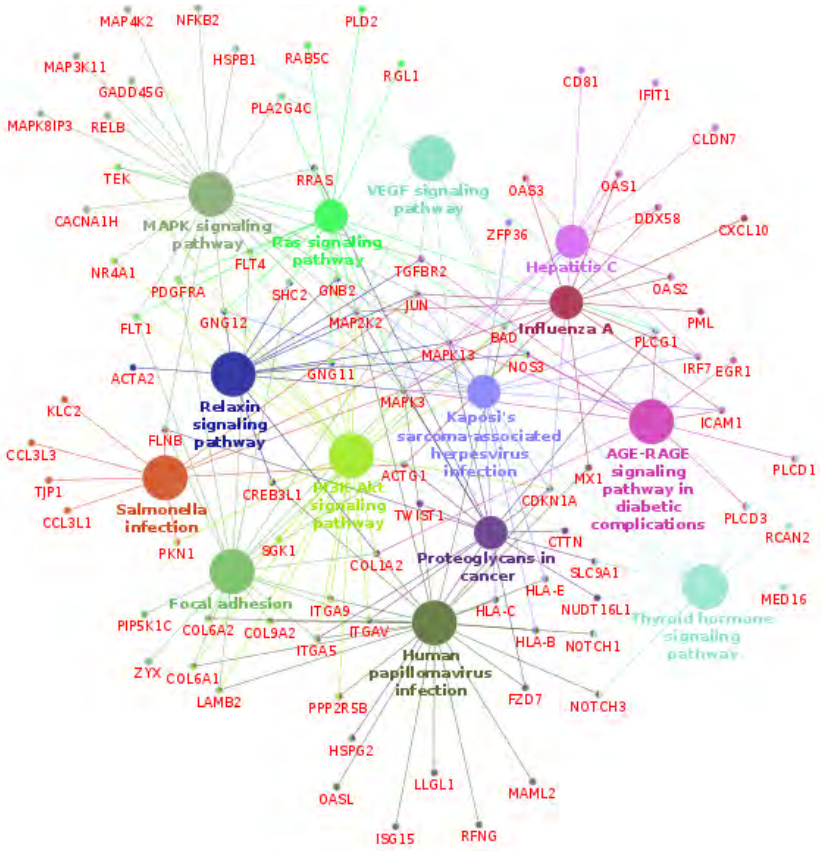


Figure 3.29: RUNX1-ETO induction downregulates myelopoiesis, cell cycle, DNA replication / repair genes and upregulates genes from multiple signalling pathways

Network diagram of KEGG pathways from genes that are at least 2-fold downregulated (A) and upregulated (B) after 24-hour expression of RUNX1-ETO with 5 ng/ml Dox in CD45+CD34+RUNX1C+ human ES-differentiated progenitors.

3.16 RUNX1-ETO induction abrogates the RUNX1-mediated gene expression programme by interfering with RUNX1 binding

Whilst induction of RUNX1-ETO using 10 ng/ml Dox yielded the strongest oncogenic effect, it also resulted in a moderate block of blood formation and affected the vascular structure, as shown above in Figure 3.12. Induction with 5 ng/ml Dox, which yielded levels of *RUNX1-ETO* transcripts comparable to those from endogenous *RUNX1*, resulted in an increase in the number of immature blood progenitors in culture as well as other RUNX1-ETO-dependent features such as a reduction in colony formation and arrest on the G1 phase, whilst allowing proper vasculogenesis and generation of blood progenitors. Moreover, gene expression analyses showed that induction with both 5 and 10 ng/ml Dox resulted in dysregulation of the same cellular pathways. For these reasons, 5 ng/ml Dox was the selected concentration for subsequent studies on the RUNX1-ETO-mediated chromatin reprogramming and effect on transcription factor and histone binding. Chromatin immunoprecipitation followed by next-generation sequencing (ChIP-Seq) allows the analysis of the binding sites of modified histones and DNA-binding factors. Therefore, we evaluated open chromatin regions and protein binding upon RUNX1-ETO induction by performing ATAC-Seq and ChIP-Seq experiments, respectively (Figure 3.30A). Alignment statistics, including the number of total and aligned reads, the percentage of reads aligned and the number of peaks for each corresponding ATAC and ChIP-Seq libraries are shown in Table 3.2 and Table 3.3.

ATAC-Seq alignment statistics				
Sample ID	**Total reads**	**Aligned reads**	**% aligned**	**No. of peaks**
RUNX1c_pos_0Dox	25355305	25274660	99.68	36298
RUNX1c_pos_3Dox	19585104	19535210	99.75	19307
RUNX1c_pos_5Dox	17577061	17546151	99.82	26532
RUNX1c_pos_10Dox	14514152	14453558	99.58	24044

Table 3.2: Read alignment statistics of ATAC-Seq libraries from RUNX1C+ cells upon RUNX1-ETO induction

ATAC-Seq statistics in the RUNX1C+ (RUNX1x_pos) samples treated with 3, 5 and 10 ng/ml Dox and untreated.

ChIP-Seq alignment statistics				
Sample ID	**Total reads**	**Aligned reads**	**% aligned**	**No. of peaks**
HA-RUNX1-Eto_0Dox	8562202	8521125	99.52	49
HA-RUNX1-Eto_5Dox	20172416	20100121	99.64	9675
RUNX1_0Dox	44712626	44616395	99.78	8403
RUNX1_5Dox	49277921	49167956	99.78	1161
H3K27ac_0Dox	24767587	24606924	99.35	36768
H3K27ac_5Dox	25320888	25144300	99.3	36928
H3K4me3_0Dox	23410265	23292055	99.5	30243
H3K4me3_5Dox	17278581	17196663	99.53	27754

Table 3.3: Read alignment statistics of ChIP-Seq libraries in RUNX1C+ cells upon RUNX1-ETO induction

Statistics for each ChIP-Seq library (HA (RUNX1-ETO), RUNX1, H3K27ac, H3K4me3) performed in RUNX1C+ samples untreated or upon RUNX1-ETO induction with 5 ng/ml Dox for 24h.

Induction of RUNX1-ETO resulted in a dramatic shift of the accessible chromatin landscape, as it involved the closure of 5419 sites and the gain of accessibility in 4112 new sites (Figure 3.30B, ATAC-Seq columns). Lost sites were enriched in binding motifs for several haematopoietic transcription factor families including PU.1 (but not other transcription factors belonging to the ETS family), RUNX, GATA, and C/EBP family members (Figure 3.30B, motifs and Figure 3.30C, 0 Dox specific sites). As expected, lost sites were associated with downregulated gene expression (Figure 3.30B gene expression column). Interestingly, different levels of RUNX1-ETO

induction (3 and 10 ng/ml Dox) also resulted on loss of accessibility in sites enriched for RUNX1, GATA, PU.1 and C/EBP binding motifs (Supplementary Figure 11A and B, motif densities).

Most of the upregulated genes as a result of RUNX1-ETO induction correlate with a gain of new accessible chromatin sites (Figure 3.30B, gene expression column). Even though this stands true upon RUNX1-ETO induction with 10 ng/ml Dox, new sites resulting from induction with 3 ng/ml Dox do not significantly affect differential gene expression (Supplementary Figure 11B). These 5-Dox specific new sites present enrichment in binding motifs for members of the ETS transcription factor family (Figure 3.30B, motifs and Figure 3.30C, 5 Dox specific sites). Binding motifs for CTCF and AP-1 family members was enriched across the gained (red) and lost (blue) sites as well as those that remained unchanged (grey zone) (Figure 3.30B, motifs).

These observations were confirmed by ChIP-Seq experiments, showing a reduction of RUNX1 binding across all the RUNX1-ETO-bound sites as well as slight decrease in the abundance of the active histone marks H3K27ac and H3K4me3 (Figure 3.30B, ChIP-Seq columns). Interestingly, newly gained open chromatin sites were not associated with RUNX1-ETO binding but presented AP-1 binding motifs and were associated with upregulation of gene expression. RUNX1-ETO-dependent interference with RUNX1 binding can be easily observed on the average profiles plotted on RUNX1 ChIP binding peaks (Figure 3.30D): RUNX1-ETO (green) displaces RUNX1 (blue) from a subset of RUNX1-binding sites (0 dox specific sites). In some other sites (common sites), RUNX1-ETO causes a reduction of RUNX1 binding (red vs blue). In both scenarios, RUNX1-ETO binding to those sites is higher

than of RUNX1 under uninduced conditions (green vs red). This result is exemplified by the representative gene loci *RASSF5* (Figure 3.31), *RUNX1* or *SPI1* (Supplementary Figure 12). In the *RASSF5* gene example, RUNX1 binds at the TSS and within the gene body, most likely representing distal regulatory elements, in the absence of RUNX1-ETO (No Dox conditions). These sites correlate with open chromatin regions (ATAC – 0 Dox) enriched with active H3K27ac marks. Induction of RUNX1-ETO results in its binding at those sites accompanied by a RUNX1-ETO-level dependent reduction (5 ng/ml Dox) or abolition (10 ng/ml Dox) of RUNX1 binding. RUNX1-ETO-bound sites retain the H3K27ac marks but display slightly less open chromatin profile.

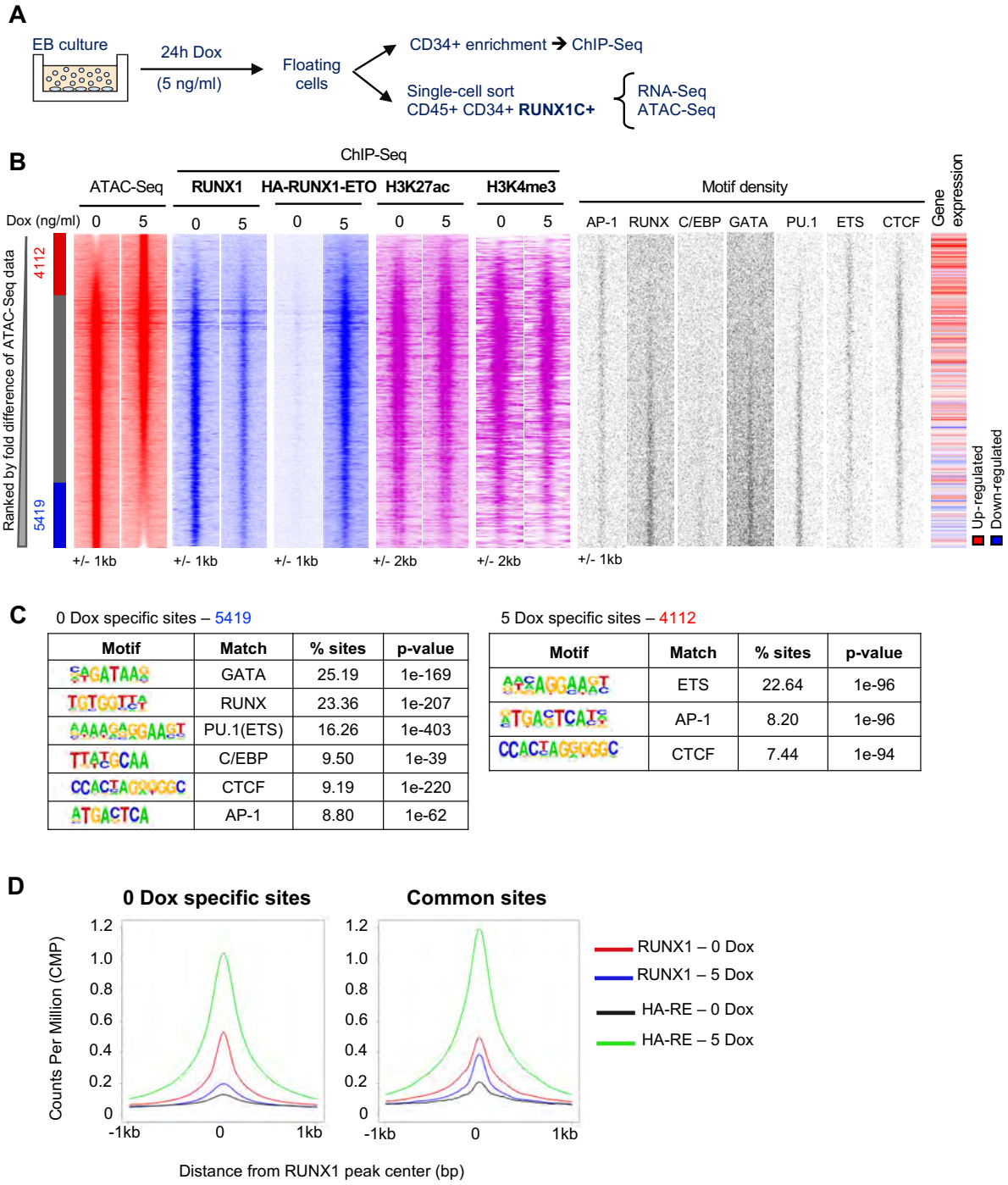


Figure 3.30: RUNX1-ETO induction causes extensive global chromatin reorganisation and blocks the binding of RUNX1

(A) Outline of the experimental strategy used to evaluate the immediate molecular changes in chromatin accessibility and protein binding upon RUNX1-ETO induction in formed progenitors of the *in vitro* haematopoietic differentiation. Dox (5 ng/ml) was added during 24h in cultures at days d20-d21. CD45+ CD34+ RUNX1C+ progenitors were subsequently sorted and subjected to gene expression (RNA-Seq) and chromatin accessibility (ATAC-Seq). Analysis of transcription-factor and histone binding (by Chromatin Immunoprecipitation, ChIP-Seq) was performed on CD34+ enriched populations and in non-adherent progenitors >30% CD34+, respectively.

(B) Heatmaps depicting accessible chromatin regions ranked by fold difference between the 0 and 5 Dox RUNX1C+ treated samples. ATAC-Seq peaks were considered sample-specific when displaying a greater than 2-fold enrichment compared to the other sample. Sample-specific sites and number of peaks are indicated alongside, being: red the 5-Dox specific, blue the 0-Dox specific and grey the shared peaks. ChIP-Seq enrichment for RUNX1, HA-RUNX1-ETO, H3K27ac and H3K4me3 in each sample, motif density plots and gene expression at these sites are ranked along the same coordinates as the ATAC-Seq peaks.

(C) Motif enrichment analysis in 0 and 5 Dox-specific peaks.

(D) Average profiles for RUNX1 and RUNX1-ETO ChIP-Seq data centred on RUNX1 binding peaks (+/- 1000bp from peak centre) in the 0 Dox-specific and common peaks.

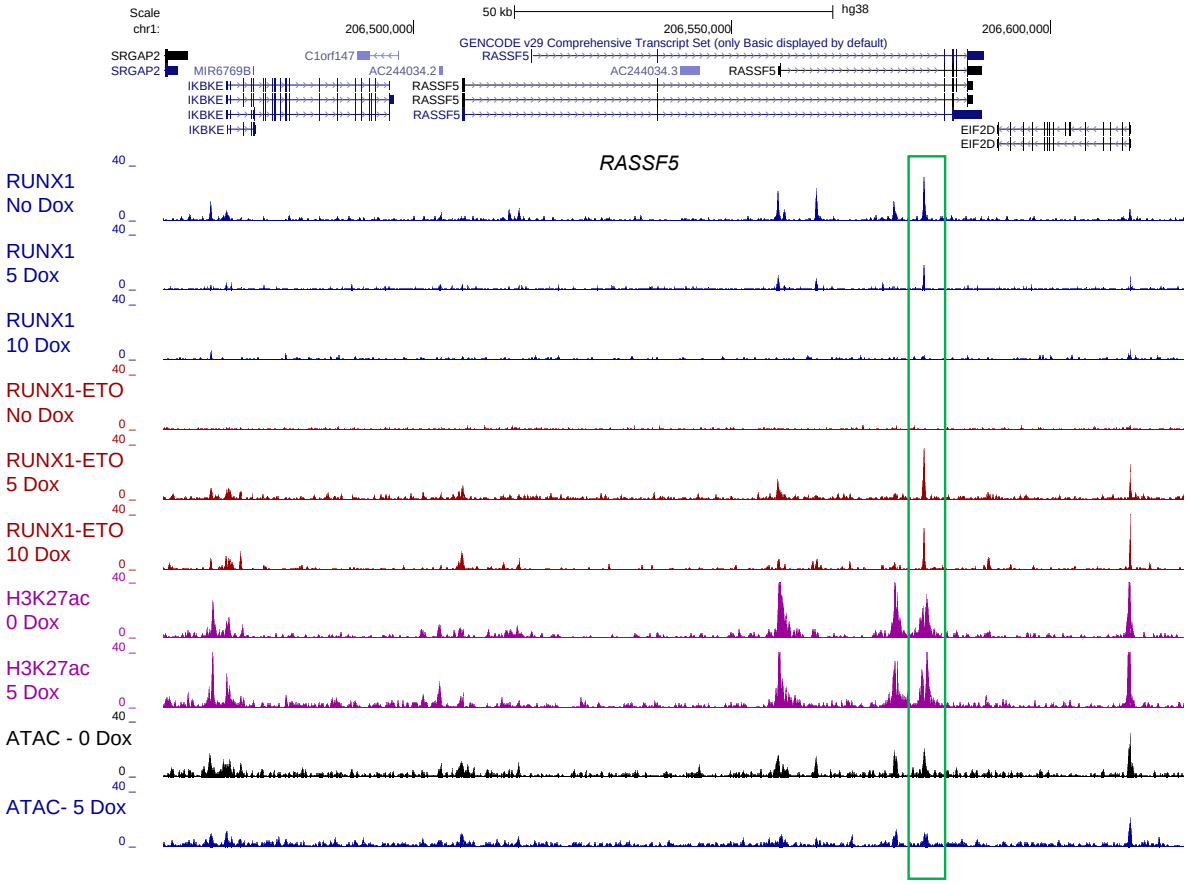


Figure 3.31: Individual examples at the *RASSF5* locus showing RUNX1-ETO-dependent displacement of RUNX1 and reduced chromatin accessibility

Genome browser screenshot depicting RUNX1, HA-RUNX1-ETO, H3K27ac ChIP-Seq and ATAC-Seq tracks for the indicated samples at *RASSF5* representative gene. Green box highlights an example of displacement of RUNX1 binding and reduction of chromatin accessibility upon RUNX1-ETO induction.

In order to dissect RUNX1 and RUNX1-ETO binding, ChIP-Seq data was compared at regulatory regions ranked by fold change of differential RUNX1 binding. Induction of RUNX1-ETO displaced RUNX1 binding at both promoters and enhancers from 5068 and 2104 sites, respectively, that remained bound by RUNX1-ETO only (Figure 3.32 A and B). In a small proportion of the sites (1271 at promoters and 446 at enhancers), RUNX1-ETO did not abrogate but reduced RUNX1 binding. RUNX1-bound promoters were highly enriched in H3K27ac and H3K4me3, but RUNX1-ETO did not affect enrichment on these active marks in neither the 0-dox specific nor the common sites (Figure 3.32B, promoters). However, reduction of these active marks upon RUNX1-ETO induction was most pronounced on distal elements, with H3K27ac being the most affected one (Figure 3.32B, distal elements). Not surprisingly, we observed a reduction of chromatin accessibility in lost RUNX1-bound sites mainly in distal elements, whilst promoter sites were less affected.

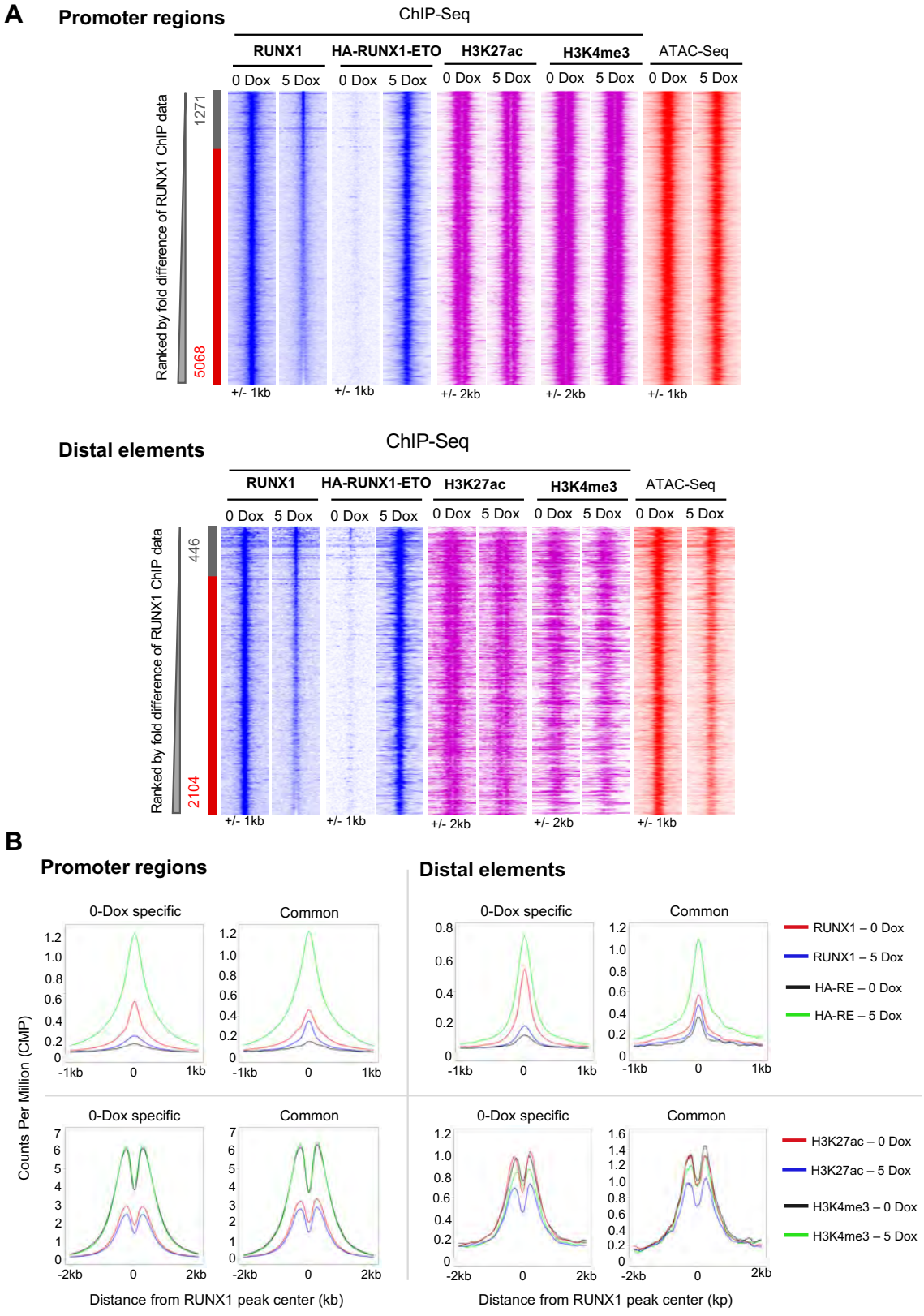


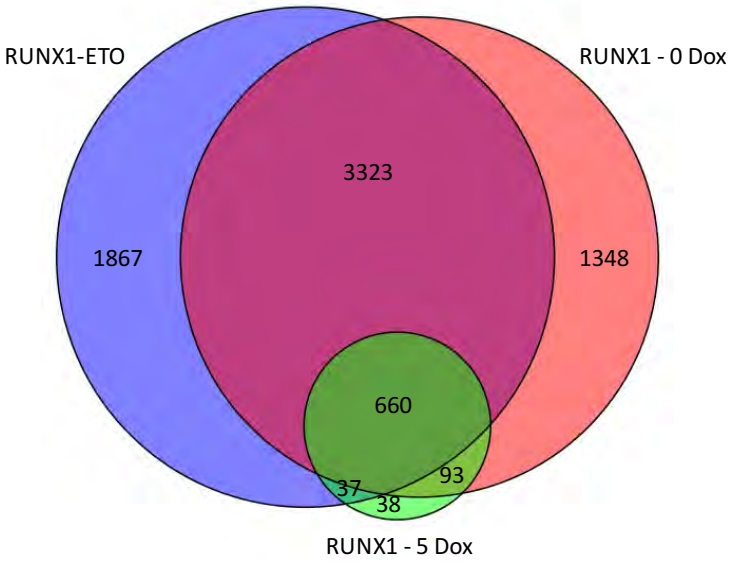
Figure 3.32: Loss of RUNX1 binding and active histone marks upon RUNX1-ETO induction is more pronounced at distal elements than at promoters

(A) Comparison of RUNX1 binding at promoter (above) and distal element (below) regions from ChIP-Seq in 0 and 5 Dox-treated CD34-enriched populations ranked by fold difference of RUNX1 ChIP data, considering peaks with enrichments greater than 2-fold between samples to be specific. Sample-specific sites and number of peaks are indicated alongside, being: red the 0-Dox specific and grey the shared peaks. ChIP-Seq enrichment for HA-RUNX1-ETO, H3K27ac and H3K4me3 in each sample and chromatin accessibility peaks are plotted along the same coordinates as the RUNX1 ChIP-Seq promoter and distal element peaks.

(B) Average profiles for transcription factor (Top panels) and histone modification (Bottom panels) ChIP-Seq data centred on RUNX1 promoter (left) or RUNX1 distal element (right) -binding peaks in the 0 Dox-specific and common peaks.

We next analysed the overall pattern of RUNX1-ETO and RUNX1 binding in both wild-type and in the presence of RUNX1-ETO (Figure 3.33 A). RUNX1-ETO binding shows a large overlap with that of RUNX1 in uninduced conditions (3323 sites) and, as expected, most of the genes that can be bound by both factors lose the RUNX1 binding after induction, resulting in 660 overlapping sites only. Interestingly, RUNX1-ETO is able to bind sites that were not previously bound by RUNX1. We next combined gene expression and transcription factor binding data and observed that around 40% of both up and downregulated genes were RUNX1-ETO target genes (Figure 3.33B). Despite the majority of dysregulated RUNX1 targets that lose binding upon induction being also RUNX1-ETO target genes, there is a small proportion that do not appear to be targeted by RUNX1-ETO (Figure 3.33B, light colours), suggesting indirect mechanisms of gene regulation. The full list of up and downregulated RUNX1-ETO and RUNX1 target genes can be found in Supplementary Table 4.

A **Total number of RUNX1-ETO, RUNX1 (0 Dox) and RUNX1 (5 Dox) target genes**



B **Percentage of genes responding that are targets**

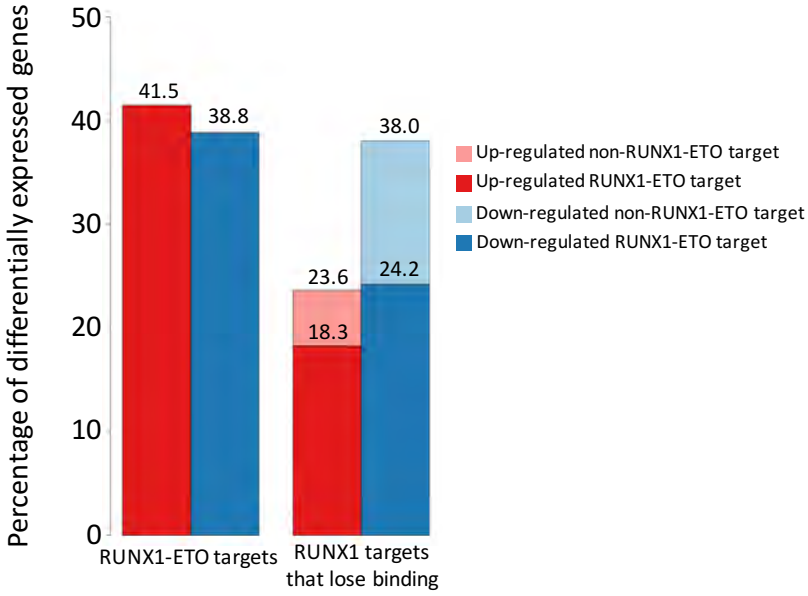


Figure 3.33: Many but not all of the dysregulated RUNX1 targets that lose binding upon induction are also RUNX1-ETO target

(A) Venn diagram of RUNX1-ETO and RUNX1 ChIP data showing overlap of binding events between the total number of RUNX1-ETO, RUNX1 (in 0 Dox uninduced cells) and RUNX1 (after 5 Dox induction) target genes.

(B) Graph depicting the percentage of differentially expressed (up- or down-regulated) genes that respond to RUNX1-ETO induction and are either RUNX1-ETO or RUNX1 targets.

3.17 RUNX1-ETO induction extinguishes many of the adult HSC/myeloid chromatin accessibility pattern and accounts for a large part of the altered network in t(8;21) AML patients

ES-cell-derived definitive haematopoietic cells generated using our culture conditions have a similar transcriptional profile to those generated within the human AGM and are capable of short-term reconstitution in immunocompromised mice (Ng et al., 2016), indicating that our cultures are able to generate progenitor cells with multipotent capacity. Therefore, we aimed to identify whether our ES cell-derived progenitors shared a similar chromatin structure to that of wild-type stem and progenitor blood cells. For this purpose, we compared our ATAC-Seq data from un-induced and induced RUNX1C+ cells to that generated from highly purified human hematopoietic precursor populations as well as monocytic cells (Corces et al., 2016) (Figure 3.34A). We found that the accessible chromatin profile in our wild-type CD45+CD34+RUNX1C+ population strongly resembles that of adult HSCs and MPP populations. However, it differs to that of monocytes, which present a large portion of closed sites. Intriguingly, RUNX1-ETO induction completely shifted the HSC-specific chromatin accessibility pattern, resulting in loss of open chromatin regions specific for early multipotent progenitors and in appearance of new accessible sites that are absent in all the wild-type blood cells. Given the strong difference of our induced progenitors as compared to the wild-type counterparts, we next evaluated the similarities in differential gene expression between our RUNX1-ETO-induced RUNX1C+ progenitors and t(8;21) AML patient cells, as compared to wild-type CD34+ stem/progenitor cells (Figure 3.34B). We observed that up and downregulated RUNX1-ETO target genes in our RUNX1C+ cells are dysregulated in

the same fashion in t(8;21) patient cells. This result indicates that, despite the additional mutations present in t(8;21) patient cells, RUNX1-ETO may account for a large portion of the t(8;21) altered network in leukaemic cells.

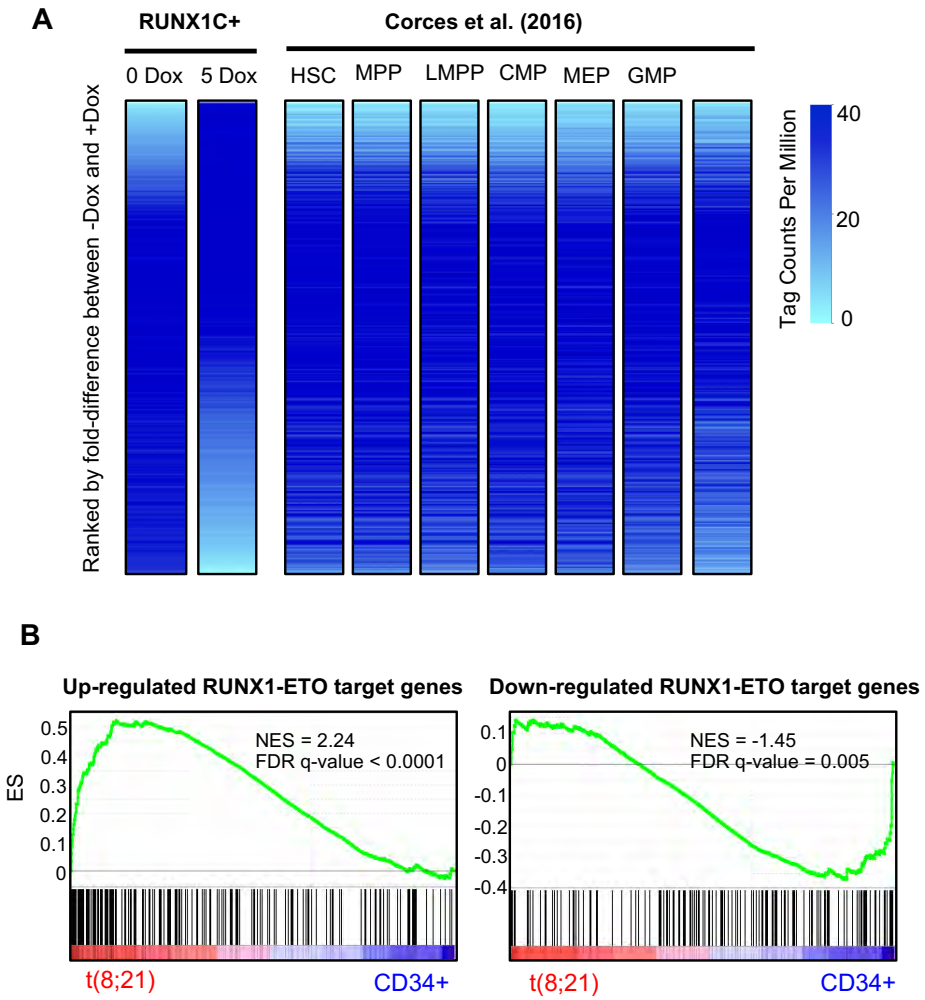


Figure 3.34: RUNX1-ETO-induced human ESC-derived progenitors lack many accessible chromatin sites of adult HSC/myeloid cells and share the same dysregulated RUNX1-ETO targets as in cells from t(8;21) AML patients

(A) Comparison of chromatin accessibility in RUNX1C+ cells (0 and 5 Dox-treated samples) to myeloid progenitor cell types from Corces et al (2016). Heatmaps show ATAC-Seq tag-counts ranked by fold difference between the 0 and 5 Dox treated RUNX1C+ samples. ATAC-Seq tag counts from distinct myeloid progenitor cell types (Corces et al., 2016) are ranked along the same coordinates as the 0 Dox ATAC-Seq peaks. Colour intensity reflects tag counts per million, with red representing closed chromatin.

(B) Gene Set Enrichment Analysis (GSEA) for correlation of upregulated (upper panel) and downregulated (lower panel) RUNX1-ETO target genes between CD45+ CD34+ RUNX1C+ sorted cells upon 24-hour RUNX1-ETO induction (5 ng/ml Dox) and the gene expression profile of the RUNX1-ETO targets in t(8;21) patients. ES: Enrichment Score, NES: Normalized Enrichment Score, FDR: False discovery rate.

3.18 Induction of RUNX1-ETO results in the emergence of a new subpopulation of cells that present a block at the G1 phase

Gene expression analysis on uninduced bulk cell populations showed that RUNX1C+ cells expressed genes and associated signalling pathways from distinct haematopoietic lineages, whilst the RUNX1C- population expressed mainly monocyte and macrophage-related genes, hence indicating that multipotent progenitors might be found in the RUNX1C+ population. In order to identify the nature of the precursor cells present within the RUNX1C+ cell population, we performed single cell (sc) RNA-Seq in wild-type and induced (5 ng/ml Dox for 24 hours) CD45+CD34+RUNX1C+ purified cells (Figure 3.35A). A total number of 7,135 cells were sequenced, being 3,321 from the uninduced and 3,814 from the induced conditions. We obtained a read depth (median reads per cell) of 75,452 and 67,098 reads in each condition, respectively, with approximately 3,100 genes sequenced per cell. Alignment statistics for the two scRNA-Seq datasets are shown in Table 3.4.

scRNA-Seq alignment statistics						
Sample ID	Total reads	% aligned	Total number of cells	Total number of cells after QC	Mean reads per cell	Median genes detected per cell
scRNAseq_0Dox	251032072	97.8	3327	3141	75452	3149
scRNAseq_5Dox	255914998	97.7	3814	3669	67098	3166

Table 3.4: Read alignment statistics of scRNA-Seq datasets (0 and 5 Dox) in CD45+CD34+RUNX1C+ purified cell populations

Statistics for scRNA-Seq data from CD45+CD34+RUNX1C+ samples untreated (0Dox) or upon RUNX1-ETO induction for 24h (5 Dox).

Clusters representing distinct cell populations were visualised upon analysis of the combined datasets (uninduced and 5 ng/ml Dox-treated) with a t-distributed stochastic neighbour embedding (t-SNE) algorithm, which constructs a probability distribution over pairs of points (individual gene expression) and embeds the data points in a low-dimensional map based on their similarities (Supplementary Figure 13A). Thereafter, the nature of each cell cluster was identified by comparison of their individual gene expression to previously known lineage marker genes. We identified eight progenitor populations including early and maturing erythroid precursors, GMPs, monocytes, eosinophils and immature populations resembling stem and CD34+ progenitor cells (Supplementary Figure 13A), which had a cell-type specific gene expression pattern (Supplementary Figure 13B). A complete list of the marker genes used to classify each scRNA-seq cell population can be found in Supplementary Dataset 1.

In order to recognize whether RUNX1-ETO induction affected the density of the cell populations, we performed the clustering analysis on the two datasets individually (Figure 3.35B). Strikingly, induction of RUNX1-ETO resulted in the strong enrichment of a cell population, herein referred to as 5-Dox enriched cluster (green), which was barely detectable in untreated conditions. Comparison of the number of sequenced cells belonging to each cluster clearly depicts an equal contribution within each population in both uninduced and 5 Dox conditions (Figure 3.35C). Notably, a higher proportion of cells contributed to 5-Dox enriched cluster (green) upon induction, as compared to cell contribution across the other clusters (Figure 3.35C).

We next evaluated the total number of deregulated genes and we observed that the 5-Dox enriched cluster represented the most responsive cell population to RUNX1-

ETO induction (Figure 3.35D). Other populations showing a moderate transcriptional response to RUNX1-ETO induction included the stem-like population followed by cells from the myeloid lineage, but, overall, gene expression changes within the whole RUNX1C+ population were mainly represented by those occurring in the 5-Dox enriched cluster. Conversely, CD34+ and erythroid progenitors did not undergo many gene expression changes upon induction.

We showed that long-term induction of RUNX1-ETO in *in vitro* haematopoietic progenitors results in an arrest in the G1 phase of the cell cycle, preventing the cells to enter the DNA replication phase. The next question was whether the RUNX1-ETO-mediated G1 block held true in all the distinct cell populations or whether this was a cell-type specific phenotype. We therefore analysed changes in expression of genes regulated during cell cycle within our scRNA-Seq clusters (Supplementary Table 5). We observed that the 5-Dox enriched population presented a strong arrest in the G1 phase, as the majority of the cells in the induced dataset lacked expression of genes regulated during the G2-M and S phases of the cell cycle (Figure 3.35D). Cell cycle arrest was also observed, to a lesser degree, in the stem/progenitor, GMP-like and eosinophil populations, whilst the erythroid, monocytic and CD34+ populations remained unaffected, in agreement with their lack of RUNX1-ETO responsive genes.

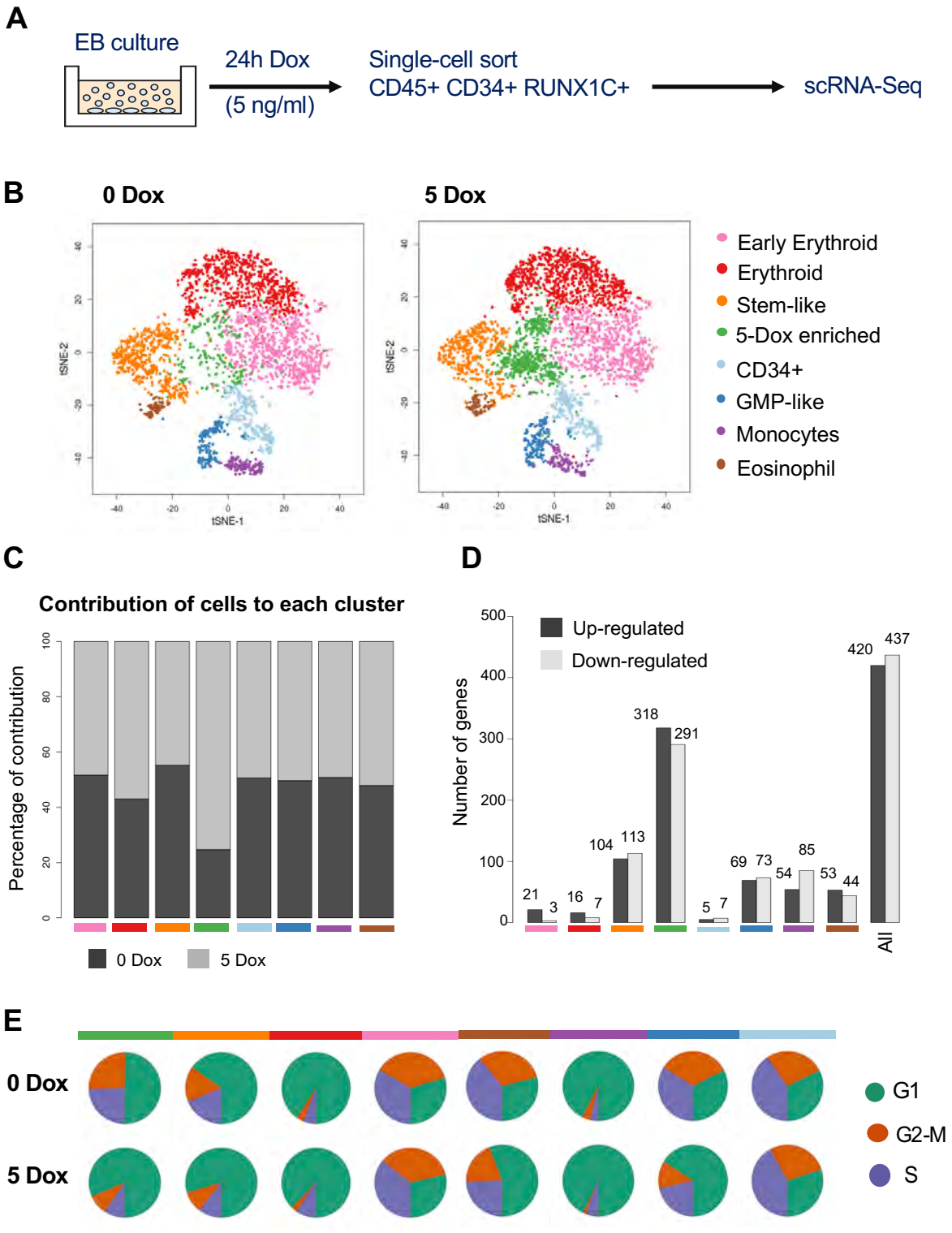


Figure 3.35: RUNX1-ETO induction expands a cell population arrested in the G1 phase of the cell cycle and results in a cell-specific transcriptional response

(A) Diagram of the sorting strategy for scRNA-Seq performed upon 24h Dox induction of RUNX1-ETO with 5 ng/ml Dox.

(B) Two-dimensional t-SNE maps displaying sequenced CD45+CD34+RUNX1C+ single cells upon 0 and 5 ng/ml Dox treatment (24h), respectively. Colours represent the different clusters identified based on expression of known marker genes.

(C) Proportion bars showing the percentage of contribution of cells from each 0 and 5 ng/ml Dox dataset to each individual cell cluster. Cell cluster colour coding as in (B).

(D) Number of up- and down- regulated genes in each cell cluster upon treatment with 5 ng/ml Dox. Cell cluster colour coding as in (B).

(E) Pie plots displaying the proportion of cells in each cell cycle phase (G1, G2-M and S) within each cluster, as identified by expression of cell cycle responsive genes. Cell colour coding as in (B).

We then characterized the cell type-specific response to induction of RUNX1-ETO by evaluating whether single genes were differentially regulated in each cell cluster. We therefore projected the expression levels of individual haematopoietic genes on the cell cluster map (Figure 3.36). Not surprisingly, the majority of differentially expressed genes between the two datasets (0 and 5 Dox) occurred in the 5-Dox enriched cells. The 5-Dox enriched population presented reduced expression of *Spi1* (PU.1) expression, high expression of *SOX4* and abrogation of *CEBPA, GFI1B* and *GATA2*, as compared to the other cell clusters.

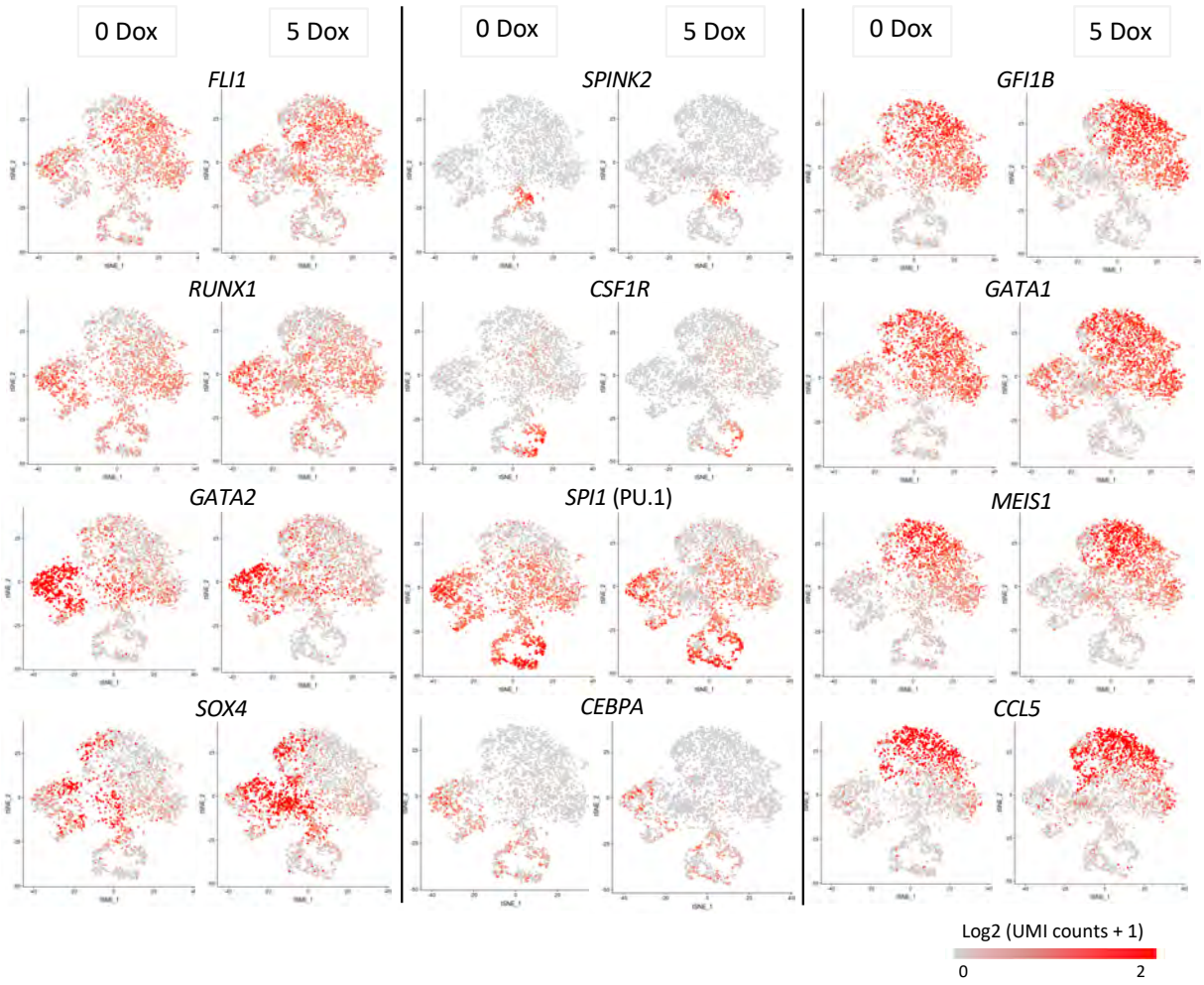


Figure 3.36: RUNX1-ETO induction leads to cell-type specific changes in gene expression mainly in the 5-Dox enriched cell cluster

Expression of individual marker genes projected on the t-SNE maps of both untreated (0 Dox) and treated (5 Dox; 5ng/ml for 24h) scRNA populations. Colour intensity represents number of transcripts sequenced in Log2 of unique molecular identifier (UMI) counts +1.

In order to precisely identify the most RUNX1-ETO-responsive genes and compare their transcriptional behaviour between cell clusters, we quantified differential gene expression in single cells from each cell cluster (Figure 3.37). We quantitatively confirmed the upregulation of *SOX4* as well as downregulation of *GATA2* within the 5-Dox enriched and the stem-cell like populations upon induction of RUNX1-ETO. Expression of the erythroid lineage genes *GFI1B* and *GATA1* appeared reduced in

the 5-Dox enriched and myeloid populations but not in the erythroid cell clusters. Two of the most striking changes were the downregulation of *Spi1* (PU.1) and the complete abrogation of *CEBPA* expression, which occurred only in the 5-Dox enriched cells. Expression of the mature myeloid gene *CSF1R* was downregulated in the myeloid lineages and completely shut down in cells of the 5-Dox enriched cluster. The fact that myeloid regulators were less perturbed in monocytic cells than in the 5-Dox enriched population suggests that cells that have passed a certain differentiation stage may become less sensitive to perturbation. Expression of HSC-specific and early haematopoietic genes, such as *FLI1*, *MEIS1* and *SPINK2*, remained unaffected in all cell clusters, with the latter being downregulated in the GMP population only. In the same way, expression of the master haematopoietic regulator RUNX1 did not change upon induction of RUNX1-ETO, which suggests the existence of a specific window for RUNX1-ETO perturbation during differentiation. Interestingly, the 5-Dox enriched population upregulated the inflammatory chemokine gene *CCL5*. In agreement with the prior analyses, the erythroid populations did not show much response to RUNX1-ETO induction.

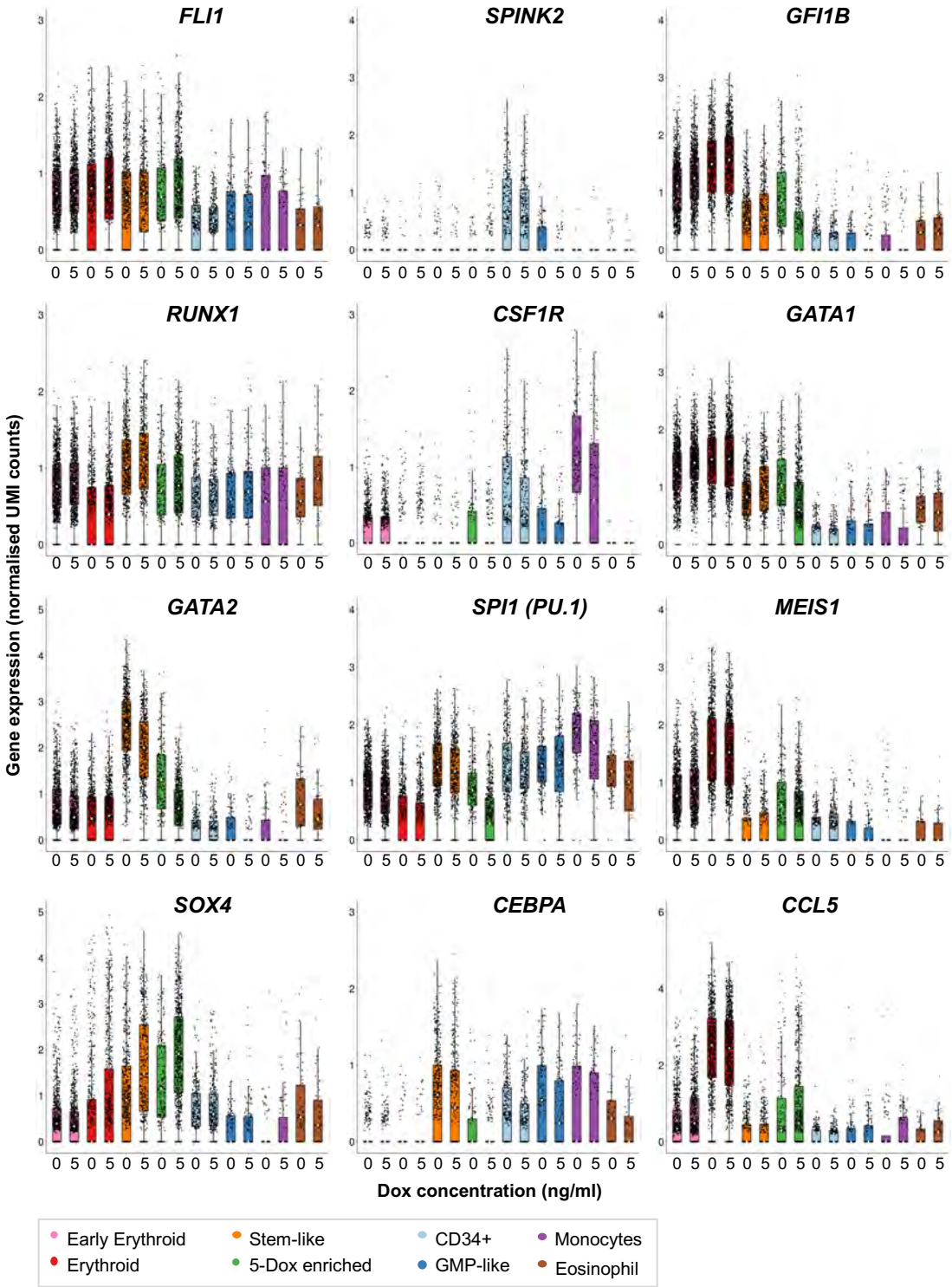


Figure 3.37: RUNX1-ETO induction of results in the upregulation of *SOX4* and downregulation of regulators of myelopoiesis in the 5-Dox enriched population

Box plots indicating expression levels of the individual marker genes in Figure 3.36 in the different ESC derived cell populations (colour coded) and both datasets (0 and 5 Dox). Black dots represent transcript levels in an individual single cell. Boxes and bars represent the mean and standard deviation, respectively.

In prior analyses, comparison of dysregulated RUNX1-ETO targets in our RUNX1-ETO-induced human ESC-derived progenitors to differential gene expression in t(8;21) AML patients (as compared to CD34+ wild-type counterparts), showed that both RUNX1-ETO-expressing cell populations share a similar expression profile. In our scRNA-Seq analyses we identified the 5-Dox enriched cell population as the most responsive to RUNX1-ETO induction, hence most likely to represent the specific RUNX1-ETO target cell population. We therefore wanted to compare RUNX1-ETO target genes that respond to induction in the 5-Dox enriched cell cluster to the gene expression profile in t(8;21) leukaemic patient samples. Gene Set Enrichment Analyses (GSEA) in these cell populations showed a strong correlation, with most of the up and downregulated RUNX1-ETO target genes in the 5-Dox enriched population also being deregulated in t(8;21) patients in the same fashion (Figure 3.38A). The pattern of differentially expressed genes in the 5-Dox enriched population is similar to that of the overall CD45+CD34+RUNX1C+ population, when compared to data from our bulk RNA-Seq analysis (Figure 3.38B). These results confirm the prior observation that RUNX1-ETO deregulates similar targets in our selected subpopulation of *in vitro* progenitors as in t(8;21) patient cells and therefore strengthens the argument of the 5-Dox enriched population representing the early progenitor cell that acquires the translocation event.

Gene expression within the bulk progenitor population identified a downregulation of cell cycle and DNA replication as well as an upregulation of signalling pathways in response of RUNX1-ETO induction. In order to determine the RUNX1-ETO-deregulated pathways within the 5-Dox enriched cell cluster, we again performed KEGG pathway analysis in the differentially expressed genes within this specific

population (Supplementary Figure 14). Deregulated pathways were consistent with the prior analyses in the bulk population, since induction of RUNX1-ETO resulted in downregulation genes with roles in regulating cell cycle (such as *CCND3* and *CDK1*) and myelopoiesis (such as *MYC*). Interestingly, it also led to the downregulation of spliceosome and ribosomal genes. We also observed the upregulation of genes involved in signalling pathways, such as *MAPK13*, *JAK1* and *CDKN1A*. All genes associated to the KEGG pathways are listed in Supplementary Dataset 1.

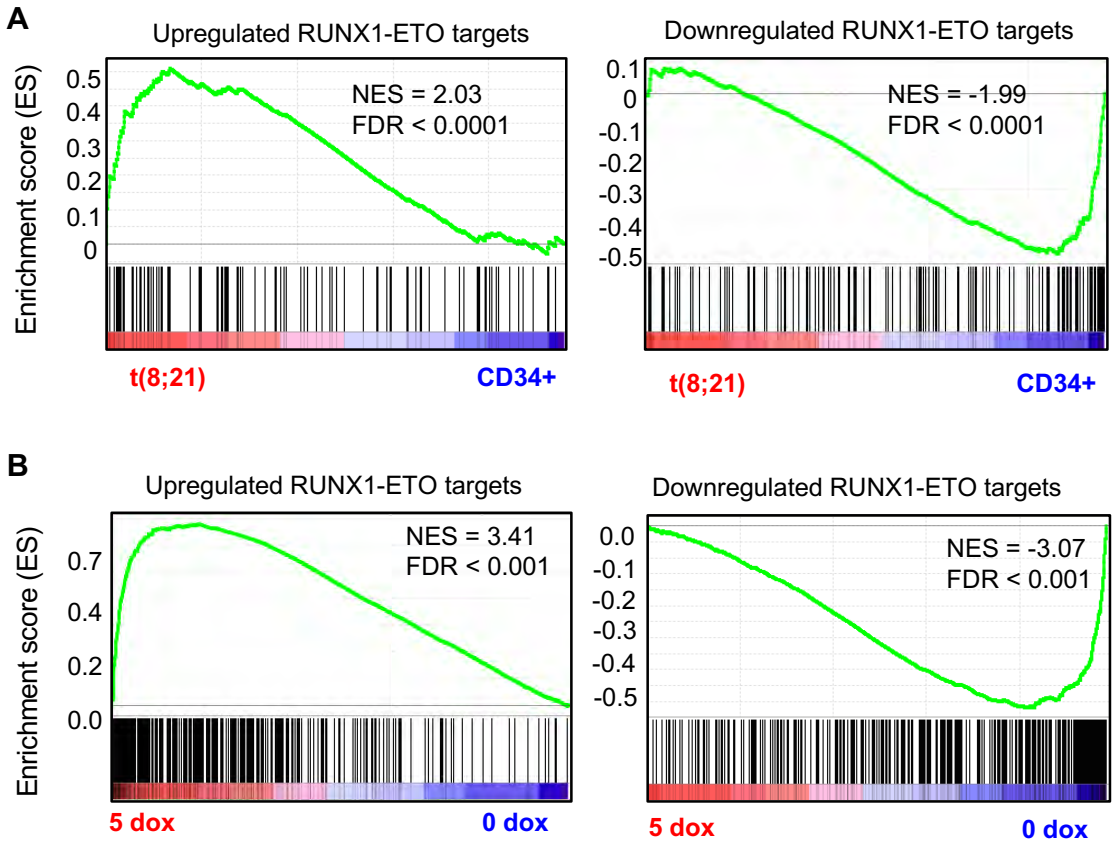


Figure 3.38: A large portion of the RUNX1-ETO-targeted transcriptional network in t(8;21) AML patients might be already deregulated in the pre-leukaemic population

Gene Set Enrichment Analysis for correlation of upregulated and downregulated gene signatures between the 5-Dox enriched single-cell sorted cells and (A) the gene expression profile of t(8;21) patient cells, and (B) induced CD45+CD34+RUNX1C+ bulk populations (5 ng/ml Dox for 24 hours). ES, Enrichment Score; NES, Normalized Enrichment Score. FDR, False discovery rate.

3.19 RUNX1-ETO induction disturbs the development of stem/progenitor and myeloid cells but not the erythroid differentiation trajectory

Following identification of the progenitor cell types composing the CD45+CD34+RUNX1C+ ES-cell derived population, we aimed to infer the developmental trajectories of the populations conforming these cells. In addition, we wanted to evaluate whether RUNX1-ETO induction interferes with the normal distribution of the distinct cell populations in the haematopoietic paths and to identify the position of the 5-Dox enriched cell cluster within the trajectories. We therefore performed a pseudo-time trajectory analysis, in both uninduced (0 Dox) and induced (5 Dox) scRNA-Seq datasets, by ordering the cells based on comparison of the complete set of cell marker genes identified (Figure 3.39). In uninduced conditions, progenitor cell populations are clearly distributed across the differentiation trajectories; with progenitors committed to the erythroid and myeloid lineages branching off the stem-like (orange) population and the most differentiated progenitor populations, monocytes (purple) and more mature/committed erythroid (red), at the tip of each lineage-specific branch (Figure 3.39, top panels). Intriguingly, induction of RUNX1-ETO not only resulted in the emergence of the 5-Dox enriched population but completely altered the differentiation trajectories (Figure 3.39, bottom panels). Consistent with the cell-specific responsive gene expression, the erythroid branch remained unaffected. However, induction of RUNX1-ETO distorted both the stem-like and myeloid branches, since cells belonging to those lineages appeared scattered all over the trajectory. Moreover, the 5-Dox induced population presented an additional small ramification within the trajectories that branched off before the erythroid lineages.

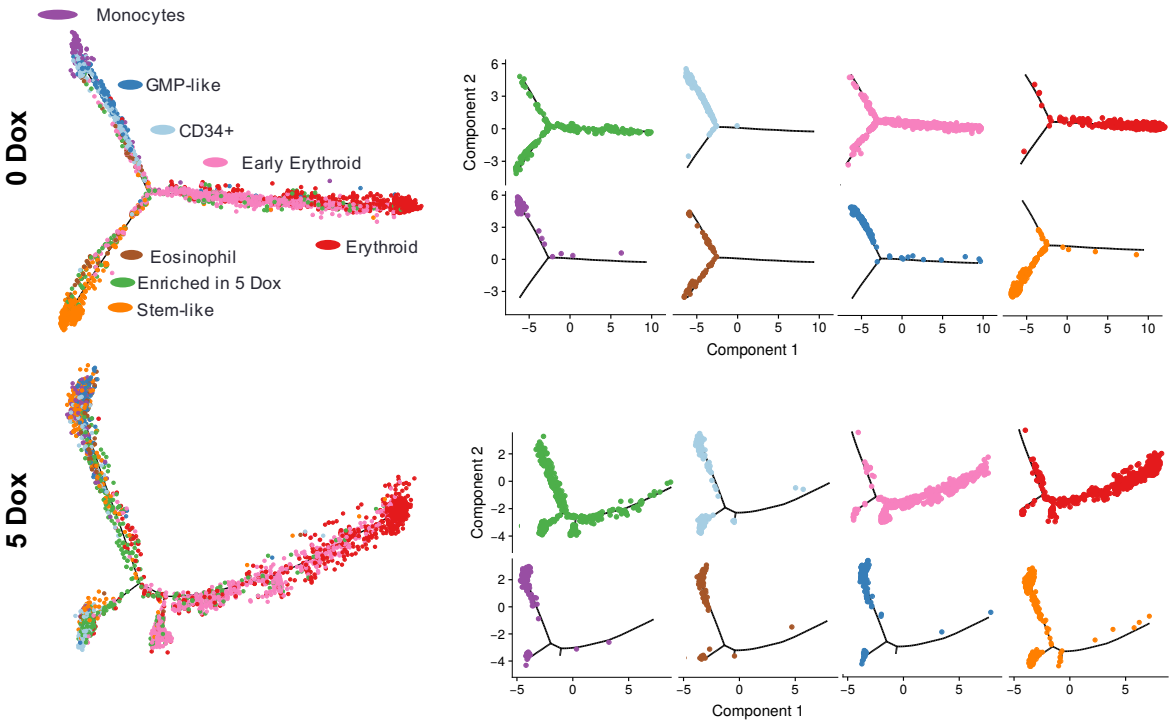


Figure 3.39: RUNX1-ETO induction distorts the myeloid but not the erythroid differentiation trajectory

Trajectory analysis using Monocle algorithm of the 0 dox (top panel) and 5 Dox (bottom panel) sorted cell populations plotted according to each cell cluster (same colour coding, as indicated in the 0 Dox panel). Right panels show the distribution within the trajectories of each cell clusters individually.

In order to better identify the shift in the position of the RUNX1-ETO-responding cells within the differentiation branches, we projected on the trajectories the expression of the individual marker genes shown in Figure 3.36, and also plotted them individually in selected cell clusters (Figure 3.40). Agreeing with prior analyses (Figure 3.37), we observed an upregulation of *SOX4* in cells of the stem-like and 5-Dox induced populations, which appeared scattered on the trajectories, but not on cells of other lineages such as monocytes. Not surprisingly, cells from the stem-like and 5-Dox enriched populations also presented downregulated expression of the lineage-specific *GATA2* and *SPI1* (PU.1), which also appeared disorganized across the differentiation paths. Despite the lack of changes in *RUNX1* and *FLI1* expression

levels (Figure 3.37), cells expressing these genes appeared shifted within the differentiation trajectory. Additionally, we did not observe changes in the cell distribution expressing genes of mature myeloid cells, such as *CSF1R*, suggesting a developmental window receptive to RUNX1-ETO transcriptional dysregulation.

Altogether, our single cell gene expression analyses show a distortion of the stem-progenitor cell programme upon RUNX1-ETO induction, with the emergence of a specific target cell population representing multipotent progenitor or a very early myeloid cell with a diffuse distribution across the differentiation trajectories. The enriched population responds to RUNX1-ETO induction by dysregulating genes crucial for stemness and early myeloid development, but not those for erythroid commitment and/or maturation, and, in agreement with preliminary assays, causes an arrest in the G1 phase of the cell cycle and a block in differentiation.

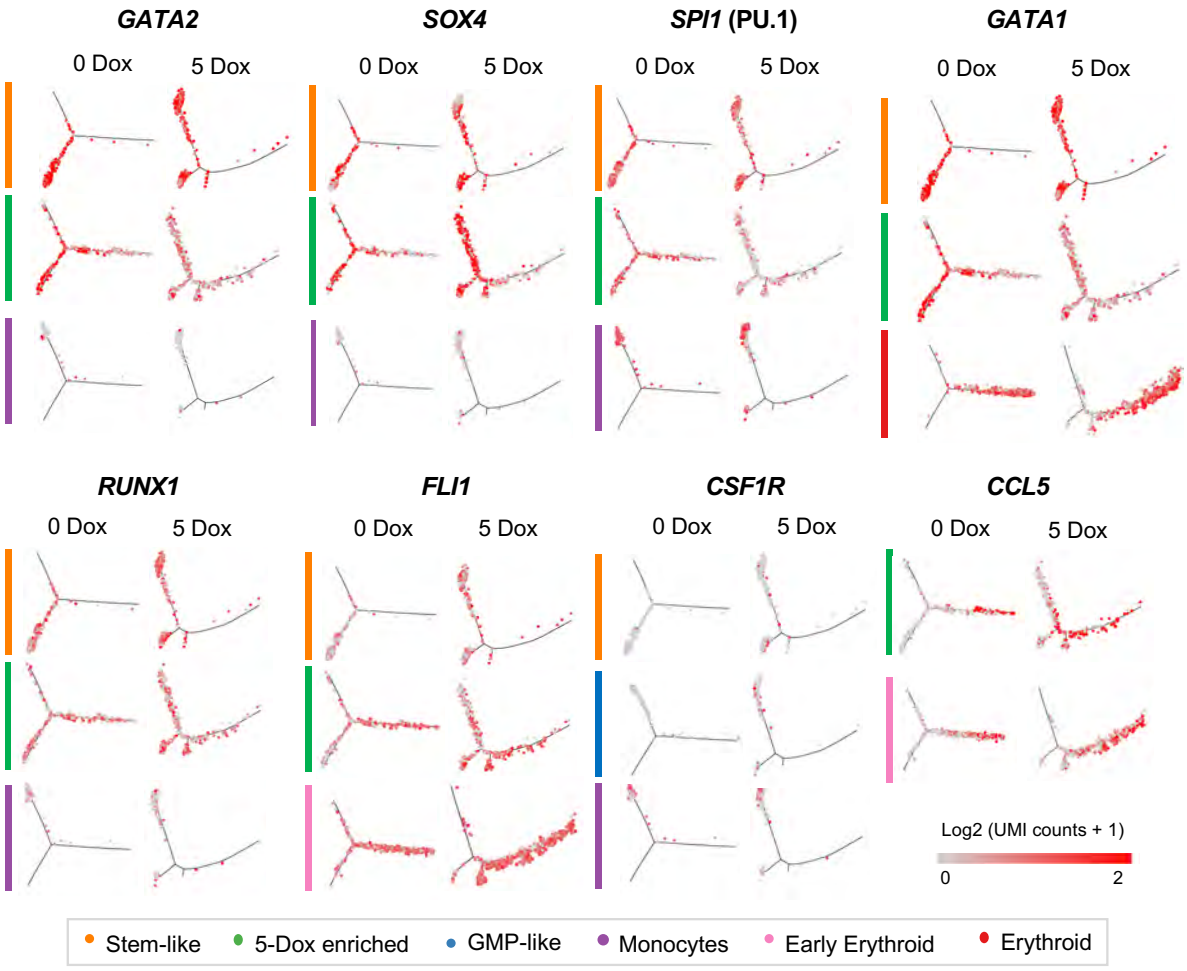


Figure 3.40: Induction of RUNX1-ETO dysregulates genes involved in stem/progenitor development

Expression of individual marker genes projected on the trajectories plotted according to each cell cluster in (A), indicated with the respective colour-coded vertical bar besides each set of two panels (0 and 5 Dox). Each dot represents a cell expressing that individual gene and red colour intensity indicates number of transcripts sequenced in Log2 of unique molecular identifier (UMI) counts +1.

3.20 Introduction of mutated KIT (N822K) and K-RAS (G12D)

Human t(8;21) acute myeloid leukaemia is a multi-step process that occurs upon acquisition of self-renewal-conferring mutations within cells harbouring the RUNX1-ETO translocation (Higuchi et al., 2002; Kuchenbauer et al., 2005; Rhoades et al., 2000; Shima et al., 2014; Yuan et al., 2001). Additional mutations that are known to co-operate with RUNX1-ETO in transformation include activated *KIT (N822K)* or *K-RAS (G12D)* oncogenes (Higuchi et al., 2002). Moreover, K-RAS (G12D) has been shown to cooperate with RUNX1-ETO to initiate leukaemogenesis in a mouse model of human leukaemia (Zhao et al., 2014).

Aiming to further dissect the molecular features involved in the establishment of overt AML, *KIT (N822K)* and *K-RAS (G12D)* cDNA was introduced into the RUNX1-ETO cell line. Constructs were targeted into the safe harbour of GAPDH locus (Kao et al., 2016). In order to control KIT/K-RAS protein level independently of the Dox-inducible RUNX1-ETO, a modified *E.coli* DHFR module was introduced as a destabilization domain (DD) that targets the protein to the proteasome for degradation in the absence of the ligand Trimethoprim (TMP) (Figure 3.41A) (Banaszynski et al., 2006; Iwamoto et al., 2010). Both *KIT* and *K-RAS* showed TMP-independent expression across different clones, as compared to an untargeted control (Figure 3.41B). Given that KIT is a surface receptor, flow cytometric analysis was performed upon addition of 10 μM TMP for 2 days to evaluate the functionality of the induction system. Surface KIT was upregulated in three independent KIT-targeted clones upon addition of the ligand (Figure 3.41C). However, low levels of KIT were also observed in the absence of the ligand as compared to an untargeted control. This could be due to the protein transiently reaching the cell surface before being degraded. K-RAS protein

was successfully detected in undifferentiated cells upon TMP treatment (Figure 3.41D).

Two clones from both KIT and K-RAS cell lines were differentiated to haematopoietic progenitors and KIT and K-RAS protein levels were evaluated. K-RAS was successfully detected in blood progenitors upon TMP-induction independently of RUNX1-ETO Dox-dependent induction (Figure 3.41D). Note that K-RAS could not be detected in samples corresponding to 0, 3 and 5 ng/ml Dox treatment, most likely due to insufficient protein loading or suboptimal transfer from the gel to the membrane, since the GAPDH loading control band was not visible on the gel. Unfortunately, KIT could not be detected neither through flow cytometry nor Western blotting assay in the differentiated progenitors.

In order to discard a shut-down of the locus due to possible toxicity from the mutated KIT, targeted clones were subjected to Geneticin treatment to re-select for resistance cells, hence harbouring the transgene and displaying open chromatin in the locus. Experiments were repeated using the re-selected clones and KIT was again detected by flow cytometry on ES cell stage. However, protein upregulation could not be detected in haematopoietic progenitors. Possible explanations for this could be: (i) a change on the chromatin state during differentiation due to some specific sequences of the transgenes, (ii) a detrimental effect of the expressed transgene on the cells causing a forced closure of the locus, or (iii) the requirement for stage-specific splicing sequences for proper translation of the transgene.

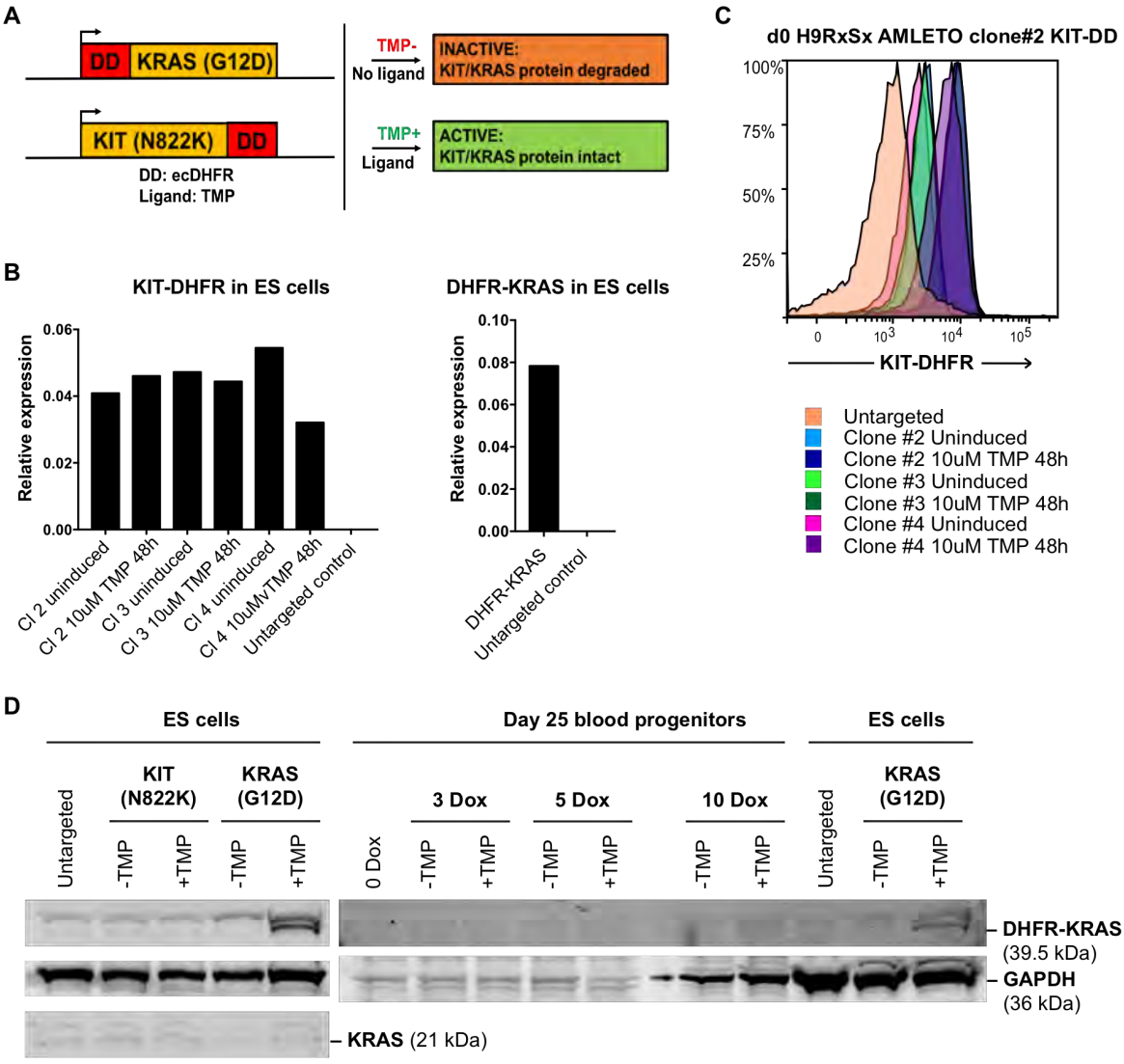


Figure 3.41: K-RAS(G12D) and KIT(N822K) expression in inducible RUNX1-ETO lines

(A) Diagram depicting the induction system of K-RAS (G12D) and KIT (N822K) targeted into the GAPDH locus of the RUNX1-ETO-targeted human H9 ES cells. An E.Coli DHFR, used as a Destabilization Domain (DD), was fused to the oncogenes prior to the locus targeting. In the absence of the ligand Trimethoprim (TMP), proteins are directed to the proteasome and broken down.

(B) Transgene expression from total RNA upon 2 days of TMP treatment, with relative expression to GAPDH. K-RAS expression shown corresponds to clone #2.

(C) Flow cytometric detection of KIT performed on untargeted RUNX1-ETO clone#2 cell line, and three independent KIT-targeted clones (2, 3 and 4) on both un-induced and induced conditions. Induction was performed with a 2-day 10 µM TMP treatment.

(D) DHFR-K-RAS (39.5 KDa) protein detection by Western blotting in both hES and differentiated haematopoietic progenitors at 48 hours after 10 µM TMP treatment. GAPDH (36 KDa) was used as a loading control. Untargeted parental cell line and KIT-targeted ES cells samples were used as a negative control. Dox treatments for RUNX1-ETO expression are an independent event to induction of K-RAS. The anti-K-RAS antibody also detected endogenous K-RAS (21 KDa).

4 DISCUSSION

4.1 A novel inducible RUNX1-ETO system representing an advantageous human model to recapitulate early oncogenic events

Transcripts of the t(8;21) translocation and *RUNX1-ETO*-expressing clones can be detected *in utero* and in post-natal blood samples, respectively, of asymptomatic patients, suggesting that cells that acquired the mutation might form a pre-leukaemic, non-proliferative clonal reservoir (Wiemels et al., 2010). Molecular events that occur in pre-leukaemic haematopoietic stem cells and progenitors in the developing embryo may not manifest in the same fashion as when the same oncogenic event occurs in adult counterparts (Barrett et al., 2016). A possible explanation for the distinct oncogenic mechanisms in both embryonic and adult progenitors may be the gene regulatory networks regulating the development and maintenance of mutated progenitors, which differ at the different developmental stages. AML patients may present minimal residual disease after chemotherapy, retaining HSCs harbouring the original t(8;21) translocation as the single oncogenic event that are the basis for relapse upon acquisition of secondary mutations (Shima et al., 2014). However, it is still not known how RUNX1-ETO reshapes the transcriptional network of these HSCs harbouring the translocation in remission. Therefore, we generated a model system of inducible *RUNX1-ETO* on *in vitro* generated embryonic haematopoietic progenitors to: (i) help to elucidate the changes that occur when the translocation is acquired during early development, unveiling the first steps of the leukaemic process, and (ii) improve the understanding of the molecular mechanisms implicated in the maintenance of the *RUNX1-ETO*-expressing progenitor cells, which may represent both the pre-leukaemic clone and/or that cellular reservoir present during remission.

However, it is important to highlight a caveat of our model system; despite the observed homogeneous induction of *RUNX1-ETO* in clones of undifferentiated ES cells, it is uncertain whether the oncogene induction works equally in all cell types during the haematopoietic differentiation protocol. This scenario would skew some of the data and therefore has to be taken into consideration for subsequent interpretations, as uninduced cells types could be the responsible of the observed effect after Dox addition to our cultures.

Until recently, the majority of *in vitro* human haematopoietic differentiation protocols were biased towards embryonic yolk sac haematopoiesis and were unable to generate HSC-like cells (reviewed in Ackermann et al. 2015). Aiming to attain the ideal *in vitro* model for the mechanistic study of HSC development and the modelling of childhood leukaemia, the laboratory of Prof A Elefanty has developed a protocol for the generation of CD34+RUNX1C+ definitive haematopoietic progenitors from *HOXA+SOX17+* haemogenic endothelium, hence recapitulating the molecular events leading to definitive haematopoiesis within the AGM (Ng et al., 2016). Although cells derived from this protocol have not shown evidence of long-term repopulation capacity, they show similar expression of cell surface markers, signalling molecules and transcriptional profiles to corresponding cells sorted from human 5-week AGM. However, similarities in the chromatin structure between our *in vitro* generated progenitors and stem and haematopoietic progenitors were not determined. Here, we further characterized the nature of the CD45+CD34+ populations generated in our cultures based on their expression of *RUNX1C* (GFP+). Expression of the *RUNX1C* isoform is restricted to haematopoietic cells and marks CD34+ cells with high clonogenic activity and with ability to home to the bone marrow (Ng et al., 2016).

Moreover, *RUNX1C* is the dominant isoform in foetal liver blood progenitors and is expressed in developing T-cells (Sroczynska et al., 2009; Telfer and Rothenberg, 2001). Here we show that our *in vitro* CD45+CD34+RUNX1C+ progenitors share an accessible chromatin pattern with human adult HSC and multipotent progenitor cells (Corces et al., 2016), but not with monocytes. This finding strengthens the significance of our differentiation model over other systems and sheds light onto the nature of the CD45+CD34+RUNX1C+ progenitor population, resembling adult immature progenitors.

The comparison of the transcriptional profiles of our *in vitro* generated RUNX1C+ and RUNX1C- populations of CD45+CD34+ progenitors showed over 500 differential expressed genes, demonstrating the different nature of the two cell populations. It is important to take into consideration the delay between the expression of the RUNX1C-reporter gene, GFP, and the formation of the detectable fluorescent chromophore (Tsien, 1998). Therefore, the earliest cells acquiring RUNX1C+ expression may have been considered as RUNX1C- during our GFP-based cell sort due to the time required for GFP maturation. However, these cells would represent a very small percentage of the whole sorted population, not materially affecting the outcome of the analysis. Having a closer look at the differentially expressed genes, we identified expression of HSC-specific genes (*SOX4, GATA2, LYL1, MYB, NFE2, MEIS1* and *MEIS2*) in the RUNX1C+ population (Orlic et al., 1995; Zhang et al., 2014) as well as genes regulating erythroid development (*GATA1*, *GFI1* and *KLF1*). GATA2 is important for HSC generation, function and maintenance (De Pater et al., 2013). HSC function, at both foetal and adult developmental stages, is also dependent on expression of *LYL1* (Capron et al., 2006; Souroullas et al., 2009). The

homeobox gene *MEIS2* is expressed in HSC and in maturing erythrocytes, but not in granulocyte-monocyte progenitors (GMP) (Miller et al., 2016; Wang et al., 2018). Whilst G*ATA2* is expressed in HSCs and multilineage progenitors, GATA1 marks erythroid, megakaryocyte, eosinophil, mast and dendritic cells and it is crucial for erythroid differentiation (Weiss et al., 1994). The switch from *GATA2* to *GATA1* expression regulates expression of early and late erythroid genes, with GATA1 essential for terminal erythroid differentiation (Grass et al., 2003). GFI1 has roles in regulating the formation of haematopoietic progenitors during the EHT, HSC self-renewal and erythropoiesis (Kim et al., 2014b; Lancrin et al., 2012; Zeng et al., 2004). KLF1 regulates definitive erythropoiesis and expression of β-globin genes (Hodge et al., 2006). Therefore, based on expression of lineage marker genes, the RUNX1C+ population may include multipotent progenitors as well as cells from the early myeloid and erythroid lineages. Conversely, the RUNX1C- population displayed expression of *CD14, KLF4, MS4A7, VCAN, FGL2, FCGR3A, CSF1R*, which are genes expressed in cord blood monocytic cell populations (Frankenberger et al., 2012; Zhao et al., 2019). RUNX1C- cells expressed genes essential for monocytic differentiation, such as *KLF4* and *IRF8* (Kurotaki et al., 2013) together with primitive macrophage genes, such as *CEBPB, IL10* and *IL13RA1* (Ruffell et al., 2009), lymphoid lineage genes including *NKG7*, which marks natural killer cells, and *IL7R*, which is expressed in T cells (Zhao et al., 2019). Thus, the RUNX1C- population may include common lymphoid progenitors as well as mature monocytes and macrophages. We found that RUNX1C+ cells expressed genes involved in pathways from distinct haematopoietic lineages, whilst cells lacking *RUNX1C* expression were mainly enriched in pathways related to distinct lymphoid cell functions as well as

complement and coagulation cascades, indicating presence of more mature megakaryocytes. Adding to the transcriptional differences between RUNX1C- and RUNX1C+ cells, these populations also differed in their chromatin structure, confirming that *RUNX1C* expression marks distinct haematopoietic lineages. Transcription factor motif analyses within accessible chromatin regions showed enrichment in GATA and AP-1 binding motifs within the RUNX1C+ specific sites, which correlated with upregulated gene expression.

Altogether, analyses of differentially enriched pathways and transcription factor binding motifs suggest that RUNX1C+ marks populations of multipotent progenitors with multilineage capacity and early myeloid- and erythroid-committed progenitors, whilst RUNX1C- progenitors may represent progenitors committed to the lymphoid, monocyte/macrophage and megakaryocyte/erythrocyte lineages. While we cannot rule out the possibility of RUNX1C+ also including a mixed population of more committed progenitors (see single cell analysis), our molecular analyses are in concordance with phenotypic observations from Prof Elefanty's laboratory of progenitors generated using this differentiation system (data not shown). They observed that *RUNX1C* was expressed in myelo-erythroid progenitors and was downregulated very early during commitment to the lymphoid lineage and during late erythroid differentiation. In addition, RUNX1C has been described as a master regulator of human haematopoietic development through upregulation of pro-inflammatory signalling pathways, which have been previously described to modulate HSC emergence (Navarro-Montero et al., 2017). Therefore, our comparison suggests that the RUNX1C+ population contains the target cells of interest to study the effects

of RUNX1-ETO induction, representing a multipotent progenitor with capacity to develop into cells from the myeloid and erythroid lineage.

4.2 The transcriptional response to RUNX1-ETO induction is dose- and cell type-specific

Overexpression of *RUNX1-ETO* in our *in vitro* human differentiation cultures using high levels of Dox (100-500 ng/ml) abrogates blood formation, as previously demonstrated in the murine ES cell differentiation system (Regha et al., 2015), and here we also show that it disrupts vasculogenesis. Small numbers of multilineage haematopoietic progenitors have been detected in the foetal liver of *RUNX1-ETO* knock-in mice (Okuda et al., 1998; Yergeau et al., 1997), suggesting that *RUNX1-ETO* could be expressed at levels that allow blood formation. Moreover, the low levels of *RUNX1-ETO* transcripts remaining in the blood of t(8;21) AML patients in long-term remission (Miyamoto et al., 2000), suggest that these residual levels of RUNX1-ETO are compatible with near normal cellular function. A more recent study has shown in a human iPSC model system that RUNX1-ETO overexpression results in cell death (Mandoli et al., 2016), again suggesting that the leukaemogenic effect of RUNX1-ETO is dose-sensitive. Therefore, failure of blood formation in our cultures upon induction of RUNX1-ETO reflects impaired haematopoietic differentiation resulting from high levels of *RUNX1-ETO* expression. Prior studies have demonstrated the requirement for a balanced expression of *RUNX1-ETO* and endogenous *RUNX1* to allow cell growth and survival as well as to maintain the RUNX1-ETO-driven leukaemic phenotype (Ben-Ami et al., 2013; Loke et al., 2017; Ptasinska et al., 2014). After the EHT, induction of RUNX1-ETO at levels approximating those of *RUNX1* expression (in response to 5 ng/ml Dox) allowed haematopoiesis to progress without morphological distortion of the vasculature or significant disruption of blood generation. However, similar induction of RUNX1-ETO

to the same level (5 ng/ml Dox) before the EHT caused a marked disorganization of the vascular structures and reduced formation of haematopoietic progenitors. This result confirms, in the human system, the RUNX1-ETO-dependent block of the EHT previously shown in the mouse (Regha et al., 2015). GFI1 and GFI1B downregulate endothelial differentiation and promote the morphological changes of emerging haematopoietic progenitors during intra-embryonic EHT, and are required for the production of HSCs and haematopoietic progenitor cells in the murine AGM (Lancrin et al., 2012; Thambyrajah et al., 2015). Given that *GFI1* is a RUNX1 target (Lancrin et al., 2012), induction of RUNX1-ETO might interfere with the RUNX1-mediated upregulation of *GFI1* expression, preventing the downregulation of the endothelial program and the detachment of blood cells from the endothelium, which could explain the disorganization of the vasculature and the lack of formed progenitors in culture.

In agreement with previous literature (Mandoli et al., 2016), we observed a dose-dependent RUNX1-ETO driven decrease in cells expressing markers of more mature myeloid cells, such as granulocytic CD16+ cells. Moreover, balanced RUNX1-ETO induction also promoted an increase in numbers of a cell population presenting a CD34+CD38-/loCD90+ signature. These surface markers are expressed, together with *CD45* and *RUNX1*, in embryonic cell populations containing the first few definitive human HSC (Ivanovs et al., 2014). Altogether, we have shown a RUNX1-ETO-mediated differentiation block that results in the accumulation of early progenitor cells that are unable to progress to more mature lineages. Our results confirm, in the human system, the partial block in myeloid differentiation upon induction of *RUNX1-ETO* previously seen in *in vivo* and *in vitro* mouse studies

(Regha et al., 2015; Rhoades et al., 2000). Our observations invite us to hypothesize that, during embryogenesis, the translocation event might occur after the EHT in a subpopulation of progenitors where RUNX1-ETO might remain expressed at low levels, allowing normal cell function and cell differentiation until the acquisition of collaborative molecular aberrations.

Prior studies in the mouse ES cell differentiation model have demonstrated that the RUNX1-ETO-mediated transcriptional dysregulation depends on the target cell that acquires the oncogenic event (Regha et al., 2015). Therefore, understanding the molecular outcome of RUNX1-ETO induction in the distinct cell populations obtained in our *in vitro* system is important to identify the target cell that more closely resembles the embryonic progenitors that acquire the translocation event and establish the pre-leukaemic clone. Similar to the murine studies, which showed that induction of RUNX1-ETO elicited a different response before and after the EHT, we observed distinct genomic responses to RUNX1-ETO induction in RUNX1C+ and the RUNX1C- cells. We found distinct accessible chromatin profiles in both induced populations, although the resulting gained and lost sites were enriched in similar transcription factor binding motifs, including loss of RUNX1, PU.1 and C/EBP and gain of AP.1 binding motifs. This observation suggests that RUNX1-ETO-might mediate transcriptional deregulation through a common mechanism in both populations. Analysis of pathways associated with the differentially expressed genes highlighted a cell-specific outcome of RUNX1-ETO induction; RUNX1C+ cells upregulated genes from several signalling pathways and downregulated genes associated to regulation of cell cycle, DNA replication and repair and haematopoietic lineages, whilst RUNX1C- cells only upregulated a small subset of those genes

observed in RUNX1C+ cells and downregulated genes associated to several haematopoietic functions. Thus, based on comparison of the differential gene expression, the RUNX1C+ population was identified as the most responsive to RUNX1-ETO induction. Overall, our results indicate that the resulting gene expression profile upon RUNX1-ETO induction is cell-type specific, but that the underlying epigenetic reprogramming mechanisms might be common in different cell populations.

Intriguingly, our RUNX1-ETO-induced progenitors presented a different global transcriptional profile compared to that of *RUNX1-ETO*-transduced CD34+ cord blood cells (Lin et al., 2017b) and t(8;21) patient samples (Ptasinska et al., 2012). Discrepancies in gene expression patterns between our induced progenitors and the other *RUNX1-ETO*-expressing cell populations could be explained by the different nature of the cells, whereby our progenitors represent a more immature phenotype and a different developmental stage. Moreover, distinct transcriptional profiles may also be indicative of the presence of secondary mutations in the patient samples. Importantly, when focussing on the expression of known RUNX1-ETO targets in our induced RUNX1C+ populations their expression overlapped with those in *RUNX1-ETO*-transduced CD34+ cord blood cells. This reflects basic similarities in the core transcriptional profile directly deregulated by RUNX1-ETO even in these developmentally disparate model systems. It is important to note that our *in vitro* derived immature progenitors might express genes directly bound by RUNX1-ETO not present in the data sets used for the comparison. Also, comparison data were obtained from more differentiated progenitors and hence these RUNX1-ETO targets might be underrepresented in our samples. Finally, all transduced cells are likely to

have undergone extensive clonal outgrowth and selection for appropriate RUNX1-ETO expression levels.

Altogether, our results show that the RUNX1C+ cell population generated in our *in vitro* system represents an immature embryonic progenitor that shares differentially expressed RUNX1-ETO targets with other *RUNX1-ETO*-expressing systems. Therefore, the RUNX1C+ cells were the population of choice for subsequent evaluation of the phenotypic and molecular changes driven by expression of *RUNX1-ETO*.

4.3 Balanced levels of RUNX1-ETO result in a reversible differentiation and growth arrest of human *in vitro* haematopoietic progenitors

In line with the observed RUNX1-ETO-mediated differentiation block, 7-day induced progenitors were unable to form colonies in the ongoing presence of Dox, as shown by CFU assays. Strikingly, the latent clonogenic capacity of these RUNX1-ETO-expressing cells was revealed following Dox removal, allowing cells to proliferate and differentiate, indicating reversibility of the differentiation block. These data suggest that expression of RUNX1-ETO leads to an accumulation of cell populations harbouring high clonogenic capacity that may remain in a quiescent stage, thus preventing its progression to further differentiated lineages. In line with our observations, a partial and reversible myeloid differentiation block at an early stage has also been observed in *in vitro* assays of mouse bone marrow cells upon removal of *RUNX1-ETO* expression (Rhoades et al., 2000). Our findings are also corroborated by a recent study that demonstrated that HSCs from RUNX1$^{ETO/+}$ mice present increased self-renewal and a quiescent transcriptional profile and phenotype, as compared to HSC from control mice (Di Genua et al., 2019).

Balanced long-term expression of *RUNX1-ETO* – using 3 and 5 ng/ml Dox – promoted the survival but not proliferation of a subset of cells during replating assays, allowing their maintenance in culture longer than 100 days. These results are again consistent with the work of *Rhoades et al.* in *RUNX1-ETO*-expressing murine bone marrow cells, showing an increase of self-renewal but not proliferation during replating assays (Rhoades et al., 2000). Conversely, previous studies in human CD34+ cord blood cells constitutively expressing *RUNX1-ETO* showed enhanced *in vitro* proliferation and maintenance of self-renewal and differentiation capacity

(Mulloy et al., 2002, 2003). However, in our experimental system none of the balanced levels of RUNX1-ETO induction showed cell expansion. The different outcome can be explained by adaptation of the CD34+ cord blood cells to cell culture, resulting from the outgrowth and associated selection of the most proliferative cells.

Induction of RUNX1-ETO in our cultures also resulted in a profound decrease in DNA-synthesis activity due to an arrest in the G1 phase of the cell cycle, as measured by BrDU incorporation. This finding confirms what has been recently shown in the mouse (Di Genua et al., 2019), indicating that RUNX1-ETO might confer a quiescent phenotype through an arrest in cell cycle. Prior studies have shown a decrease in the G1 to S phase transition upon RUNX1-ETO knockdown, along with impairment of engraftment, clonal expansion and proliferation capacity (Martinez et al., 2004; Martinez Soria et al., 2009). However, in both these studies the RUNX1-ETO knock-down: (i) was performed in the Kasumi-1 cell line, which harbours additional mutations to the t(8;21) translocation, and (ii) did not yield a complete depletion of *RUNX1-ETO* expression. *Martinez et al*. (Martinez et al., 2004) described induced cellular senescence in Kasumi-1 cells with reduced RUNX1-ETO levels. Interestingly, a role for RUNX1-ETO in inducing senescence-like growth arrest dependent on p53 has been previously described in primary fibroblasts (Wolyniec et al., 2009). Moreover, a follow up study also described a RUNX1-ETO-driven senescence-associated secretory phenotype that promotes the escape from senescence with subsequent immortalization of a subset of *RUNX1-ETO*-expressing murine fibroblasts (Anderson et al., 2018). It would therefore be interesting to

evaluate, in future experiments, whether RUNX1-ETO induces senescence -like growth arrest and associated secretory phenotype in our *in vitro* system.

A dual role of RUNX1-ETO in blocking differentiation and arresting cell growth has been previously described (Burel et al., 2001). In this study, RUNX1-ETO-dependent growth arrest resulted in apoptosis. However, RUNX1-ETO induction did not cause cell death via apoptosis in our *in vitro* human progenitors. It has been previously shown that elevated levels of *RUNX1-ETO* expression result in RUNX1-ETO-mediated apoptosis (Ben-Ami et al., 2013; Mandoli et al., 2016), explaining the increased cell death observed in the former study, as they used non-physiological levels of RUNX1-ETO. In their study, *Burel et al.* used the U937 cell line, which was generated from monocytic cells isolated from a lymphoma patient and hence represents a different type of myeloid leukaemia. Moreover, the U937 cell line most likely carries additional mutations, accounting for distinct results as compared to our study.

Taken together, we have shown that a balanced level of RUNX1-ETO induction in our *in vitro* generated human haematopoietic progenitors: (i) does not disrupt blood formation but blocks differentiation at an early stage, (ii) does not increase cell death via apoptosis, (iii) maintains cells with higher clonogenic capacity in a quiescent stage, (iv) confers survival but not proliferation to a subset of blood progenitors and (v) results in an arrest of cells in the G1 phase of the cell cycle. The combined observation of these phenotypes in *in vitro RUNX1-ETO*-expressing human progenitor cells has been only possible in our differentiation system but not in other human pre-leukaemic models, as discussed above, which strengthens the significance of our experimental approach as a *bona-fide* pre-leukaemic model.

There are several potential reasons that could explain the discrepancies of our data with previous findings. The first reason may be differences in the expression levels of *RUNX1-ETO* used. We expressed *RUNX1-ETO* at a level equivalent to that of the endogenous *RUNX1*, whilst other studies induced RUNX1-ETO at higher levels (Mulloy et al., 2002, 2003). Secondly, the different nature of the cells where the oncogene is expressed. Our system uses definitive multipotent progenitor cells at an embryonic stage of development (Ng et al., 2016), whilst previous experiments used either a later source of multipotent progenitors – CD34+ human cord blood – or mouse or human progenitors resembling those of generated in the yolk-sac, which are unlikely to represent the cell normally targeted by RUNX1-ETO (Mandoli et al., 2016; Mulloy et al., 2002, 2003; Regha et al., 2015). Lastly, the fact that other systems were based on ectopic oncogene expression that relied on the subsequent outgrowth and proliferation of the cells in the culture dish, implying the requirement for a selection step and hence making it difficult to distinguish the *RUNX1-ETO*-driven phenotype from that resulting from additional adaptations to cell culture (Mulloy et al., 2002, 2003).

4.4 RUNX1-ETO induction leads to dose-dependent deregulation of genes associated to distinct pathways

RUNX1-ETO-expressing RUNX1C+ cells presented different patterns of transcriptional response depending on the oncogene dosage. Whilst the total number of dysregulated genes positively correlated with the levels of RUNX1-ETO induction, individual gene response was highly heterogeneous, which highlights the importance of inducing the oncogene at the right level. One individual gene example that appeared consistently upregulated at all three levels of RUNX1-ETO induction (3, 5, 10 ng/ml Dox) was *ERG*, a member of the ETS family of transcription factors. This finding is consistent with prior studies in our lab using Kasumi-1 cell lines showing *ERG* downregulation upon RUNX1-ETO knockdown (Ptasinska et al., 2014). ERG is important for the emergence of definitive HSCs as well as for enhancing the repopulating capacity of adult HSCs (Loughran et al., 2008; Ng et al., 2011) and is included in the gene expression signature of both HSC and LSCs (Eppert et al., 2011). Therefore, RUNX1-ETO might upregulate *ERG* as a mechanism to promote "stemness" and to prevent apoptosis within the pre-leukaemic clone. Another individual gene exemplifying the heterogeneity of the RUNX1-ETO response is *GFI1*, which is both a RUNX1 and a RUNX1-ETO target in our progenitor cells and appears upregulated only upon RUNX1-ETO induction with 10 ng/ml, even though it is also bound by RUNX1-ETO when induced with 5 ng/ml Dox. GFI1 displays context dependent roles but is also required for leukaemic maintenance, since reduced *GFI1* expression impairs initiation and progression of RUNX1-ETO-driven leukaemogenesis (Marneth et al., 2018). Therefore, high levels of GFI1 might play a

role during disease development by maintaining a pool of cells expressing increased levels of *RUNX1-ETO* transcripts.

Genes belonging to other common pathways responded to RUNX1-ETO in a similar fashion. For example, genes associated with the RAP1 signalling pathway were downregulated only in response to RUNX1-ETO induction with 10 ng/ml dox, thus indicating that the RUNX1-ETO dosage might dictate the resulting aberrant cell phenotype. This finding invites us to hypothesise that tightly regulated variation of the expression levels of the fusion gene through disease progression might be crucial for AML development, maintenance and relapse.

For the purposes of this work, we focused our molecular studies on induction of RUNX1-ETO using 5 ng/ml Dox in RUNX1C+ progenitors. We observed that RUNX1-ETO deregulates the same target genes in our progenitors as in cells from t(8;21) AML patients (Assi et al., 2019), which not only adds relevance to our study but suggests that the initial oncogenic event might account for a large part of the altered network in leukaemic blasts. Further evaluation of the RUNX1-ETO-responsive gene expression confirmed the downregulation of myelopoiesis, cell cycle, DNA replication and DNA repair genes. This molecular response is consistent with the RUNX1-ETO-dependent accumulation of progenitors at an immature stage, the absence of significant proliferative advantage conferred by RUNX1-ETO alone, and the block of transition to S phase. In addition, RUNX1-ETO induction resulted in upregulation of genes from multiple signalling pathways, including the MAPK and VEGF pathways, as previously observed in patient samples (Zaidi et al., 2009). Activation of signalling pathway genes may occur either as a survival response adopted by cells attempting to exit the RUNX1-ETO-dependent cell cycle arrest, as a direct RUNX1-ETO-

mediated mechanism to activate expression of early response genes necessary for leukaemic development, or as a combination of both.

Several studies provide supporting evidence for our findings, highlighting a balanced interplay between the RUNX1-ETO-dependent activated signalling pathway, cell cycle block and defective DNA repair. We found upregulation of the MAP kinase family members p38 (*MAPK13*) and ERK1 (*MAPK3*) as well as of the AP-1 members *JUN* and *JUNB*, which promote gene expression in response to a stimulus during early development and in distinct cellular contexts. *MAPK13* and *JUNB* were bound by RUNX1-ETO in our progenitors, suggesting the presence of direct mechanisms for the activation of signalling and early response genes. MAPK signalling upregulates expression of AP-1 transcription factors (Whitmarsh and Davis, 1996), which in turn upregulate *CCND2* expression (Mathas et al., 2002) resulting in cell cycle activation (Zhang et al., 2018). A recent study in Kasumi-1 cells has also shown a RUNX1-ETO-dependent induction of *JUN*, with subsequent upregulation of genes involved in the control of G1 progression, including *CDK6* and *CCND1* (Martinez-Soria et al., 2019). Upregulation of *CDK6* and *CCND1* was dependent on binding of both AP-1 and RUNX1-ETO with higher expression levels of these two genes found in cells from t(8;21) AML patients, compared with cells from other AMLs. Conversely, we found downregulation of *CDK6* and *CCND1* in a Dox-dependent manner, which was not dependent on direct RUNX1-ETO binding. In line with this, we observed strong upregulation of the cell-cycle inhibitor p21 (*CDKN1A*), as previously described (Berg et al., 2008; Peterson et al., 2007). *CDKN1A* was bound by RUNX1 in uninduced conditions but RUNX1-ETO bound in its place upon induction. The RUNX1-ETO-driven upregulation of p21 (*CDKN1A*) and downregulation of *CDK6* are

consistent with the observed G1 cell cycle arrest in our *in vitro* progenitors. It has been previously shown that expression of p21 is crucial for maintaining self-renewal of LSCs through restricting cell cycle and limiting DNA damage (Viale et al., 2009). This study demonstrated that RUNX1-ETO expression in murine HSCs induces DNA damage and activates a p21-dependent cell cycle block that, in turn, results in repair of the damaged DNA, leading to the accumulation of genomically unstable HSCs. Maintenance of DNA integrity by different pathways is regulated during the cell cycle (Rothkamm et al., 2003; Takata et al., 1998), as exemplified by suppression of the BRCA1-driven DNA end resection and homologous recombination of DNA double strand breaks during the G1 phase (Escribano-Díaz et al., 2013). We found downregulation of cell cycle genes, such as *BUB1*, *CCNE2* and *CCNA1,* which were also bound by RUNX1-ETO. RUNX1-ETO induction also resulted in the downregulation of genes involved in DNA double strand break repair upon induction, agreeing with studies using *RUNX1-ETO*-expressing cell lines (Forster et al., 2016). We show that these DNA repair genes (such as *BRCA1*, *BRCA2*, *CHEK1* and *CHEK2*) were not targeted by *RUNX1-ETO*, and some of these genes (such as *RAD54L*, *RAD51AP1* and *BLM*) displayed RUNX1 binding in uninduced conditions, which was then lost upon induction without gain of RUNX1-ETO binding. These findings agree with the idea of RUNX1-ETO mediating a cell cycle arrest through direct binding and transcriptional repression of cell cycle regulators, which may lead to the indirect downregulation of DNA repair genes. In turn, reduced DNA repair could result in genomic instability and hence the accumulation of secondary mutations. It has been shown that *RUNX1-ETO*-expressing murine bone marrow cells display downregulation of genes involved in DNA repair via homologous

recombination, resulting in increased DNA damage (Esposito et al., 2015). RUNX1-ETO predisposes cells to acquisition of genetic alterations, spontaneously and after treatment with DNA damage-inducing agents, with a resulting mutator phenotype that positively correlates with its expression levels (Forster et al., 2016). In our data, however, expression of RUNX1-ETO is low as compared to other overexpression systems, which might prevent the excessive accumulation of DNA damage in our cells. Interestingly, we found upregulated expression of the Growth arrest and DNA-damage-inducible protein 45 gamma (*GADD45G*) upon RUNX1-ETO binding. As its name indicates, GADD45G acts as a tumour suppressor by responding to DNA damage stress via activation of the S-phase checkpoint, hence inhibiting cell cycle and resulting in suppression of cell growth and colony formation (Vairapandi et al., 2002; Ying et al., 2005), thus supporting the phenotype observed in our cultures.

Taken together, our data suggests that RUNX1-ETO regulates the interplay between cell cycle, DNA repair and cell signalling in order to ensure long-term survival of the pre-leukaemic clone and sensitise these cells to the acquisition of secondary mutations during for leukaemic development. Although seeming contradictory, the balance between the restricted DNA repair and the cell cycle arrest might favour a mutator phenotype, leading to the accumulation of cells carrying the translocation and variable levels of DNA damage (Rossi et al., 2007). In turn, low levels of *RUNX1-ETO* expression might prevent the excessive accumulation of DNA damage, which could result in exhaustion for the pre-leukaemic clone (Cheng et al., 2000), via the upregulation of *GADD45G* and hence cell growth restriction. The RUNX1-ETO-mediated activation of signalling pathways could establish a receptive environment for the acquired cooperating mutations, which, over time, could lead to by-pass of cell

cycle arrest, and subsequent proliferation. Importantly, re-activation of the cell cycle might then result in repair of DNA damage, leading to AML progression and therapeutic resistance (Bullinger et al., 2007; Liddiard et al., 2010).

4.5 RUNX1-ETO abrogates the RUNX1-mediated transcriptional program by interfering with RUNX1 binding

Transcriptional changes upon RUNX1-ETO induction were the result of extensive global reorganisation of the chromatin structure, including both loss and gain of accessible chromatin sites. In uninduced conditions, RUNX1-bound sites were enriched in H3K27ac and H3K4me3 active marks in both promoters and distal elements. Although H3K4me3 is widely known to mark active promoters, our data shows enrichment at distal elements as well, which might mark active enhancers correlating with high enhancer driven transcription (Henriques et al., 2018). Upon induction, RUNX1-ETO displaced RUNX1 from many of its binding sites and remained bound at the same site, resulting in reduced accessibility of those sites, which correlated with downregulated gene expression. H3K27ac and H3K4me3 were reduced mainly at distal regulatory elements, confirming the direct interference of RUNX1-ETO with the binding of RUNX1. Given that the combination of active distal regulatory elements defines cellular identity (Corces et al., 2016; Heinz et al., 2010), our data suggest that RUNX1-ETO may corrupt lineage specificity and reprogram the cellular identity of the target cell. Depletion of RUNX1 binding by RUNX1-ETO might be not only the cause for the block in myeloid differentiation, but for the disruption of many other crucial cellular functions that are transcriptionally regulated by RUNX1. This can be exemplified by the upregulation of Early growth response gene (*EGR1)*, which maintains quiescence and long-term self-renewal of HSCs (Min et al., 2008) and is bound by RUNX1-ETO in our *in vitro* progenitors as well as in t(8;21) cell lines (Fu et al., 2014). Here we also show that RUNX1-ETO upregulates *EGR1* expression by displacing RUNX1 and binding in its place. It has been previously shown that

presence of RUNX1 prevents RUNX1-ETO-mediated apoptosis by upregulating genes involved in cell cycle checkpoints (Ben-Ami et al., 2013; Loke et al., 2017; Mandoli et al., 2016). We showed that, whilst RUNX1-ETO partially abolishes the RUNX1-mediated program, it also upregulates genes inhibiting cell cycle, such as *GADD45G*. This could be an alternative mechanism to prevent apoptosis in compensation for the lack of RUNX1-driven transcriptional program, hence promoting the survival of the pre-leukaemic clone. Other studies show that absence of RUNX1 increases the number of quiescent HSC (Ichikawa et al., 2008) but it also reduces the replating capacity of bone marrow cells as compared to their *RUNX1-ETO*-expressing conditions (Ichikawa et al., 2004). These observations agree with our data and suggest that the combined suppression of the RUNX1-mediated gene expression program and upregulation of several pathways preventing cell proliferation might be key mechanisms in regulating the homeostasis of the *RUNX1-ETO*-expressing pre-leukaemic clone.

RUNX1-ETO was also bound to many chromatin sites that displayed reduced levels of RUNX1 binding and remained accessible. These sites correlated with upregulated gene expression, suggesting that RUNX1-ETO might also promote exchange of factors to dictate gene expression. Accordingly, our lab has previously shown that 60% of the RUNX1-ETO sites also display RUNX1 binding (Ptasinska et al., 2012). These sites could represent co-existence of RUNX1 and RUNX1-ETO binding within one individual cell or exclusive binding in different cells within our population, hence appearing as “common” binding events in the overall bulk ChIP data. Heterogeneous binding within our cell population may represent several scenarios: (i) cell-type specific regulation at those sites involving complementary transcription factors

hindering the RUNX1-ETO-mediated displacement of RUNX1 in some cell populations, (ii) binding of the RUNX1-ETO co-factor complex in adjacent motifs thus not overlapping to those where RUNX1 is bound, and (iii) limited duration of Dox treatment (24-hour) not allowing enough time for RUNX1-ETO to disrupt RUNX1 binding at specific sites equally within the cells of our population.

We also observed RUNX1-ETO binding at genes that were not previously bound by RUNX1. A possible explanation could be due to RUNX1-ETO binding to non-RUNX1 motifs indirectly, by means of other factors contained within the RUNX1-ETO cofactor complex. This is exemplified by E-box-mediated recruitment of RUNX1-ETO to a selection of genes, enabling RUNX1-ETO binding independently of its RUNT domain (Sun et al., 2013). Likewise, we could also see that, upon induction, RUNX1 binding was lost in sites that were not subsequently bound by RUNX1-ETO. It has been previously shown that RUNX1 is able to bind DNA via PARP-dependent poly(ADP-ribosyl)ation (PARylation), hence in the absence of motifs for binding through the RUNT Domain (Tay et al., 2018). This scenario could explain the lack of RUNX1-ETO binding at such sites. That said, RUNX1 PARylation is associated with DNA damage response and enables RUNX1 binding to DNA repair structures. The fact that we observe that these RUNX1 targets appear deregulated upon induction points out to an indirect mechanism for the loss of RUNX1 binding, such as via RUNX1-ETO-driven signalling changes or dysregulated effector proteins.

Induction of RUNX1-ETO also resulted in the appearance of new accessible sites that correlated with upregulated gene expression but were not bound by RUNX1-ETO. We observed widespread enrichment of binding motifs for members for the AP-1 transcription factor family through all the detected chromatin sites, including the

newly accessible sites that lacked RUNX1-ETO binding. This is consistent with prior observations of extensive binding of AP-1 members across the genome (Gomes et al., 2018; Obier et al., 2016; Ptasinska et al., 2019). Together with the observed upregulation of AP-1 genes through direct RUNX1-ETO binding, and maybe through activated signalling, we propose that transcriptional activation at those sites might be mediated through AP-1 binding. The presence of AP-1 is essential for the engraftment of t(8;21) AML patient cells in mouse xenotransplants (Assi et al., 2019), indicating that the subset of genes upregulated by AP-1 upon RUNX1-ETO induction might be essential for leukaemic development. In addition, the transcription factors GATA2 and GFI1 can cooperatively bind AP-1 motifs together with the AP-1 family member FOS (Gomes et al., 2018), hence adding more complexity to the indirect activator phenotype resulting from RUNX1-ETO induction.

In conclusion, we show that RUNX1-ETO abrogates a large part of the RUNX1-mediated transcriptional program, leading to repression of those genes. Expression of RUNX1-ETO also enhances the accessibility of other sites that do not themselves bind RUNX1-ETO but present gene upregulation, most likely through AP-1 dependent transcriptional activation. We believe that RUNX1-ETO might compensate for the absence of the RUNX1-dependent gene expression program by upregulating signalling pathways and downregulating genes that promote entry to cell cycle, thus allowing cell survival.

4.6 *RUNX1-ETO* expression blocks the differentiation of a specific cell population

Differential gene expression analyses in *RUNX1-ETO*-expressing RUNX1C+ and RUNX1C- populations showed a cell-type specific response. In order to fully decipher the molecular response upon acquisition of the first oncogenic event, it is essential to understand the nature of the cells susceptible to transformation and, ultimately, to identify the cell of origin where the translocation occurs. Our single cell gene expression data demonstrated that the CD45+CD34+RUNX1C+ population is composed of precursors from diverse blood lineages as well as multipotent cell progenitors. This data agrees with the gene expression and associated pathways identified in whole transcriptome analysis of the bulk population. Our current efforts are focused on further characterization of the identity of these multipotent progenitors. Recent studies based on single cell transcriptomic analyses of embryonic progenitors have shed new light onto differentiation trajectories and defined cell populations with multi-lineage potential that were previously unknown, as exemplified by the erythroid-megakaryocyte-primed MPP (EMPP) (Drissen et al., 2019) or the megakaryocyte-erythroid-mast cell progenitor (MEMP) (Popescu et al., 2019). Genes encoding transcription factors and receptors are dynamically regulated during differentiation of the more immature HSC/MPP populations to the earliest lineage specific progenitors. Therefore, comparison of selected individual marker genes to the embryonic populations described in these studies can help elucidate the identity of our “stem-like” and “CD34+” populations. We believe that our CD34+ population might represent the earliest neutrophil-myeloid progenitors, as it expresses *CD34*, *SPINK2*, and myeloid markers including *MPO* and *LYZ*. Our GMP-

like population expresses *LYZ*, *MPO* and *AZU1*, which appear expressed in neutrophil-myeloid progenitor populations, hence confirming its identity as GMPs. Finally, our stem-like population shares marker genes with those present in eosinophil progenitors but also expresses genes marking the megakaryocyte-erythroid-mast cell lineage, including *EPOR,* low levels of *GATA1,* and high *GATA2* expression (Popescu et al., 2019). These similarities suggest that the Stem-like population might represent embryonic EMPPs or MEMPs recently described in those studies, positioning this population as one of the most multipotent of our single cell analysis.

Induction of RUNX1-ETO resulted in enrichment of a population, referred as 5-Dox enriched cluster, which was barely present in the uninduced conditions. The 5-Dox enriched cluster was arrested in the G1 phase of the cell cycle, demonstrating that the RUNX1-ETO-driven cell cycle arrest is cell type specific. Our preliminary results in bulk populations suggested that this cell cycle arrest is also reversible. We found that 5-Dox enriched cells downregulated multiple cell cycle genes, confirming what was found in the bulk population. However, DNA repair pathways were not affected in the 5-Dox population, suggesting that deregulation of the cell cycle might be the main RUNX1-ETO-driven mechanism to promote survival of the pre-leukaemic clone, whilst downregulation of DNA repair genes might be a downstream effect.

RUNX1-ETO induction strongly distorted the myeloid but not the erythroid differentiation trajectories, indicating that RUNX1-ETO produces a cell-type specific block by hindering cell fate decisions at a specific differentiation stage. Accordingly, not all the single cell populations were susceptible to induction of the oncogene. The most responsive populations included stem-like and the 5-Dox enriched cells, whilst

early and maturing erythroid cell populations did not suffer many changes. One of the most dramatic gene expression changes was a severe downregulation of *SPI1* (PU.1) expression only in the 5-Dox enriched cells. PU.1 is an essential transcription factor for myelopoiesis (McKercher et al., 1996; Scott et al., 1994), thus its downregulation confirms the phenotypic differentiation block seen in our *in vitro* cultures. Conversely, levels of PU.1 remained unaffected in other myeloid lineages (CD34+, GMP-like, eosinophil and monocytic) upon RUNX1-ETO induction. This finding suggests that mature cells that have passed a certain differentiation state lose sensitivity to RUNX1-ETO-mediated perturbation, supporting the concept of a developmental window of leukaemogenic RUNX1-ETO response. This cell-stage specific differential response could be explained by activation of additional distal regulatory elements that do not require the presence of RUNX1, as shown for the autoregulation of the *SPI1* (PU.1) locus (Leddin et al., 2011), hence acting independently of the RUNX1-ETO-mediated abrogation of RUNX1 binding. Another mechanism could be the priming of promoters by other transcription factors, hence maintaining active gene expression. This is exemplified by binding of the C/EBP family of transcription factors to regulatory elements, forming stable complexes that maintain *SPI1* expression even in the absence of RUNX1 (Hoogenkamp et al., 2007). Accordingly, other myeloid genes were also significantly downregulated predominantly in the 5-Dox enriched population, including *CEBPA* and *CSF1R*, as well as the early erythro-myeloid regulator *GATA2*, confirming the differentiation block in this population at a multipotent progenitor stage. Induction of RUNX1-ETO also resulted in the upregulation of the stem cell regulator *SOX4* in the most undifferentiated populations, including the stem-like and the 5-Dox enriched clusters,

consistent with the idea of RUNX1-ETO promoting self-renewal of a pre-leukaemic clone at an immature myeloid progenitor stage. Downregulation of *CEBPA* might be responsible for the higher expression of *SOX4*, as described in HSCs from *CEBPA*-null mice and in patients with abnormal CEBPA function (Zhang et al., 2013a).

The enriched 5-Dox population upregulated expression of the inflammatory chemokine ligand *CCL5* (RANTES), which may promote cancer cell proliferation as well as the formation of a microenvironment favouring immune evasion (reviewed in Aldinucci and Colombatti, 2014). In addition, 5-Dox enriched cells expressed interferon (IFN)-stimulated genes, such as *IFIT1* and *STAT1,* and it was the only population displaying expression of the IFN-inducible gene *MX1*. Induction of genes related to the IFN/STAT pathway has been described as a feature of chemo-sensitive cancers linked to DNA damage-response (Legrier et al., 2016). However, IFN-stimulated genes, including *IFT1* and *MX1*, are associated with immune adaptive resistance of cancer cells and hence are an indicator of relapse (Benci et al., 2016). Innate immune signalling pathways, including the MAPK signalling pathway, can be activated in response to cellular stress mechanisms, such as DNA damage, and also contribute to adaptive resistance in AML (Melgar et al., 2019), again highlighting the interplay of different deregulated pathways in the RUNX1-ETO oncogenic phenotype. In addition, induction of RUNX1-ETO results in the upregulation of the Aryl Hydrocarbon Receptor (*AHR*) in the 5-Dox enriched population as well as in the other two populations of multipotent progenitors (stem-like and CD34+). Activation of *AHR* expression is used in some protocols to induce *ex vivo* human HSC expansion (Boitano et al., 2010), indicating that it may also promote the expansion of the *RUNX1-ETO*-expressing pre-leukaemic pool.

Taken together, RUNX1-ETO distorts the stem progenitor cell programme with the presumed RUNX1-ETO-target cell representing a multipotent progenitor or early myeloid cell that appears blocked in differentiation. This enriched subpopulation presents similar deregulated RUNX1-ETO targets to cells from t(8;21) patients and, therefore, strengthens the argument of the 5-Dox enriched population representing the early progenitor cell acquiring the primary leukaemic mutation. Over the past twenty years, numerous studies have suggested the existence of a rare population of stem cells or HSCs as the cell of origin for leukaemic transformation (Bonnet and Dick, 1997; Lapidot et al., 1994) and of relapse (Shima et al., 2014). Moreover, both HSCs and LSCs share a common gene expression signature with genes involved in "stemness" functions such as self-renewal and quiescence (Eppert et al., 2011). Given that the 5-dox enriched population (i) was constituted by a rare number of cells in uninduced conditions, (ii) had the highest transcriptional response to RUNX1-ETO induction, (iii) was arrested at the G1 phase of cell cycle, (iv) suffered upregulation of stemness and downregulation of early myeloid development genes and (v) shared a large proportion of deregulated RUNX1-ETO target genes with cells from t(8;21) AML patients, we infer that this population might represent an *in vitro* counterpart of embryonic HSCs that mimics the expansion of a quiescent pre-leukaemic clone upon induction of RUNX1-ETO.

4.7 Summary

In summary, we developed a human ES cell differentiation system able to generate definitive haematopoiesis to study the early events occurring in t(8;21) AML. We demonstrated that induction of RUNX1-ETO results in a dose-dependent and cell-type specific transcriptional response, with the presumed target cells being multipotent progenitors that appear to be blocked in differentiation and arrested in the G1 phase of cell cycle. We have found that RUNX1-ETO reprograms the epigenome in part by directly interfering with the RUNX1-mediated transcriptional program, resulting in repression of genes involved in the regulation of cell cycle, DNA repair and myelopoiesis (Figure 4.1). Abrogation of RUNX1 binding was accompanied by a reduction of the active histone marks H3K27ac and H3K4me3 mainly at distal regulatory elements, suggesting that RUNX1-ETO might corrupt cellular identity and providing a molecular explanation for the observed distortion of the differentiation trajectories. Induction of RUNX1-ETO also leads to upregulated expression of early response genes and genes associated with signalling pathways. Whilst the precise mechanisms leading to RUNX1-ETO-mediated transcriptional activation remain unclear, our data support a mechanism involving the upregulation of AP-1 members that would subsequently bind and upregulate target genes, independently of RUNX1 and RUNX1-ETO binding at those sites. We hypothesize that the RUNX1-ETO-mediated abrogation of RUNX1 binding and the associated downregulation of target genes might be a mechanism to promote quiescence, as previously described in RUNX1-deficient HSCs (Ichikawa et al., 2008). We believe that the combined downregulation of cell cycle and DNA repair genes – resulting from the suppressed RUNX1 program – and upregulation of signalling pathways are a mechanism to

control the homeostasis of the *RUNX1-ETO*-expressing clone by (i) preventing its exhaustion, (ii) promoting its long-term survival and (iii) enhancing the accumulation of genetic instability, thus favouring a mutator phenotype.

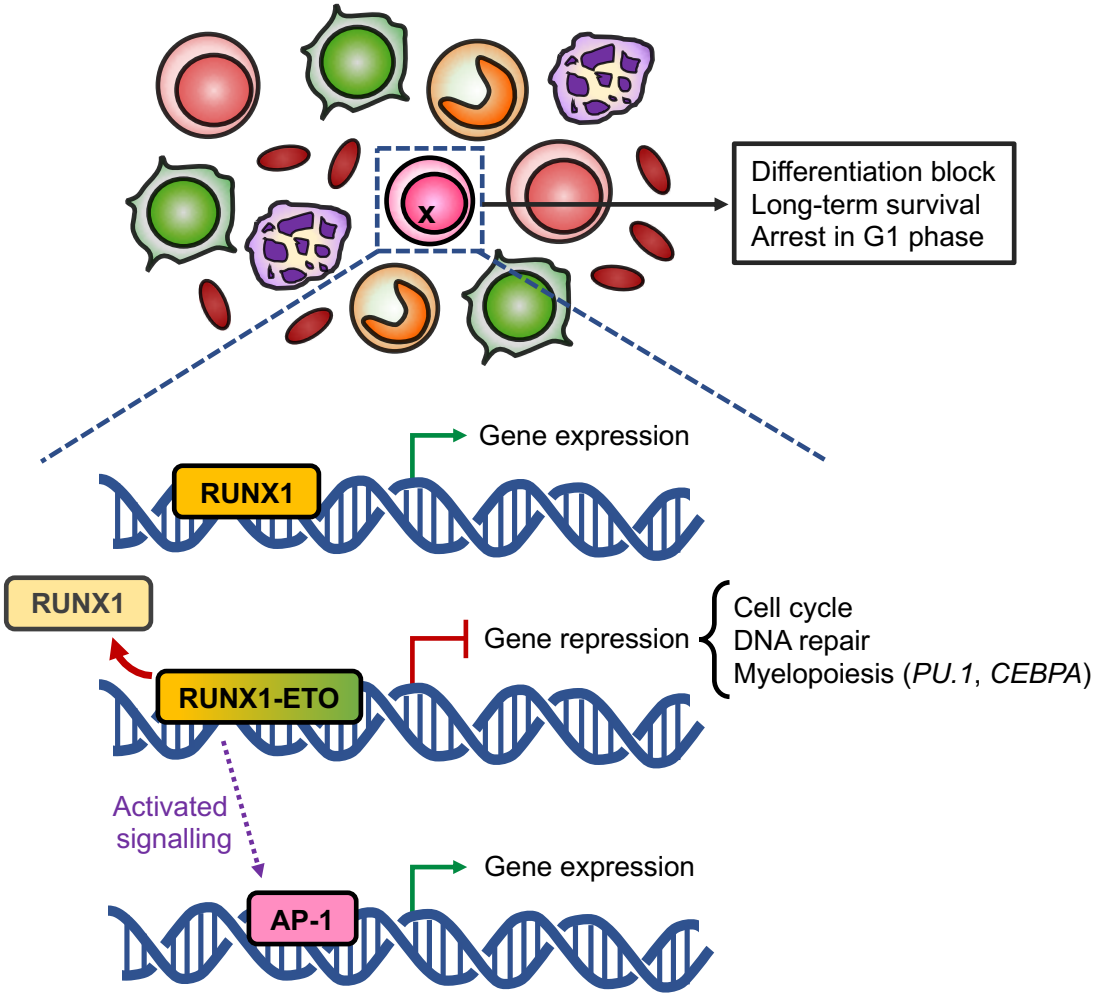


Figure 4.1: RUNX1-ETO-mediated epigenetic reprogramming of a sub-population of *in vitro* human definitive haematopoietic progenitors

Graphical summary depicting the RUNX1-ETO-mediated reprogramming of the chromatin landscape and growth of a sub-population of *in vitro* human progenitors, inferred to represent a multipotent progenitor. Presence of RUNX1-ETO results in a cell-type specific differentiation block, arrest in the G1 phase of cell cycle and promotes long-term survival. At the molecular level, RUNX1-ETO abrogates the RUNX1-mediated transcriptional program through interference with RUNX1 binding at target genes, resulting in the downregulation of genes involved in the control of cell cycle, DNA repair and myeloid lineage differentiation. RUNX1-ETO induction also results in the activation of several signaling pathways and of early response genes, such as AP-1 family members, which could be responsible for the upregulation of gene expression independently of RUNX1 and RUNX1-ETO binding. Upregulation of the binding of AP-1 members through activated signaling (purple dashed line) and direct binding of AP-1 on upregulated target genes has been shown in the t(8;21) Kasumi-1 cell line (Ptasinska et al., 2019), but has not been verified in our system yet.

4.8 Future plans and directions

Current therapeutic efforts are centred on the development of drugs that target oncogenic events occurring late during leukaemic development, such as activated FLT3 or RAS. However, therapeutic strategies targeting only late collaborating mutations in t(8;21) leukaemia may not be efficient in eliminating the pre-leukaemogenic reservoir, therefore failing to eradicate these mutated HSCs, favouring their clonal evolution during remission and supporting subsequent relapse. For this reason, future approaches should focus in ways to target both the factors contained within the RUNX1-ETO complex and the pathways directly deregulated by RUNX1-ETO. For example, we observed expression of IFN-stimulated genes and activation of signalling pathways related to the innate immune response in the *RUNX1-ETO*-responsive population, which could be a hallmark for the relapse observed in t(8;21) AML patients. Our *RUNX1-ETO*-expressing system can therefore be further exploited for the investigation of genes and pathways associated with the innate immune stress response and for the identification of druggable targets, which would permit the evaluation of inhibitor molecules for the development of treatment strategies to overcome adaptive resistance in AML. In order to target these resistant cells, it would be necessary to further elucidate the role of the preleukaemic clone in disease evolution and relapse. Success of drug therapies will also rely on the identification of molecular and functional differences between wild-type HSCs and those harbouring the translocation. For example, understanding the mechanisms behind the mutator phenotype associated with *RUNX1-ETO* expression could aid in developing targeted therapies to prevent the acquisition of secondary collaborative events during remission. Therefore, future studies should focus on elucidating key

RUNX1-ETO mechanisms sustaining the survival of the preleukaemic clone, with experiments including:

(i) Further evaluation of the RUNX1-ETO-mediated differentiation block, aiming to understand the nature of the blocked progenitors. Experiments would be performed by testing the potential of the induced cells, including a wider range of cytokines in our CFU assays, as well as by surface marker profiling of the colonies formed upon removal of Dox.

(ii) Identification of determinant factors in recruiting co-regulators to the RUNX1-ETO co-factor complex, addressing whether these work in a co-operational manner in specific subsets of RUNX1-ETO target genes.

(iii) Comprehensive analysis of the G1 arrest and downregulation of DNA repair genes in order to address whether *RUNX1-ETO*-expressing cells are going into quiescence or senescence. It would also be interesting to identify whether cells arrested in cell cycle are being immortalized via a RUNX1-ETO-mediated senescence-associated secretory phenotype, as described previously (Anderson et al., 2018).

(iv) Evaluation of the molecular reversibility of the RUNX1-ETO-mediated transcriptional changes, which would be crucial to identify essential druggable molecules and pathways to abrogate the *RUNX1-ETO*-dependent leukaemogenic phenotype.

(v) Transplantation experiments to evaluate the ability of our RUNX1-ETO-induced progenitors (in particular of the 5-Dox enriched cells) to home and repopulate bone marrow. Addition of collaborative mutations and improvement of the host

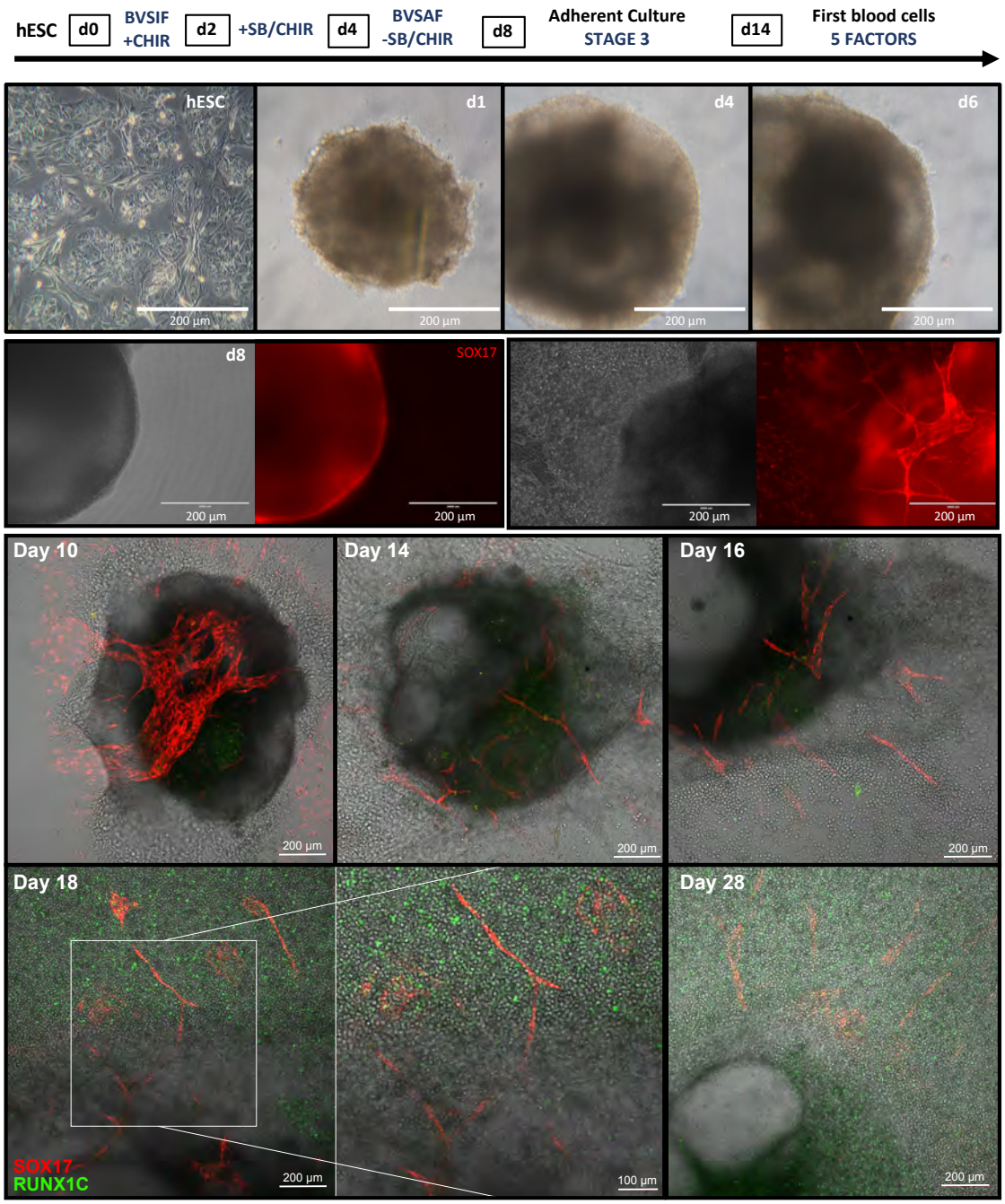
environment within the recipient mice might enhance the engraftment capabilities as well as the putative development of a myelodysplastic syndrome or leukaemic phenotype, as recently shown in mice (Di Genua et al., 2019).

Patients display distinct clinical presentations and treatment responses depending on the order at which mutations are acquired, and addition of a secondary mutational event in *RUNX1-ETO*-expressing mouse haematopoietic progenitors is sufficient to cause *in vivo* overt leukaemia (Nick et al., 2012; Schessl et al., 2005; Wang et al., 2011b; Zhao et al., 2014). However, leukaemic transformation after a single second hit has not been achieved in studies using primary human samples, indicating that human cells might be more resistant to transformation (Chou et al., 2011a; Goyama et al., 2016; Wichmann et al., 2015). Therefore, this system is also a valuable tool to attempt the step-wise modelling of t(8;21) leukaemia, to understand changes in the transcriptional network upon acquisition of each oncogenic event and to evaluate the nature of the second hit required to kickstart the leukemogenicity of the *RUNX1-ETO*-expressing clone. Recent mouse studies show that the presence of mutated K-RAS is incompatible with the pre-leukaemic clone, arguing for a cell stage-specific activation of pathways leading to leukemogenesis (Di Genua et al., 2019). However, *ex vivo* studies in human CD34+ progenitors show that activator mutations in signalling pathway genes co-operate with the presence of a truncated version of RUNX1-ETO, promoting clonal selection and expansion via re-activation of homologous recombination DNA repair pathways (Wichmann et al., 2015). In addition, t(8;21) AML patients undergo disease relapse due to acquisition of new signalling pathway mutations (Shima et al., 2014), suggesting that mechanisms driving leukaemogenesis might differ between species.

Aiming to evaluate whether the second hit is an activated signaling pathway mutation, we introduced an inducible K-RAS(G12D) into our inducible RUNX1-ETO cell lines. Making use of our double targeted cell line, we will be able to compare the growth, immunophenotype, clonogenic potential, and transcriptional profile to those progenitors expressing *RUNX1-ETO* only. These data would inform us of differentially expressed pathways and therefore we would be able to compare therapeutic targets in the early and later stages of leukaemic development. Moreover, it would be valuable to determine whether these dual oncogene-bearing lines can engraft immunodeficient mice.

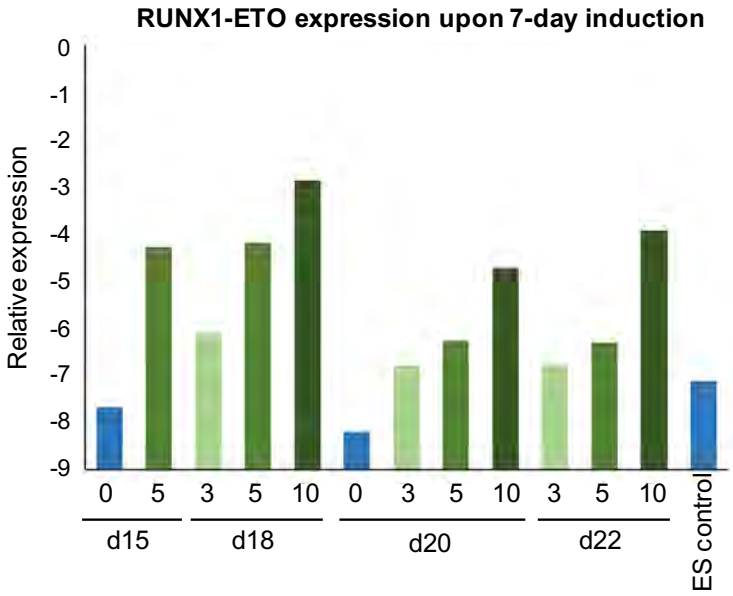
There are therefore several key questions that remain unanswered; would the re-establishment of the RUNX1-mediated transcriptional program reverse or aggravate the oncogenic phenotype? Is the second hit a mutation promoting a proliferative leukaemic phenotype? What are the key RUNX1-ETO-driven features that hinder the elimination of the pre-leukaemic clone by use of current therapies? Overall, what is clear is that, despite many studies on the molecular pathophysiology of t(8;21) AML in several settings and species, the fine molecular details underlying the survival of the pre-leukaemic clone and subsequent disease development are still open to question.

5 SUPLEMENTARY FIGURES



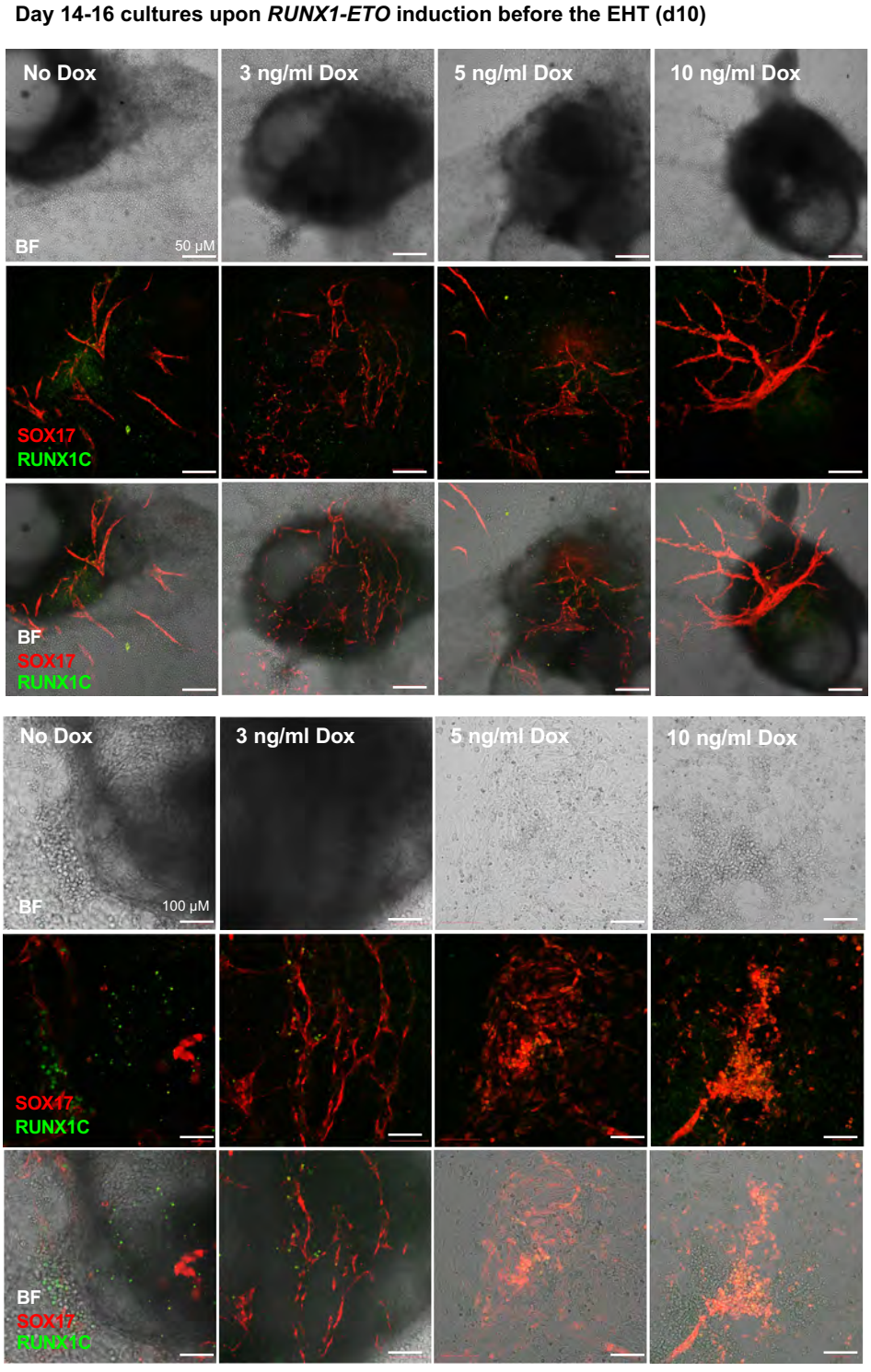
Supplementary Figure 1: Time course of *in vitro* human definitive haematopoietic differentiation as spin EBs

Epi-fluorescence (hESC, d1 to d8) and confocal images (d10-d28) representing a time course of the *in vitro* human definitive haematopoietic differentiation as spin EB cultures. Piror to harvesting, hESC are at 50% confluency. EBs appear as dark opaque round structures. Fluorescence and bright field channels are merged. Scale bar: 200 µm. Day 18 zoom in at 100um. SOX17 (mCHERRY, red) expression marks arterial structures and RUNX1C (GFP, green) marks haematopoietic progenitors.



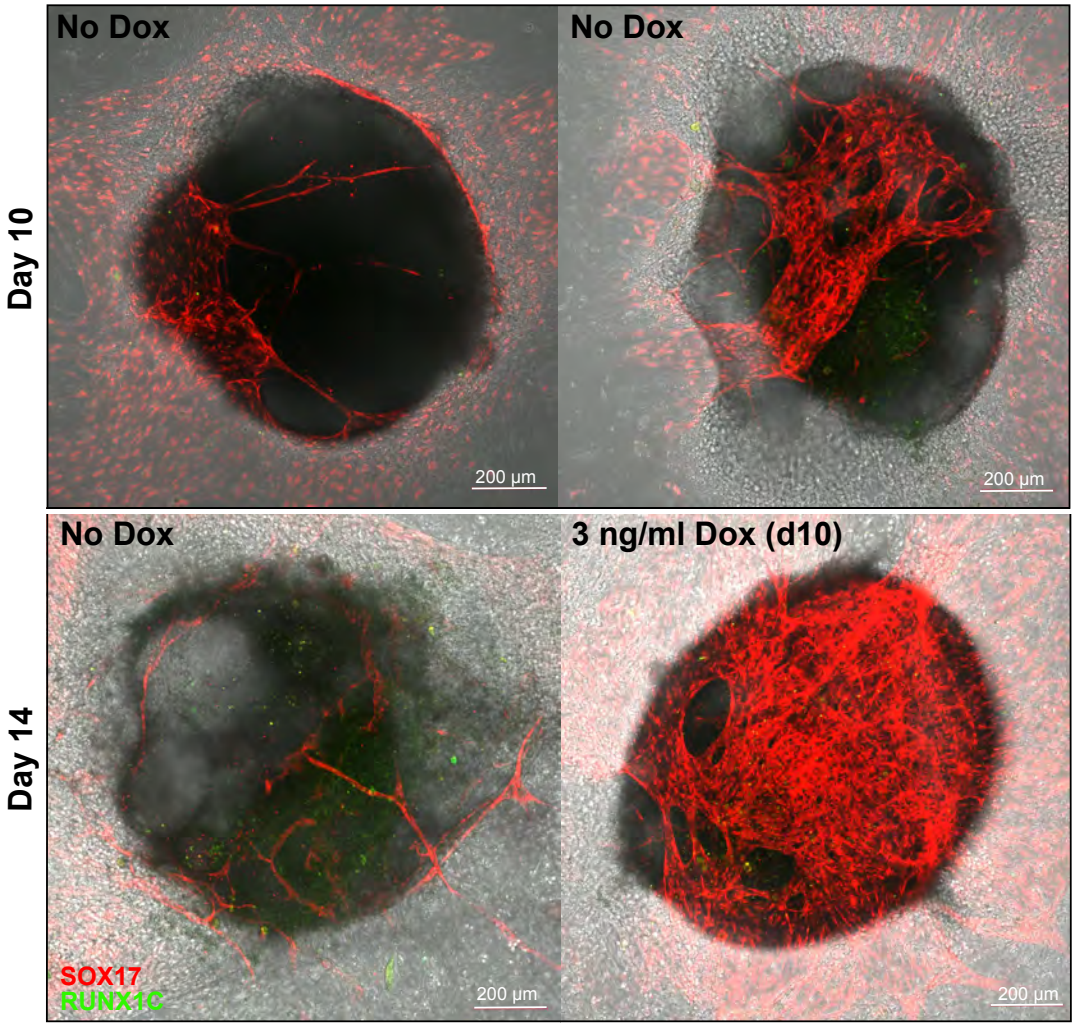
Supplementary Figure 2: *RUNX1-ETO* expression levels are strictly dependent on Dox dosage regardless of the induction time point during differentiation

RUNX1-ETO expression from total RNA at 7 days after dox treatment (Dox refeed every 2-3 days), with relative expression to GAPDH. The figure corresponds to one experiment. X-axis labels indicate the amount of Dox added (ng/ml) and the day of addition during the differentiation. “ES neg control” corresponds to an untargeted and undifferentiated ES cell line.



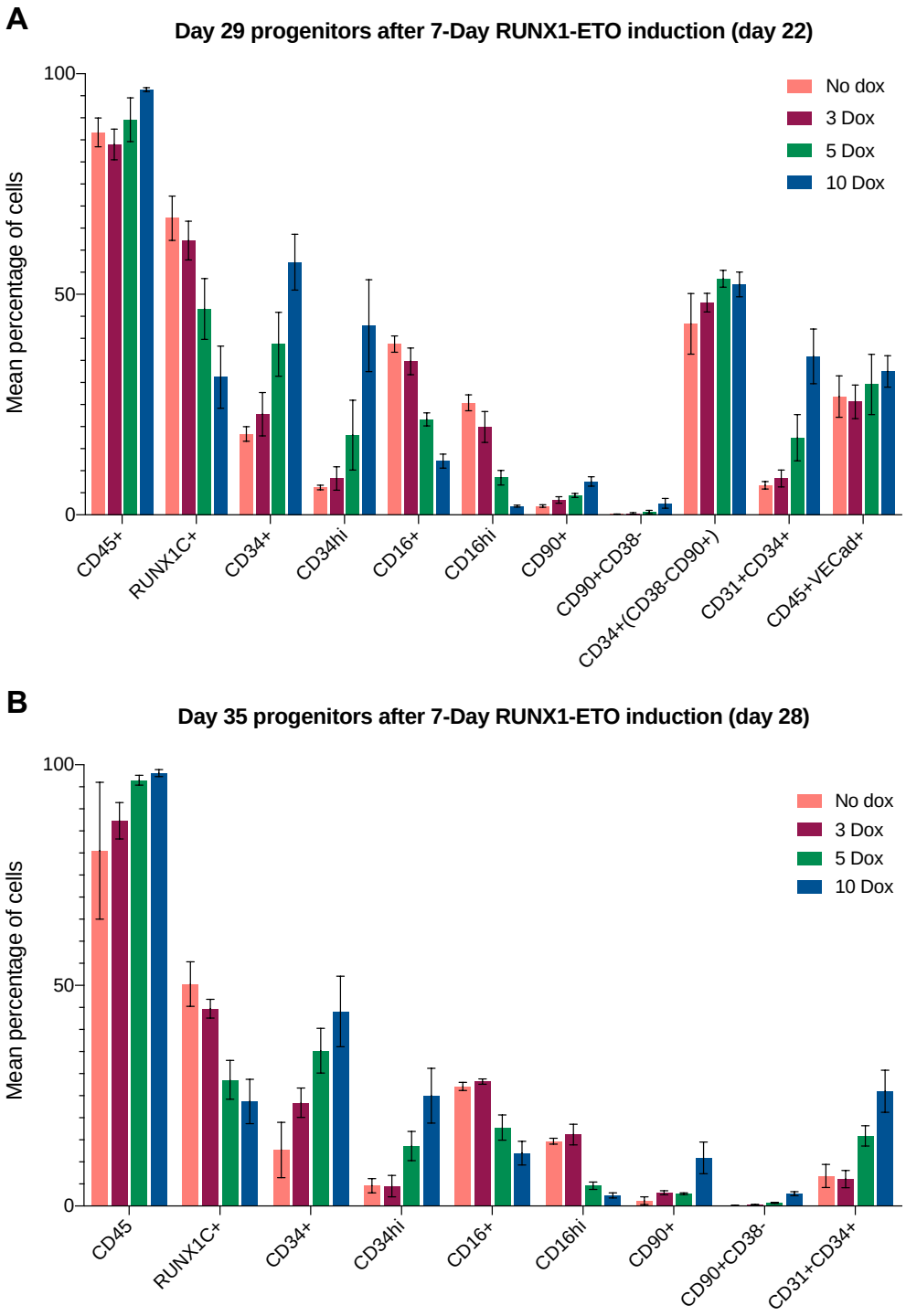
Supplementary Figure 3: RUNX1-ETO induction at balanced levels before the EHT transition disrupts the vascular organization and blocks blood formation

Confocal images of combined Z-stack layers of cultures at d14-16 of differentiation upon RUNX1-ETO induction from d10 (before the EHT) using 3, 5 or 10 ng/m Dox. Each column shows brightfield, fluorescence and merged field channels for each Dox concentration. Arrows show aberrant vascular structures (top panel) and *RUNX1C*+ *SOX17*+ co-expressing progenitors (bottom panel). Scale bars: 50 µm (top panel) and 100 µm (bottom panel). SOX17 (mCHERRY, red) and RUNX1C (GFP, green).



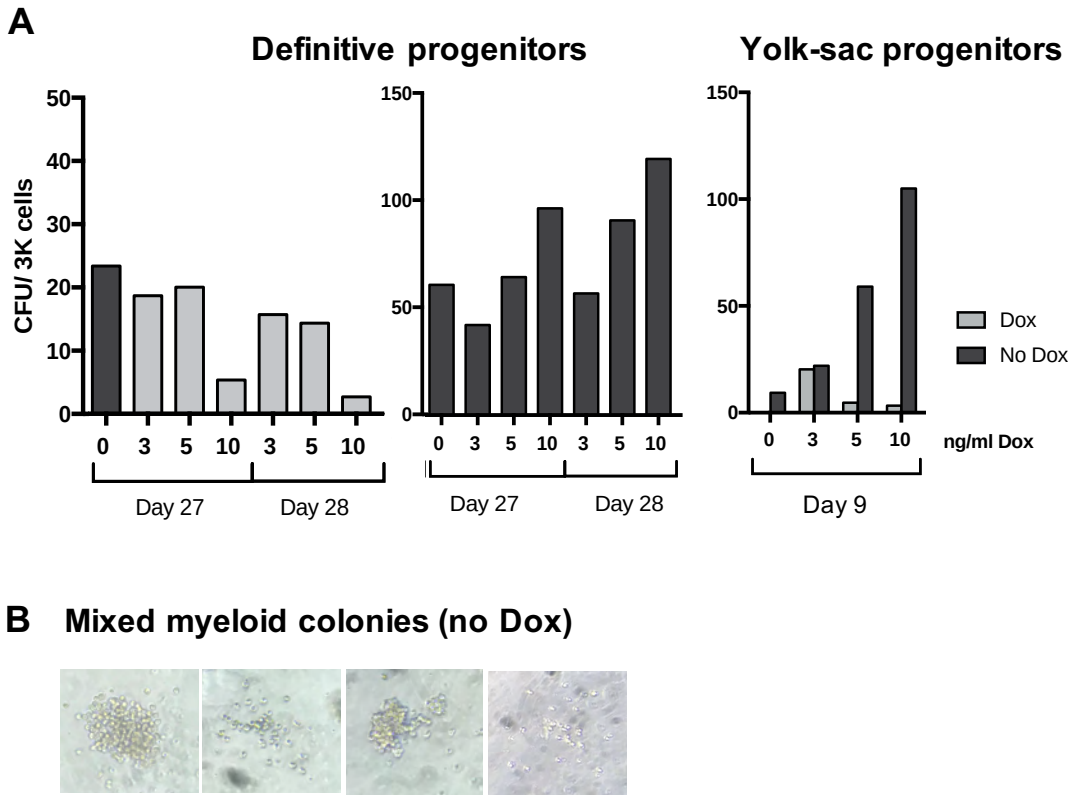
Supplementary Figure 4: Low RUNX1-ETO expression before the EHT transition (d10-12) disrupts the vascular organization and blocks blood formation

Confocal images of combined Z-stack layers of cultures at d14 of differentiation with RUNX1-ETO induction from d10 (before the EHT) using 3 ng/ml Dox. Brightfield and fluorescence field channels are merged. Scale bars: 200 µm, as indicated. SOX17 (mCHERRY, red) and RUNX1C (GFP, green).



Supplementary Figure 5: *RUNX1-ETO*-expressing cultures retain markers of immature myeloid progenitors

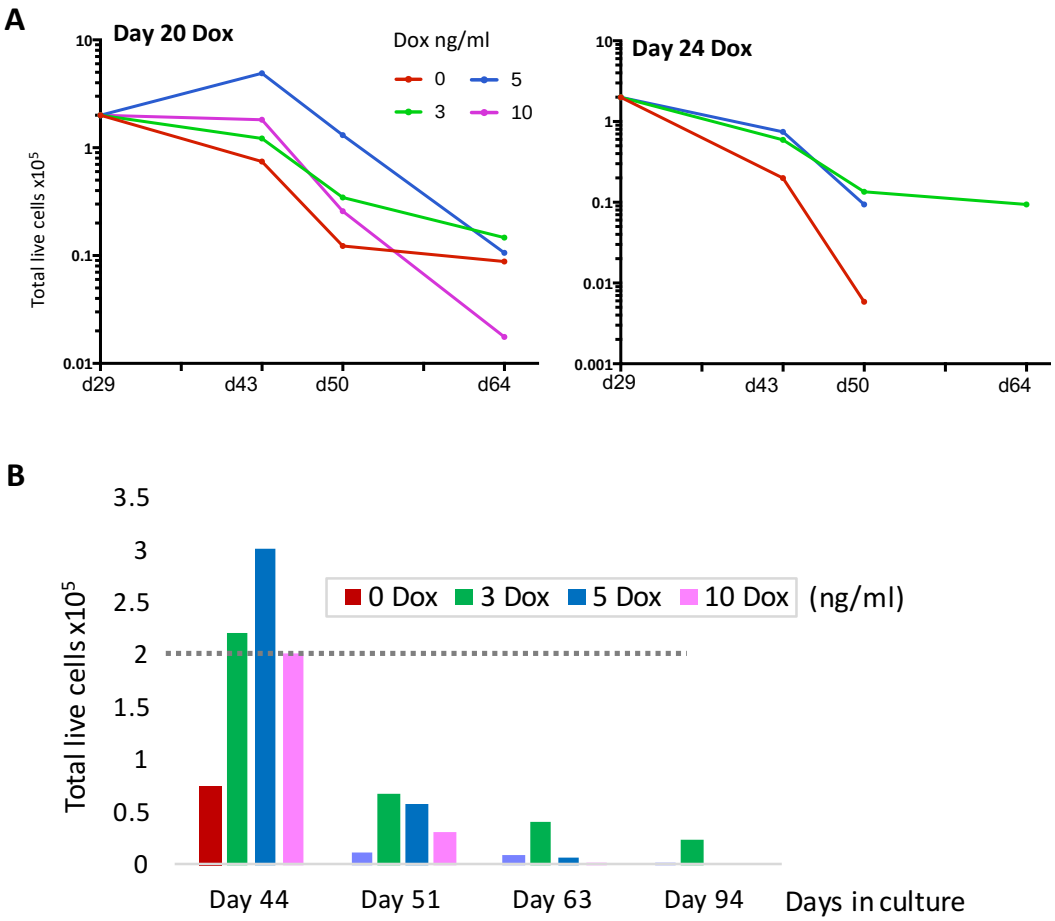
Average data of flow cytometry analysis of the suspension cells of d29 (A) and d35 (B) haematopoietic progenitors upon RUNX1-ETO induction for 7 days. Bar diagrams show the mean percentage of cells expressing selected surface markers representing data from five (A) and three (B) independent biological replicates. Error bars represent Standard Error of the Mean (SEM).



Supplementary Figure 6: *RUNX1-ETO* expression maintains clonogenic cells in a quiescent stage regardless of the origin of the progenitor cell

(A) Colony-forming unit assays of definitive progenitors from the floating fraction of EB cultures treated with Dox during 7 days at different stages during differentiation, and of yolk-sac-type progenitors treated with Dox from day 9 until different days of differentiation. Progenitors were plated in presence (light grey) or absence of Dox (dark grey) at a concentration of 3,000 live cells/well and in triplicate. Individual graphs correspond to different biological replicates.

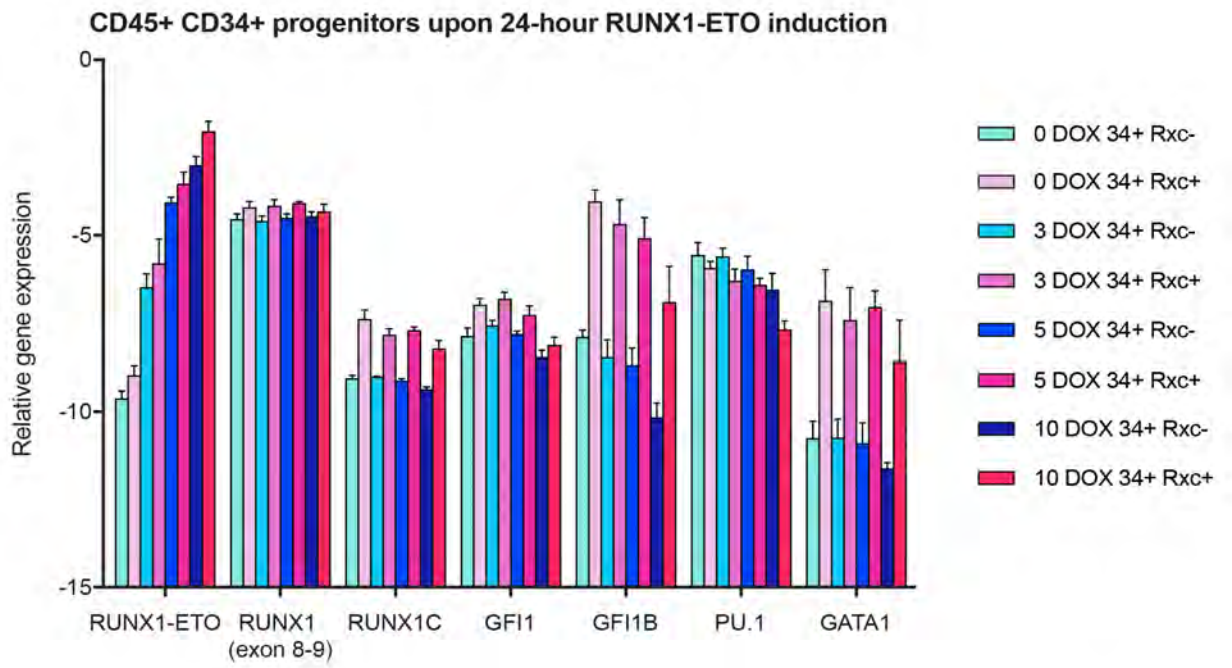
(B) Brightfield images representative of the colonies generated in CFU assays from wild-type in vitro cultures.



Supplementary Figure 7: *RUNX1-ETO* expressed at low levels increases the survival of a subset of progenitor cells

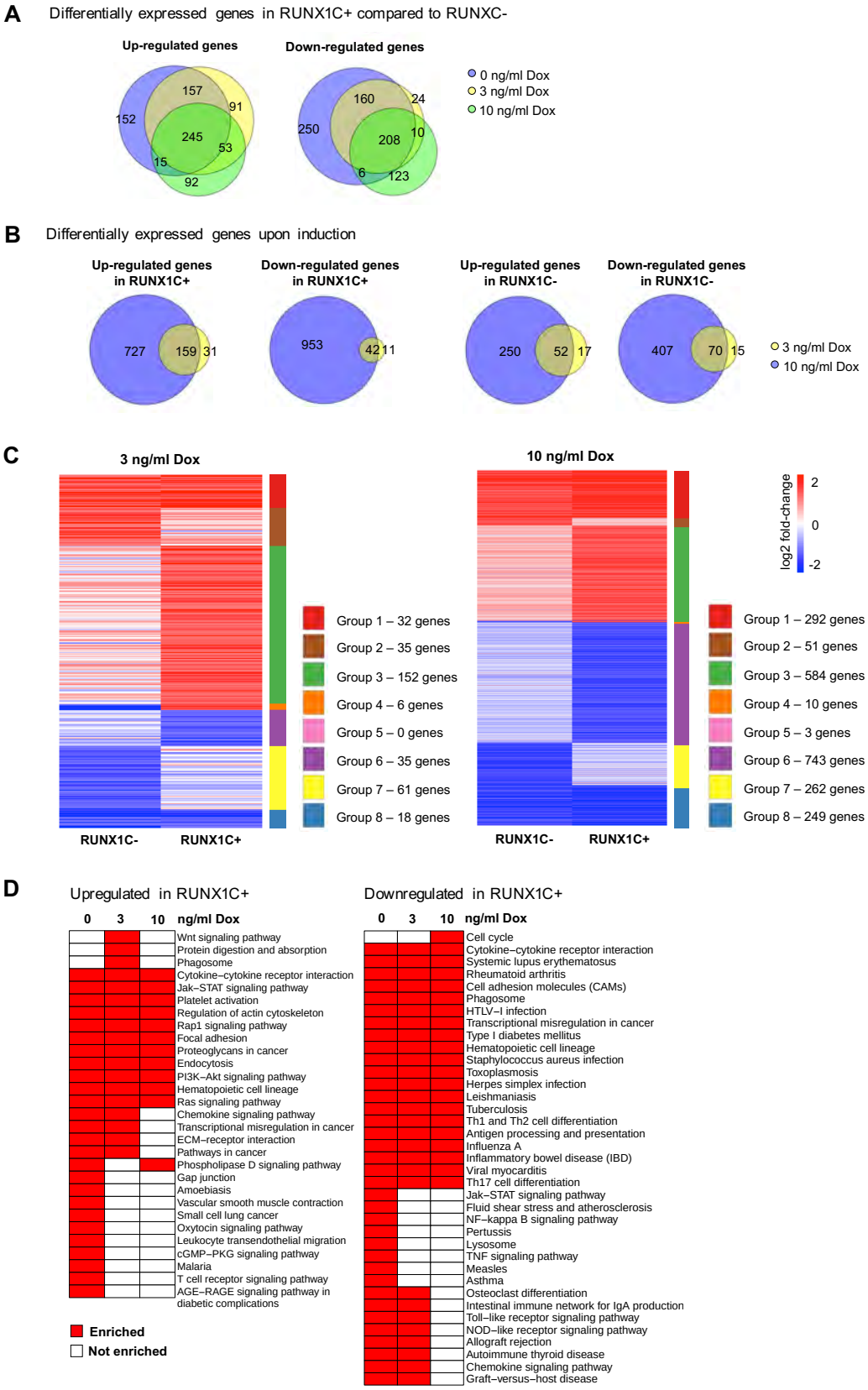
(A) Additional examples of replating assays of haematopoietic progenitors from cultures treated at d20 or d24 with different Dox concentrations, showing two representatives each of three independent experiments. Floating haematopoietic cells were plated at 2 x10^5 cells/well in the correspondent Dox concentration and cell numbers were measured at three time points as indicated. On d24 graph, only 3-dox treated cells were able to survive over 28 days on the replating assays.

(B) Absolute values (x10^5) of the total live cells counted during replating assays (same data represented in log scale in figure 3.15). Dotted line represents the number of progenitors plated at the start of the replating assay.



Supplementary Figure 8: *RUNX1-ETO* induction for 24h causes a dose-dependent downregulation of haematopoietic genes

Gene expression from total RNA of CD45+ CD34+ RUNX1C-/+ sorted EB cultures upon induction of RUNX1-ETO for 24 hours at day 20 of differentiation with 0, 3, 5 or 10 ng/ml Dox. Gene expression was analyzed using Taqman primers and normalized to that of GAPDH. Error bars represent the s.e.m. of three independent experiments. RUNX1 (exon 8-9) primers bind downstream of the highly-conserved RUNT domain, thus detecting all RUNX1 isoforms but not the RUNX1-ETO oncoprotein.



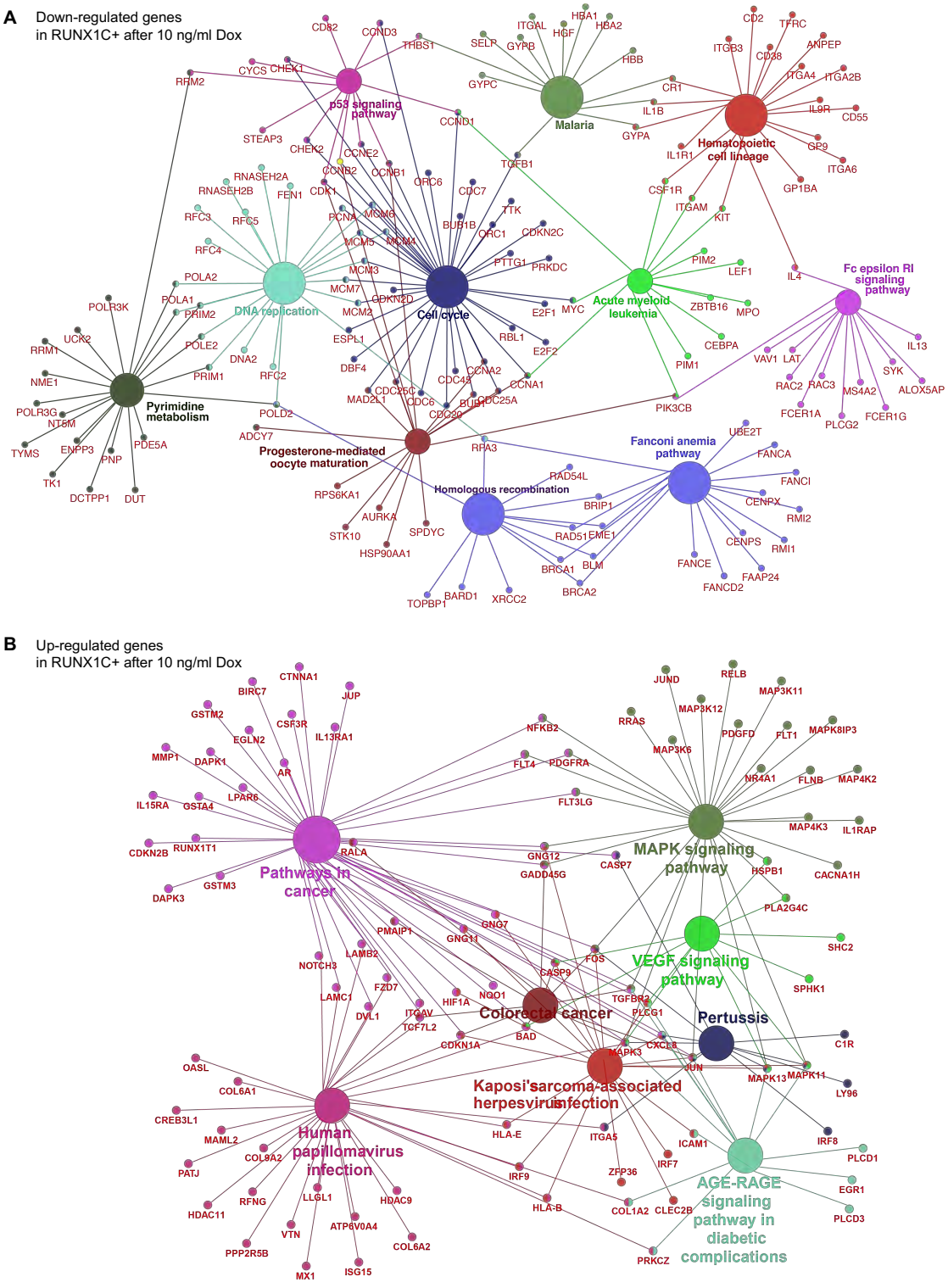
Supplementary Figure 9: Different levels of RUNX1-ETO dysregulates a common subset of genes but this differs depending on the type of progenitor cell

(A) Venn diagrams showing the numbers and overlap of differentially expressed genes (upregulated: top, downregulated: bottom) in CD45+CD34+RUNX1C+ compared to CD45+CD34+RUNX1C- upon treatment with 0, 3 or 10 ng/ml Dox for 24 hours.

(B) Venn diagrams showing number and overlap of differentially expressed genes (up/down-regulated) upon treatment with 3 or 10 ng/ml Dox for 24 hours in both CD45+CD34+RUNX1C+ (right two diagrams) and CD45+CD34+RUNX1C- (left two diagrams) populations.

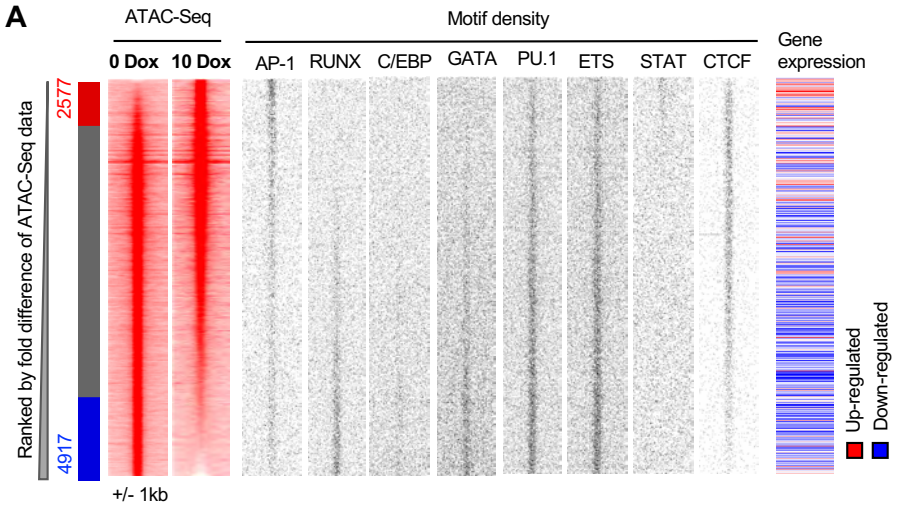
(C) Heatmap representation of differentially expressed genes upon treatment with 3 ng/ml Dox (left) or 10 ng/ml Dox (right) in both CD45+CD34+RUNX1C- and CD45+CD34+RUNX1C+ populations. Differentially expressed genes were clustered in 8 different groups depending on their response to RUNX1-ETO induction in each cell population or in both. Groups represent genes: upregulated in both (1, red), upregulated in RUNX1C- cells but not in RUNX1C+ (2, brown), upregulated in RUNX1C+ cells but not in RUNX1C- (3, green), downregulated in RUNX1C- cells but upregulated in RUNX1C+ (4, orange), upregulated in RUNX1C- cells but downregulated in RUNX1C+ (5, pink), downregulated in RUNX1C+ but not in RUNX1C- (6, purple), downregulated in RUNX1C- but not in RUNX1C+ (7, yellow) and downregulated in both (8, blue).

(D) Enriched KEGG pathways for genes upregulated (right) and downregulated (left) in CD45+CD34+RUNX1C+ cells compared to CD45+CD34+RUNX1C- cells upon treatment with 3 and 10 ng/ml and in uninduced conditions (0 Dox).



Supplementary Figure 10: Up and downregulated pathways upon RUNX1-ETO induction with 10 ng/ml Dox

KEGG pathways highlighting genes and pathways that are at least 2-fold (A) downregulated or (B) upregulated after 24-hour expression of RUNX1-ETO with 10 ng/ml Dox in CD45+CD34+RUNX1C+ hESC-differentiated progenitors.

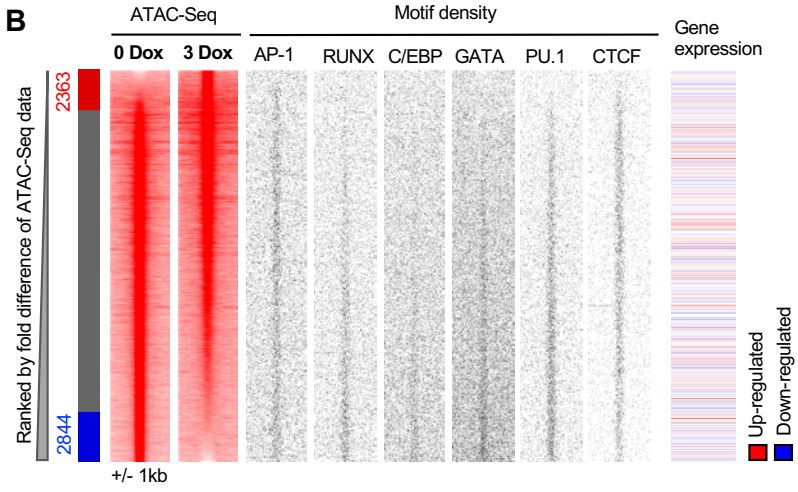


10 Dox specific sites – 2577

Motif	Match	% sites	p-value
	ETS	24.06	1e-72
	AP-1	16.14	1e-232
	CTCF	11.02	1e-161
	STAT	8.54	1e-38

0 Dox specific sites – 4917

Motif	Match	% sites	p-value
	GATA	24.08	1e-168
	PU.1(ETS)	20.99	1e-422
	RUNX	20.70	1e-255
	C/EBP	12.24	1e-55
	AP-1	9.52	1e-63
	CTCF	7.81	1e-151



3 Dox specific sites – 2363

Motif	Match	% sites	p-value
	CTCF	11.47	1e-38
	AP-1	4.87	1e-18

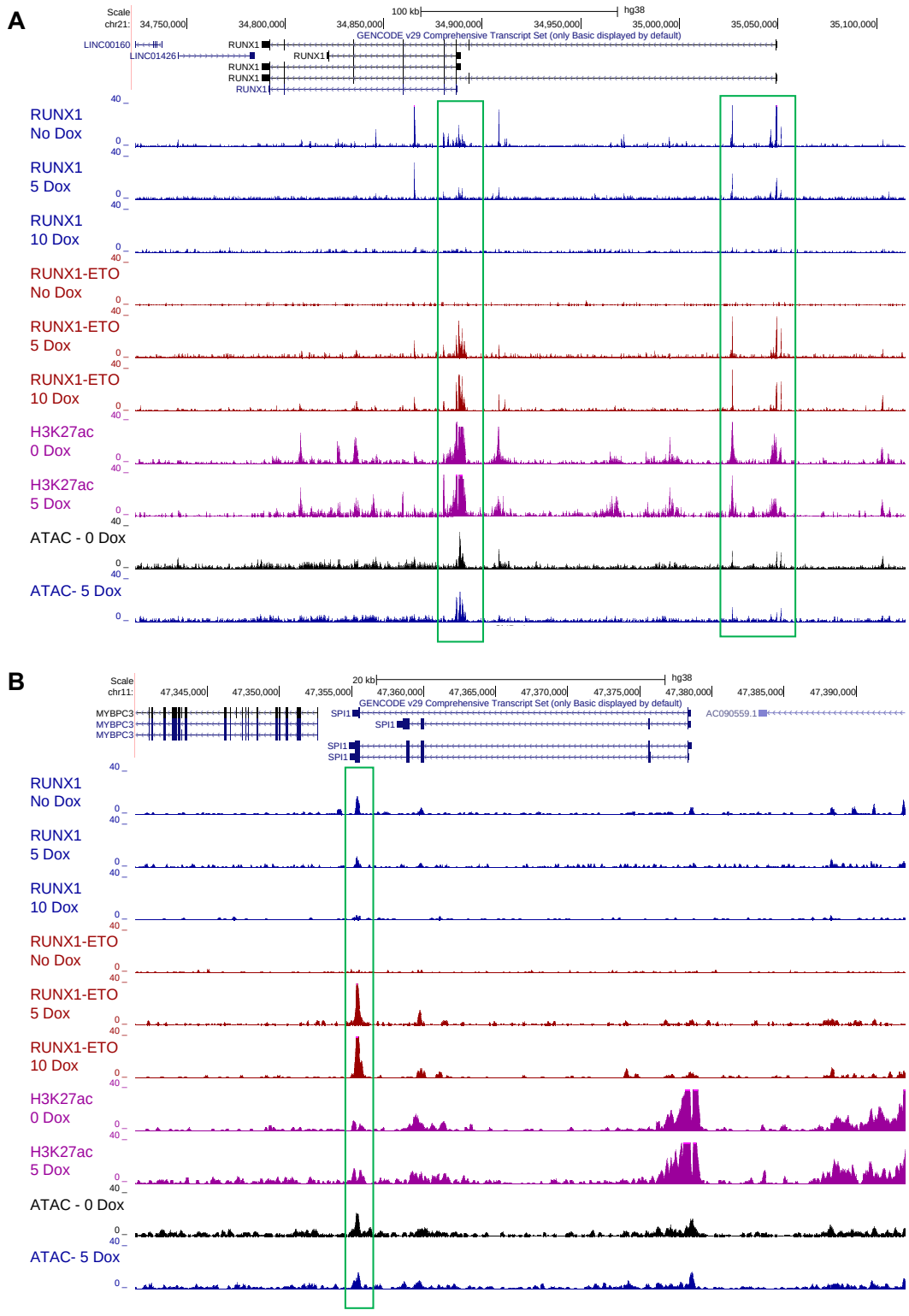
0 Dox specific sites – 2844

Motif	Match	% sites	p-value
	GATA	26.90	1e-94
	PU.1	18.11	1e-228
	RUNX	14.73	1e-94
	C/EBP	14.59	1e-40
	AP-1	12.20	1e-40
	CTCF	9.42	1e-142

Supplementary Figure 11: Induction of RUNX1-ETO results in loss of RUNX1, GATA, PU.1 and C/EBP accessible sites

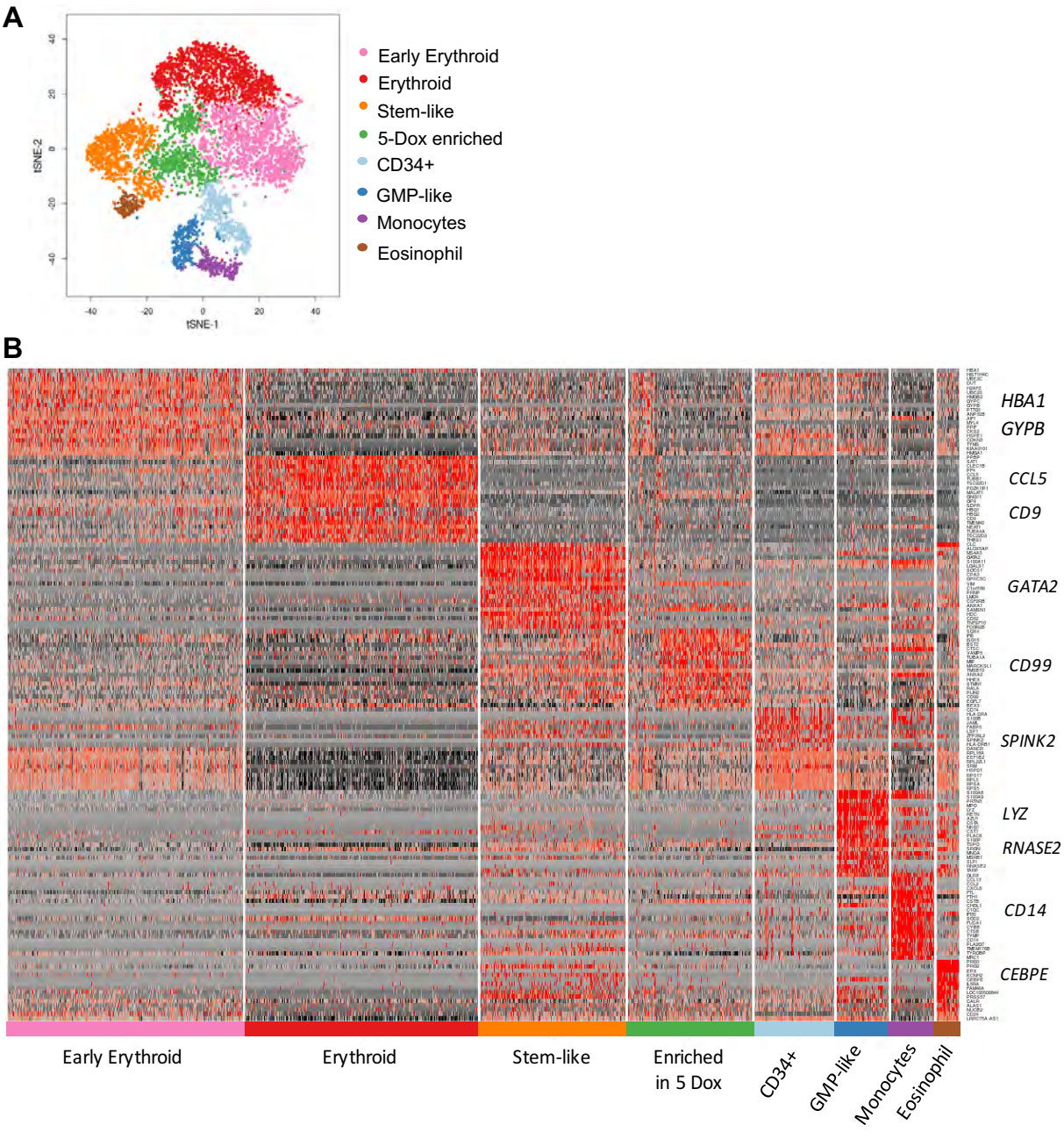
Heat maps depicting accessible chromatin regions ranked by fold difference between (A) the 0 and 10 ng/ml Dox samples and (B) the 0 and 3 ng/ml Dox samples. ATAC-Seq peaks were considered

sample-specific when displaying a greater than 2-fold enrichment compared to the other sample. Sample-specific sites and number of peaks are indicated alongside, being: red the Dox specific, blue the 0-Dox specific and grey the shared peaks. Motif density plots and gene expression at these sites are ranked along the same coordinates as the ATAC-Seq peaks. Below: Motif enrichment analysis in (A) 0 and 10 ng/ml Dox or (B) 0 and 3 ng/ml Dox -specific peaks.



Supplementary Figure 12: Individual gene examples showing RUNX1-ETO-dependent displacement of RUNX1 with associated reduction in chromatin accessibility

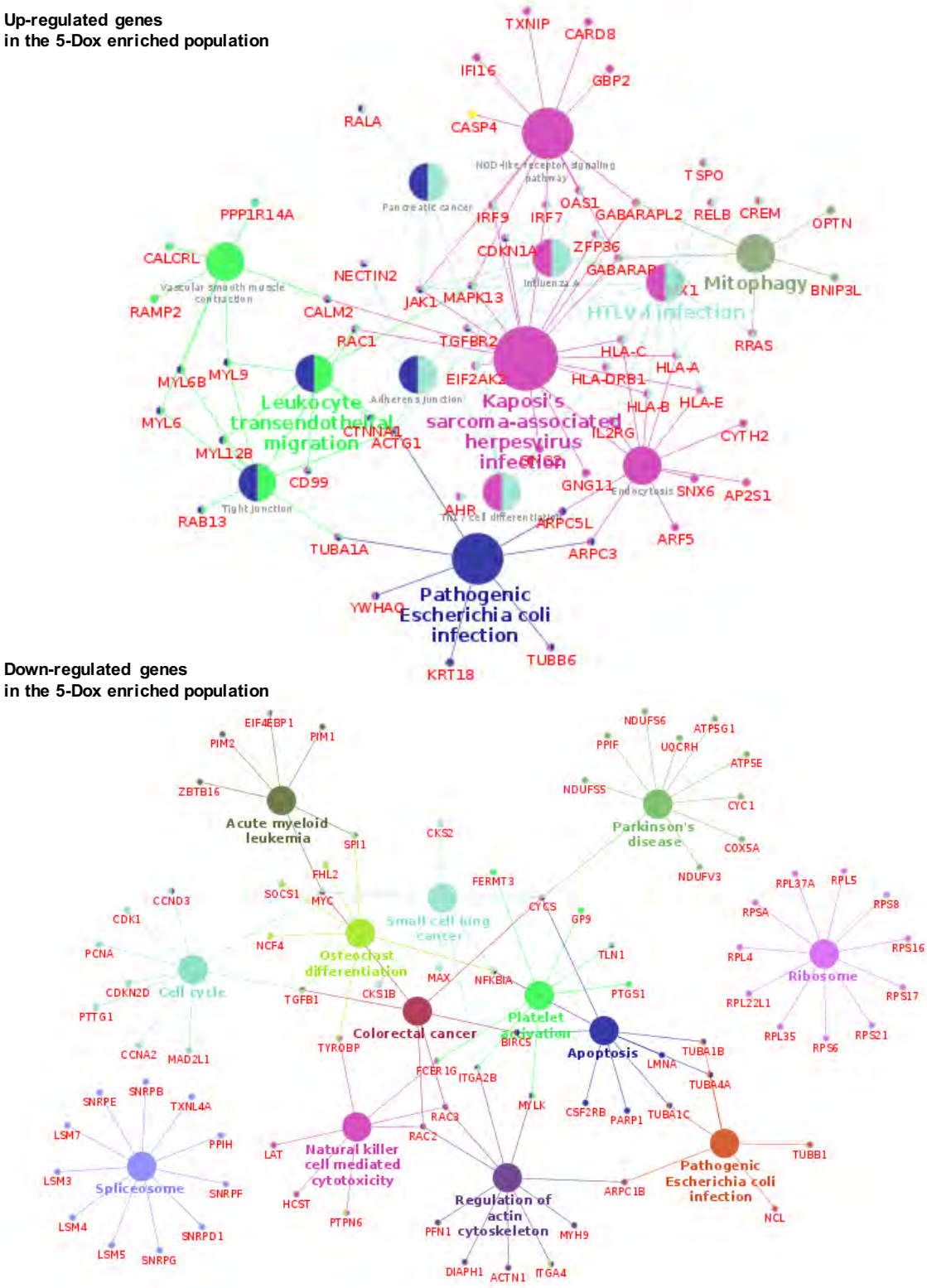
Genome browser screenshot depicting RUNX1, HA-RUNX1-ETO, H3K27ac ChIP-Seq and ATAC-Seq tracks for the indicated samples at (A) *RUNX1* and (B) *SPI1* representative gene loci. Green boxes show displacement of RUNX1 and reduction of accessible chromatin upon RUNX1-ETO induction.



Supplementary Figure 13: The CD45+CD34+RUNX1C+ population contains precursors from distinct blood lineages as well as multipotent cell progenitors

A) Two-dimensional t-SNE map displaying a total number of 7,135 CD45+CD34+RUNX1C+ sorted single cells from the combined data of 0 and 5 Dox treated cells including identified cell populations based on expression of known marker genes.

B) Heatmap showing expression of the top 20 marker genes specific to each cluster (same colour coding as in (B)). Representative genes from each cluster are indicated.



Supplementary Figure 14: RUNX1-ETO-deregulated pathways in the 5-Dox enriched single cells are representative for those observed within the induced bulk population

Network diagram of KEGG pathways for upregulated (above) and downregulated (below) genes in the 5-Dox enriched cell cluster of CD45+CD34+RUNX1C+ sorted single cells upon 24-hours 5 Dox treatment.

6 SUPLEMENTARY TABLES AND DATASETS

RUNX1C- specific			
GeneID	RUNX1C+	RUNX1C-	logFC
CCL13	2.8876	169.7916	5.4572
CD1C	0.8294	66.1279	5.1975
C1QA	0.8258	60.1952	5.0668
ASGR2	3.9370	161.4937	5.0406
MRC1	6.8812	257.9699	5.0382
CLEC10A	1.1271	65.5592	4.9677
C1QC	1.5475	76.4195	4.9256
MPEG1	0.9700	58.0315	4.9052
MS4A7	0.8214	52.9438	4.8883
C1QB	0.4992	42.3162	4.8526
HLA-DRA	7.6657	230.9140	4.7421
CD14	0.9925	51.1300	4.7095
S100B	13.2541	370.0203	4.7021
LY86	3.6811	112.6247	4.6013
FGL2	1.1510	51.1035	4.5983
JCHAIN	7.3962	198.8754	4.5732
STEAP4	2.2489	72.7786	4.5052
CCR6	1.8984	60.8520	4.4155
LUZP6	0.3894	28.3049	4.3986
IRF8	1.5533	52.5322	4.3900
TIFAB	0.2911	25.8938	4.3806
KCTD12	1.3244	46.0588	4.3396
FGD2	1.0607	40.5671	4.3342
CLEC7A	2.9326	74.7159	4.2670
HLA-DRB5	15.5822	304.2367	4.2022
GAS6	1.3554	41.0838	4.1592
CYBB	10.1549	197.3484	4.1523
MS4A6A	4.3473	89.3500	4.0787
RNASE1	4.8113	96.8727	4.0740
DCANP1	0.3273	20.3700	4.0091
AGR2	0.0363	15.6422	4.0053
IGSF6	3.8040	76.1071	4.0046
MMP9	0.4428	22.0772	3.9995
LGMN	2.8123	59.6595	3.9920
TMEM236	0.7530	26.5215	3.9727
ACY3	0.1637	17.1191	3.9607
DPP4	1.6829	38.5125	3.8804
CD180	0.7566	24.2885	3.8476
CXorf21	5.9388	97.7852	3.8315
CD74	67.5525	940.5089	3.7797
CX3CR1	3.2869	57.6799	3.7749
CIITA	1.4253	31.7827	3.7567
CLEC4A	2.6730	48.4261	3.7502
CSF1R	20.9445	286.2944	3.7106
RASSF4	4.1597	66.1799	3.7027
HLA-DMB	0.2392	14.9101	3.6825
SAMHD1	8.1186	114.8835	3.6677
FGL1	0.2004	13.9677	3.6403
IL1RN	5.3865	78.4092	3.6362
SLCO2B1	0.8152	20.6994	3.5794
HLA-DPA1	0.8482	20.8888	3.5660
STAB1	6.3658	85.7480	3.5579
GPR183	3.5632	51.8735	3.5344
CD163	0.6730	18.1309	3.5154
APCDD1	0.2294	12.6884	3.4769
IL7R	2.0860	32.6217	3.4456
CD1E	0.2520	12.3896	3.4188
CCR5	0.5443	15.3831	3.4072
IL13RA1	4.7034	59.0314	3.3958
NME1-NME2	2.0647	31.0184	3.3851
HFE	0.5443	14.9832	3.3715
RNASE6	16.0558	173.5263	3.3551
JAG1	2.2843	32.3618	3.3446
TLR10	0.2801	11.9804	3.3421
CD300E	0.4279	13.4297	3.3370
CD36	14.7887	157.6992	3.3293
CCR2	4.4802	53.1563	3.3048

RUNX1C- specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
AOAH	1.6205	24.2833	3.2703
MAF	2.5947	33.5749	3.2658
RBM47	0.4813	13.1359	3.2545
CD86	8.8782	91.6162	3.2289
TLR2	3.4616	40.7607	3.2265
NAIP	0.0065	8.3185	3.2108
ENPP2	6.0048	63.4453	3.2017
ABI3	0.1957	9.9868	3.1999
MTUS1	0.5942	13.6348	3.1985
TYMP	4.3381	47.8204	3.1931
FCGR1A	1.7139	23.8017	3.1920
CCR7	0.2315	10.1001	3.1720
GIMAP4	1.2879	19.4203	3.1579
A2M	1.5168	20.7652	3.1124
PTPRO	0.6138	12.7957	3.0957
HLA-DMA	1.5632	20.4594	3.0656
SGSH	0.9456	15.2641	3.0634
RGL1	1.5095	19.8529	3.0548
SDC2	0.3223	9.9775	3.0534
LSP1	21.1577	181.9020	3.0452
MX1	0.7311	12.9624	3.0118
EPB41L3	1.6643	20.1521	2.9890
CSF2RA	6.0563	53.9480	2.9611
BLNK	0.1027	7.5199	2.9499
CCL24	1.0165	14.4864	2.9410
OLFML2B	0.9832	14.1962	2.9378
ODF3B	0.5785	11.0399	2.9312
SPP1	9.3731	77.5519	2.9208
PADI2	1.2452	15.9675	2.9179
OGFRL1	6.2997	53.8151	2.9087
FOLR2	1.5263	17.7382	2.8909
CASP1	5.4252	46.2983	2.8800
PLBD1	1.0780	14.2621	2.8767
MOB3B	0.7255	11.6039	2.8688
ACPP	1.4938	17.0157	2.8528
P2RY13	0.6753	10.9489	2.8344
NRP1	4.6262	38.8605	2.8247
CD163L1	0.2831	7.9888	2.8084
LPAR6	20.1844	145.9210	2.7940
TGFBI	2.1542	20.8632	2.7931
B3GNT5	13.4922	98.5663	2.7804
FLVCR2	0.7991	11.2997	2.7733
NLRC4	1.1247	13.4959	2.7703
BCL6	0.7681	10.9070	2.7515
JAML	76.7661	514.5234	2.7288
RIN2	0.5202	8.9756	2.7142
PARM1	0.2048	6.9012	2.7133
SETBP1	0.5730	9.2745	2.7075
MS4A4A	0.1420	6.4362	2.7030
P2RY6	0.3068	7.4907	2.6998
IFI30	8.2767	59.1654	2.6973
FAM198B	3.3329	26.9624	2.6901
MS4A14	0.0713	5.8819	2.6835
CCR1	6.0937	44.3408	2.6762
GAS7	1.6254	15.5604	2.6572
TMEM176B	6.6799	46.7833	2.6373
IRF4	0.1464	6.0900	2.6287
TLR6	2.0141	17.6147	2.6267
SCN4B	1.1682	12.3210	2.6191
S100Z	3.8733	28.8520	2.6149
PDK4	1.3438	13.3381	2.6129
TRIM71	1.6908	15.4227	2.6096
TLR1	3.1784	24.3784	2.6026
TLR7	0.1295	5.8438	2.5992
SLC15A3	0.9428	10.7232	2.5932
KCNE5	0.6312	8.7665	2.5819
AMOT	2.8347	21.9116	2.5789

RUNX1C- specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
GGT5	8.2407	54.1799	2.5781
CES1	0.3273	6.9034	2.5740
PLA2G7	3.6853	26.6610	2.5617
EPHB3	0.3626	6.9422	2.5431
ADAM28	4.8370	32.9885	2.5418
NCF2	24.7128	148.5181	2.5398
GLIPR1	10.2546	64.3448	2.5376
URGCP-MRPS24	0.5652	7.9922	2.5223
LY96	3.6721	25.4297	2.5000
FHDC1	0.3229	6.4109	2.4860
TLR8	0.1837	5.5935	2.4777
F13A1	40.9262	232.3278	2.4764
HNMT	3.4006	23.4604	2.4747
PRAM1	3.3670	23.2532	2.4735
POU2F2	1.3206	11.8041	2.4641
SERPINF1	1.0500	10.3082	2.4637
IFI16	31.3585	176.1627	2.4529
EVI2A	8.9035	52.9008	2.4443
SLA	12.3262	70.3708	2.4211
HLX	1.3941	11.5541	2.3906
CMKLR1	0.7943	8.4018	2.3895
ADAMDEC1	0.0725	4.6091	2.3868
CD1B	0.0230	4.3445	2.3852
TMEM71	5.3441	32.0140	2.3796
RTN1	0.3480	5.9963	2.3758
CLCN5	1.9979	14.5501	2.3749
SCN3A	0.3386	5.8708	2.3597
ACP5	0.5779	7.0602	2.3528
IL6R	4.0044	24.5076	2.3496
IL21R	5.3284	31.1267	2.3439
TNS3	0.7617	7.9386	2.3431
TNFRSF1B	11.2411	60.7732	2.3352
MARCH1	1.7013	12.5897	2.3308
GNGT2	0.1623	4.8464	2.3306
CLEC5A	2.0970	14.4546	2.3191
ADAP2	0.2155	5.0597	2.3177
HMOX1	4.4943	26.2917	2.3125
HLA-DQB1	0.4515	6.2095	2.3124
APBB3	6.6369	36.8557	2.3095
MYOF	0.6618	7.2208	2.3065
TMEM176A	3.0792	19.1772	2.3064
PRDM1	0.6892	7.3294	2.3019
CLNK	0.1369	4.5815	2.2955
RHBDF2	0.8229	7.8799	2.2843
TNFSF18	0.0566	4.1450	2.2837
IGLON5	1.7454	12.1596	2.2610
TMEM144	2.5276	15.8702	2.2577
CTSS	11.1114	56.5810	2.2492
CCDC170	1.4418	10.5023	2.2359
SLFN12L	0.1605	4.4664	2.2358
TNP1	0.0065	3.7369	2.2347
FCGR1B	0.6690	6.7710	2.2191
OAS1	2.1062	13.4519	2.2180
KYNU	0.2481	4.7919	2.2144
FCGR3A	0.9365	7.8985	2.2001
FPR1	1.3473	9.7083	2.1897
HLA-B	0.9198	7.7165	2.1828
NLRP3	3.5490	19.5947	2.1786
SYT17	0.1355	4.1347	2.1769
SLC37A2	0.1952	4.4013	2.1760
ASIC1	0.7449	6.7731	2.1554
SLC16A10	0.5461	5.8638	2.1504
HLA-DOA	0.0378	3.5910	2.1453
PLD4	9.0088	42.9952	2.1361
STAMBPL1	2.0009	12.1646	2.1332
HDAC9	1.8561	11.4960	2.1294
CDKN1A	18.8473	85.6317	2.1260

RUNX1C- specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
MILR1	14.2724	65.3402	2.1190
NLRP1	5.2463	25.9830	2.1110
GK	11.3089	51.8031	2.1009
IPCEF1	3.8166	19.6545	2.1004
NAAA	3.1884	16.8835	2.0942
PLD2	1.0533	7.7612	2.0932
SIGLEC1	0.0984	3.6652	2.0866
LRRK2	0.4510	5.1459	2.0826
SLC11A1	2.9469	15.6795	2.0793
CD1A	0.0476	3.3510	2.0543
PDE1B	5.0058	23.7428	2.0426
CCDC85A	0.2557	4.1401	2.0333
ANKEF1	0.4338	4.8673	2.0328
NSUN7	0.6639	5.7924	2.0293
VCAN	2.5495	13.3154	2.0119
MAFB	0.1290	3.5521	2.0115
CAMK2D	0.6884	5.8038	2.0107
GPNMB	0.1441	3.6102	2.0106
NRG1	0.3074	4.2585	2.0079
CYTH4	17.1772	71.5870	1.9976
PLEKHG5	0.6789	5.6971	1.9960
GPRIN3	0.5480	5.1606	1.9926
TRIM36	0.3238	4.2197	1.9792
FFAR2	0.1690	3.5977	1.9757
C9orf72	7.3704	31.8974	1.9746
MYO1B	1.3157	8.0757	1.9706
FCN1	6.7389	29.3054	1.9694
TRIM22	7.0022	30.0795	1.9575
XAF1	0.6838	5.5392	1.9574
SNAP25	0.5626	5.0624	1.9559
PILRA	11.2724	46.5549	1.9542
IGFBP5	8.0348	33.9551	1.9519
TNFSF8	0.2353	3.7546	1.9444
IRF5	2.3543	11.8250	1.9349
HAVCR2	4.9626	21.7853	1.9341
BTLA	0.1086	3.2122	1.9258
RAB31	25.3531	98.9233	1.9228
SLC2A5	1.0188	6.6416	1.9204
NKG7	45.7796	174.6392	1.9087
IFNGR1	30.2466	115.4885	1.8984
SIRPB2	2.8259	13.2008	1.8921
S100A9	230.0380	853.6254	1.8872
LRRC8C	2.0986	10.4592	1.8868
PLEKHO1	15.0284	58.0848	1.8822
ADRB3	0.0725	2.9516	1.8814
BMF	12.1994	47.5235	1.8782
SLC7A7	1.4050	7.8297	1.8763
P2RY12	0.2119	3.4419	1.8739
HRH1	0.3304	3.8615	1.8696
ATP10A	0.4644	4.3341	1.8650
KCNA5	0.1571	3.1946	1.8580
TMOD2	1.0623	6.4632	1.8555
CLDN4	0.1182	3.0334	1.8509
HLA-DRB1	0.0065	2.6224	1.8477
P3H2	0.9528	6.0109	1.8440
RAP2A	16.4150	61.4066	1.8414
DUSP27	0.8308	5.5511	1.8393
SLC18B1	1.0839	6.4463	1.8373
DDIT4L	0.2998	3.6376	1.8350
KL	0.3493	3.7638	1.8199
SNX18	3.6612	15.4177	1.8165
RNASE4	0.7015	4.9473	1.8054
SECTM1	0.3546	3.7097	1.7978
CD300LB	0.5550	4.4026	1.7968
ATP2B1	5.2371	20.6510	1.7955
SLC7A2	1.4020	7.3191	1.7922
HLA-DPB1	0.5321	4.2955	1.7893

RUNX1C- specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
IL1R2	0.0866	2.7458	1.7854
SORBS3	0.6391	4.6447	1.7840
ALOX15B	0.0643	2.6414	1.7746
OLR1	0.3274	3.5415	1.7746
RASD1	0.2819	3.3786	1.7722
HRH2	3.1466	13.1575	1.7716
RNF19B	5.6128	21.5064	1.7670
CD1D	0.1490	2.9026	1.7640
MMP12	0.0065	2.4027	1.7574
UST	1.0957	6.0839	1.7571
SGPL1	3.4649	14.0822	1.7562
GADD45B	3.6204	14.5501	1.7508
CPVL	12.1932	43.0660	1.7399
CD80	0.4194	3.7366	1.7385
AXL	0.1702	2.8951	1.7349
SLC26A11	0.6173	4.3678	1.7307
SFMBT2	3.0290	12.3688	1.7304
CYSLTR1	14.2427	49.4740	1.7274
DPYD	2.1483	9.4229	1.7271
CST3	33.9564	114.6243	1.7258
LACC1	4.1165	15.8879	1.7228
TFEC	1.4258	7.0007	1.7216
OAS2	0.2831	3.1991	1.7105
FUCA1	15.8185	53.9377	1.7077
PKIB	46.6342	154.0756	1.7029
LRRC25	2.9803	11.8733	1.6934
OSBPL11	5.9445	21.4187	1.6908
TBC1D9	2.7014	10.9486	1.6907
TMEM200A	0.9148	5.1630	1.6864
SLAMF8	0.2737	3.0932	1.6842
CTSC	119.3962	385.6146	1.6831
MALT1	3.6557	13.9266	1.6808
CTSB	39.4222	128.2130	1.6765
SERPINA1	0.4601	3.6455	1.6697
PAK1	17.8109	58.5571	1.6627
GPR84	0.1819	2.7395	1.6617
CRIP3	0.2830	3.0582	1.6613
SPNS3	3.9221	14.4837	1.6534
FABP3	0.3778	3.3312	1.6524
ADAMTS10	2.6268	10.4000	1.6523
DSE	6.3350	22.0519	1.6520
GPC4	0.1040	2.4558	1.6463
HCK	48.9878	155.2923	1.6446
ICAM1	5.7920	20.1884	1.6414
TNFRSF25	0.6452	4.1313	1.6410
GFRA2	0.0946	2.4101	1.6394
KDM7A	4.6073	16.4615	1.6388
IMPA2	2.6521	10.3527	1.6362
WDFY4	4.1096	14.8353	1.6319
VSIG4	0.1170	2.4585	1.6306
TRPM2	0.5228	3.7083	1.6285
TBC1D14	15.7787	50.7298	1.6244
ASPH	9.6223	31.6930	1.6219
DOCK4	0.4775	3.5378	1.6189
PPARG	0.8957	4.8172	1.6176
FGD6	0.5815	3.8479	1.6161
SOCS3	8.7451	28.8700	1.6159
PCNX2	0.4493	3.4403	1.6153
CACNB4	0.3558	3.1496	1.6138
SMARCA1	6.8302	22.9053	1.6102
MEF2C	25.0302	78.0666	1.6029
ARPIN	2.9789	11.0425	1.5977
NOD2	3.3686	12.1897	1.5942
TNFSF13B	1.2483	5.7629	1.5888
ZNF366	0.2044	2.6143	1.5854
RUNX3	5.4560	18.3621	1.5845
ALDH7A1	3.8379	13.5062	1.5842

RUNX1C- specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
FCGRT	21.5756	66.6162	1.5826
IL10RA	29.6444	90.7321	1.5818
TIAM1	2.3763	9.1068	1.5818
ADAM9	2.0118	7.9371	1.5692
LAP3	30.0218	90.5197	1.5608
CD93	8.5874	27.2171	1.5573
SAT1	333.8391	981.7407	1.5533
GPR82	0.1094	2.2512	1.5512
EDNRB	0.2609	2.6911	1.5496
CEBPD	3.1659	11.0787	1.5358
PDE4B	2.0970	7.9684	1.5340
RPGRIP1	0.2226	2.5370	1.5326
DTNA	1.8264	7.1740	1.5321
GPR68	1.1796	5.2956	1.5303
GLRX	38.1914	111.4727	1.5210
SORL1	4.3952	14.4682	1.5195
ZNF217	16.9862	50.4762	1.5170
CD101	2.8766	10.0604	1.5125
SIPA1L1	2.9887	10.3772	1.5122
GPR35	4.3383	14.1460	1.5045
PTAFR	28.2361	81.9331	1.5042
SASH3	29.8112	85.9270	1.4964
DENND3	21.7381	63.1356	1.4960
RNF144B	0.3379	2.7690	1.4942
SCIMP	0.1758	2.3097	1.4930
SLC8A1	0.0708	2.0130	1.4925
PLSCR1	36.4718	103.9023	1.4852
GPR171	0.2059	2.3612	1.4789
TTC39C	2.5522	8.8804	1.4759
TRPS1	0.5293	3.2455	1.4731
TICAM2	0.7306	3.7834	1.4668
FAM170A	0.0065	1.7764	1.4639
SLC12A9	5.4700	16.8427	1.4635
PLAU	4.1101	13.0867	1.4629
MICAL1	14.8386	42.6354	1.4621
SOD2	25.7165	72.0879	1.4519
OTUD1	1.2650	5.1961	1.4518
ARHGAP26	2.5880	8.7826	1.4470
IL4I1	0.2286	2.3340	1.4403
GPR162	3.2296	10.4239	1.4335
TFEB	0.9413	4.2396	1.4325
TOX	3.1145	10.1018	1.4320
TREM2	1.2061	4.9524	1.4320
UNC93B1	4.1326	12.7724	1.4240
FARP1	0.8343	3.8993	1.4173
LYZ	141.7378	379.9348	1.4162
VSIG10	2.1001	7.2565	1.4132
CEBPB	1.7379	6.2910	1.4131
ABCA6	0.1420	2.0312	1.4084
TTC7A	4.2226	12.8512	1.4072
CTSO	0.2245	2.2437	1.4054
ARL4C	2.0374	7.0422	1.4048
ABCG2	0.0769	1.8509	1.4045
FZD2	0.3961	2.6915	1.4028
LILRB3	0.0801	1.8493	1.3995
PTPRK	0.0065	1.6521	1.3979
SKIDA1	0.2703	2.3472	1.3978
TIGAR	0.7125	3.5057	1.3956
PXDC1	0.3524	2.5572	1.3952
GAA	4.9673	14.6696	1.3928
RFTN1	12.0211	33.0423	1.3865
VENTX	0.9450	4.0838	1.3861
DHTKD1	10.0446	27.8231	1.3839
SNX10	10.0190	27.7225	1.3822
HP	7.1337	20.1951	1.3817
OLIG1	0.7883	3.6496	1.3785
SLC7A8	1.8644	6.4417	1.3774

RUNX1C- specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
ITGB7	3.2309	9.9897	1.3771
SYS1-DBNDD2	0.0400	1.6992	1.3759
MYCL	1.2470	4.8297	1.3754
EEPD1	0.9355	4.0111	1.3724
PSTPIP1	3.2323	9.8979	1.3646
PARP9	7.2338	20.1904	1.3638
RBKS	0.9508	4.0177	1.3630
IL1R1	13.2232	35.5825	1.3629
SRA1	11.0816	30.0603	1.3623
TSPAN33	6.6017	18.5425	1.3622
KLF4	0.3765	2.5375	1.3617
PLXNB2	9.5881	26.1504	1.3585
GSTA4	6.1171	17.2250	1.3566
TRIM25	14.3106	38.1700	1.3552
ABCA1	0.4430	2.6888	1.3540
CD200	3.4035	10.2378	1.3516
TTYH2	1.3278	4.9303	1.3491
TNFRSF13C	0.8222	3.6341	1.3466
KIAA0930	5.7788	16.2264	1.3455
CREM	7.5745	20.7566	1.3433
LRRK1	1.0993	4.3248	1.3428
CHCHD7	30.1587	77.7159	1.3370
TMEM106A	6.8319	18.7425	1.3339
C15orf38-AP3S2	0.0321	1.6015	1.3338
IFI35	6.4387	17.6889	1.3291
CXCL16	0.9889	3.9960	1.3288
EFHD2	14.1153	36.8871	1.3257
PLB1	3.1214	9.3229	1.3246
MEFV	0.5967	2.9970	1.3238
EMILIN2	13.4069	35.0324	1.3225
SLAMF7	0.1359	1.8330	1.3185
TBC1D12	2.0170	6.5099	1.3157
PIK3CD	12.3399	32.1872	1.3149
CHI3L2	0.2042	1.9916	1.3129
CD300C	12.6444	32.8791	1.3121
EPSTI1	1.3955	4.9472	1.3118
LAMP5	0.2877	2.1914	1.3094
CLECL1	1.1884	4.4217	1.3089
RASSF2	15.9241	40.8798	1.3072
RXFP1	3.0625	9.0410	1.3055
SGK1	3.6151	10.3935	1.3038
SLC38A1	27.4840	69.3063	1.3035
PLXDC2	2.1093	6.6657	1.3018
MEIKIN	0.2751	2.1424	1.3012
PRR16	0.0541	1.5948	1.2997
RNF149	18.0269	45.7826	1.2979
RAB7B	5.8130	15.7493	1.2977
ZFP36	19.9226	50.3838	1.2963
TVP23A	0.3291	2.2636	1.2960
IRF7	0.9413	3.7630	1.2948
RGS2	8.9346	23.3731	1.2947
GRN	22.2964	56.0951	1.2933
SPATA6	0.3444	2.2913	1.2917
DMXL2	12.2995	31.5366	1.2907
NPL	7.0602	18.7070	1.2898
C19orf38	9.4307	24.4564	1.2872
DHRS3	0.4046	2.4270	1.2868
PTGER4	0.5944	2.8855	1.2851
RCBTB2	9.7494	25.1610	1.2832
SH3TC1	4.0669	11.3228	1.2822
SOX2	0.0065	1.4471	1.2817
MAN2B1	20.7869	51.6771	1.2737
CEP112	0.3253	2.2040	1.2735
ST5	0.5280	2.6881	1.2712
MAML2	0.6154	2.8947	1.2696
AKNA	13.3133	33.3982	1.2650
PARP14	6.0605	15.9604	1.2643

RUNX1C- specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
FOSL2	4.8823	13.1280	1.2641
RETN	9.1797	23.4482	1.2640
GIMAP8	0.9428	3.6594	1.2620
CASP5	0.0065	1.4111	1.2604
KCNMB4	0.9170	3.5770	1.2556
PIGR	0.0065	1.4012	1.2545
FAM20A	0.1790	1.8125	1.2543
FES	25.7229	62.7083	1.2534
DPYSL2	9.7263	24.5611	1.2528
TYROBP	117.7270	281.4525	1.2504
MXD1	10.2013	25.5231	1.2436
SPECC1	4.3630	11.6869	1.2422
IL17RB	4.9198	13.0036	1.2422
MB21D2	0.5715	2.7148	1.2412
AFF3	0.1214	1.6497	1.2405
SBNO2	1.3891	4.6314	1.2370
ARHGEF40	4.0187	10.8141	1.2351
C15orf48	0.0322	1.4290	1.2347
DAPP1	12.5258	30.8114	1.2338
CBLB	3.5094	9.6056	1.2338
COLGALT1	21.5382	52.0068	1.2338
DIP2B	13.6222	33.3753	1.2332
RAB43	1.7800	5.5332	1.2327
SIPA1L2	0.9361	3.5313	1.2267
MAP3K8	8.9964	22.3455	1.2237
HEXB	31.6994	75.2841	1.2221
CREBL2	6.7793	17.1473	1.2220
PCM1	24.5534	58.6024	1.2219
MARCKS	7.5912	19.0231	1.2207
DHX58	0.5914	2.7078	1.2203
MPPE1	6.1812	15.7094	1.2184
C22orf42	0.0731	1.4813	1.2093
VDR	2.0745	6.0850	1.2044
CLEC2D	1.3494	4.4122	1.2039
SMCO4	3.0551	8.3298	1.2021
MMP19	1.4739	4.6898	1.2016
TMEM243	4.9230	12.5840	1.1975
IL10	0.2890	1.9490	1.1940
TXLNB	0.5426	2.5245	1.1920
SERPING1	25.6230	59.7892	1.1911
IGFLR1	4.5019	11.5529	1.1900
LILRA3	0.0065	1.2952	1.1893
ATP6V1G2-DDX39B	0.6457	2.7525	1.1892
LITAF	25.5466	59.4290	1.1867
FMNL2	2.3010	6.5049	1.1850
DTX4	0.2484	1.8368	1.1842
ANXA5	33.5634	77.3732	1.1811
ZFP36L2	31.0888	71.7420	1.1807
ITGAL	23.0119	53.3653	1.1789
AIG1	4.9770	12.5152	1.1771
ZFP36L1	8.7816	21.1126	1.1767
FBXO21	2.8712	7.7488	1.1763
STK17B	14.6417	34.2692	1.1730
NINJ1	11.7201	27.6624	1.1721
CPNE3	21.0387	48.6214	1.1709
RAPGEF1	9.6797	23.0092	1.1687
STARD13	2.5228	6.9016	1.1654
MPZL1	6.1113	14.9312	1.1637
ZNF532	1.3087	4.1659	1.1620
HOXA9	2.9441	7.8013	1.1580
DUSP3	25.9813	59.0384	1.1539
WDR91	3.6178	9.2689	1.1530
DDX60	0.9374	3.2983	1.1497
CARD11	1.6643	4.9092	1.1492
TESC	37.5560	84.5047	1.1490
TPK1	1.9035	5.4313	1.1473
C2orf91	0.0280	1.2760	1.1467

RUNX1C- specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
OASL	0.2194	1.6986	1.1460
SLITRK4	0.6419	2.6285	1.1440
EVI2B	27.1952	61.2867	1.1435
MX2	3.2307	8.3294	1.1409
MVP	10.1901	23.6756	1.1409
FOS	30.9171	69.2302	1.1378
FAM219B	12.3021	28.2502	1.1368
LAT2	29.6567	66.3231	1.1349
KIAA1551	21.4939	48.3626	1.1339
KLRF1	2.0567	5.7053	1.1333
CYP2S1	1.7517	5.0284	1.1315
MKNK1	15.1045	34.2563	1.1304
TNNI2	0.9164	3.1870	1.1275
SNX8	4.4767	10.9584	1.1266
IFIH1	2.6106	6.8668	1.1235
SH3BP5	5.6492	13.4692	1.1217
BCL11A	1.8471	5.1873	1.1198
AHRR	7.1087	16.6014	1.1181
PXN	13.2216	29.8259	1.1161
SPINK2	7.2935	16.9638	1.1150
ARHGEF10L	0.2441	1.6889	1.1119
NUAK2	0.1481	1.4814	1.1118
HADH	27.5588	60.6539	1.1103
KCNJ2	0.3443	1.9015	1.1099
MNDA	5.7076	13.4669	1.1089
NOXA1	1.9277	5.3136	1.1087
NMRK1	5.4478	12.8725	1.1054
MYO1E	1.9893	5.4309	1.1052
PTPN1	39.0586	85.0649	1.1033
RAPH1	0.5475	2.3241	1.1030
RILPL1	1.2858	3.9094	1.1029
RNF125	4.3817	10.5269	1.0989
IFIT1	0.6754	2.5862	1.0980
TRIM26	0.0065	1.1528	1.0969
PRKCA	9.1663	20.7069	1.0944
NCF1	4.7575	11.2896	1.0939
RNF43	0.1507	1.4519	1.0913
OAS3	3.0669	7.6625	1.0909
CEP68	5.4603	12.7377	1.0885
PREX1	14.3886	31.7193	1.0883
IRF2BPL	3.5235	8.6123	1.0875
HIP1	22.4084	48.7039	1.0863
SBF2	4.1068	9.7916	1.0794
FCMR	0.0316	1.1793	1.0790
DISC1	2.8958	7.2208	1.0774
ITPR2	7.1842	16.2645	1.0769
ASS1	1.7285	4.7544	1.0766
ITGB2	57.9809	123.2379	1.0748
CEACAM4	0.0065	1.1197	1.0745
BCL3	4.0206	9.5681	1.0738
DDTL	0.0065	1.1175	1.0731
BTN2A2	2.9888	7.3704	1.0693
CREB5	0.3415	1.8137	1.0687
SEMA3C	1.2544	3.7282	1.0686
SMIM12	4.1577	9.8077	1.0673
SERPINB8	3.5400	8.5075	1.0664
TMEM51	1.7645	4.7892	1.0663
TMEM86A	0.7032	2.5658	1.0660
YPEL2	2.4143	6.1418	1.0647
SLC35F6	2.1925	5.6759	1.0643
DNASE1L3	0.0065	1.1040	1.0638
TRIM38	8.6086	19.0772	1.0632
LAIR1	3.3673	8.1232	1.0628
HAAO	3.0046	7.3638	1.0625
BRI3	14.2396	30.8146	1.0619
CCDC146	0.9143	2.9902	1.0597
FXYD6	0.5966	2.3271	1.0593

RUNX1C- specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
CYBRD1	12.9053	27.8983	1.0554
MEF2A	12.0783	26.1787	1.0553
SPATS2L	1.6475	4.5013	1.0551
SETD7	2.6423	6.5605	1.0536
CDK18	1.2826	3.7368	1.0532
KCNE3	24.7085	52.2877	1.0516
USP28	5.2182	11.8858	1.0512
ABCG1	0.5813	2.2747	1.0503
EBI3	0.2441	1.5748	1.0494
FAM177B	0.0239	1.1167	1.0477
PIK3AP1	19.9394	42.2852	1.0477
S100A8	307.0215	635.2106	1.0465
ITPRIPL2	4.9687	11.3132	1.0447
ADGRG5	10.7950	23.3316	1.0447
RUBCN	2.5623	6.3269	1.0404
DIAPH2	6.5961	14.6141	1.0395
GNPTAB	30.3642	63.4627	1.0393
PTPRE	12.0748	25.8633	1.0388
RCAN1	4.1065	9.4860	1.0381
SOAT1	5.4732	12.2825	1.0370
VNN2	0.7741	2.6402	1.0369
TLR4	3.5043	8.2385	1.0363
GLIPR2	11.0974	23.8005	1.0357
PALD1	3.6122	8.4552	1.0356
FTL	1829.0759	3750.1983	1.0354
MLKL	4.1620	9.5688	1.0338
TNFSF13	5.7512	12.8202	1.0336
PCED1B	1.0469	3.1883	1.0329
SAT2	16.0282	33.7160	1.0277
TNFRSF10C	1.9256	4.9589	1.0263
CARD16	3.5538	8.2638	1.0245
CAMSAP2	0.6071	2.2666	1.0233
SNX7	0.3556	1.7552	1.0232
CSRP2	1.2756	3.6199	1.0216
SCARB2	6.9595	15.1319	1.0192
SEC61A2	2.4550	5.9940	1.0175
DEXI	8.7215	18.6755	1.0172
CXCL1	1.4863	4.0309	1.0168
GALNT3	2.1910	5.4536	1.0161
CCDC50	6.5588	14.2859	1.0160
ELK3	9.3228	19.8586	1.0148
BEST1	0.9281	2.8894	1.0124
ANKRD44	7.8833	16.9187	1.0123
CABYR	0.1485	1.3165	1.0121
CPQ	6.4954	14.1140	1.0118
CC2D2A	0.4487	1.9206	1.0115
QKI	17.5279	36.3456	1.0112
TRIM34	5.2261	11.5470	1.0109
TMPRSS13	0.9302	2.8891	1.0107
HLA-DRB3	0.0065	1.0266	1.0098
KCNK13	0.0306	1.0748	1.0094
GRAMD1B	0.3741	1.7630	1.0078
SESN1	8.8287	18.7323	1.0055
CLDND1	67.9722	137.2576	1.0033
CD274	1.2197	3.4487	1.0030
GHRL	0.6540	2.3114	1.0015
TBC1D8	3.5184	8.0431	1.0010
PDCD1LG2	0.0065	1.0143	1.0010
PIK3R5	9.1291	19.2654	1.0005

RUNX1C+ specific			
GeneID	RUNX1C+	RUNX1C-	logFC
PF4	407.1894	12.6352	-4.9038
PPBP	470.9270	18.6257	-4.5877
ITGB3	120.5848	4.2863	-4.5236
F2RL2	72.1927	2.2379	-4.4986
SELP	54.3543	1.7782	-4.3165
CMTM5	81.4454	3.5022	-4.1947
HBE1	160.6262	9.1677	-3.9906
MYLK	37.8432	1.4907	-3.9630
HBD	272.1830	17.2104	-3.9070
GP9	50.7529	2.9579	-3.7088
KCNK17	27.2310	1.2755	-3.6330
GP1BA	32.5027	1.7115	-3.6271
CLEC1B	21.3150	0.8521	-3.5908
LGALSL	28.0066	1.6347	-3.4607
GFI1B	65.5486	5.1370	-3.4388
GYPE	55.0826	4.1868	-3.4346
HPSE	19.8414	1.0132	-3.3719
TUBB1	32.1975	2.2228	-3.3647
C2orf88	29.8940	2.0136	-3.3578
DNAJC6	24.9718	1.5856	-3.3283
MYL4	96.5687	8.8985	-3.3011
GMPR	57.3117	5.0067	-3.2791
THBS1	197.1660	19.9407	-3.2423
KEL	8.5983	0.0150	-3.2413
RHAG	87.6630	8.4345	-3.2323
DMTN	23.0298	1.5930	-3.2121
HBA1	36.6793	3.0818	-3.2065
GYPB	108.1702	10.8869	-3.1991
CD24	316.3218	34.1396	-3.1748
KALRN	10.9174	0.3229	-3.1713
LTBP1	41.1551	3.6986	-3.1654
HBG1	781.7916	88.0384	-3.1361
GATA1	89.2283	9.7970	-3.0630
ITGA2B	268.7705	31.3240	-3.0611
HBG2	1180.9603	142.5230	-3.0418
ESAM	35.4592	3.4701	-3.0279
BEX1	18.8467	1.4485	-3.0189
GAD1	11.2272	0.5312	-2.9973
TMEM40	10.7668	0.5047	-2.9672
ABCC3	15.8833	1.1910	-2.9459
SERPINE2	44.1786	4.9758	-2.9184
NXF3	30.5602	3.2360	-2.8973
PRICKLE1	16.3624	1.3550	-2.8822
SCGN	39.5320	4.5279	-2.8742
PRKAR2B	31.9884	3.5060	-2.8721
CKB	39.6798	4.5733	-2.8677
GJA4	9.9921	0.5238	-2.8507
CD226	19.0269	1.7998	-2.8385
TIMP3	164.5715	22.3877	-2.8236
ELOVL7	13.8463	1.1494	-2.7881
LAT	151.9159	21.6256	-2.7567
MTURN	32.7398	4.0414	-2.7426
CLCN4	10.5382	0.7860	-2.6916
SLC40A1	130.2583	19.4343	-2.6833
SFRP5	8.4525	0.4728	-2.6821
FAXDC2	15.7888	1.6183	-2.6808
MYCT1	7.6769	0.3771	-2.6555
TUBA4A	42.5565	6.0037	-2.6367
MMRN1	71.9797	10.8743	-2.6197
ITGA6	18.4259	2.1926	-2.6052
CRYM	11.6427	1.0840	-2.6009
FREM1	20.3005	2.5231	-2.5960
TEK	10.0083	0.8285	-2.5899
PDZK1IP1	7.7884	0.4765	-2.5734
SNRPN	21.2454	2.7629	-2.5636
HBA2	23.4130	3.1381	-2.5606

RUNX1C+ specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
F2R	115.3393	19.3341	-2.5164
PNCK	29.6455	4.3773	-2.5107
NRGN	15.1577	1.8556	-2.5004
TJP2	12.9362	1.4630	-2.5003
DLK1	22.9875	3.2653	-2.4916
CXCL3	15.6606	1.9932	-2.4767
LCN2	5.7308	0.2161	-2.4685
ALAS2	28.6642	4.3736	-2.4648
SLC37A1	28.2145	4.3774	-2.4417
KLF1	13.1505	1.6105	-2.4384
PEAR1	17.1927	2.4326	-2.4060
MFSD2B	8.7463	0.8438	-2.4022
GRAP2	17.1399	2.4508	-2.3942
C1orf116	5.9604	0.3347	-2.3827
HGD	6.4788	0.4366	-2.3802
ICAM2	51.8714	9.1852	-2.3760
LY6G6F	4.2518	0.0150	-2.3714
OSBP2	8.9804	0.9384	-2.3642
GFOD1	13.5125	1.8362	-2.3553
AQP3	21.2932	3.3830	-2.3466
GABRE	24.0055	3.9430	-2.3388
KDR	11.8388	1.5552	-2.3290
PDLIM1	91.1233	17.4899	-2.3168
HBZ	13.0103	1.8145	-2.3155
SEPT5	10.2704	1.2751	-2.3085
CCL5	8.1992	0.8736	-2.2957
IGF2	17.6137	2.7979	-2.2931
SLC6A8	64.8717	12.4979	-2.2869
ANK1	9.2659	1.1343	-2.2660
COL18A1	8.9271	1.0729	-2.2597
RAB27B	104.3083	21.3538	-2.2360
DEFA1B	13.3332	2.0625	-2.2266
PLEK	227.5691	48.4728	-2.2079
FAM171A1	28.0334	5.2869	-2.2073
CH25H	10.7326	1.5438	-2.2055
DEFA1	13.5161	2.1510	-2.2038
UBASH3B	23.2405	4.3767	-2.1726
CCDC173	11.9823	1.8810	-2.1719
DEFA3	16.0960	2.8025	-2.1687
ABCC4	162.9254	36.1643	-2.1410
EFS	6.2502	0.6523	-2.1336
F2RL3	11.2052	1.7850	-2.1317
HEPH	24.1746	4.7960	-2.1188
SLC22A17	8.1689	1.1228	-2.1108
PDE3A	6.5319	0.7459	-2.1090
BCL6B	12.1491	2.0576	-2.1045
EFHC2	14.3224	2.5711	-2.1012
EHD3	13.3034	2.3457	-2.0960
CXCL2	15.8236	3.0219	-2.0645
ASIC4	6.6721	0.8379	-2.0616
CD84	50.0329	11.2271	-2.0613
FUT1	6.7679	0.8972	-2.0337
PCYT1B	13.1163	2.4504	-2.0325
GLOD5	3.3972	0.0760	-2.0309
TREML1	4.1283	0.2775	-2.0051
LXN	76.7412	18.3989	-2.0027
HEMGN	12.2949	2.3259	-1.9991
TBPL1	49.9254	11.8272	-1.9892
OR1J2	10.4860	1.9066	-1.9825
TFR2	10.2430	1.8634	-1.9732
EXOC3L2	4.7701	0.4893	-1.9540
PCSK6	10.8581	2.0817	-1.9441
TRIM58	9.4129	1.7092	-1.9424
GYPA	10.5451	2.0240	-1.9328
ARHGAP6	15.7267	3.3838	-1.9319
CNST	37.3677	9.0655	-1.9305

RUNX1C+ specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
LRRC8B	23.7230	5.5184	-1.9233
SMIM1	8.2543	1.4596	-1.9117
TPM1	33.4898	8.2355	-1.9009
CDKN2D	21.5082	5.0396	-1.8979
TSC22D1	64.8483	16.7016	-1.8953
PBX1	7.0490	1.1866	-1.8801
BCL2L1	56.1378	14.5438	-1.8781
SPDYC	3.9043	0.3377	-1.8743
KCNK5	14.5948	3.2700	-1.8688
PF4V1	3.2822	0.1780	-1.8620
CCND1	8.1951	1.5323	-1.8604
ARL15	25.2003	6.2459	-1.8543
DEFA4	7.4550	1.3641	-1.8385
SAMD14	9.6652	1.9956	-1.8320
ABCB6	12.4563	2.7806	-1.8316
NES	5.1994	0.7436	-1.8300
CD2	7.7503	1.4677	-1.8262
SLFN14	4.5390	0.5622	-1.8260
SLC35D3	2.8610	0.0965	-1.8161
LYVE1	7.8263	1.5115	-1.8133
GADD45A	38.1597	10.1676	-1.8101
TMEM98	17.8026	4.3820	-1.8047
ENDOD1	6.8663	1.2719	-1.7918
TOM1L1	4.4779	0.5884	-1.7860
PRKCQ	28.5925	7.5811	-1.7860
ICAM4	17.1494	4.2751	-1.7826
SYTL4	26.9915	7.1727	-1.7761
ADRA2A	3.2528	0.2473	-1.7696
RAB38	19.0048	4.8802	-1.7664
IGFBP3	6.1023	1.0944	-1.7617
ECE1	11.1720	2.6002	-1.7574
TSPAN32	34.9946	9.6635	-1.7551
KCNH2	27.2019	7.3650	-1.7534
TMEM163	6.2128	1.1502	-1.7461
KIFC3	8.8737	1.9464	-1.7447
LPAR5	17.9532	4.6738	-1.7401
PDGFC	21.0323	5.6012	-1.7388
ZNF185	5.5951	0.9776	-1.7377
PKHD1L1	2.7125	0.1198	-1.7292
MPL	10.9605	2.6083	-1.7289
TSPAN9	5.8434	1.0652	-1.7284
TMSB15A	8.9346	2.0289	-1.7137
ITGB5	10.5419	2.5259	-1.7108
ASAP2	5.9827	1.1367	-1.7084
RAB6B	3.9778	0.5235	-1.7081
SLC39A4	3.5457	0.3914	-1.7080
RAP1GAP	5.8253	1.1032	-1.6983
SLC24A3	15.6967	4.1541	-1.6958
RBM38	19.4085	5.3021	-1.6953
STXBP5	96.4275	29.2885	-1.6856
DNM3	4.0440	0.5692	-1.6846
FHL2	29.1336	8.4121	-1.6788
SLC27A2	21.4914	6.0945	-1.6646
TGFB1I1	7.7731	1.7714	-1.6625
ATP2A3	66.7198	20.5638	-1.6510
RHCE	3.8043	0.5302	-1.6506
GATA2	178.8254	56.2938	-1.6501
CTDSPL	14.8864	4.0712	-1.6474
LRRN2	3.1281	0.3186	-1.6464
YPEL4	11.2945	2.9313	-1.6449
MEX3B	18.7037	5.3084	-1.6431
HYAL3	12.2633	3.2759	-1.6331
PLA2G3	10.3088	2.6469	-1.6327
CCR4	2.7907	0.2226	-1.6325
PTGDR2	24.8490	7.3501	-1.6303
C20orf203	3.7679	0.5402	-1.6302

RUNX1C+ specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
KCNQ4	2.7807	0.2243	-1.6267
MAGIX	3.8195	0.5716	-1.6167
CD82	49.8164	15.5844	-1.6155
HBB	3.9749	0.6236	-1.6154
GCSAML	122.5330	39.3320	-1.6149
STRA6	4.6049	0.8320	-1.6132
ZBTB16	10.0033	2.6016	-1.6112
CXCL5	3.0638	0.3313	-1.6100
GYPC	56.4666	17.8475	-1.6084
SERPINE1	3.5858	0.5049	-1.6075
EPDR1	12.4639	3.4250	-1.6053
ITGB1BP2	6.5100	1.4725	-1.6029
PDE5A	3.0462	0.3373	-1.5973
MFSD6	9.1879	2.3703	-1.5959
STARD8	15.7627	4.5467	-1.5955
RASGRP3	20.0224	5.9674	-1.5932
OSBPL6	5.2327	1.0680	-1.5916
PTGS1	158.2573	51.8931	-1.5902
YOD1	18.9514	5.6279	-1.5899
GNG11	55.6855	17.8546	-1.5881
FRRS1	31.8031	9.9505	-1.5828
RGS6	3.9999	0.6706	-1.5816
SORBS1	6.8153	1.6358	-1.5681
ID1	4.6744	0.9139	-1.5679
CD55	55.7449	18.1840	-1.5646
P2RY1	17.7406	5.3377	-1.5641
ACSM3	8.5690	2.2548	-1.5558
CRISP3	16.9297	5.1237	-1.5499
PBK	9.6308	2.6417	-1.5456
PIK3CB	46.6277	15.3206	-1.5451
SEC14L5	2.4352	0.1787	-1.5432
MYH10	15.1623	4.5552	-1.5407
TEAD2	3.9934	0.7180	-1.5393
ST6GALNAC1	4.4379	0.8740	-1.5369
DBN1	19.4734	6.0620	-1.5356
SLAMF6	8.5290	2.2957	-1.5317
FKBP10	22.9937	7.3066	-1.5303
CCDC68	3.2179	0.4649	-1.5257
LRRC32	2.5031	0.2179	-1.5243
OLFM2	3.3001	0.4982	-1.5212
EHD2	5.6033	1.3012	-1.5208
NTRK1	10.5071	3.0168	-1.5184
MYL9	50.1445	16.8972	-1.5148
CD3D	16.6317	5.1807	-1.5123
PRKCB	35.2632	11.7871	-1.5038
COL4A1	5.0358	1.1348	-1.4994
TSPAN18	2.8429	0.3625	-1.4960
SEMA7A	42.4014	14.4306	-1.4919
ACTN1	57.3195	19.7488	-1.4909
CYP11A1	9.2483	2.6518	-1.4887
SELENBP1	17.7062	5.6677	-1.4883
SLC39A3	27.0865	9.0328	-1.4852
C6orf223	37.9487	12.9279	-1.4836
ETS1	16.1891	5.1495	-1.4830
IRX3	1.9091	0.0440	-1.4785
TNIK	12.5452	3.8705	-1.4756
TUBAL3	2.3868	0.2187	-1.4745
ZFPM2	3.2158	0.5174	-1.4742
FAM117A	12.6843	3.9285	-1.4733
CNRIP1	27.6454	9.3252	-1.4721
AQP1	3.2877	0.5456	-1.4720
GBP4	3.1471	0.4957	-1.4713
MYEOV	2.0600	0.1058	-1.4684
LYL1	25.3562	8.5271	-1.4680
ST8SIA6	23.8789	8.0157	-1.4644
FUT8	22.5738	7.5451	-1.4640

RUNX1C+ specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
GDPD1	6.6347	1.7679	-1.4638
TESPA1	27.8623	9.4940	-1.4596
ACKR2	3.9366	0.8036	-1.4526
SLC45A3	47.2568	16.7290	-1.4446
NT5M	5.9160	1.5458	-1.4418
VANGL2	9.7370	2.9550	-1.4409
LTC4S	13.3708	4.2948	-1.4405
EGF	3.0227	0.4841	-1.4386
PLPP1	66.0007	23.7449	-1.4370
SERPINI1	6.9589	1.9416	-1.4360
TRPC6	2.2756	0.2121	-1.4343
TCN1	17.6329	5.8965	-1.4339
RUFY1	53.3757	19.1266	-1.4339
STON2	13.4989	4.4072	-1.4230
WNT11	2.9125	0.4598	-1.4223
ARHGEF12	14.0058	4.6020	-1.4215
PLCB4	2.4816	0.3026	-1.4183
FOXN2	30.8825	10.9309	-1.4181
GJB5	2.1468	0.1820	-1.4127
STON1	7.7402	2.2837	-1.4123
PDZD8	19.4137	6.6837	-1.4097
SKAP1	26.0177	9.2340	-1.4005
LIMS1	64.3620	23.7977	-1.3982
PRG2	1793.3210	681.2815	-1.3950
CCNI2	9.4382	2.9708	-1.3944
CACNA2D3	4.0290	0.9142	-1.3935
HDC	259.8718	98.3455	-1.3928
AREG	8.5163	2.6249	-1.3925
IL13	11.0595	3.6072	-1.3882
SNCG	7.7336	2.3395	-1.3870
SH2D2A	11.9966	3.9806	-1.3837
UROS	53.6602	19.9774	-1.3817
ABHD2	34.4801	12.6279	-1.3804
DNAJC12	11.7466	3.9192	-1.3736
MFAP3L	1.6252	0.0150	-1.3710
TMOD1	5.8293	1.6425	-1.3698
EFNB2	2.4131	0.3206	-1.3698
SLC22A23	4.6502	1.1865	-1.3697
MEIS2	15.1025	5.2324	-1.3694
AKAP12	24.5956	8.9315	-1.3658
ANGPT1	57.5136	21.7617	-1.3622
UGT1A6	8.5830	2.7311	-1.3609
ZFPM1	2.4944	0.3617	-1.3596
NT5C3A	36.6954	13.7013	-1.3584
NDFIP2	13.8482	4.7970	-1.3569
CMAS	66.1422	25.2885	-1.3528
SIGLEC8	4.0249	0.9675	-1.3527
TMEM45A	16.6755	5.9222	-1.3524
CASS4	14.2230	4.9648	-1.3517
GALNT10	17.5857	6.2835	-1.3515
LRTM1	1.5861	0.0150	-1.3493
LIN28A	2.0782	0.2083	-1.3491
IL2RB	6.7770	2.0587	-1.3463
MARCKSL1	50.5862	19.2977	-1.3457
CRYBA4	1.8003	0.1053	-1.3411
RNF144A	33.9889	12.8229	-1.3398
RILP	1.5572	0.0150	-1.3331
AFAP1	2.0553	0.2190	-1.3256
CHST2	4.9671	1.3823	-1.3247
UGT2B28	2.9593	0.5812	-1.3243
SOCS1	9.5729	3.2247	-1.3235
CR1	16.8989	6.1581	-1.3222
MRAP2	2.1089	0.2444	-1.3210
LIPG	4.8416	1.3418	-1.3187
STAR	6.8775	2.1627	-1.3166
CD9	28.0642	10.7065	-1.3119

RUNX1C+ specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
MRC2	7.3981	2.3840	-1.3113
SPINK4	2.3993	0.3704	-1.3106
PIEZO2	5.5900	1.6587	-1.3096
MINPP1	20.4883	7.6808	-1.3076
EMILIN1	24.2800	9.2235	-1.3061
CLC	1747.7061	707.7230	-1.3030
SV2A	16.0350	5.9122	-1.3013
MYO1D	12.4836	4.4723	-1.3010
C19orf33	2.3556	0.3633	-1.2995
MS4A3	177.9294	71.6934	-1.2995
MRVI1	4.8497	1.3789	-1.2981
OSM	12.1067	4.3316	-1.2977
HIPK2	9.7784	3.3863	-1.2971
GSE1	23.6500	9.0363	-1.2964
CCND3	171.6093	69.3412	-1.2951
ICAM5	4.4148	1.2071	-1.2948
AVPR1A	1.7384	0.1176	-1.2930
FKBP1B	12.3982	4.4712	-1.2921
MAP1A	9.4737	3.2940	-1.2864
KIF21B	18.2344	6.8929	-1.2851
PSRC1	28.9128	11.2801	-1.2844
IL1RL1	107.5314	43.5580	-1.2844
DLC1	7.7022	2.5740	-1.2838
PDLIM7	32.4029	12.7206	-1.2836
NLK	32.8073	12.8941	-1.2829
MANSC1	3.8156	0.9800	-1.2822
BAMBI	8.9558	3.0977	-1.2807
DNAAF3	2.2594	0.3422	-1.2800
CLCNKB	1.6435	0.0905	-1.2774
UGT2B11	2.8226	0.5809	-1.2738
CTTN	3.7979	0.9870	-1.2718
GNAZ	1.7041	0.1213	-1.2699
GPRC5B	5.3874	1.6491	-1.2697
AMHR2	4.1380	1.1319	-1.2691
MEIS3	3.7297	0.9662	-1.2664
PTK2	18.4538	7.0934	-1.2652
APOBEC3B	29.2050	11.5777	-1.2639
TTC7B	23.6508	9.2661	-1.2637
RAB44	15.5699	5.9064	-1.2626
ST3GAL4	47.3443	19.1632	-1.2616
S100A16	5.4133	1.6749	-1.2616
SLC43A3	91.5514	37.6669	-1.2592
LRP12	20.2728	7.8924	-1.2584
NEXN	1.7910	0.1671	-1.2578
GPRC5C	45.5171	18.4561	-1.2575
LIF	34.2605	13.7536	-1.2570
SSX2IP	8.9730	3.1776	-1.2553
SIGLEC12	29.1021	11.6372	-1.2522
GJA5	2.0334	0.2738	-1.2518
IL9R	11.6481	4.3262	-1.2477
NET1	37.9138	15.3872	-1.2477
VCL	85.9621	35.8236	-1.2398
SH3BGRL2	2.3889	0.4359	-1.2389
ALOX12	4.5288	1.3436	-1.2382
WT1	7.1443	2.4593	-1.2353
CA2	8.7641	3.1549	-1.2327
SPX	1.6921	0.1460	-1.2321
SYP	5.3341	1.6994	-1.2305
GBP5	3.4852	0.9126	-1.2296
IGFBP4	12.5759	4.8028	-1.2262
CPA3	630.8627	269.1059	-1.2261
LIPC	1.4079	0.0315	-1.2230
BASP1	34.0894	14.0331	-1.2229
CMIP	49.8152	20.8073	-1.2204
SLC6A9	2.6175	0.5527	-1.2202
OPN3	62.5212	26.2876	-1.2190

RUNX1C+ specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
CAMK1	39.5861	16.4406	-1.2185
RBPMS2	1.8961	0.2485	-1.2139
TGFB1	60.3693	25.4741	-1.2129
IL32	18.5808	7.4530	-1.2119
ABLIM1	5.0806	1.6251	-1.2118
NFATC1	8.3043	3.0231	-1.2096
LINGO1	2.1396	0.3602	-1.2068
AKAP2	23.9061	9.8015	-1.2053
PLCH1	5.0644	1.6332	-1.2035
ETV4	4.1996	1.2581	-1.2033
EPOR	7.0360	2.4911	-1.2028
TREML2	26.8227	11.0907	-1.2024
LNX1	2.4814	0.5134	-1.2018
RGS18	86.3322	37.0307	-1.1993
SHC1	58.6164	25.0080	-1.1968
CLU	19.8272	8.1149	-1.1922
CD8B	7.0293	2.5141	-1.1921
ARL4A	14.6465	5.8550	-1.1906
TFRC	102.2147	44.2340	-1.1902
LGALS12	4.1895	1.2777	-1.1880
SPN	84.0086	36.3379	-1.1870
GUCD1	23.6265	9.8297	-1.1852
IL4	7.4611	2.7236	-1.1841
TXK	4.2535	1.3132	-1.1834
SEMA3F	3.0560	0.7899	-1.1802
IL5RA	8.4519	3.1720	-1.1798
STXBP6	4.2690	1.3268	-1.1792
FLNA	98.3984	42.9354	-1.1778
SOX4	24.9479	10.4704	-1.1777
CKAP4	19.5152	8.0786	-1.1762
LDLRAD3	7.5707	2.7933	-1.1760
MED12L	1.5982	0.1503	-1.1755
ADORA2A	3.4480	0.9714	-1.1740
CMPK2	24.7663	10.4461	-1.1706
C1orf61	1.7569	0.2252	-1.1700
MGAT3	13.7830	5.5814	-1.1675
KRT79	26.7759	11.3659	-1.1675
MLLT11	53.9281	23.4734	-1.1663
ADGRG1	15.7089	6.4462	-1.1660
TGFBR3	4.4993	1.4545	-1.1638
STRIP2	4.5733	1.4912	-1.1617
MPP7	2.2158	0.4375	-1.1616
AQP10	2.7228	0.6647	-1.1611
IL33	1.8146	0.2590	-1.1607
INO80C	30.9180	13.2776	-1.1606
BPI	8.8096	3.3947	-1.1584
SLC2A14	8.4821	3.2563	-1.1556
ATP6V0A2	34.5030	14.9501	-1.1544
PRF1	6.7707	2.4939	-1.1532
LMO4	46.2660	20.2613	-1.1526
S1PR1	7.1161	2.6550	-1.1509
EMID1	4.8982	1.6573	-1.1503
HRC	5.6832	2.0144	-1.1487
CHRNE	2.1760	0.4351	-1.1461
FRY	5.7756	2.0670	-1.1435
RAMP1	4.2615	1.3890	-1.1391
ECSCR	4.1199	1.3257	-1.1384

RUNX1C+ specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
CDH5	3.1790	0.9038	-1.1343
ARAP3	8.1675	3.1806	-1.1328
MICALL2	7.2190	2.7541	-1.1305
HES6	9.3554	3.7333	-1.1295
DLG5	1.2196	0.0150	-1.1289
CCRL2	12.3866	5.1243	-1.1282
LDOC1	5.2050	1.8391	-1.1280
TMBIM1	83.0022	37.4763	-1.1265
SERPINB9	8.6840	3.4420	-1.1244
HES1	2.8261	0.7560	-1.1236
CDR2	14.7141	6.2233	-1.1213
PHLDB1	4.6703	1.6072	-1.1209
RAB33A	53.6513	24.1668	-1.1187
PRTN3	11.8724	4.9285	-1.1185
ADGRF5	1.6669	0.2288	-1.1179
TEC	18.9082	8.1747	-1.1176
KLHDC8B	7.0302	2.7012	-1.1174
NEDD8-MDP1	1.2008	0.0150	-1.1166
MAP1B	4.5019	1.5377	-1.1164
UNC13D	45.3171	20.3954	-1.1142
SLC18A2	64.2029	29.1550	-1.1125
CSF2RB	96.5008	44.1193	-1.1117
YES1	19.3933	8.4422	-1.1109
SMYD3	61.5458	27.9645	-1.1106
MAPRE2	23.3334	10.2817	-1.1090
NBEAL2	35.1329	15.7853	-1.1061
APOBEC3D	8.9108	3.6058	-1.1056
VSIG2	1.3404	0.0881	-1.1050
NOSTRIN	2.1908	0.4845	-1.1039
ADARB1	5.0479	1.8147	-1.1034
ZNF683	2.8409	0.7888	-1.1025
CLEC11A	78.9024	36.2921	-1.0994
MPP1	86.5976	39.9220	-1.0980
ARHGEF6	166.2099	77.1173	-1.0979
ACOT11	7.5226	2.9866	-1.0961
GCKR	1.1678	0.0150	-1.0948
FHOD1	24.8878	11.1223	-1.0946
LAMB2	10.0479	4.1747	-1.0942
OLFM4	2.6032	0.6891	-1.0931
C10orf82	6.8247	2.6686	-1.0928
PLXNA4	1.3336	0.0941	-1.0928
GDF11	11.4233	4.8285	-1.0918
TMEM136	1.7795	0.3079	-1.0876
CCDC28B	15.3257	6.6862	-1.0868
FERMT3	93.1384	43.3218	-1.0868
TPST2	48.4021	22.2642	-1.0865
DAPK2	2.6416	0.7168	-1.0849
GNG8	1.9609	0.3970	-1.0837
EPAS1	14.9046	6.5048	-1.0836
SPTB	2.1009	0.4672	-1.0796
EFNA1	2.2685	0.5469	-1.0792
PPP1R15A	38.7695	17.8225	-1.0792
DAPK1	25.6077	11.6790	-1.0694
EGFL7	63.5077	29.8033	-1.0664
DAG1	8.8522	3.7047	-1.0664
SKI	12.5721	5.4839	-1.0657
CSGALNACT1	1.6899	0.2862	-1.0644

RUNX1C+ specific (continuation)			
GeneID	RUNX1C+	RUNX1C-	logFC
NRIP3	7.6018	3.1170	-1.0630
PTX3	5.9018	2.3075	-1.0612
TUBA8	1.7750	0.3306	-1.0604
COL24A1	13.5044	5.9620	-1.0589
AZU1	7.0051	2.8515	-1.0555
EMB	13.5118	5.9834	-1.0552
SVIP	14.9244	6.6695	-1.0540
TUSC3	9.4122	4.0151	-1.0539
PLIN2	141.6868	67.7298	-1.0538
NFE2	92.6239	44.1391	-1.0525
C12orf57	17.3557	7.8554	-1.0516
SH3GL3	1.7352	0.3197	-1.0515
MYZAP	1.1471	0.0372	-1.0497
LCP2	82.7068	39.4442	-1.0494
DOCK6	3.0334	0.9491	-1.0492
NUDT4	29.3808	13.6989	-1.0474
LURAP1L	1.3260	0.1257	-1.0470
MYB	41.3491	19.4978	-1.0469
PRKAB1	30.2153	14.1101	-1.0467
NMT2	10.3018	4.4733	-1.0461
NFIB	1.3203	0.1241	-1.0455
PAPSS1	87.2055	41.7599	-1.0446
UBE2S	41.4885	19.6075	-1.0439
AKR1C1	4.1250	1.4926	-1.0399
NCKAP1	5.8468	2.3308	-1.0396
SCARF1	3.3785	1.1331	-1.0375
CEACAM1	1.7112	0.3225	-1.0357
FAM83A	2.3917	0.6561	-1.0342
SULT1A3	14.8904	6.7619	-1.0337
CORO6	4.3782	1.6311	-1.0314
MBOAT1	13.3406	6.0169	-1.0312
CYTL1	64.4705	31.0908	-1.0287
XK	3.1024	1.0163	-1.0248
MAX	97.8121	47.6550	-1.0221
TP53INP1	21.8238	10.2765	-1.0172
MAGI1	2.1334	0.5487	-1.0166
SIGLEC6	34.2205	16.4109	-1.0164
FGFR2	4.8079	1.8746	-1.0146
CDKN2C	4.8257	1.8867	-1.0130
PPIF	97.1352	47.6771	-1.0115
PPFIBP1	13.4868	6.2026	-1.0082
ZNF792	10.0144	4.4762	-1.0081
FAM20B	26.5721	12.7109	-1.0079
PLEKHA4	3.3046	1.1416	-1.0072
TCTEX1D1	4.4888	1.7317	-1.0067
CHML	9.3314	4.1439	-1.0061
SPRY1	4.7978	1.8868	-1.0060
SYNJ2	3.2068	1.0949	-1.0059
BTK	227.7614	112.9251	-1.0058
SLC48A1	12.2376	5.5973	-1.0047
AGAP1	3.2975	1.1433	-1.0036
FAM102A	8.5986	3.7915	-1.0024
LIPE	2.1527	0.5749	-1.0014
LRRC8D	24.2424	11.6108	-1.0012

Supplementary Table 1: Gene expression levels and differential gene expression in the RUNX1C- and RUNX1C+ cell populations

Tables including expression levels of the differentially expressed genes between RUNX1C- (blue) and RUNX1C+ (orange) specific genes (no Dox treatment), ranked by log2 fold (LogFC) of the FPKM +1 (fragments per kilobase of transcripts per million mapped reads) values of genes differentially expressed (two-fold change) in uninduced RUNX1C- and RUNX1C+ (CD45+ CD34+) populations.

RUNX1C+					
3 Dox		5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated	Upregulated	Downregulated
CYP1A1	ENPP3	CYP1A1	GAR1	CYP1A1	WT1
SPINK4	HPGD	RUNX1T1	HPGD	RUNX1T1	CNRIP1
TMEM176A	CEACAM4	CLIP3	GPR34	PLD2	TFRC
TMEM176B		GPR32	IL4	SMARCD3	CTSG
DPP4		TMEM176A	C8orf59	PLA2G4C	KIT
		PLD2	CKS2	NAAA	ITGA4
		TMEM120A	ITGB3BP	PRAME	S100A12
		SMARCD3	PFDN4	KREMEN1	SULF2
		TNFSF12	MSMO1	GBP2	C11orf21
		SIDT2	CETN3	ARMCX2	SPN
		KLK13	RIDA	CLIP3	LIF
		HLA-B	MAD2L1	IRF6	NFE2
		RNF213	AP1S2	LEFTY2	KRT1
		MAPK13	GOLT1B	SMARCA1	PLA2G3
		COLQ	CENPK	PIK3IP1	SLC43A3
		KREMEN1	HMGCS1	TMEM120A	FABP5
		SPINT1	AK6	MAPK13	LMO4
		PRAME	KRT1	ZG16B	IL4
		SPTAN1	CNIH1	SLC35F3	SLC48A1
		PLA2G4C	G2E3	KLK13	SPIN4
		FLNB	LMO4	HLA-B	PSMG1
		SLC44A2	S100A12	THRSP	ODC1
		CLSTN1	NDFIP2	KIAA0040	DNMT3B
		SLC26A11	SEC61G	GSTA4	ADAMTS14
		AXL	CENPH	RAB3B	PTAFR
		NLRP1	FANCL	TFPI2	MYB
		ACTA2	CYP1B1	COLQ	MYBL2
		TIMP2	C3orf58	FAM114A1	CKS2
		VAT1	SNRPG	ASAP3	RRS1
		PLSCR3	FAM72C	FMNL2	ZFP36L2
		CDKN1A	ITGA4	RNASE4	MAD2L1
		ZG16B	PARPBP	MTMR11	ZNF730
		CTTN	UGT1A6	LURAP1	CENPH
		ARSD	TCTEX1D1	ITGAV	INSIG1
		SLC35F3	RFXAP	BCL6	FADS1
		SYNPO	SGO1	NLRP1	SLC27A2
		PIK3IP1	HSPB11	EPHX2	TESC
		SPTBN1	KIT	P2RX4	DUSP10
		KDELR1	CENPQ	CES4A	GSE1
		EPHX1	LRRCC1	CDKN1A	CHAC2
		IGLON5	SNRPE	SLC26A11	SPINK2
		LRP1	CDK1	RASSF8	FGR
		CYP2S1	SPIN4	AXL	KCNK5

RUNX1C+			
5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated
GBP2	HMMR	OPTN	ECE2
MARK4	TGDS	FBXL2	FST
ECM1	METTL5	IL1RAP	OSM
SERPINB9	LSM5	NABP1	KCNK17
P2RX4	HSPE1	CTTN	MYH10
GDF3	NUF2	PDE1B	BIRC5
KIAA0040	DMC1	ARRDC4	NPM3
NATD1	CHAC2	SOCS2	RAP1GAP2
TRIM16	MND1	PHC1	CACNA2D3
ARMCX2	FAM72A	LGALS3	STAR
TWIST1	ASPM	DNAJB5	SLA
PLEKHG2	RPA3	TSPAN7	CAMK1
CPT1C	FAM72B	FBLIM1	MYC
CD99	SULF2	FAM166B	NKG7
PLCD1	TNFSF10	PLIN5	STEAP3
ITGA5	C3orf14	GSTM3	SPR
CASS4	ERH	SERPINB9	MFSD2B
CES4A	MS4A2	ARSD	AQP3
PDE1B	YEATS4	TESK2	NME1
MX1	LIF	FMNL3	CYP1B1
VAMP5	HMGB1	PPP2R5B	MARVELD1
C19orf38	EXOSC8	C1orf115	MCEMP1
LEFTY2	GAPT	SPINT1	SLC7A1
COL1A2	DUSP10	TDRP	DHRS9
PLIN5	HPGDS	H6PD	RNASEH2A
FGD5	CTSG	SPTAN1	SELPLG
BMF	KIF20B	IGLON5	ANK1
LDOC1	TIMM17A	BEX2	ADCY7
SMARCA1	ZNF302	RGS9	PRELID1
OAS3	SKA2	TCEA3	GCSAML
TBC1D16	ZNF85	VNN2	MRTO4
IGF2R	COMMD8	SYNPO	GATM
SERPING1	NFYB	TRAK2	PRPS2
FBLIM1	CCNA2	TRIM16	TYMS
LGALS3	SGO2	CMYA5	GMPR
DPP4	ZNF107	GSTM2	MBP
TUBB6	DEPDC1	CCDC85A	MTHFD1L
PINK1	ZNF138	CTC1	S100A8
ASAP3	UGGT2	CPT1C	PCNA
ST5	GGH	FRMD6	DUT
MAPK3	WT1	HFE	HSPE1
GPR153	SPC25	IRF9	HPGD
RELB	ACTR6	PLEKHB1	PFAS
PLCD3	CCDC58	STAG3	TREML2
MTMR11	ZNHIT6	ERMN	OGG1
SHF	SMC4	DDX17	CCNI2
TMEM176B	CEP70	ICOS	SLC50A1
SGSH	C20orf203	EGR1	VAV1
CREB3L1	FAM72D	VAMP5	GFI1B
SLC9A1	TDRD9	RGPD5	ADGRG5
MMRN2	TFRC	TJP2	SNRPF
HIP1R	PTPN22	CALCOCO1	DSCC1
DOCK6	ZNF92	GJA5	ABCB6
LRRC8A	ZNF730	ST5	PLCB2
SPINK4	CYCS	RGPD6	FEN1
SNX15	CCDC18	IL11	OPN3
LGALS3BP	RHEBL1	TMEM45B	ITGAM
TESK2	PSMG1	IGF2R	LAT2

RUNX1C+			
5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated
MAP4	C4orf46	AHR	MCM4
TSPAN7	GLMN	FSTL1	CDKN2D
WFDC1	GPR65	RCAN2	DMTN
BRSK1	CKLF	TMEM178A	ELOVL6
MRAP2	CKAP2	SHF	C4orf46
NECAP2	CALB2	NPHP3	ID2
DNAJB5	CENPE	BEX3	MYL4
TJP2	PPM1K	CCDC71L	TRIP13
NCK2	DEK	ECHDC2	FAM216A
LRG1	RPS18	PCYOX1	MYCN
PLEKHB1	GCSAML	DNALI1	ADORA2B
OPTN	CNRIP1	TMEM176A	CCNA2
TAGLN	PLA2G4A	SGK1	S100A9
ACSF2	SLF1	WSB1	IGFBP7
IFI6	DBF4	OAS3	TMTC4
PLCG1	ACYP1	APOD	RAB37
RNASE4	KNTC1	ACTA2	ELF4
ICAM1	XRCC4	CD99	WDR12
EHD2	FAM102B	FNDC3A	C20orf203
APOD	MPHOSPH6	CLIP4	ALDH1B1
CDC42BPB	MS4A3	TNFRSF9	ADGRE2
BEX3	DSCC1	IFI16	NRARP
TBC1D8	ATAD5	ENPP4	LST1
NFKB2	DLGAP5	TWIST1	STK10
PPP2R5B	SMIM10	NATD1	PTPN22
FMNL2	ZFP36L2	ULBP2	FAM72C
NDRG1	NAE1	SIDT2	AURKB
THRSP	LSM3	GDF3	WNT5B
MYO7A	PBK	FGD5	HBB
PLTP	OIP5	DPP4	CDC45
RASSF8	H2AFZ	PHF1	SMIM10
MARCKSL1	ZNF680	EPHX1	MAGEF1
UCHL1	GYPA	PLAU	CHAF1A
MINK1	SLCO4C1	KLK6	RUVBL1
PTPRF	IL13	MAGED2	GMNN
CPAMD8	HMGB3	ABAT	PARVG
GAD1	DIAPH3	SMIM10L2A	BNIP3
FZD7	EEF1E1	FLNB	RNASEH2B
ACADVL	LPXN	LAMC1	CD2
EGR1	MRPS18C	SPTBN1	LRRCC1
HLA-E	INSIG2	FABP3	SLC8A3
DAPK1	ZNF273	NRBP2	HMBS
DCHS1	CENPA	MRAP2	CALB2
BCAM	JCHAIN	CD79B	CDCA5
GJA5	HPF1	COL1A2	HMGB3
TCEA3	OSM	RCAN3	DHFR
PCED1B	S100A8	TBC1D16	UGT1A6
FCGR1A	FST	LRP1	EXOSC8
CETP	LST1	LRG1	METTL1
SASH1	PMS1	CLDND1	MRPL3
SOCS2	PFDN6	RGPD8	MCM7
ZFP36	TOP2A	SERPINB4	GGH
AMIGO2	HIST1H4C	CLSTN1	F2RL3
CSTB	KRT72	RETSAT	CENPK
TNNT1	CDKN3	MAP4K3	HJURP
LURAP1	CCNI2	TNFSF12	SIGLEC10
SLC25A23	ECT2	HIP1R	SLC25A15
TGFBR2	PCNA	GRAMD1C	TNFAIP8L2

RUNX1C+			
5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated
CD109	MCEMP1	AMIGO2	SLA2
RHBDD2	STAR	MX1	BUB1
IL27RA	HAT1	UNC13B	MCM2
PXDN	GEN1	IFI6	CST7
PTPRU	CD52	ECM1	RMI2
HIF3A	IMMP1L	SLC44A2	NINJ2
AEBP1	MTBP	ATP6V1B2	ACAT2
NRBP2	DDIAS	CREBRF	C6orf223
COL6A2	CEP55	KLF8	ST8SIA6
MAP4K2	GATM	SUN1	CDKN3
CLDN7	NDC1	ACAD11	MS4A3
BCL6	PRIM1	TWSG1	OIP5
MAGED2	TPRKB	IRF8	FADS2
RHOC	HIST1H1B	FZD7	POLR3K
AHR	NKG7	GCH1	EFHC2
CD79B	TTC32	ADAMTS9	SGO1
PCDHGB7	RNASEH2B	CA13	ARRB2
ISG20	ZBED5	NEO1	RRM2
NOTCH3	ARHGAP11A	CETP	DMC1
BAIAP3	MORN2	TUFT1	ORC1
MYOF	KIF14	CXCL10	CHI3L1
RRAS	RFC3	GNG7	PADI4
TFPI2	FBXO43	YPEL5	SH2B3
UNC13B	ENPP3	SBF2	CENPF
PLVAP	ZNF804A	JAM3	SLC43A1
CCDC85A	RMI1	UCHL1	ABCE1
FAM131A	PLK4	TUBB6	CDC20
DLGAP4	NCAPG	UBE2H	PROK2
PLEKHA4	ACAT2	TAGLN	CD244
FLT4	HMGN5	SASH1	HMMR
SEMA6C	DHRS9	ARMC9	H2AFZ
TMEM45B	CDC7	PLCD1	IL1B
ZNF581	MGST1	NECAP2	TMEM97
LRP4	PLAC8	BSDC1	TMEM246
HOMER3	KIF11	CBLN3	HK3
GRIK5	SPINK2	BRSK1	PLEKHA2
MVB12A	ESCO2	WFDC1	CENPU
DENND1C	COQ3	F2RL1	FAAP24
IFI35	ADCYAP1	ZFYVE16	ZNF367
ST6GAL1	NDUFA5	CD109	FAM72A
BCL6B	OMA1	CPAMD8	HSPD1
RCAN3	C11orf21	CECR2	FAM102B
MFGE8	ZNF683	KLHL6	CBFA2T3
C1orf115	RAD51AP1	KIF13B	CYTH4
RBPMS	HIST1H2BO	NDRG1	HMGN5
IFIT3	ID2	SAT2	RNF125
MYO1C	ZWINT	P2RX7	POLD2
RSPH9	RFESD	TUSC3	SLC40A1
JUN	UQCRB	CYTH3	SPC25
CTNNA1	CENPU	NFKB2	PLAC8
NEO1	CARHSP1	IFIT3	ASF1B
SGK1	GDPD1	B3GNT9	PRF1
ICAM2	TMEM156	IFI35	OAF
BEX2	CENPI	TMEM65	NOD2
LPIN3	CCDC15	PLSCR3	POP5
FABP3	HBB	SERPINB3	CDCA7
ZSWIM4	ZNF675	PINK1	RFC3
CD83	NDC80	CBX7	IL13

RUNX1C+			
5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated
PCDHGB6	PRIM2	SEPT10	C3orf58
RASIP1	SMC2	EZH1	MTFR2
ITGA9	HIST1H2AJ	AP1S1	PRKDC
P2RX7	TMSB15A	PLEKHH2	ITGAL
RGS12	C5orf34	PLCG1	NUP210
COL27A1	DEPDC1B	SLC37A3	APOBEC3B
RCAN2	CENPW	PDE4DIP	CCND3
PROCR	B3GNT5	QPCT	CENPW
CD81	C12orf60	SSBP2	SLC11A1
KLC2	TTK	PCDHGB7	KNTC1
PTPRS	BNIP3	ANXA2	PPIF
HSPB8	CENPF	TM4SF1	AARS
ADAMTS9	AIF1	RBFOX2	MTHFD1
DNALI1	FRA10AC1	SPOCD1	CCNF
IFI16	CRLS1	MAP4	TDRD9
ATN1	CATSPER1	HAP1	PUM3
ACVRL1	KIF18A	ACADVL	AIF1
GRASP	ITGAM	LIFR	FHL2
OAS1	RMI2	S100Z	ANGPT1
NECTIN2	RACGAP1	IL27RA	CAPN11
GRTP1	FABP5	TBC1D8	FKBP4
C2orf66	MNS1	GNG12	RCL1
MFAP2	DSN1	SETD7	IL9R
SLC5A10	IL18RAP	CADM4	ENO3
COL6A1	EMB	ARMCX3	CENPA
CACNA1H	HIST1H2BH	ZNF251	ZWINT
CERS4	HIST1H4A	SERINC1	MND1
IFI27	IL9R	KIAA0895	LAPTM5
PLOD1	CDC25C	GPR32	NCAPH
ID3	PRG2	CASP7	MS4A2
SHE	RFC4	ABCB4	CCNB1
HELZ2	DEPDC4	HOOK1	PDGFRB
SLC7A10	JAZF1	MYOF	GINS1
ATP9A	PTTG1	SWT1	PLCH1
TM4SF1	DNA2	TAGLN3	PSAT1
PML	HELLS	KCTD21	SH3BP2
CTSC	BUB1	LDOC1	DCTPP1
CAPN12	BUB1B	H1F0	PCYT1B
SFN	SLA	HIF1A	PASK
GLDC	IGFBP5	RELB	HIRIP3
KCTD17	HIST1H4E	PRKCZ	SNX20
PPP1R14A	IGFBP7	MAML2	MLC1
KREMEN2	LMNB1	ZNF608	FCN1
KRT18	NEIL3	SNX33	NUDT1
PRKAR1B	MEIOB	PLEKHG2	PLEKHG3
SPOCD1	GZMB	CLDN7	BCOR
CADM4	PRF1	COX6B2	NCR3LG1
OASL	DCLRE1A	B2M	GNA15
PHLDB1	OPN3	IKZF4	GYPE
NPTX1	S100A9	VAT1	RCOR2
EBI3	APOC2	IFIT1	GALNT6
GPX4	MAGOHB	MALT1	KIF18B
NACAD	FAM200B	TMC7	CCNB2
IFIT1	GPR85	SLC35B4	UBE2C
TMIGD2	ADAMTS3	NRN1	HMGB1
CD37	ANKRD18B	UBE2L6	FAM72B
TUBB3	RGS5	SLC17A5	PA2G4
RSPO4	SFR1	CDKN2B	HMGCS1

RUNX1C+			
5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated
MAML2	ZNF90	PCED1B	PIM2
ITGAV	MED31	SOAT1	MGST1
NUMA1	MPC1	AR	RPA3
FCGR1B	CCDC136	LY96	TMEM163
LRP10	MCM8	LTA4H	CMSS1
CDC34	CLEC11A	DOCK6	FANCA
FAM222B	GPAT3	BBS1	TPM1
MSRA	DHFR	GPX3	TAGLN2
MVP	IDI1	JUN	ORC6
SAT2	SKA3	EIF1	UGGT2
VPS9D1	CPA3	ARRDC3	LCP1
MED25	SLCO5A1	SLC41A1	SIGLEC8
ATP6V0C	SKA1	PDGFD	PNP
GIMAP4	HIST1H2BM	KIAA1671	HSPA4L
ERG	MTFR2	CDH24	MCM10
GIMAP8	EAF2	ACSF2	HP
NRN1	RNASE2	CC2D2A	CPXM1
TAGLN3	XRCC2	BTN2A2	RCC1
ZYX	ZDHHC13	EIF4A2	CACNA2D4
CTSB	FBXO4	PPFIBP1	MARS2
ARHGEF15	ZGRF1	CASP4	BIN2
TMEM74B	CLC	EPS8	PGAM5
AP1B1	ZBED8	B3GLCT	ADAMTS1
EGR2	POLQ	ITGA5	BAG2
SERTAD1	PPIH	ADD3	POLA2
SOX7	FANCB	TGFBR2	CR1
TAX1BP3	RPL22L1	GIMAP4	CCND1
RNASEK	MYB	PHTF2	CDC6
CXCL10	DNAH14	IL13RA1	SLCO4C1
KLHL3	HSPA4L	CLIP1	CPT1A
MIIP	PLA2G3	CTTNBP2NL	CKAP2
RPS2	COMMD7	MYO9A	CHEK2
CACFD1	SMIM11B	GPR153	RAD54L
C8orf58	SLC48A1	EPB41L1	SLC7A5
TMEM178A	BLM	TVP23C	CXXC5
SPARC	GYPE	HIF3A	LGALS12
LTBR	CCNA1	SPINK4	NT5M
EVC	RNASE3	GPR50	ULK3
ULBP2	CR1	HLA-E	CCDC169
PLXNA3	KIF15	SERINC3	PARPBP
TEAD2	GINS4	RBPMS	RPS6KA1
COL9A2	METTL21A	PARP14	HAUS8
VGF	ALOX5AP	ZDBF2	ZNF804A
SHANK3	UBE2T	FBXO21	SNRNP25
HSF1	WDHD1	C1R	MCM5
QPCT	RBPJ	LZTS3	GINS2
TUBB4A	LYZ	NCOA3	CDC25A
NUCB1	SUGT1	DTX3	TIMM21
CH25H	BORA	TMX3	MKI67
VIM	HCST	NTAN1	SLC1A4
MAP1LC3A	CEP152	SHE	MUC1
PDGFRA	CDC45	CNPY4	DDIAS
SEMA6B	NUSAP1	CREM	CCDC15
CRYM	GYPB	RSPH9	TTC27
PKN1	EFHC2	ICAM1	ISOC1
EPS8	SASS6	GLDC	MCM3
TGFB1I1	AP5M1	IFIT2	TOMM40
IFIT2	HBD	ISG20	RHEBL1

RUNX1C+			
5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated
CEACAM1	RAD54L	HID1	RNF24
TRIM22	CCDC169	EGR2	RFC4
FAM89B	AQP3	CTSC	THBD
ACTG1	ADCY7	AIFM2	KIF21B
JUP	CKAP2L	PLEKHA4	DIS3L
SEMA4B	CLECL1	FAM131A	ASIC4
NR4A1	RRM2	PCMTD1	FAM72D
TMEM115	CXorf21	HSPB8	NDC1
FOXP4	CYP11A1	FNDC4	CMTM5
GRINA	HIST1H2BI	CTNNA1	RNASE3
FLT1	FANCI	C19orf38	HEMGN
PALM	ZNF563	GIMAP6	GNL3
SHISA5	CAMK1	CAMSAP2	XRCC2
YPEL3	HP	CD53	ADGRG1
TSTD1	HIST1H1D	FCGR1A	NDFIP2
LIN7B	PTX3	LPIN3	NANP
TRIOBP	RFX8	SGSH	BUB1B
GSN	ZNF93	RGS17	SLC38A5
SH3D21	ZNF239	YPEL2	FANCG
ECHDC3	AURKA	SNX18	SKA3
LMF1	ATP10D	FLT1	CTSW
HIST1H1C	CMC1	TRIM22	TOP2B
ANXA2	CD200R1L	TULP3	HBD
KLHDC8B	PSAT1	CD83	BRIX1
DDX58	CCNE2	AMOTL1	FAM57A
AES	ASNS	C11orf95	GYPA
NOS3	APOBEC3B	PCMTD2	SIGLEC9
LAMB2	MS4A6A	PLCD3	PTPRE
SYNGR1	BRCA1	EVI2B	DCLRE1B
APOE	CD38	XPR1	GTF3A
IGSF8	HIST1H2AH	DDX58	CSF1R
SLC27A1	POC1B-GALNT4	ZFP36	CMIP
UBXN6	CHEK1	ANKRD12	TOR3A
VASN	TPM1	SPACA6	CSE1L
FAM171A2	HEMGN	SCG5	TIMP3
HYAL2	POLE2	FBLN5	LMNB1
MGRN1	TCN1	BAIAP3	ADCYAP1
SNAPC2	DEFA4	CASP9	SHMT1
LYPLA2	EXO1	ING4	GINS4
RGL1	HGF	CYSTM1	PRG2
ACVR1B	E2F8	FERMT2	DTL
DPYSL4	LGALS12	GALT	CATSPER1
ZNF593	PIGW	CCL28	GYPB
TUSC3	L2HGDH	WDR45	RANBP1
BCL7C	HIST1H1E	TP53INP2	IARS
DLK1	MYL4	USP27X	GPAT3
AP2A2	SLC31A2	CAPN12	IGLL1
TNFAIP8L3	CAPG	HLA-DRB5	IMP4
NPDC1	CD300A	LRP4	FBL
FAM69B	HDC	ZNF827	MIPEP
TYMP	ZNF69	FAM49A	ESPL1
TMEM8A	ZNF695	SPARC	PRR11
MAP1S	CST7	SULT1C4	KIF22
ZNF358	UQCRHL	CAST	ERVMER34-1
MAP2K2	CXCL2	TFPI	ANKLE1
TSR3	HIST1H4B	PARP9	GBGT1
PLD3	PLA2G4B	ST6GAL1	RNASE2
ZNF219	HIST1H3F	IL2RG	E2F1

RUNX1C+			
5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated
HIST1H2BK	HBG1	TMEM74B	RRM1
EEF2	C1orf162	TMEM176B	ALYREF
FCGRT	CXCL3	CCDC173	PLA2G2A
GRN	OTULIN	LGMN	MYLK
TEK	SPP1	IFI27	ARHGAP22
PHLDA1	TUBA4A	OASL	MAP1A
PGLS		BTN2A1	TEC
SPATA2L		MYO1E	SMC2
PITPNM1		C19orf66	GAL
DBN1		OAS1	TNFSF10
PIM3		ATP9A	AP1S2
OAS2		SNX15	NUF2
ARVCF		C2orf66	FAM178B
TRPM4		SNX16	SLC27A5
IRF2BP1		CXCL8	SCFD2
FOSL2		TSLP	ATP2B4
HLX		PATJ	FOXM1
GAA		LGALS3BP	SSRP1
CLPP		TOB2	CENPP
PRR12		JAK1	GPT2
ASMTL		FAM89A	CD300C
CCL3L1		CEACAM1	S100A4
LRRN3		GIMAP8	HPDL
TMEM134		DDX60L	ENPP3
IRF7		KLHL3	AHSP
AKAP12		SOX8	C1QBP
NUDT16L1		TCEAL9	TUBA4A
SHC2		CSTB	KIF2C
HIST1H2AC		PCNX2	EXO1
DMPK		LPAR6	DLGAP5
AP2A1		STAT4	MNS1
PNPLA2		COL6A2	ICOSLG
ENC1		SAMD9L	S100A2
REXO1		GRASP	NCAPD2
ATG2A		CCNG1	JADE1
TMEM234		RGS12	TK1
ATF5		TMEM98	PRIM1
GNG12		BCAM	PSMC3IP
CPTP		PLK2	MTURN
KDM6B		GSDMB	SPAG5
ULK1		SP110	CENPV
ASPG		MAP3K12	PTX3
GNG11		PGAP1	LRR1
TJP1		DHRS1	PHKA1
RFNG		IFT81	IL18RAP
MXRA8		SLC35G2	NOP16
MED16		SLC16A4	KIF15
HSPB1		STX12	TPX2
UBALD1		APOL2	DHRS11
LUZP6		FCHSD1	ASNS
SPATS2L		MLF1	MEIOB
HMG20B		DDX60	POLQ
RAB5C		L1TD1	HPGDS
HIST1H2BD		PHLDA1	PHGDH
RGN		LRRC70	UBE2T
GNB2		MAPK3	CDC7
BAD		CREBL2	PTGIR
MEGF8		INTS6L	FUCA2

RUNX1C+		
5 Dox	10 Dox	
Upregulated	Upregulated	Downregulated
HLA-C	SCARB2	MFSD3
MXD4	NFKBIZ	SH3BGRL3
APBA3	GRTP1	BRI3BP
MAP3K11	HECW2	TRIM58
PPDPF	ABCD4	DNAJC6
NBL1	COL27A1	HYAL3
LLGL1	SGPL1	ST3GAL4
SYS1-DBNDD2	FCGR1B	TOP2A
CCL3L3	SYP	CD55
PIP5K1C	ICAM2	SRM
NCKAP5L	PTPRU	KCTD15
UCKL1	AEBP1	PRPS1
WDR13	PDGFRA	ZNF683
PAFAH1B3	CFH	TALDO1
CCDC106	PTPRF	TUBG1
GUK1	GABARAPL2	EHD3
MEGF6	ARHGEF15	BOLA3
AHRR	NOTCH3	IL21R
MDK	IFIT5	IMPA2
GADD45GIP1	GABARAP	SUSD1
AGPAT2	IFIH1	CHAF1B
GADD45G	PTPRS	RMND5B
ZNF580	DAPK1	PLEK
ISG15	FAM171B	PLK1
KLHL23	MDFIC	PTTG1
RAB35	SEMA6C	PLK4
ALKBH7	RSPO4	CPNE7
OGFR	CLEC2B	CDC25C
C3orf18	COL6A1	ITGB7
ARHGDIA	ITPKC	CHKA
NOTCH1	PDK4	MYO19
STUB1	HBP1	PMS1
SPSB3	BLVRA	GPR85
STMN3	STAP2	MCM6
FKBP8	PCDHGB6	SNTB1
BIRC7	KCTD17	RFC5
NR1H2	CH25H	ZNF273
TNRC18	DSEL	WDR76
MAPK8IP3	CUL7	CYP11A1
RAVER1	AHCYL2	ACOT11
UBTD1	CACNA1H	CDK1
HSPG2	MAP2K4	ADA
SOD2	DTX3L	ARHGAP11A
PLXNB2	SLC5A12	ANKRD18B
HIST2H2AA3	DYNLT3	PUS7
STRN4	NPTX1	RAB11FIP5
CORO1B	DAB2	YOD1
GIPC1	C5orf15	ENOSF1
CIC	TRIM34	CPA3
EPN1	PROS1	FANCI
	GJA1	DEPDC1
	HIST1H1C	C15orf39
	FHL1	E2F2
	TLR7	SKA1
	LMBR1L	GCAT
	LBH	PRKCQ
	CD70	SYCE2
	SAT1	NCAPG2

RUNX1C+ 10 Dox	
Upregulated	Downregulated
PDCD4	GYPC
SEMA3A	MSRB1
TSPAN6	ANGPT2
FSBP	RRP9
SIPA1L1	PIK3R6
EML2	EME1
BATF2	BRCA1
ENC1	CYCS
CBLB	LAT
TBC1D15	NLRC3
MYO6	KIF11
RBMS2	PDLIM1
RNASE1	KLF1
SLC22A4	GP1BA
RHBDD2	PDIA5
TNNT1	KCNH2
APOL3	SRPRB
MPZL1	NDUFV3
NFAT5	PHF19
MFAP2	DEPDC1B
PAPSS2	SFRP5
CREB3L1	HSPH1
C1orf54	SLC4A8
RAB13	CD82
BCL6B	DENND2D
CAB39L	MGAT3
NQO1	TIMELESS
KLC4	PRIM2
GAP43	HDC
ATP6V0A4	ALAS2
MMRN2	RAD51AP1
MCOLN3	CLC
ASPHD1	GSR
STAT1	E2F8
DNMBP	NETO2
RGL1	DENND4B
CDC42EP3	RAD51
SAMSN1	ALB
EBI3	FANCD2
ERG	NFKBIA
RPGRIP1	C21orf58
HDAC9	TMC8
LOXL2	ADAMTS3
ADAM9	PIF1
RRAS	PPIH
LRRN3	TUBB1
GAD1	CD300A
SLFN12L	SLC6A9
HOMER3	ZNF93
APP	BRIP1
CALCRL	PDSS1
ACVR1B	LRRC8D
IL15	PIK3CB
CSRNP2	HS6ST1
SLC25A42	LCN2
FGD6	ORAI3
SH3D21	ACOT7
EPHA4	FAM83D
TPST1	SH2D2A
RGN	VRK1
UGCG	TSPAN32
SLC35D2	RAVER2

RUNX1C+ 10 Dox	
Upregulated	Downregulated
OAS2	HIST1H1B
GPR137B	ZBTB16
CALD1	METTL21A
CA2	CKLF
BST2	ANPEP
TNFRSF10D	CD84
IL6ST	SDF2L1
C1S	DEFA4
MAPK11	IGFBP5
APBB1	UBASH3B
DDX5	ZNF239
GNG11	P2RY2
MOB3B	SPX
MERTK	RASGRP2
LIMS4	CDKN2C
GIMAP2	CDCA8
ST8SIA4	RAB27B
LIPH	ZNF90
MMP1	ABCC3
XAF1	SS18L2
ANG	THBS1
LYSMD1	OXCT1
FADS3	PQLC2
KIAA1191	RAB24
HRC	LSS
C1orf61	ESCO2
AP1AR	NCAPG
CLK1	KRT72
RAB40B	WDHD1
PLIN2	GATA2
TJP1	AGAP2
KRT18	LY6G6F
SERINC5	CLSPN
SERPINH1	CCNA1
SFN	MRPS26
NACAD	CLEC11A
ARC	HIST1H2BH
WDR19	SLC9A3R1
HAPLN3	KCNQ5
RBM24	NCAPD3
ARMCX6	CXCL2
TAX1BP3	FLOT2
MCL1	ANGPTL6
APOE	MANSC1
KLC2	ITGA6
RALA	NDC80
SOD2	PKM
GDF2	LTBP1
PLVAP	CDCA2
CACHD1	LYZ
SLC7A10	LFNG
AKT3	EIF4EBP1
TCP11L2	NRGN
OLAH	SLC25A45
VIM	TICRR
PDE4B	MPP7
FAM222B	BLM
SUMO1	PCK2
TEAD2	GFI1
LPP	VOPP1
TXNIP	FCER1G
RAI14	C1orf116

RUNX1C+ 10 Dox	
Upregulated	Downregulated
TCF7L2	TFAP4
AKAP12	MRM1
ZNF581	CHEK1
HIST1H2BD	NEIL3
KLHDC8B	SYTL3
C22orf31	SCGN
INHBA	JCHAIN
A2M	PPM1M
SERPINB8	DNA2
CACFD1	EAF2
STIM2	SLC39A3
ALS2CL	HCST
UNC119	CBX5
PHLDB1	ITGB3
PROCR	GLA
MAP1LC3A	RABEPK
FOS	CYC1
HOXB7	HIST1H4C
SALL2	UBASH3A
EHD2	ERCC6L
CXCL11	PDP2
VPS9D1	ZNF485
ACVRL1	POLE2
CRHBP	SPNS3
GNG2	GPR174
SLC5A10	UHRF1
GPRIN3	S100A6
CRYBG3	C2orf88
CTSB	LEF1
SPATS2L	ANLN
HIST1H3E	KCNQ4
NISCH	RELT
PRUNE2	MPC1
SERTAD1	CCR1
SEMA6B	NAGLU
CAP2	DTYMK
TMIGD2	HIST1H2BM
COL9A2	PYCR1
NINL	GJB5
PNRC1	MZB1
HERC6	KRT79
CDRT4	PSD4
ZSWIM4	CENPI
ABCA9	KIF18A
CDR2L	BIN1
TNFAIP8L3	C19orf48
ASPG	ATAD2
DPYSL4	PTPN6
PVRIG	NUP35
HELZ2	KIF4A
ABTB1	MICAL2
SHISA5	RAC2
LIMCH1	TFR2
HIST1H2AC	GJA4
GFRA3	SERPINE2
NPR2	GINS3
FSTL3	POC1A
TSTD1	MAGIX
SULF1	PF4
GPX4	WDR4
VGF	C5orf34
AES	MCM8

RUNX1C+ 10 Dox	
Upregulated	Downregulated
RAMP2	CXCL3
GIMAP5	ARHGEF39
MVP	AURKA
PLXNA3	TLCD1
NCAM2	SUSD3
KIF12	S100B
MXRA8	GAB3
IFI44L	CCNE2
TMEM243	TTF2
MIIP	LAGE3
LIN7B	TTK
ARVCF	RXRA
C11orf52	GPR65
TYMP	KIF23
TEK	TSPOAP1
IFI27L2	NR2C2AP
TMEM234	TIPIN
SLC2A3	NRROS
SOX7	POC1B-GALNT4
TUBB2B	SELP
INPP5E	MELK
C3orf18	GRAP2
PRKAR1B	FANCB
AP2A2	APOC2
ANKRD37	FCER1A
BIRC7	CEP152
HOXB2	GPR132
KREMEN2	ADAM15
DMPK	MTSS1L
LOX	SLC35D3
KLHL23	GTSE1
IRF7	GATA1
IGSF8	RETN
TMEFF1	WNT11
AKR1C3	PIM1
SHROOM4	CYBB
NR4A1	HELLS
TRPM4	FAM53B
ENDOV	HES6
KLF6	CD320
BAIAP2	CRLS1
GADD45G	TRAIP
CA9	CKAP2L
CDC42EP5	PIGW
SNAPC2	HAUS7
EMC10	TCN1
PALM	SH2D3C
HCFC1R1	PRAM1
SHC2	SPDYC
TUBB3	GRK6
UCKL1	HSH2D
GRN	HBG2
MST1	PRICKLE1
WDR13	PTGES2
RFNG	CENPM
LIMS3	WDR62
MEGF6	CCDC86
	RFESD
	CBSL
	OSBPL6
	SH3BP1
	SLC1A5

RUNX1C+ 10 Dox Downregulated				
DOK2	RCCD1	DERL3	SLC25A10	PKMYT1
LGALSL	HGD	SAPCD2	HBA1	TREML1
MAGOHB	ADGRG3	APOBR	SLC2A6	RUVBL2
ZNF69	ITGA2B	RELL2	EXOSC5	YDJC
TNFAIP2	MPO	ZNF695	NCF4	EGLN3
LTC4S	SEPT1	CRYBA4	KATNB1	SLC17A9
RECQL4	FBXO41	CAPG	ITGB1BP2	NINJ1
SPC24	FARSB	PTPN18	NME2	SYTL1
CD200R1L	DNAH14	GP9	HIST1H4B	FERMT3
IL1RN	SLC6A8	LSM4	GNG8	CDT1
MEX3B	CD34	C1orf162	HBZ	EMID1
RAC3	MYEOV	CEP128	SLC16A3	PIGY
CENPN	CBS	SLC19A1	PDE5A	GPR35
ECI1	HGF	AGTRAP	CCDC78	CARD9
XK	DZIP1	VPS53	LSP1	ICAM4
IFRD2	CLCN4	QTRT1	ISOC2	FLNA
DAPP1	TYROBP	HERC2	ATP2A3	OBSL1
CD38	DEAF1	TSPAN33	SOCS1	INF2
COMMD7	NHP2	F13A1	PRTN3	PPBP
ITGB2	NT5DC3	NBEAL2	FASN	SPP1
HBE1	CEBPE	VARS	DDIT4	BCAT1
HBG1	C5AR1	TGFB1	EMILIN1	ABCA7
SFXN2	JDP2	GALNT14	SLCO5A1	TPSB2
				SPATC1L

RUNX1C-					
3 Dox		5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated	Upregulated	Downregulated
CYP1A1	ENPP3	CYP1A1	CLEC4A	CYP1A1	TIFAB
SPINK4	HPGD	RUNX1T1	LPAR6	RUNX1T1	MYL4
RUNX1T1	CEACAM4	EHD2	HPGD	SLC35F3	IL10RA
GAD1		SLC35F3	CEBPA	GLDC	HPGD
SLC35F3		DOCK6	CD200R1	TM4SF1	CAMK1
		SPINK4	GPR34	DOCK6	CLEC4A
		MARCKSL1	RNASE6	ZG16B	CD1E
		GLDC	LST1	ARMCX2	CD200R1
		CLIP3	ENPP3	PRAME	ITGAL
		ZG16B	CYBRD1	TFPI2	LST1
		GPR153	LPXN	GBP2	ADCY7
		ID1	HPGDS	SPINK4	LAT2
		AEBP1	CD36	EHD2	NCF2
		FLNB	CD1E	IL11	CTSG
		FLT4	ADAMDEC1	TJP2	SNX20
		PTPRS	RAP2A	FLNB	IL1R1
		TRIM16	CXorf21	TRIM16	DCANP1
		PRAF2	LRRK2	DNALI1	BLNK
		PRAME	NHLRC3	CLIP3	CLEC4F
		EGR1	CLEC4F	SALL2	SLA
		VGF	CAMK1	EGR1	LPXN
		TJP2	PTPRO	FSTL1	CD300LB
		TRIB3	HFE	PLIN5	PPM1M
		PLIN5	CD1B	JAM3	TSPAN33
		PXDN	NCF2	RAB38	ADAMDEC1
		PLEKHA4	TIFAB	THRSP	RAP1GAP2
		CREB3L1	NDRG2	PTPRS	TMPRSS13
		PTPRF	CLCN5	KREMEN1	NOD2
		PCSK6	TMEM144	PPFIBP1	LIF
		GPR32	C8orf59	LURAP1	ASGR2
		FGD5	ZNF780A	MTMR11	HRH2
		CCDC8	SERPINF1	KLK13	MBP
		NFIC	CKLF	FGD5	FGD2
		SMARCD3	AP1S2	PCSK6	ENPP3
		LDOC1	KRT1	ENPP4	S100B
		SPTBN1	GPAT3	COLQ	WNT5B
		HIF3A	C9orf72	CPAMD8	TESC
		FKBP10	LYPLAL1	PLEKHA4	VOPP1
		CA4	GPR82	TUSC3	GPAT3
		FGFRL1	MARCH1	HRC	PIM2
		ARMCX2	HMGB3	CCDC8	FST
		BCL6B	LY86	MAGED2	FGL2
		SYNPO	GCNT1	TMEM45B	IL18RAP
		GAD1	IL10	HIF3A	LRRK2
		TM4SF1	AGR2	ASAP3	KIT
		PRKAR1B	PKIB	LGALS3	CHI3L1
		LGALS3	CD1C	PLEKHB1	NFE2
		KLK13	HLA-DMB	CD109	LY86
		CH25H	HBD	SMARCD3	CLEC10A
		MPL	MSMO1	RSPO4	SULF2

RUNX1C-			
5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated
KREMEN1	IL1R1	BCL6B	HLA-DMB
KHSRP	LMO4	C1orf115	HPGDS
CPAMD8	DHRS9	UCHL1	PILRA
MFAP2	CASP1	CA13	GLIPR2
GBP2	JCHAIN	AEBP1	BIN2
CTTN	GLIPR1	KIAA0040	FCN1
CKB	DCANP1	SOCS2	SYTL3
KANK2	MS4A2	CH25H	CCR6
SHANK3	GATM	CA4	CD1C
COLQ	HLA-DPA1	FLT4	S100A9
UCHL1	HSPB11	CTTN	IL18R1
COL6A2	S100A8	TUFT1	PRKCD
DOK4	PTGER3	SLC2A10	KRT1
COL1A2	CYP1B1	KANK2	DHRS9
RSPO4	GPR65	MPL	SIGLEC6
MRPL12	LIF	SMIM10L2A	CD48
ZNF593	ZNF720	TFPI	HBG1
GRIK5	MYCL	MCOLN3	JAML
SLC25A23	S100A12	TSPAN9	NDRG2
ARAP3	CD207	TDRP	AGR2
DLK1	AIF1	CD79B	TNFRSF1B
DBN1	FST	DNAJB5	CD1B
FBLN1	FGL2	GSTM3	IRF8
TFPI2	IL10RA	PLCG1	HLA-DPA1
SALL2	CD300LB	LPIN3	GATM
CD79B	S100A4	GPR153	MAGEF1
DNAAF5	IL18RAP	GJA5	PARM1
PPRC1	FGL1	ULBP2	HMGB3
LGALS3BP	CLEC2D	CMYA5	TRAF1
ADGRL1	TMEM236	PTPRF	MYCL
MMRN2	CD48	QPCT	FHDC1
THRSP	KIT	CPT1C	CD244
SPINT1	SEC61G	RAB3B	SLC40A1
TMEM45B	NLRC4	KLK6	CLEC2D
BAIAP3	BLNK	UNC13B	HBD
HIP1R	MARCKS	MARCKSL1	CPA3
LAMB2	LAMP5	HAP1	PDGFB
MAGED2	CCNA1	LDOC1	CYTH4
VAT1	HBG1	CADM4	MPEG1
ZNHIT2	CSRP2	TSPAN7	PTAFR
MTMR11	MS4A14	OPTN	FCER1A
NPDC1	SNAP25	GNG12	EPB41L3
CNFN	HLA-DMA	FRMD6	SELPLG
FAM69B	SLA	SPTBN1	PLB1
PEX10	KCNMA1	SMARCA1	DAPP1
LOXL1	VOPP1	TCEA3	S100A8
PLCG1	SLF1	MFAP2	NLRP3
ASAP3	FPR1	RGS9	FGR
RRP12	P2RY14	AMOTL1	CXorf21
CPT1C	TMEM60	CCL5	STAB1
CD99	CATSPER1	SEMA6C	HDC
SFN	CSF2RA	EPHX2	IRF4
LURAP1	MEIKIN	FKBP10	RNASE6
DAPK1	S100B	ENAH	SLC1A4
TSPAN9	NDFIP2	NEO1	GPR68
NPTX1	FPGT	CASP7	TMEM236
CADM4	RWDD2A	BEX2	SLC43A3

RUNX1C-			
5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated
NES	LACC1	TRAK2	FGL1
ITGA5	CCR6	SV2A	HLA-DMA
P3H4	HCST	PXDN	RAB32
DUSP2	HLA-DQB1	SERPINB9	NCF1
CASS4	PARM1	FSD1	SH3BP2
TEAD2	IL13RA1	SYNPO	OGG1
TWIST1	ALOX5AP	TEAD2	SCN4B
UBALD1	TMEM71	SLC41A1	ADAMTS14
SRM	DAPP1	TCEAL9	CSF1R
SOX8	FCER1A	CCDC173	HTRA3
SERPINB9	CTSG	FAM166B	TSC22D3
GRWD1	APOC2	SULT1C4	CNRIP1
TCEA3	ZNF366	ITGAV	RWDD2A
MDK	ADCYAP1	GNG11	S100A12
RAB5C	PLBD1	C11orf95	CIITA
KREMEN2	IL18R1	RCAN3	PTPRO
CLPTM1	MYL4	HIP1R	SLCO3A1
LAPTM4B	GK	COL1A2	CSF2RA
SLC9A1	MILR1	STON1	PLCB2
B4GALT2	HLA-DRA	CETP	NLRC3
CBX2	SYTL3	IFIT2	LYZ
GPX4	LYZ	COX6B2	KCNE5
HMGA1	CPA3	RCAN2	CATSPER1
IRF2BP1	SFXN3	C16orf45	IL10
HRC	GAPT	F2R	SNAP25
KEL	FAM102B	ATXN7L2	MARCH1
REXO1	TNFSF13B	ANXA2	SIGLEC10
COL6A1	TLR6	DOK4	CACNA2D4
MCAM	TSPAN33	PFKM	PTGER3
BRSK1	FAM198B	MAGED4B	CD300LF
OLFM2	S100A9	IL1RAP	CYP1B1
SPTAN1	CD1A	CEACAM1	DUSP10
PLVAP	CHI3L1	BEX3	RNASE2
MARK4	KDM7A	LAPTM4B	OLFML2B
CACNA1H	CLDN12	PTK2	CLCN5
MYBBP1A	FCER1G	APBB1	JAG1
RELB	CCDC170	VAMP5	APCDD1
C16orf74	GALNT3	CYTL1	IGFBP7
LPIN3	CD300LF	SPINT1	HLA-DPB1
NCDN	ACTR6	TWIST1	FPR1
FSTL1	JAZF1	TRIB3	RPS6KA1
PHLDB1	OGFRL1	CDH24	FAM198B
LRG1	KL	KIAA0895	HP
TUSC3	MNDA	COL6A2	SLC25A45
ECM1	CLNK	NPTX1	IGFBP5
PPFIBP1	CD300E	CREB3L1	LAMP5
TMIGD2	IGSF6	BAIAP3	PARVG
TTLL12	LY96	LEFTY2	PLEKHG3
PPP1R14B	SIGLEC6	GAD1	LAPTM5
ALS2CL	ADAM28	ALS2CL	MS4A2
BCAM	CYP11A1	TMEM98	GALNT3
NAB2	ANXA5	GCH1	CCR2
RPS2	GLIPR2	GPR50	SORL1
CCDC86	BNIP3	HSPB8	RNF125
CRTC1	FAM170A	FBLIM1	CLNK
C1orf115	GLRX	DPPA4	CEACAM4
GRASP	GPR68	NES	ACPP

RUNX1C-			
5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated
DNAJB5	TLR1	PRKCZ	ST8SIA6
BCL7C	IL7R	RASSF8	CD300C
GALNT2	TRIM36	JUN	GCNT1
PMPCA	HIST1H4E	AHR	GPR82
GJA5	RGS5	KDELR3	TNFAIP8L2
FBLIM1	MYOM1	LAMB2	NLRC4
VAMP5	FUCA1	CALD1	IL9R
SMARCA4	ADCY7	VGF	MEFV
CHERP	SULF2	NPR2	ADCYAP1
PUSL1	DUSP10	TESK2	ZNF366
SLC25A10	TFEC	MBOAT2	CCNA1
PPP1R14A	TRNT1	ECM1	ITGB7
PLK3	KLHL2	LRRC70	OLIG1
HOMER3	CCR5	KCNA6	MYOM1
UNC13B	MAF	IKZF4	IL21R
PRMT1	INSIG2	ARHGEF12	ELF4
GRTP1	CLEC7A	SPARC	NKG7
SETD1A	FGR	TMEM74B	PIK3R5
ZNF598	PTPN22	LOXHD1	JCHAIN
POP7	ABCG2	CLIP4	CD300E
CLUH	CETN3	SFN	ATP2B4
MICAL3	SOX2	CNFN	TMEM144
INPP5J	CALB2	TAGLN3	AOAH
TSTD1	PLA2G7	PLEKHH1	PLAUR
RTKN	KLRF1	MRAP2	CCDC170
JUN	FGD2	LIFR	VDR
IL11RA	TGDS	PCDHGB6	GNGT2
RRAS	CCR2	GRTP1	PRDM1
NOTCH3	FCN1	RALA	PRF1
GDF3	CLECL1	CRYM	MS4A3
MRAP2	ARL4C	ZNF286A	TNFSF13B
NPW	GNGT2	ZNF608	GLIPR1
ARVCF	STAMBPL1	LRG1	CES1
RPTOR	CENPE	PHLDB1	CMTM5
ICAM2	PDK4	RBFOX2	RASSF4
WFDC1	C1orf162	GRIK5	HCK
ULBP2	MPEG1	BATF2	ADGRG5
SOX18	TLR10	SLC37A3	ARL4C
NOC4L	IRF4	FHL1	LILRB3
C8orf82	IFNGR1	DAPK1	CD207
PLPPR3	INPP5F	PROS1	NRG1
FSD1	CLEC10A	LOX	PROK2
NEO1	KAT2B	SOX8	IRF5
COL9A2	RPS18	ETS1	IL1RL1
SLC27A4	GPR171	ITGA5	PKIB
TUBB6	BANK1	TRO	RASSF5
IFRD2	ZBED5	RGPD5	C1orf162
CITED4	CDK1	CACNA1H	ALOX5AP
FGFR4	MAOA	BOK	SIRPB2
MYC	NRG1	TMC7	SEMA4A
STUB1	TPM1	TMEM136	PTPN22
DNLZ	SDC2	L1TD1	CST7
PLSCR3	TNFSF18	KLHDC8B	IL13
ATP6V0C	PTAFR	KEL	HLA-DRA
SLC9A3R2	PLAC8	SPTAN1	KCNMA1
TFAP4	B3GNT5	PLS3	CENPF
DGCR6	CARD16	FLT1	RAB37
PALM	TICAM2	BRSK1	KLF13

RUNX1C-			
5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated
GAMT	SLC37A2	AMIGO2	IGSF6
CCDC173	RNASE3	PTPRG	CTTNBP2
GET4	NMRK1	LIPG	HLA-DRB5
LRRC8A	GPNMB	AIF1L	HVCN1
BEX2	TLR8	ERG	C11orf21
CLPP	MS4A3	MLF1	HBG2
ARC	IL13	COL6A1	MILR1
MFGE8	RNASE2	IFIT3	FAM102B
RAVER1	SORL1	NRBP2	GPR65
TNNT1	MCEMP1	SHE	DTNA
PPDPF	HDC	CXCL10	OGFRL1
EEF2	OLIG1	PTPRU	TMEM173
CCDC85B	NPL	MAGED4	IL1RN
ATP6V0E2	IL4	SERPINB4	SLC27A2
HCFC1	JAG1	DTX3	HCST
ASPG	CASP5	GFRA3	MUC1
DHX37	DDIT4L	CD3D	CD1A
THOP1	PILRA	PRKAR1B	DOK2
NME4	MRC1	SCG5	INSIG1
FOXP4	FAM72B	IL11RA	P2RY6
KLC2	CENPK	PDE4DIP	CBFA2T3
TUBB3	ENPP2	LGALS3BP	MEIKIN
HELZ2	FAM219B	AKAP6	LACC1
TMEM120A	FHDC1	SHF	ITGAM
DDX54	OLFML2B	TM4SF18	P2RY14
ISYNA1	HMMR	APOD	SFXN3
IGFBP2	SLC18A2	CAPN12	BNIP3
RNF126	HRH1	B3GLCT	CLU
HYAL2	FAM72D	LRP4	ZFP36L2
TRIM28	NUF2	CALCRL	INPP5F
KLHDC8B	A2M	ID1	TREML2
ATAD3B	RASSF4	ICAM2	C9orf72
PEMT	GPR183	PLIN2	CST3
FARSA	FAM72A	GPR32	DNMT3B
CRYM	MS4A6A	F2RL1	TICAM2
ALKBH7	NME8	MXRA8	JAZF1
HSPB1	PPBP	FERMT2	GFRA2
LOXL2	ZBED8	WFDC1	MRC1
SPATA2L	SGO2	PCDHGB7	NFIL3
JAG2	UGT1A6	SHANK3	CYP11A1
KATNB1	NKG7	ATP6V0A4	ARHGAP6
MEGF6	ADAMTS3	INPP5J	GK
C19orf24	IL17RB	NACAD	PLAC8
GFRA3	AOAH	CDKN2B	SLC7A8
LEFTY2	RRM2B	TMIGD2	BANK1
PLIN2	MAP3K8	SLC35G2	GFI1B
DOT1L	TCN1	ASPG	FUT7
CETP	STEAP4	PDGFD	GSE1
TM7SF2	C5orf34	GDF3	TRIM36
GTF2IRD1	HLA-DRB1	ZNF521	STK17B
FAM171A2	P3H2	ADAMTS9	C5AR1
XYLT2	ZNF33A	SSBP2	CALB2
CD248	VCAN	CIB2	CD180
CEACAM1	EGLN3	NFAT5	MNDA
GUCD1	CR1	C1orf54	AIF1
SEMA6B	GYPE	GNB4	MS4A14
HSPB8	ZNF680	FABP3	PADI2
PKN1	TNFSF10	AKAP12	CENPW

RUNX1C-			
5 Dox		10 Dox	
Upregulated	Downregulated	Upregulated	Downregulated
NCLN	PRCP	FAM171B	KL
SNAPC2	PGM3	AKR1C3	FAM170A
PAK4	SLC4A8	PLVAP	ITGA4
CDC34	SLC16A10	DBN1	RAB31
POLRMT	ATP1B1	TSPAN6	VCAN
PNPLA2	AHSP	TSTD1	RNASE3
MICALL2	SLAMF8	BLVRA	GAS6
MTA1	GRAP2	SASH1	CCR5
HSPG2	ITGB7	TNNT1	MPO
SREBF1	ITGAM	RELB	MAF
DLG5	CD27	RRAS	CLC
RAD23A	CCNI2	AS3MT	FUCA1
APOE	HIST1H4A	COL9A2	CLEC7A
RPL8	RGS2	NOTCH3	SETBP1
CD81	IGFBP5	TXNIP	GPR34
GADD45GIP1	CRLF2	SLC22A17	GYPE
TAGLN3	KYNU	ENC1	IGLL1
TGFB1I1	FCGR3B	MECOM	AHSP
WIZ	CLC	SLC16A4	NDFIP2
ZNF581	C11orf21	ARC	TNFAIP2
LRWD1	HBG2	BCAM	FCER1G
GIPC1	ZFP62	PLSCR3	CD300A
ACAP3	PIGP	NAB2	CCND3
PRR12	MMP12	APOE	ANXA5
RAB40C	PLN	CKB	TCN1
MRPS2	CENPQ	XPR1	SDC2
FSCN1	CCL13	IFIT1	SLC37A2
MRPS34	MS4A7	AP1AR	TXNDC5
RHBDD2	CD80	RAMP2	POU2F2
SLC22A17	ZNF641	MMP1	BMP6
BIRC7	HP	HOMER3	NME8
SOX7	CLDN4	OLAH	SOX2
SARS2	ACP5	KCTD17	SIGLEC9
MED16	CMKLR1	AKT3	HMOX1
SURF6	CCL24	ARRDC3	RMND5B
ATAD3A	TGFBI	TRPV3	MPZL3
MATK	RGL1	IFI6	RHOB
ZNF276	MS4A4A	KCTD14	TSPOAP1
GPSM1	GYPB	GUCD1	PTPRJ
DUS3L	IL1R2	BIRC7	OSM
NOC2L	C1QA	PPP2R5B	SPINK2
MCRIP2	F13A1	NPDC1	PADI4
ID3	KRT79	C12orf57	CCR1
JUP	GOLGA8H	GZMM	F2RL3
NRBP2	SSR3	SOX7	SLC45A4
WDR18	MTUS1	CACFD1	BTLA
EBI3	CD274	SVOPL	CR1
SLC2A4RG	ITGB3	LRP1	HRH1
MRPL38	CST7	PALM	GAB3
MAP1S	HIST1H4H	MTPN	IL4
SNTA1	HBE1	NLGN2	TLR8
PUF60	SLCO5A1	IGSF8	IL17RB
PIAS4	PGPEP1	ISYNA1	SLC11A1
COMT	KCNE3	NR4A1	MGAT3

RUNX1C-		
5 Dox		10 Dox
Upregulated	Downregulated	Downregulated
ATG2A	CAPG	B3GNT5
CCDC106	BLOC1S2	PLEK
LMF1	CD163L1	RTN1
DVL1	FOLR2	PPBP
LAMA5	RNASE1	KCNH2
TSR3	NRP1	GLRX
EMC10	ATP6V1G2-DDX39B	HGF
FCHO1	SUGT1	JDP2
HPDL	LRRC8C	ALDH3B1
SHF		LRRC25
RPS19BP1		GPR183
NACC1		TYROBP
DNPH1		ADGRE2
SLC25A22		ENPP2
SYNGR1		PLP2
CAPN15		IL7R
NR4A1		GCSAML
MOSPD3		APBB3
FAM207A		MICAL2
BEX1		APOC2
FBXL19		RASD1
NDUFAF3		PLA2G7
AGRN		TNFSF10
SSBP4		TDRD9
EHD1		GNA15
PLXNA3		MCEMP1
SGTA		C6orf223
LYL1		RXRA
EIF4EBP1		GPR132
MAP4K2		PIK3R6
ARHGAP23		CYBB
CIB2		CCNI2
MLLT1		GATA2
LRP3		ACP5
STRN4		ITGB3
INPP5E		EGLN3
IGSF8		SASH3
HSF1		TNFSF18
LONP1		A2M
LRP1		LAT
ASMTL		ID2
PGLS		WT1
MRPL4		PSTPIP1
FUOM		LTBP1
AKAP12		TFEC
EPN1		PLEKHG5
PKD1		PAQR5
SLC25A6		LGALS12
H2AFX		ABI3
CERCAM		PTGDR2
CD320		C15orf39
FBXW5		TUBB1
PKMYT1		FHL2
TPRN		ANPEP

RUNX1C-						
5 Dox Upregulated				10 Dox Downregulated		
CAPN12	VASN	FAAP20	RBM38	CHDH	SLC4A8	ATP1B1
TMEM201	TNRC18	RBM42	MZT2B	HBE1	CD14	TNFRSF18
HAGHL	SAMD1	MBD3	RPL18A	MAP3K8	FAM178B	ADAM8
MFSD12	NELFA	ARRDC1	MLST8	KRT79	UGT1A6	CMKLR1
SLC52A2	EXOSC4	MGRN1	RNH1	LGMN	CPT1A	LTC4S
CACFD1	DDX49	GZMM	PITPNM1	FPR2	HK3	MZB1
MIIP	AP2A1	RPS15	LMF2	VENTX	ADAMTS3	SLCO2B1
ZNF358	SCRIB	PACS2	MEGF8	GRAP2	RAC2	PRAM1
AURKAIP1	CROCC	APBA3	RAB11B	GPR171	RETN	TNP1
PIGQ	CDT1	FLYWCH1	TMEM259	PTPRE	F13A1	IGFLR1
BSG	TCF3	ADRM1	C19orf25	CLECL1	SOCS1	MTUS1
FASN	MFSD10	THEM6	SPSB3	CNN2	SIGLEC1	C1QB
FAM234A	TMEM161A	HIST2H2AA4	BAD	TUBA4A	LFNG	COMMD7
R3HDM4	STK11	CPTP	CIC	CEBPA	CTSW	RUNX3
RPLP2	MKNK2	TRMT2A	FBLN2	DMTN	MMP12	GOLGA8H
STMN3	AHDC1	WDR24	ISG15	CD82	HBA2	LSP1
ATF5	GNB1L	GNB2	SLC19A1	NFATC2	S100A6	MS4A7
FAAP100	SART1	COX8A	GALK1	CD74	ITGA2B	ALOX15
FAM173A	ZDHHC8	HIST2H2AA3	ZGPAT	CLDN4	TNFRSF25	PSAT1
HMG20B	LRRC61	FKBP8	H1FX	SRGN	RELT	GPR35
C12orf57	ISOC2	SNCG	ZNF524	SLC24A3	TSPAN32	BPI
ZSWIM4	RPUSD1	BTBD2	OGFR	MICAL1	SORBS3	KCNE3
MAP2K2	NCKAP5L	LTBP4	RGS19	BHLHE40	MYO1G	CCL13
POLD1	RFX1	CEP131	JUND	GGT5	MMP9	APOBR
HSPBP1	PIP5K1C	RAI1	RPLP1	P3H2	SH2D3C	FES
LLGL1	TBL3	TRIP6	SCARF1	CCR7	SLC38A1	CD163L1
LIN7B	UBXN6	INTS1	MIB2	SERPINB10	ITGB2	HS6ST1
C9orf16	CSNK1G2	ZNF414	TIMM13	MS4A6A	SUSD3	NRP1
CTU2	PTOV1	MAP3K11	EGLN2	UNC93B1	OXER1	CD274
PAFAH1B3	PKD1P1	DGCR6L	NDUFS7	HPR	IL1R2	RGS2
EGFL7	E4F1	RHOT2	PTPRCAP	HDAC5	SH3BP1	EFHD2
NECAB3	TMEM134	A1BG	CFD	TLR10	SLC16A10	CAPG
NT5C	ARMC5	NUBP2	TSSC4	C1QC	HBA1	PRTN3
SPNS1	ILVBL	APRT	UQCC3	C1QA	PRG2	SLC7A5
RUVBL2	LRFN4	PFKL	TRPM4	DDIT4L	THBS1	KYNU
TSPO	CACTIN	PDLIM7	SBF1	SECTM1	ADAMTS10	ASNS
ARHGDIA	ALDH16A1	C1orf159	MIF	CRLF2	RGL1	SYTL1
DECR2	SLC1A5	MCRIP1	TSPAN4	TGFBI	CD80	TPSAB1
C7orf50	ZYX	DPP7	FBRS	SPNS3	PF4	PLAT
LY6E	FBXL14	RPS6KA4	NTHL1	GYPB	CCL24	HLA-DRB1
MRPL41	PIEZO1	BCAR1	TSC22D4			
SHC2	PCNX3	NR1H2	SRSF2			
BOP1	RFNG	GSTP1	MRPL28			
GTPBP6	ORAI1	MVB12A	BRAT1			
SF3A2	TELO2	JUNB				
VPS37B	AGPAT2	WDR13				
UBE2S	GPS1	CCDC124				
FLYWCH2	TMEM8A	CPSF1				
PIM3	RAC3	NME3				
VPS51	THAP7	SLC39A4				

Supplementary Table 2: Differentially expressed genes upon RUNX1-ETO induction (3, 5 or 10 ng/ml Dox) in RUNX1C+ and RUNX1C- cell populations

Upregulated (green) and downregulated (red) genes upon 24h treatment with 3, 5 and 10 ng/ml Doxvin RUNX1C+ (white top panels) and RUNX1C- (blue bottom panels) cell populations.

Cluster 1						
A2M	BAIAP3	CD70	DNMBP	GAP43	HSPB8	LGALS3
ABAT	BATF2	CD79B	DOCK6	GATS	ICAM1	LGMN
ABCA9	BBS1	CD83	DPYSL4	GBP2	ICAM2	LIMCH1
ABCB4	BCL6	CD99	DTX3L	GCH1	ICOS	LIMS4
ABCD4	BEX2	CDC42EP3	EBI3	GDF2	IFI16	LIPH
ABTB1	BEX3	CDC42EP5	EGR1	GDF3	IFI27	LMBR1L
ACADVL	BIRC7	CDH24	EGR2	GIMAP4	IFI27L2	LOX
ACSF2	BLVRA	CDKN1A	EIF1	GIMAP5	IFI35	LOXL2
ACVR1B	BRSK1	CDKN2B	ENC1	GIMAP6	IFI6	LPIN3
ADAMTS9	BSDC1	CDR2L	ENDOV	GIMAP8	IFIH1	LRG1
ADD3	BST2	CEACAM1	EPHA4	GJA5	IFIT1	LRP4
AHCYL2	BTN2A1	CECR2	EPHX2	GLDC	IFIT2	LRRN3
AHR	BTN2A2	CES4A	EPS8	GNG11	IFIT3	LTA4H
AIFM2	C10orf10	CETP	ERMN	GNG12	IFIT5	LURAP1
AKAP12	C11orf52	CFH	EVI2B	GNG2	IGF2R	LZTS3
AKR1C3	C11orf95	CH25H	EZH1	GNG7	IGLON5	MAGED2
AKT3	C19orf66	CLDN7	F2RL1	GPR137B	IKZF4	MALT1
ALS2CL	C1R	CLIP1	FABP3	GPR153	IL11	MAML2
AMIGO2	C1S	CLIP3	FAM114A1	GPRIN3	IL13RA1	MAP2K4
AMOTL1	C1orf115	CMYA5	FAM131A	GRAMD1C	IL1RAP	MAP3K12
ANG	C1orf61	COL1A2	FAM166B	GRASP	IL27RA	MAP4
ANXA2	CA13	COL27A1	FAM171B	GRTP1	IL2RG	MAPK11
AP1S1	CA2	COL6A1	FAM49A	GSTA4	IL6ST	MAPK13
APOD	CA9	COL6A2	FAM89A	GSTM2	IRF6	MFAP2
APOE	CAB39L	COLQ	FBLIM1	GSTM3	IRF8	MMP1
APOL2	CACHD1	COX6B2	FBLN5	GUCY1A3	IRF9	MOB3B
APOL3	CADM4	CPAMD8	FBXL2	H1F0	ISG20	MPZL1
APP	CALCOCO1	CPT1C	FBXO21	H6PD	ITGAV	MRAP2
AR	CALCRL	CRHBP	FCGR1A	HAP1	JAK1	MST1
ARC	CALD1	CSRNP2	FCGR1B	HDAC9	JAM3	MTMR11
ARHGEF15	CAMSAP2	CTC1	FCHSD1	HECW2	JUN	MX1
ARMCX2	CAPN12	CTNNA1	FERMT2	HFE	KCTD17	MXRA8
ARMCX6	CASP4	CTSC	FGD5	HID1	KCTD21	MYO1E
ARRDC3	CASP7	CTTN	FGD6	HIF1A	KIAA0040	MYO9A
ARRDC4	CASP9	CTTNBP2NL	FHL1	HIF3A	KIAA1671	MYOF
ARSD	CAST	CXCL10	FLT1	HIP1R	KIF13B	NAAA
ASAP3	CBLB	CXCL8	FMNL2	HIST1H1C	KLC4	NATD1
ASPG	CBLN3	CYP1A1	FMNL3	HIST1H2AC	KLF6	NCOA3
ASPHD1	CC2D2A	CYSTM1	FNDC3A	HIST1H2BD	KLHDC8B	NDRG1
ATP6V0A4	CCDC173	DAB2	FSTL1	HLA-B	KLHL3	NECAP2
ATP6V1B2	CCDC71L	DDX58	FSTL3	HLA-DRB5	KLK13	NFAT5
ATP9A	CCDC85A	DDX60	FZD7	HLA-E	KREMEN1	NFKB2
AXL	CCL28	DHRS1	GABARAP	HOXB2	LAMC1	NFKBIZ
B3GLCT	CD109	DNAJB5	GABARAPL2	HOXB7	LBH	NINL
B3GNT9	CD53	DNALI1	GALT	HRC	LEFTY2	NISCH

Cluster 1 (continued)					
NLRP1	PIK3IP1	RCAN3	SGK1	SSFA2	TMEM98
NPTX1	PINK1	RETSAT	SGPL1	ST5	TNFAIP8L3
NQO1	PLA2G4C	RGL1	SH3D21	ST6GAL1	TNFRSF9
NRBP2	PLAU	RGN	SHE	STAG3	TOB2
NRN1	PLCG1	RGS12	SHF	STAT1	TP53INP2
NTAN1	PLD2	RGS9	SHROOM4	STIM2	TRAK2
OAS1	PLEKHA4	RNASE1	SIPA1L1	STX12	TRIM16
OAS3	PLEKHB1	RNASE4	SLC22A4	SULF1	TRIM22
OASL	PLIN2	RPGRIP1	SLC25A42	SUN1	TSPAN7
OLAH	PLIN5	RSPH9	SLC26A11	SYNPO	TULP3
OPTN	PLK2	RSPO4	SLC2A3	TAGLN	TUSC3
P2RX4	PNRC1	RUNX1T1	SLC35B4	TAGLN3	TWIST1
P2RX7	PPFIBP1	S100Z	SLC35D2	TAX1BP3	TXNIP
PAPSS2	PPP2R5B	SALL2	SLC35F3	TBC1D8	UBE2H
PARP14	PRAME	SAMD9L	SLC37A3	TCEA3	UBE2L6
PARP9	PRKCZ	SAMSN1	SLFN12L	TCF7L2	UCHL1
PATJ	PROS1	SASH1	SMARCA1	TESK2	ULBP2
PCDHGB6	PTPRS	SAT2	SMARCD3	TFPI2	UNC13B
PCDHGB7	QPCT	SBF2	SMIM10L2A	TGFBR2	USP27X
PCED1B	RAB13	SCARB2	SNX18	THRSP	VAMP5
PCMTD2	RAB3B	SCG5	SNX33	TJP2	VIM
PCNX2	RAB40B	SEMA3A	SOCS2	TLR7	VNN2
PDE1B	RAI14	SEMA6C	SOX8	TM4SF1	WDR19
PDE4B	RALA	SERINC5	SPACA6	TMEM120A	WFDC1
PDE4DIP	RAMP2	SERPINB3	SPARC	TMEM178A	XAF1
PDGFD	RASSF8	SERPINB4	SPATS2L	TMEM2	YPEL2
PDGFRA	RBFOX2	SERPINB8	SPINT1	TMEM234	ZG16B
PHC1	RBM24	SERPINB9	SPOCD1	TMEM243	ZNF251
PHF1	RBPMS	SERPINH1	SPTAN1	TMEM45B	ZNF608
PHLDA1	RCAN2	SETD7	SSBP2	TMEM74B	ZNF827

Cluster 2						
AARS	C5AR1	CLCN4	FADS1	GYPE	LCN2	MYLK
ABCE1	C6orf223	CLSPN	FAM102B	H2AFZ	LCP1	MYO19
ACAT2	CACNA2D3	CMSS1	FAM105A	HAUS8	LEF1	MZB1
ADAMTS1	CACNA2D4	CNRIP1	FAM178B	HBB	LFNG	NANP
ADAMTS14	CAMK1	CPA3	FAM212B	HBD	LGALS12	NCAPD3
ADCY7	CAPN11	CPT1A	FAM216A	HBE1	LGALSL	NCAPG2
ADGRE2	CATSPER1	CR1	FAM46C	HBG1	LIF	NCAPH
ADGRG3	CBX5	CRYBA4	FAM57A	HBG2	LMNB1	NCR3LG1
ADGRG5	CCDC169	CSE1L	FAM83D	HCST	LMO4	NDC1
ADORA2B	CCDC78	CTSG	FANCA	HDC	LRR1	NDUFV3
AGAP2	CCNA1	CXCL2	FANCD2	HEMGN	LRRC8D	NETO2
AHSP	CCNA2	CXCL3	FANCG	HGD	LST1	NFE2
AIF1	CCNB1	CYP11A1	FANCI	HIRIP3	LTC4S	NINJ2
ALB	CCNB2	DAPP1	FARSB	HJURP	LY6G6F	NKG7
ALDH1B1	CCND1	DCLRE1B	FBL	HMGB3	MAGEF1	NLRC3
ANGPT1	CCNE2	DDIAS	FCER1A	HMGN5	MAGIX	NME1
ANKLE1	CCNI2	DDIT4	FCER1G	HSH2D	MANSC1	NME2
ANKRD18B	CD2	DENND2D	FEN1	HSPA4L	MBP	NOD2
ANLN	CD244	DERL3	FGR	HSPD1	MCM10	NOP16
APOBEC3B	CD300A	DHFR	FKBP4	HSPE1	MCM3	NPM3
APOBR	CD300C	DHRS11	FST	HSPH1	MCM4	NR2C2AP
AQP3	CD84	DHRS9	FUCA2	IARS	MCM6	NRARP
ARHGEF39	CDC25A	DIS3L	GAB3	ID2	MEIOB	NT5DC3
ASIC4	CDC45	DNAJC6	GALNT14	IGFBP7	METTL1	NUDT1
ASNS	CDC6	DNMT3B	GALNT6	IL13	METTL21A	ODC1
ATAD2	CDCA2	DSCC1	GATM	IL9R	MEX3B	OIP5
ATP2B4	CDCA7	DTL	GCSAML	INSIG1	MFSD3	OPN3
ATP5G1	CDCA8	DUT	GGH	ISOC1	MGST1	ORAI3
AURKA	CDKN2C	DZIP1	GINS1	ITGA4	MIPEP	ORC1
AURKB	CDKN2D	E2F2	GINS2	ITGA6	MKI67	ORC6
BAG2	CDKN3	E2F8	GINS3	ITGAL	MNS1	OSBPL6
BIRC5	CENPF	EFHC2	GINS4	ITGAM	MPP7	OSM
BLM	CENPH	ELOVL6	GJB5	ITGB3	MRM1	OXCT1
BNIP3	CENPN	EME1	GLA	JADE1	MRPL3	PA2G4
BOLA3	CENPP	ENO3	GMNN	KCNK17	MRTO4	PADI4
BRCA1	CENPU	ENOSF1	GNG8	KCNK5	MS4A3	PARVG
BRIP1	CENPV	ERCC6L	GNL3	KCNQ5	MTFR2	PASK
BRIX1	CENPW	ERVMER34-1	GPAT3	KIF15	MTHFD1	PCNA
BUB1	CEP128	ESPL1	GPR132	KIF18B	MTHFD1L	PCYT1B
BUB1B	CHAC2	EXO1	GPT2	KIF2C	MTURN	PDE5A
C11orf21	CHAF1B	EXOSC5	GRAP2	KIF4A	MUC1	PDIA5
C1orf162	CHEK1	F13A1	GSE1	KIT	MYB	PDSS1
C21orf58	CHEK2	F2RL3	GSR	KNTC1	MYCN	PHF10
C2orf88	CHKA	FAAP24	GTF3A	KRT79	MYEOV	PHGDH
C4orf46	CLC	FABP5	GYPB	LAT2	MYL4	PHKA1

Cluster 2 (continued)						
PIF1	PRKCQ	RCCD1	SIGLEC8	SPAG5	THBD	UBASH3B
PIK3CB	PRKDC	RCL1	SIGLEC9	SPC24	TICRR	UBE2C
PIK3R6	PRPS1	RFC3	SKA1	SPC25	TIMELESS	UBE2T
PIM1	PRPS2	RFC4	SKA3	SPDYC	TIMM21	VOPP1
PLA2G3	PSAT1	RFC5	SLA	SPIN4	TIPIN	VRK1
PLAC8	PSMC3IP	RMI2	SLC11A1	SPINK2	TLCD1	WDHD1
PLCB2	PSMG1	RMND5B	SLC1A4	SPN	TMEM163	WDR12
PLCH1	PTAFR	RNASEH2B	SLC25A15	SPR	TMEM246	WDR4
PLEK	PTGIR	RNF125	SLC27A2	SPX	TMEM97	WDR76
PLEKHA2	PTPN18	RRM1	SLC27A5	SRPRB	TMTC4	WNT5B
PNP	PTPN22	RRM2	SLC35D3	SS18L2	TNFAIP8L2	WT1
POC1A	PTPRE	RRS1	SLC38A5	SSRP1	TOP2B	XK
POC1B-GALNT4	PTTG1	RUVBL1	SLC40A1	ST3GAL4	TOR3A	XRCC2
POLA2	PTX3	RXRA	SLC43A3	ST8SIA6	TPM1	YOD1
POLE2	PUM3	S100A12	SLC48A1	STAR	TPX2	ZFP36L2
POLQ	PUS7	S100A2	SLC4A8	STEAP3	TRAIP	ZNF239
POLR3K	RAB24	S100A4	SLC50A1	SULF2	TREML1	ZNF367
POP5	RAB27B	S100A6	SLC6A9	SUSD1	TREML2	ZNF485
PPBP	RAB37	SCFD2	SLC8A3	SYCE2	TRIM58	ZNF69
PPIF	RABEPK	SELP	SMC2	SYTL3	TRIP13	ZNF730
PPIH	RAD51	SELPLG	SMIM10	TAGLN2	TTC27	ZNF804A
PPM1M	RAD54L	SEPT1	SNRNP25	TCN1	TTF2	ZNF93
PRF1	RAP1GAP2	SERPINE2	SNRPF	TDRD9	TUBA4A	ZWINT
PRG2	RAVER2	SFXN2	SNTB1	TEC	TYMS	
PRICKLE1	RCC1	SHMT1	SNX20	TFRC	UBASH3A	

Cluster 3						
ABCA7	CDC20	FLOT2	JDP2	NHP2	RASGRP2	SPNS3
ABCB6	CDCA5	FOXM1	KATNB1	NINJ1	RCOR2	SRM
ABCC3	CDT1	GAL	KCNH2	NRGN	RECQL4	STK10
ACOT11	CEBPE	GATA1	KCNQ4	NRROS	RELL2	SUSD3
ACOT7	CENPM	GATA2	KCTD15	NT5M	RELT	SYTL1
ADA	CHAF1A	GBGT1	KIF21B	NUP210	RNASEH2A	TALDO1
ADAM15	CMIP	GCAT	KIF22	OAF	RNF24	TESC
ADGRG1	CMTM5	GFI1	KLF1	OBSL1	RPS6KA1	TFAP4
AGTRAP	CPNE7	GFI1B	LAGE3	OGG1	RRP9	TFR2
ALAS2	CPXM1	GJA4	LAPTM5	PCK2	RUVBL2	TGFB1
ALYREF	CSF1R	GMPR	LAT	PDGFRB	SAPCD2	THBS1
ANGPTL6	CTSW	GNA15	LSM4	PDLIM1	SCGN	TIMP3
ANK1	CXXC5	GP1BA	LSP1	PF4	SDF2L1	TK1
ANPEP	CYC1	GP9	LSS	PFAS	SFRP5	TMC8
ARHGAP22	CYTH4	GPR35	LTBP1	PGAM5	SH2B3	TNFAIP2
ARRB2	DCTPP1	GRK6	MAP1A	PHF19	SH2D2A	TNFRSF1B
ASF1B	DEAF1	GTSE1	MARS2	PIGY	SH2D3C	TOMM40
ATP2A3	DENND4B	GYPC	MARVELD1	PIM2	SH3BGRL3	TPSB2
BCOR	DMTN	HAUS7	MCM2	PKM	SH3BP1	TSPAN32
BIN1	DOK2	HBA1	MCM5	PKMYT1	SH3BP2	TSPAN33
BIN2	DTYMK	HBZ	MCM7	PLEKHG3	SIGLEC10	TUBB1
BRI3BP	E2F1	HERC2	MFSD2B	PLK1	SLA2	TUBG1
BYSL	ECE2	HES6	MGAT3	POLD2	SLC16A3	TYROBP
C15orf39	ECI1	HMBS	MICAL2	PQLC2	SLC17A9	UHRF1
C19orf48	EHD3	HPDL	MLC1	PRAM1	SLC19A1	ULK3
C1QBP	EIF4EBP1	HS6ST1	MPO	PRELID1	SLC1A5	VARS
C1orf116	ELF4	HYAL3	MRPS26	PRTN3	SLC25A10	VAV1
CARD9	EMID1	ICAM4	MSRB1	PSD4	SLC25A45	WDR62
CBFA2T3	EMILIN1	ICOSLG	MTSS1L	PTGES2	SLC2A6	WNT11
CBS	FADS2	IFRD2	MYBL2	PTPN6	SLC39A3	YDJC
CBSL	FAM53B	IL21R	MYC	PYCR1	SLC43A1	ZBTB16
CCDC86	FASN	IMP4	MYH10	QTRT1	SLC6A8	
CCND3	FBXO41	IMPA2	NAGLU	RAB11FIP5	SLC7A1	
CCNF	FERMT3	INF2	NBEAL2	RAC2	SLC7A5	
CD320	FHL2	ISOC2	NCAPD2	RAC3	SLC9A3R1	
CD34	FKBP1B	ITGA2B	NCF4	RANBP1	SOCS1	
CD82	FLNA	ITGB2	NFKBIA	RASA3	SPATC1L	

Cluster 4						
ACTG1	CORO1B	GIPC1	LAMB2	NBL1	PPDPF	TMEM115
AGPAT2	CPTP	GNB2	LRP10	NCK2	PPP1R14A	TMEM134
AHRR	CYP2S1	GPR32	LRRC8A	NCKAP5L	PRR12	TMEM8A
ALKBH7	DBN1	GPX1	LTBR	NECTIN2	RAB35	TNRC18
AP1B1	DCHS1	GRINA	MAP1S	NOS3	RAB5C	TRIOBP
AP2A1	DENND1C	GSN	MAP2K2	NOTCH1	RAVER1	TSR3
APBA3	DLK1	GUK1	MAP3K11	NPDC1	REXO1	TUBB3
ARHGDIA	EEF2	HIST2H2AA3	MAP4K2	NR1H2	RNASEK	TUBB4A
ASMTL	EHD2	HLA-C	MAPK8IP3	NUCB1	RPS2	UBALD1
ATG2A	EVC	HLX	MARCKSL1	OGFR	SERPING1	UBTD1
ATN1	FAM171A2	HMG20B	MARK4	PAFAH1B3	SHANK3	UBXN6
ATP6V0C	FAM69B	HSF1	MDK	PALM	SLC25A23	ULK1
BCL7C	FAM89B	HSPB1	MED16	PGLS	SLC9A1	VASN
BSG	FCGRT	HSPG2	MEGF8	PIM3	SPATA2L	YPEL3
C8orf58	FKBP8	HYAL2	MGRN1	PIP5K1C	SPSB3	ZNF219
CASS4	FLT4	ID3	MSRA	PKN1	STMN3	ZNF358
CD81	FOSL2	IRF2BP1	MVB12A	PLOD1	STRN4	ZNF580
CDC34	FOXP4	JUP	MXD4	PLXNB2	STUB1	ZNF593
CIC	GAA	KDM6B	MYO1C	PML	SYS1-DBNDD2	ZYX
CLPP	GADD45GIP1	KREMEN2	MYO7A	PNPLA2	TGFB1I1	

Cluster 5						
ACYP1	CENPK	EEF1E1	HIST1H2BO	MCM8	PLK4	TIMM17A
ALOX5AP	CENPQ	EMB	HIST1H4E	METTL5	PMS1	TNFSF10
AP1S2	CEP152	EXOSC8	HMGB1	MND1	PPM1K	TOP2A
APOC2	CKAP2	FAM72A	HMGCS1	MPHOSPH6	PRIM1	TPRKB
ASPM	CKAP2L	FAM72B	HMMR	MRPS18C	RAD51AP1	TTK
ATAD5	CKLF	FAM72C	HPF1	MS4A6A	RFESD	UGGT2
ATP10D	CKS2	FAM72D	HPGDS	MTBP	RFXAP	UGT1A6
B3GNT5	CLEC11A	FANCB	ITGB3BP	NAE1	RGS5	UQCRHL
BORA	COQ3	FBXO4	JAZF1	NCAPG	RMI1	XRCC4
C3orf14	CRLS1	FBXO43	JCHAIN	NDFIP2	RPA3	ZBED5
C3orf58	CXorf21	FRA10AC1	KIF14	NFYB	RPL22L1	ZDHHC13
CARHSP1	CYCS	GAPT	KIF18A	NUF2	SASS6	ZNF107
CCDC18	DBF4	GAR1	KIF20B	NUP35	SFR1	ZNF138
CCDC58	DCLRE1A	GDPD1	L2HGDH	NUSAP1	SGO2	ZNF273
CD38	DEK	GEN1	LPXN	OMA1	SKA2	ZNF563
CD52	DEPDC1	GPR65	LRRCC1	PARPBP	SLCO4C1	ZNF675
CDC7	DEPDC1B	GYPA	LSM3	PBK	SLF1	ZNF85
CDK1	DNAH14	HAT1	LSM5	PFDN6	SMC4	ZNF90
CENPE	DSN1	HELLS	MAD2L1	PLA2G4A	SNRPE	
CENPI	ECT2	HGF	MAGOHB	PLA2G4B	TCTEX1D1	

Cluster 6						
ACAD11	CNPY4	GJA1	LPP	PHTF2	ST8SIA4	TSLP
ADAM9	CREBL2	GSDMB	LRRC70	PLEKHH2	STAT4	TSPAN6
ANKRD12	CREBRF	HBP1	LY96	RGPD5	SULT1C4	TVP23C
ANKRD37	CREM	HCFC1R1	LYSMD1	RGPD6	SUMO1	TWSG1
AP1AR	CRYBG3	HOOK1	MAP4K3	RGPD8	SWT1	UGCG
ARMCX3	CXCL11	IFT81	MCL1	RGS17	SYP	WSB1
B2M	DDX17	IL15	MCOLN3	SAT1	TBC1D15	XPR1
C1orf54	DDX5	ING4	MLF1	SERINC1	TCEAL9	YPEL5
C22orf31	DDX60L	INHBA	MYO6	SERINC3	TCP11L2	ZDBF2
C5orf15	DYNLT3	INTS6L	NABP1	SLC16A4	TDRP	ZFYVE16
CBX7	ECHDC2	KIAA0895	NPHP3	SLC17A5	TFPI	
CCNG1	EIF4A2	KLF8	PCMTD1	SLC35G2	TMC7	
CLDND1	ENPP4	KLHL6	PCYOX1	SLC5A12	TMEFF1	
CLEC2B	FRMD6	L1TD1	PDCD4	SNX16	TMEM65	
CLIP4	FSBP	LIFR	PDK4	SOAT1	TMX3	
CLK1	GIMAP2	LPAR6	PGAP1	SP110	TRIM34	

Cluster 7						
AEBP1	CERS4	GRIK5	LIMS3	NR4A1	PXDN	SNAPC2
APBB1	CLSTN1	GRN	LLGL1	NUMA1	RASIP1	SNX15
ARVCF	CREB3L1	HELZ2	LMF1	PHLDB1	RELB	TBC1D16
ATF5	CSTB	IGSF8	LYPLA2	PITPNM1	RFNG	TRPM4
BAD	CUL7	IRF7	MAP1LC3A	PLCD1	RHBDD2	TUBB6
BAIAP2	DLGAP4	ISG15	MAPK3	PLCD3	SEMA4B	UNC119
BCAM	DMPK	ITPKC	MED25	PLD3	SERTAD1	VGF
CACFD1	EMC10	KDELR1	MFGE8	PLEKHG2	SFN	VPS9D1
CACNA1H	FADS3	KLC2	MVP	PLSCR3	SHC2	WDR13
CCDC106	FLNB	KRT18	NACAD	PLTP	SHISA5	WDR45
CD37	GFRA3	LDOC1	NOTCH3	PRKAR1B	SLC27A1	ZFP36
CDC42BPB	GPX4	LGALS3BP	NPR2	PTPRU	SLC44A2	ZSWIM4

Cluster 8						
ADAMTS3	CD55	DUSP10	HIST1H2AJ	ITGB1BP2	PDP2	VPS53
ADCYAP1	CDC25C	EAF2	HIST1H2BH	ITGB7	PIGW	ZGRF1
ANGPT2	CENPA	ENPP3	HIST1H2BI	KIF11	PRIM2	ZNF683
AP5M1	CEP55	ESCO2	HIST1H2BM	KIF23	PRR11	ZNF695
ARHGAP11A	CLECL1	GPR174	HIST1H3F	KRT1	RACGAP1	
C5orf34	COMMD7	GPR85	HIST1H4B	KRT72	RFX8	
CALB2	CYP1B1	GZMB	HIST1H4C	MELK	RHEBL1	
CAPG	DIAPH3	HIST1H1B	HPGD	MPC1	SGO1	
CCDC136	DLGAP5	HIST1H1D	IGFBP5	MS4A2	SLC31A2	
CCDC15	DMC1	HIST1H1E	IL18RAP	NDC80	SLCO5A1	
CD200R1L	DNA2	HIST1H2AH	IL4	NEIL3	SPP1	

Cluster 9						
ACTA2	CCL3L3	FAM222B	LUZP6	PTPRF	SPINK4	TNNT1
ACVRL1	COL9A2	GAD1	MDFIC	RHOC	SPTBN1	TSTD1
AES	CRYM	GADD45G	MEGF6	RNF213	SYNGR1	TUBB2B
AP2A2	CTSB	HIST1H2BK	MIIP	RRAS	TEAD2	TYMP
BCL6B	DAPK1	HOMER3	MINK1	SEMA6B	TEK	UCKL1
BMF	DPP4	ITGA5	MMRN2	SGSH	TIMP2	VAT1
C19orf38	ECHDC3	ITGA9	NUDT16L1	SIDT2	TMEM176A	ZNF581
C2orf66	ECM1	KLHL23	PLVAP	SLC5A10	TMEM176B	
C3orf18	EPHX1	LIN7B	PLXNA3	SLC7A10	TMIGD2	
CCL3L1	ERG	LRP1	PROCR	SOX7	TNFSF12	

Cluster 10	
ACTR6	MORN2
AK6	MSMO1
C12orf60	NDUFA5
C8orf59	OTULIN
CEP70	PFDN4
CETN3	RBPJ
CMC1	RIDA
CNIH1	RPS18
COMMD8	SEC61G
DEPDC4	SMIM11B
ERH	SNRPG
FAM200B	SUGT1
FANCL	TGDS
G2E3	TMEM156
GLMN	TMSB15A
GOLT1B	TTC32
GPR34	UQCRB
HIST1H4A	YEATS4
HSPB11	ZBED8
IDI1	ZNF302
IMMP1L	ZNF680
INSIG2	ZNF92
MED31	ZNHIT6

Cluster 11	
ARMC9	KIAA1191
CAP2	KIF12
CDRT4	KLK6
CYTH3	MERTK
DSEL	NCAM2
DTX3	NEO1
EML2	OAS2
EPB41L1	PRUNE2
FNDC4	RBMS2
FOS	SLC41A1
GPR50	SOD2
GPX3	STAP2
HAPLN3	TJP1
HIST1H3E	TPST1
IFI44L	TUFT1
INPP5E	

Cluster 12
BCAT1
CCR1
CHI3L1
CST7
CYBB
DEFA4
EGLN3
FCN1
HK3
HP
IGLL1
IL1B
IL1RN
LYZ
MCEMP1
P2RY2
PLA2G2A
PROK2
RETN
RNASE2
RNASE3
S100A8
S100A9
S100B
TSPOAP1

Supplementary Table 3: List of genes included within each cluster of differential response to RUNX1-ETO induction levels

Clusters (1-12) including genes showing differential gene expression response to the level of RUNX1-ETO induction. Clusters have been generated upon co-variance analysis of gene expression RNA-Seq data by Z-score from CD45+ CD34+ RUNX1C+ sorted progenitor cells upon RUNX1-ETO induction with 3, 5 or 10 ng/ml Dox for 24h. Green: upregulated gene expression trend. Red: downregulated gene expression trend. Genes are organized in columns by alphabetical order.

RUNX1-ETO targets

UpRegulated

GeneID	logFC	GeneID	logFC	GeneID	logFC	GeneID	logFC
CYP1A1	5.951725858	GUK1	1.420799672	ATG2A	1.226079791	GADD45G	1.098848135
RUNX1T1	5.555791735	AEBP1	1.411054756	FAM171A2	1.222317039	TIMP2	1.096179037
MX1	2.862784233	NATD1	1.406899611	TNFAIP8L3	1.21733847	GSN	1.095454856
CDKN1A	2.556758886	SMARCD3	1.404390304	LRRC8A	1.215708042	C8orf58	1.093003166
BAIAP3	2.499115224	SLC5A10	1.403879003	P2RX4	1.215094413	NUDT16L1	1.09026539
CTTN	2.124154609	PKN1	1.395888507	NDRG1	1.212642804	MDK	1.082515095
HELZ2	2.066890578	TMEM120A	1.390949366	NCK2	1.209427906	TBC1D16	1.081061611
PPDPF	2.06608069	DBN1	1.384426949	EGR2	1.209167178	KDELR1	1.079943534
NPDC1	2.033081835	TESK2	1.379969454	LRP10	1.205598519	HLX	1.078953137
SPTAN1	2.028619022	SIDT2	1.377603987	DAPK1	1.191618887	PXDN	1.078408438
SGSH	2.006978866	VAT1	1.36244068	CDC34	1.185804324	CDC42BPB	1.076848268
SPTBN1	1.996253582	NFKB2	1.35314645	PIM3	1.185001926	DLGAP4	1.075584878
COL9A2	1.976270399	MAP3K11	1.352158567	LIN7B	1.184471016	GRINA	1.07504328
NR4A1	1.948063562	TUBB6	1.350501656	TNRC18	1.1755315	FOXP4	1.07300497
TUBB3	1.918033424	BCL6	1.348895606	GNB2	1.175189835	EPS8	1.0632467
KIAA0040	1.888957867	PLEKHB1	1.348872046	KLC2	1.173985349	MYO1C	1.061902813
FCGR1A	1.880275656	FCGR1B	1.34844142	PIP5K1C	1.170291559	MAP2K2	1.060263217
COL6A1	1.860041763	BCL6B	1.344361494	MARCKSL1	1.16965579	ACVR1B	1.059409644
MFAP2	1.842681216	SERTAD1	1.341173805	BMF	1.169089753	MXRA8	1.055167646
KLHL23	1.810665553	CCDC85A	1.338580397	GNG11	1.168327693	SEMA6C	1.054441523
SMARCA1	1.790814439	SYS1-DBNDD2	1.33603067	CD83	1.167416739	RAB35	1.053758741
CES4A	1.759872787	ARHGDIA	1.327108263	PLD3	1.160198859	MEGF8	1.047532708
SOX7	1.74552543	CTSC	1.326661616	KDM6B	1.157810629	AHRR	1.045376716
HSPB1	1.726998825	SGK1	1.323704768	SLC9A1	1.156839705	TRIOBP	1.043464894
IGSF8	1.72411602	ID3	1.323525532	NUMA1	1.15645644	SERPING1	1.043437519
JUN	1.710342413	CPTP	1.322930985	TGFBR2	1.151650417	MINK1	1.042850625
RASSF8	1.706647838	MAPK3	1.315358309	LMF1	1.147647299	RNF213	1.039184001
YPEL3	1.663849163	CIC	1.312206755	ITGA9	1.146332227	GNG12	1.038298534
JUP	1.61458624	FLT1	1.311089202	FAM89B	1.138398242	PLOD1	1.034809922
FAM69B	1.605126253	ULK1	1.307052989	RGL1	1.135843453	NEO1	1.028582478
CREB3L1	1.595539023	NECTIN2	1.306088981	PHLDB1	1.135251071	EPHX1	1.027771219
ICAM1	1.592096977	FAM222B	1.297930703	C10orf10	1.134220773	PGLS	1.025715899
PITPNM1	1.587745114	ATP6V0C	1.29755722	FZD7	1.13343259	NECAP2	1.02126552
ALKBH7	1.55913892	RCAN2	1.296609508	FLNB	1.130867511	RHOC	1.018206442
SHF	1.553913453	MVB12A	1.295066827	CTSB	1.129554663	ITGAV	1.017744417
MAP1LC3A	1.536893704	CACFD1	1.293164597	MXD4	1.128865219	NCKAP5L	1.017619719
GPX1	1.498244857	RASIP1	1.291106767	PLCG1	1.127526622	HMG20B	1.016987919
SHISA5	1.486476949	ZFP36	1.286247275	IRF2BP1	1.126392376	LTBR	1.016093726
AP2A2	1.479498181	NLRP1	1.281815333	MAPK8IP3	1.122611127	STRN4	1.014822805
AGPAT2	1.477118334	NRN1	1.28168554	UCKL1	1.121534623	ACTG1	1.012181475
SFN	1.475818208	MAP4K2	1.281059349	CSTB	1.121031004	EVC	1.011003321
GLDC	1.475400943	ZNF580	1.281008265	SLC25A23	1.120800351	PLTP	1.010276511
TMEM8A	1.474912283	SLC7A10	1.279593232	FAM131A	1.120645153	GIPC1	1.008012407
FMNL2	1.464932295	P2RX7	1.27667298	PHLDA1	1.119977061	MED25	1.007248945
CORO1B	1.463913139	RGS12	1.261735389	SEMA4B	1.119615119	C3orf18	1.005922391
TMEM134	1.458610969	PDGFRA	1.255738835	SLC27A1	1.118379336	REXO1	1.003098981
HSPG2	1.451910386	CLSTN1	1.25292865	AP1B1	1.118202042	SOD2	1.000567005
TMEM45B	1.436214193	ANXA2	1.248501598	FOSL2	1.11638032		
GPX4	1.429308857	ASAP3	1.24411836	NUCB1	1.115952939		
EGR1	1.428364787	PLCD1	1.239764508	STUB1	1.112144808		
NPTX1	1.424410212	KCTD17	1.238224808	ACADVL	1.110876074		
UBALD1	1.424212899	BCL7C	1.230958928	MFGE8	1.108939138		
CAPN12	1.423624419	MYOF	1.230685704	MAP4	1.107076584		
ICAM2	1.422990568	BSG	1.228334417	MGRN1	1.099711127		

RUNX1-ETO targets

DownRegulated

GeneID	logFC	GeneID	logFC	GeneID	logFC
FST	-2.086354497	AQP3	-1.231362584	CEP55	-1.06885002
DNAH14	-2.073301035	CYCS	-1.231357879	ATP10D	-1.067477982
IDI1	-1.852603127	G2E3	-1.220094206	PPM1K	-1.066138258
SULF2	-1.830877188	KIF14	-1.218765664	PPIH	-1.065452688
PLAC8	-1.767964549	SGO1	-1.216071616	OTULIN	-1.065237656
OSM	-1.763192097	HELLS	-1.206673128	CENPU	-1.064427901
CYP1B1	-1.755781734	BUB1	-1.203334673	ZDHHC13	-1.063645198
SUGT1	-1.634667892	GPAT3	-1.199823608	OIP5	-1.063208073
LIF	-1.618916388	CEP70	-1.199147789	SLC31A2	-1.062288648
MSMO1	-1.610800559	GGH	-1.193535731	ALOX5AP	-1.056657029
TNFSF10	-1.592391827	CXCL2	-1.188939187	NAE1	-1.054326538
AP1S2	-1.529774115	ITGAM	-1.187308987	EXO1	-1.054237593
ITGA4	-1.517856284	HSPB11	-1.186825081	FANCI	-1.053913702
HMGB3	-1.513139138	PFDN4	-1.186515961	METTL5	-1.051203035
PARPBP	-1.483249951	WT1	-1.176630622	B3GNT5	-1.04415946
SNRPG	-1.466455855	LSM5	-1.17304011	ANKRD18B	-1.042815997
UGGT2	-1.4574321	EFHC2	-1.172842811	SKA3	-1.039798596
RNASE2	-1.447742408	NDC1	-1.171955647	ZBED8	-1.038832663
C8orf59	-1.439501888	TUBA4A	-1.169962432	SFR1	-1.035870595
DSCC1	-1.439049906	ECT2	-1.166986889	E2F8	-1.033542012
ESCO2	-1.434348728	FRA10AC1	-1.166809379	CMC1	-1.033454447
CLEC11A	-1.415450491	OMA1	-1.16607659	CD300A	-1.033113471
CDC7	-1.410825774	MNS1	-1.160442768	SLA	-1.032549129
HIST1H4E	-1.406089437	OPN3	-1.155233148	CNIH1	-1.02936023
CKLF	-1.398634456	HPF1	-1.150132593	ZNF302	-1.029274117
SGO2	-1.392522444	PRF1	-1.145375445	CNRIP1	-1.02721622
C5orf34	-1.391136958	MED31	-1.140429146	HIST1H4A	-1.027099939
FAM102B	-1.3842082	RFXAP	-1.132553976	FABP5	-1.025482679
POLE2	-1.382974418	GDPD1	-1.127607633	LMNB1	-1.025390551
RFESD	-1.376131831	DHFR	-1.126119307	FBXO4	-1.024120058
PRIM1	-1.368756725	DUSP10	-1.124821254	TIMM17A	-1.023766199
PSAT1	-1.358748341	AP5M1	-1.122670129	CKAP2L	-1.023729058
SLCO4C1	-1.339789584	ZNF239	-1.121212651	PTTG1	-1.023339964
SPIN4	-1.323836822	GINS4	-1.119517743	RRM2	-1.021997661
TFRC	-1.323736706	SASS6	-1.114136797	TTC32	-1.019885611
PSMG1	-1.322986141	RGS5	-1.113354372	METTL21A	-1.016450296
CRLS1	-1.316239976	HMGB1	-1.107575871	GCSAML	-1.016215695
TGDS	-1.311379005	INSIG2	-1.102713369	DEK	-1.016203222
ZWINT	-1.302834763	ACTR6	-1.102483039	XRCC2	-1.011381303
ID2	-1.301742341	FBXO43	-1.101834827	CLECL1	-1.011163591
FANCB	-1.29736489	CKAP2	-1.09970209	CXCL3	-1.008380087
MRPS18C	-1.293683693	HSPA4L	-1.098823493	MYB	-1.007429257
KIT	-1.284722786	MPHOSPH6	-1.095260081	S100A4	-1.006874585
PTX3	-1.283840419	HCST	-1.093785627	HIST1H2BI	-1.006756914
ZFP36L2	-1.280506669	CCNE2	-1.091670636	RMI2	-1.006401936
PIGW	-1.259827856	CATSPER1	-1.09143749	LSM3	-1.005606419
RPL22L1	-1.247445525	ADCY7	-1.088430618	RBPJ	-1.00520292
CKS2	-1.245406045	BNIP3	-1.081421543	TPM1	-1.005014763
NFYB	-1.234876522	KIF11	-1.079207699	CCDC15	-1.004355179
ASNS	-1.233598398	SKA1	-1.078052854	GATM	-1.004024923
NEIL3	-1.233314553	DEPDC1B	-1.074853492		
RNASEH2B	-1.231698578	CCDC169	-1.071992026		

RUNX1 - 0 Dox targets

UpRegulated				DownRegulated					
GeneID	logFC	GeneID	logFC	GeneID	logFC	GeneID	logFC	GeneID	logFC
RRAS	2.665569152	HSF1	1.265398725	FST	-2.086354497	TPRKB	-1.2787987	MPHOSPH6	-1.095260081
CDKN1A	2.556758886	RGS12	1.261735389	DNAH14	-2.073301035	RAD51AP1	-1.2783923	HCST	-1.093785627
RELB	2.184489588	CLSTN1	1.25292865	IDI1	-1.852603127	HEMGN	-1.253976	ADCY7	-1.088430618
CTTN	2.124154609	HIST1H1C	1.246275496	PTPN22	-1.845156748	RPL22L1	-1.2474455	BNIP3	-1.081421543
HELZ2	2.066890578	PLCD1	1.239764508	SULF2	-1.830877188	ZNF695	-1.2407035	BORA	-1.081387614
PPDPF	2.06608069	KCTD17	1.238224808	PLAC8	-1.767964549	CENPE	-1.2398952	KIF11	-1.079207699
SPTAN1	2.028619022	BSG	1.228334417	CYP1B1	-1.755781734	ATAD5	-1.2371477	DEPDC1B	-1.074853492
SGSH	2.006978866	PAFAH1B3	1.226509788	GYPB	-1.700995172	NEIL3	-1.2333146	RHEBL1	-1.070710078
COL9A2	1.976270399	ATG2A	1.226079791	FAM72C	-1.682262367	CYCS	-1.2313579	CEP55	-1.06885002
FCGR1A	1.880275656	TNFAIP8L3	1.21733847	FAM72A	-1.672770084	ATP5G1	-1.2266383	ATP10D	-1.067477982
TYMP	1.856571315	LRRC8A	1.215708042	ITGB3BP	-1.655835311	BLM	-1.2251798	HAT1	-1.066691287
SOX7	1.74552543	P2RX4	1.215094413	FAM72B	-1.644321821	G2E3	-1.2200942	PPM1K	-1.066138258
HSPB1	1.726998825	NDRG1	1.212642804	SUGT1	-1.634667892	HIST1H2BO	-1.2187813	PPIH	-1.065452688
YPEL3	1.663849163	LRP10	1.205598519	GAR1	-1.620249478	KIF14	-1.2187657	OTULIN	-1.065237656
GAA	1.654162363	AP2A1	1.203819648	MSMO1	-1.610800559	SGO1	-1.2160716	SMC2	-1.064930401
FKBP8	1.642916748	CLPP	1.192897621	TNFSF10	-1.592391827	HELLS	-1.2066731	CENPU	-1.064427901
SERPINB9	1.640697223	CDC34	1.185804324	AP1S2	-1.529774115	BUB1	-1.2033347	ZDHHC13	-1.063645198
JUP	1.61458624	PIM3	1.185001926	MAD2L1	-1.522190002	GPAT3	-1.1998236	SLC31A2	-1.062288648
SASH1	1.597652518	TNRC18	1.1755315	ITGA4	-1.517856284	CEP70	-1.1991478	ALOX5AP	-1.056657029
ICAM1	1.592096977	GNB2	1.175189835	HMGB3	-1.513139138	GGH	-1.1935357	NAE1	-1.054326538
VPS9D1	1.565921122	BMF	1.169089753	NKG7	-1.4833186	CXCL2	-1.1889392	FANCI	-1.053913702
ALKBH7	1.55913892	CD83	1.167416739	PARPBP	-1.483249951	ITGAM	-1.187309	LGALS12	-1.051086218
SHF	1.553913453	PLEKHG2	1.158419082	MAGOHB	-1.480386531	HSPB11	-1.1868251	B3GNT5	-1.04415946
MAP1LC3A	1.536893704	KDM6B	1.157810629	SNRPG	-1.466455855	SKA2	-1.181177	ZNF563	-1.043901968
SNAPC2	1.515293352	SLC9A1	1.156839705	UGGT2	-1.4574321	CAMK1	-1.1751376	ANKRD18B	-1.042815997
GPX1	1.498244857	NUMA1	1.15645644	RNASE2	-1.447742408	FAM200B	-1.1746436	SKA3	-1.039798596
AHR	1.494081523	LMF1	1.147647299	C8orf59	-1.439501888	LSM5	-1.1730401	RACGAP1	-1.038628751
SHISA5	1.486476949	PML	1.141030616	ESCO2	-1.434348728	EFHC2	-1.1728428	SFR1	-1.035870595
AP2A2	1.479498181	FAM89B	1.138398242	HMGCS1	-1.433251619	TOP2A	-1.170856	ZGRF1	-1.035415384
SFN	1.475818208	RGL1	1.135843453	CLEC11A	-1.415450491	ECT2	-1.1669869	RAD54L	-1.035180778
GLDC	1.475400943	CTSB	1.129554663	HIST1H4E	-1.406089437	OMA1	-1.1660766	E2F8	-1.033542012
TMEM8A	1.474912283	MXD4	1.128865219	IGFBP7	-1.405448113	MNS1	-1.1604428	CD300A	-1.033113471
CORO1B	1.463913139	MAPK8IP3	1.122611127	SNRPE	-1.394243348	PLK4	-1.1581797	SLA	-1.032549129
TMEM134	1.458610969	SLC25A23	1.120800351	SGO2	-1.392522444	C3orf58	-1.1565439	CDC25C	-1.03046314
GPX4	1.429308857	PHLDA1	1.119977061	HIST1H4C	-1.391598606	OPN3	-1.1552331	CNIH1	-1.02936023
EGR1	1.428364787	SLC27A1	1.118379336	C5orf34	-1.391136958	GLMN	-1.1518212	ZNF302	-1.029274117
FCGRT	1.427084156	FOSL2	1.11638032	FAM102B	-1.3842082	HPF1	-1.1501326	HIST1H4A	-1.027099939
NPTX1	1.424410212	NUCB1	1.115952939	RFESD	-1.376131831	PRF1	-1.1453754	HIST1H2AH	-1.026856976
GUK1	1.420799672	ACADVL	1.110876074	KIF20B	-1.365799701	C4orf46	-1.1445468	LMNB1	-1.025390551
NATD1	1.406899611	SNX15	1.109192157	PSAT1	-1.358748341	NDUFA5	-1.1435585	FBXO4	-1.024120058
DBN1	1.384426949	MAP4	1.107076584	SLCO4C1	-1.339789584	MED31	-1.1404291	TIMM17A	-1.023766199
TESK2	1.379969454	RAB5C	1.102651425	LMO4	-1.332401179	GDPD1	-1.1276076	CKAP2L	-1.023729058
SPATA2L	1.365090881	MGRN1	1.099711127	EXOSC8	-1.331834958	DHFR	-1.1261193	PTTG1	-1.023339964
NFKB2	1.35314645	TIMP2	1.096179037	SPIN4	-1.323836822	DUSP10	-1.1248213	TTC32	-1.019885611
BCL6	1.348895606	GSN	1.095454856	TFRC	-1.323736706	ZNF239	-1.1212127	POC1B-GALNT4	-1.019719209
PLEKHB1	1.348872046	NUDT16L1	1.09026539	PSMG1	-1.322986141	PCNA	-1.1206343	METTL21A	-1.016450296
BCL6B	1.344361494	CDC42BPB	1.076848268	CRLS1	-1.316239976	GINS4	-1.1195177	GCSAML	-1.016215695
SERTAD1	1.341173805	DLGAP4	1.075584878	RIDA	-1.313873085	SASS6	-1.1141368	DEK	-1.016203222
EEF2	1.328735823	SH3D21	1.068085154	CENPI	-1.313817272	RGS5	-1.1133544	BUB1B	-1.013346952
ARHGDIA	1.327108263	EPS8	1.0632467	HIST1H4B	-1.306038075	HMGB1	-1.1075759	XRCC2	-1.011381303
CTSC	1.326661616	RPS2	1.062608813	NCAPG	-1.301764134	EMB	-1.1055041	RMI2	-1.006401936
SGK1	1.323704768	RAVER1	1.062467765	ID2	-1.301742341	INSIG2	-1.1027134	LSM3	-1.005606419
ID3	1.323525532	MAP2K2	1.060263217	FANCB	-1.29736489	ACTR6	-1.102483	RBPJ	-1.00520292
CIC	1.312206755	MSRA	1.058137964	MRPS18C	-1.293683693	FBXO43	-1.1018348	TPM1	-1.005014763
FLT1	1.311089202	RAB35	1.053758741	PBK	-1.288564658	SLC48A1	-1.101574	CARHSP1	-1.004710698
ULK1	1.307052989	TMEM115	1.053375473	KIT	-1.284722786	CKAP2	-1.0997021	CCDC15	-1.004355179
NECTIN2	1.306088981	MEGF8	1.047532708	PTX3	-1.283840419	CENPF	-1.0993086		
SOCS2	1.300813534	ECHDC3	1.046361952	ZFP36L2	-1.280506669	HSPA4L	-1.0988235		
FAM222B	1.297930703	MINK1	1.042850625						
ATP6V0C	1.29755722	ITGAV	1.017744417						
MVB12A	1.295066827	HMG20B	1.016987919						
CACFD1	1.293164597	STRN4	1.014822805						
HYAL2	1.292691649	ACTG1	1.012181475						
RASIP1	1.291106767	PLTP	1.010276511						
ZFP36	1.286247275	MED25	1.007248945						
NRN1	1.28168554	REXO1	1.003098981						
SLC7A10	1.279593232	SOD2	1.000567005						
ZNF219	1.268526853								

RUNX1 - 5 Dox targets

UpRegulated	
GeneID	logFC
RRAS	2.665569152
CTTN	2.124154609
SASH1	1.597652518
ICAM1	1.592096977
SNAPC2	1.515293352
AP2A2	1.479498181
NPTX1	1.424410212
GUK1	1.420799672
NFKB2	1.35314645
PLCD1	1.239764508
TNFAIP8L3	1.21733847
GNB2	1.175189835
BMF	1.169089753
IGF2R	1.072567068
EPS8	1.0632467
MAP2K2	1.060263217
TMEM115	1.053375473

DownRegulated	
GeneID	logFC
IDI1	-1.852603127
PTPN22	-1.845156748
MSMO1	-1.610800559
TNFSF10	-1.592391827
AP1S2	-1.529774115
HMGCS1	-1.433251619
CLEC11A	-1.415450491
TFRC	-1.323736706
MRPS18C	-1.293683693
GGH	-1.193535731
FAM200B	-1.17464365
RGS5	-1.113354372
CKAP2	-1.09970209
MPHOSPH6	-1.095260081
HCST	-1.093785627
ADCY7	-1.088430618
BNIP3	-1.081421543
CNIH1	-1.02936023
ZNF302	-1.029274117
TIMM17A	-1.023766199
RBPJ	-1.00520292

Genes that have lost RUNX1 binding and gained RUNX1-ETO binding

UpRegulated			
GeneID	logFC	GeneID	logFC
CDKN1A	2.5567589	RGS12	1.261735389
HELZ2	2.0668906	CLSTN1	1.25292865
PPDPF	2.0660807	KCTD17	1.238224808
SPTAN1	2.028619	BSG	1.228334417
SGSH	2.0069789	ATG2A	1.226079791
COL9A2	1.9762704	LRRC8A	1.215708042
FCGR1A	1.8802757	P2RX4	1.215094413
SOX7	1.7455254	NDRG1	1.212642804
HSPB1	1.7269988	LRP10	1.205598519
YPEL3	1.6638492	CDC34	1.185804324
JUP	1.6145862	PIM3	1.185001926
ALKBH7	1.5591389	TNRC18	1.1755315
SHF	1.5539135	CD83	1.167416739
MAP1LC3A	1.5368937	KDM6B	1.157810629
GPX1	1.4982449	SLC9A1	1.156839705
SHISA5	1.4864769	NUMA1	1.15645644
SFN	1.4758182	LMF1	1.147647299
GLDC	1.4754009	FAM89B	1.138398242
TMEM8A	1.4749123	RGL1	1.135843453
CORO1B	1.4639131	CTSB	1.129554663
TMEM134	1.458611	MXD4	1.128865219
GPX4	1.4293089	MAPK8IP3	1.122611127
EGR1	1.4283648	SLC25A23	1.120800351
NATD1	1.4068996	PHLDA1	1.119977061
DBN1	1.3844269	SLC27A1	1.118379336
TESK2	1.3799695	FOSL2	1.11638032
BCL6	1.3488956	NUCB1	1.115952939
PLEKHB1	1.348872	ACADVL	1.110876074
BCL6B	1.3443615	MAP4	1.107076584
SERTAD1	1.3411738	MGRN1	1.099711127
ARHGDIA	1.3271083	TIMP2	1.096179037
CTSC	1.3266616	GSN	1.095454856
SGK1	1.3237048	NUDT16L1	1.09026539
ID3	1.3235255	CDC42BPB	1.076848268
CIC	1.3122068	DLGAP4	1.075584878
FLT1	1.3110892	RAB35	1.053758741
ULK1	1.307053	MEGF8	1.047532708
NECTIN2	1.306089	MINK1	1.042850625
FAM222B	1.2979307	ITGAV	1.017744417
ATP6V0C	1.2975572	HMG20B	1.016987919
MVB12A	1.2950668	STRN4	1.014822805
CACFD1	1.2931646	ACTG1	1.012181475
RASIP1	1.2911068	PLTP	1.010276511
ZFP36	1.2862473	MED25	1.007248945
NRN1	1.2816855	REXO1	1.003098981
SLC7A10	1.2795932	SOD2	1.000567005

Down Regulated			
GeneID	logFC	GeneID	logFC
FST	-2.086354	HPF1	-1.1501326
DNAH14	-2.073301	PRF1	-1.1453754
SULF2	-1.830877	MED31	-1.1404291
PLAC8	-1.767965	GDPD1	-1.1276076
CYP1B1	-1.755782	DHFR	-1.1261193
SUGT1	-1.634668	DUSP10	-1.1248213
ITGA4	-1.517856	ZNF239	-1.1212127
HMGB3	-1.513139	GINS4	-1.1195177
PARPBP	-1.48325	SASS6	-1.1141368
SNRPG	-1.466456	HMGB1	-1.1075759
UGGT2	-1.457432	INSIG2	-1.1027134
RNASE2	-1.447742	ACTR6	-1.102483
C8orf59	-1.439502	FBXO43	-1.1018348
ESCO2	-1.434349	HSPA4L	-1.0988235
HIST1H4E	-1.406089	KIF11	-1.0792077
SGO2	-1.392522	DEPDC1B	-1.0748535
C5orf34	-1.391137	CEP55	-1.06885
FAM102B	-1.384208	ATP10D	-1.067478
RFESD	-1.376132	PPM1K	-1.0661383
PSAT1	-1.358748	PPIH	-1.0654527
SLCO4C1	-1.33979	OTULIN	-1.0652377
SPIN4	-1.323837	CENPU	-1.0644279
PSMG1	-1.322986	ZDHHC13	-1.0636452
CRLS1	-1.31624	SLC31A2	-1.0622886
ID2	-1.301742	ALOX5AP	-1.056657
FANCB	-1.297365	NAE1	-1.0543265
KIT	-1.284723	FANCI	-1.0539137
PTX3	-1.28384	B3GNT5	-1.0441595
ZFP36L2	-1.280507	ANKRD18B	-1.042816
RPL22L1	-1.247446	SKA3	-1.0397986
NEIL3	-1.233315	SFR1	-1.0358706
CYCS	-1.231358	E2F8	-1.033542
G2E3	-1.220094	CD300A	-1.0331135
KIF14	-1.218766	SLA	-1.0325491
SGO1	-1.216072	HIST1H4A	-1.0270999
HELLS	-1.206673	LMNB1	-1.0253906
BUB1	-1.203335	FBXO4	-1.0241201
GPAT3	-1.199824	CKAP2L	-1.0237291
CEP70	-1.199148	PTTG1	-1.02334
CXCL2	-1.188939	TTC32	-1.0198856
ITGAM	-1.187309	METTL21A	-1.0164503
HSPB11	-1.186825	GCSAML	-1.0162157
LSM5	-1.17304	DEK	-1.0162032
EFHC2	-1.172843	XRCC2	-1.0113813
ECT2	-1.166987	RMI2	-1.0064019
OMA1	-1.166077	LSM3	-1.0056064
MNS1	-1.160443	TPM1	-1.0050148
OPN3	-1.155233	CCDC15	-1.0043552

Supplementary Table 4: Up and downregulated RUNX1-ETO and RUNX1 target genes

Tables of dysregulated RUNX1-ETO and RUNX1 target genes in RUNX1C+ cells (0 and 5 ng/ml Dox), including gene name (GeneID) and ranked by log2 fold (LogFC) of the FPKM +1. RUNX1-ETO targets that lose RUNX1 binding are also listed.

Cell cycle regulated genes			
G2-M phase		S phase	
HMGB2	HJURP	MCM5	SLBP
CDK1	CDCA3	PCNA	CCNE2
NUSAP1	HN1	TYMS	UBR7
UBE2C	CDC20	FEN1	POLD3
BIRC5	TTK	MCM2	MSH2
TPX2	CDC25C	MCM4	ATAD2
TOP2A	KIF2C	RRM1	RAD51
NDC80	RANGAP1	UNG	RRM2
CKS2	NCAPD2	GINS2	CDC45
NUF2	DLGAP5	MCM6	CDC6
CKS1B	CDCA2	CDCA7	EXO1
MKI67	CDCA8	DTL	TIPIN
TMPO	ECT2	PRIM1	DSCC1
CENPF	KIF23	UHRF1	BLM
TACC3	HMMR	MLF1IP	CASP8AP2
FAM64A	AURKA	HELLS	USP1
SMC4	PSRC1	RFC2	CLSPN
CCNB2	ANLN	RPA2	POLA1
CKAP2L	LBR	NASP	CHAF1B
CKAP2	CKAP5	RAD51AP1	BRIP1
AURKB	CENPE	GMNN	E2F8
BUB1	CTCF	WDR76	
KIF11	NEK2		
ANP32E	G2E3		
TUBB4B	GAS2L3		
GTSE1	CBX5		
KIF20B	CENPA		

Supplementary Table 5: Cell cycle regulated genes used to infer cell cycle state of single cells from the scRNA-Seq derived clusters

Lists of genes regulated during G2-M and S phase of the cell cycle used to infer the cell cycle stage for each single cell of the scRNA-Seq datasets. Gene lists are provided by the CellCycleScoring function of the Seurat software.

Supplementary Dataset 1: Nafria et al., 2019

Supplementary datasets and excel files corresponding to the Supplementary tables are accessible in our paper submission to BioRxiv (doi: https://doi.org/10.1101/748921). BioarXiv will make available a link to the corresponding journal upon final publication.

Raw data from our sequencing experiments has been deposited on Gene expression Omnibus (GEO) number: GSE137673.

7 REFERENCES

Ackermann, M., Liebhaber, S., Klusmann, J., and Lachmann, N. (2015). Lost in translation: pluripotent stem cell-derived hematopoiesis. EMBO Mol Med *7*, 1388–1402.

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

Adolfsson, J., Borge, O.J., Bryder, D., Theilgaard-Mönch, K., Åstrand-Grundström, I., Sitnicka, E., Sasaki, Y., and Jacobsen, S.E.. (2001). Upregulation of Flt3 Expression within the Bone Marrow Lin−Sca1+c-kit+ Stem Cell Compartment Is Accompanied by Loss of Self-Renewal Capacity. Immunity *15*, 659–669.

Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.-J., Thoren, L.A.M., et al. (2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell *121*, 295–306.

Agrawal, M., Schwarz, P., Giaimo, B.D., Bedzhov, I., Corbacioglu, A., Weber, D., Gaidzik, V.I., Jahn, N., Rücker, F.G., Schroeder, T., et al. (2019). Functional and clinical characterization of the alternatively spliced isoform AML1-ETO9a in adult patients with translocation t(8;21)(q22;q22.1) acute myeloid leukemia (AML). Leukemia 1–5.

Ahn, E.-Y., Yan, M., Malakhova, O.A., Lo, M.-C., Boyapati, A., Ommen, H.B., Hines, R., Hokland, P., and Zhang, D.-E. (2008). Disruption of the NHR4 domain structure in AML1-ETO abrogates SON binding and promotes leukemogenesis. Proc. Natl. Acad. Sci. *105*, 17103–17108.

Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature *404*, 193–197.

Albig, W., and Doenecke, D. (1997). The human histone gene cluster at the D6S105 locus. Hum. Genet. *101*, 284–294.

Alcalay, M., Meani, N., Gelmetti, V., Fantozzi, A., Fagioli, M., Orleth, A., Riganelli, D., Sebastiani, C., Cappelli, E., Casciari, C., et al. (2003). Acute myeloid leukemia fusion proteins deregulate genes involved in stem cell maintenance and DNA repair. J. Clin. Invest. *112*, 1751–1761.

Aldinucci, D., and Colombatti, A. (2014). The inflammatory chemokine CCL5 and cancer progression. Mediators Inflamm. *2014*, 292376.

Alexander, R.D., Innocente, S.A., Barrass, J.D., and Beggs, J.D. (2010). Splicing-Dependent RNA Polymerase Pausing in Yeast. Mol. Cell *40*, 582–593.

Alhashem, Y.N., Vinjamur, D.S., Basu, M., Klingmüller, U., Gaensler, K.M.L., and Lloyd, J. a. (2011). Transcription factors KLF1 and KLF2 positively regulate embryonic and fetal β-globin genes through direct promoter binding. J. Biol. Chem. *286*, 24819–24827.

Allan, J., Hartman, P.G., Crane-Robinson, C., and Aviles, F.X. (1980). The structure of histone H1 and its location in chromatin. Nature *288*, 675–679.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex : a central integrator of transcription. Nat. Publ. Gr. *16*, 155–166.

Amann, J.M., Nip, J., Strom, D.K., Lutterbach, B., Harada, H., Lenny, N., Downing, J.R., Meyers, S., and Hiebert, S.W. (2001). ETO, a Target of t(8;21) in Acute Leukemia, Makes Distinct Contacts with Multiple Histone Deacetylases and Binds mSin3A through Its Oligomerization Domain. Mol. Cell. Biol. *21*, 6470–6483.

Anderson, G., Mackay, N., Gilroy, K., Hay, J., Borland, G., McDonald, A., Bell, M., Hassanudin, S.A., Cameron, E., Neil, J.C., et al. (2018). RUNX-mediated growth arrest and senescence are attenuated by diverse mechanisms in cells expressing RUNX1 fusion oncoproteins. J. Cell. Biochem. *119*, 2750–2762.

Araten, D.J., Krejci, O., DiTata, K., Wunderlich, M., Sanders, K.J., Zamechek, L., and Mulloy, J.C.

(2013). The rate of spontaneous mutations in human myeloid cells. Mutat. Res. Mol. Mech. Mutagen. *749*, 49–57.

Argiropoulos, B., and Humphries, R.K. (2007). Hox genes in hematopoiesis and leukemogenesis. Oncogene *26*, 6766–6776.

Aronson, B.D., Fisher, a L., Blechman, K., Caudy, M., and Gergen, J.P. (1997). Groucho-dependent and -independent repression activities of Runt domain proteins. Mol. Cell. Biol. *17*, 5581–5587.

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., et al. (2019). Subtype-specific regulatory network rewiring in acute myeloid leukemia. Nat. Genet. *51*, 151–162.

Avellino, R., Havermans, M., Erpelinck, C., Sanders, M.A., Hoogenboezem, R., van de Werken, H.J.G., Rombouts, E., van Lom, K., van Strien, P.M.H., Gebhard, C., et al. (2016). An autonomous CEBPA enhancer specific for myeloid-lineage priming and neutrophilic differentiation. Blood *127*, 2991–3003.

Azcoitia, V., Aracil, M., Martínez-A, C., and Torres, M. (2005). The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. Dev. Biol. *280*, 307–320.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., et al. (2006). Chromatin signatures of pluripotent cell lines. Nat. Cell Biol. *8*, 532–538.

Bacon, C.W., and D'Orso, I. (2019). CDK9: a signaling hub for transcriptional control. Transcription *10*, 57–75.

Banaszynski, L.A., Chen, L., Maynard-Smith, L.A., Lisa Ooi, A.G., and Wandless, T.J. (2006). A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules. Cell *126*, 995–1004.

Bannister, A.J., and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. Nature *384*, 641–643.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

Barboric, M., Lenasi, T., Chen, H., Johansen, E.B., Guo, S., and Peterlin, B.M. (2009). 7SK snRNP / P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development. Proc. Natl. Acad. Sci. U. S. A. *106*, 1–6.

Barrett, N.A., Malouf, C., Kapeni, C., Bacon, W.A., Giotopoulos, G., Jacobsen, S.E.W., Huntly, B.J., and Ottersbach, K. (2016). Mll-AF4 Confers Enhanced Self-Renewal and Lymphoid Potential during a Restricted Window in Development. Cell Rep. *16*, 1039–1054.

Barseguian, K., Lutterbach, B., Hiebert, S.W., Nickerson, J., Lian, J.B., Stein, J.L., Wijnen, A.J. van, and Stein, G.S. (2002). Multiple subnuclear targeting signals of the leukemia-related AML1/ETO and ETO repressor proteins. Proc. Natl. Acad. Sci. *99*, 15434–15439.

Becker, P.B., and Workman, J.L. (2013). Nucleosome remodeling and epigenetics. Cold Spring Harb. Perspect. Biol. *5*, a017905.

Bee, T., Ashley, E.L.K., Bickley, S.R.B., Jarratt, A., Li, P.-S., Sloane-Stanley, J., Göttgens, B., and de Bruijn, M.F.T.R. (2009). The mouse Runx1 +23 hematopoietic stem cell enhancer confers hematopoietic specificity to both Runx1 promoters. Blood *113*, 5121–5124.

Bee, T., Swiers, G., Muroi, S., Pozner, A., Nottingham, W., Santos, A.C., Li, P., Taniuchi, I., and Bruijn, M.F.T.R. De (2010). Nonredundant roles for Runx1 alternative promoters reflect their activity at discrete stages of developmental hematopoiesis. Blood *115*, 3042–3051.

Begay, V., Smink, J., and Leutz, A. (2004). Essential Requirement of CCAAT / Enhancer Binding Proteins in Embryogenesis. Mol. Cell. Biol. *24*, 9744–9751.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482–485.

Ben-Ami, O., Friedman, D., Leshkowitz, D., Goldenberg, D., Orlovsky, K., Pencovich, N., Lotem, J., Tanay, A., and Groner, Y. (2013). Addiction of t(8;21) and inv(16) Acute Myeloid Leukemia to Native RUNX1. Cell Rep. *4*, 1131–1143.

Benci, J.L., Xu, B., Qiu, Y., Wu, T.J., Dada, H., Twyman-Saint Victor, C., Cucolo, L., Lee, D.S.M., Pauken, K.E., Huang, A.C., et al. (2016). Tumor Interferon Signaling Regulates a Multigenic Resistance Program to Immune Checkpoint Blockade. Cell *167*, 1540–1554.e12.

Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A.G., Gralnick, H.R., And Sultan, C. (1985). Proposed Revised Criteria for the Classification of Acute Myeloid Leukemia. Ann. Intern. Med. *103*, 620.

Berg, T., Fliegauf, M., Burger, J., Staege, M.S., Liu, S., Martinez, N., Heidenreich, O., Burdach, S., Haferlach, T., Werner, M.H., et al. (2008). Transcriptional upregulation of p21/WAF/Cip1 in myeloid leukemic blasts expressing AML1-ETO. Haematologica *93*, 1728–1733.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315–326.

Bertrand, J.Y., Chi, N.C., Santoso, B., Teng, S., Stainier, D.Y.R., and Traver, D. (2010). Haematopoietic stem cells derive directly from aortic endothelium during development. Nature *464*, 108–111.

Bickmore, W.A., and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. Cell *152*, 1270–1284.

Biggs, J.R., Peterson, L.F., Zhang, Y., Kraft, A.S., and Zhang, D.-E. (2006). AML1/RUNX1 phosphorylation by cyclin-dependent kinases regulates the degradation of AML1/RUNX1 by the anaphase-promoting complex. Mol. Cell. Biol. *26*, 7420–7429.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics *25*, 1091–1093.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91–99.

Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG Islands Recruit a Histone H3 Lysine 36 Demethylase. Mol. Cell *38*, 179–190.

Blake, M.C., Jambou, R.C., Swick, A.G., Kahn, J.W., and Azizkhan, J.C. (1990). Transcriptional initiation is controlled by upstream GC-box interactions in a TATAA-less promoter. Mol. Cell. Biol. *10*, 6632–6641.

Bloom, K., and Joglekar, A. (2010). Towards building a chromosome segregation machine. Nature *463*, 446–456.

Bloom, W., and Bartelmez, G.W. (1940). Hematopoiesis in young human embryos. Am. J. Anat. *67*, 21–53.

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., et al. (2018). Prospective Isolation and Characterization of Genetically and Functionally Distinct AML Subclones. Cancer Cell *34*, 674–689.e8.

Boisset, J.-C., van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E., and Robin, C. (2010). In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. Nature *464*, 116–120.

Boitano, A.E., Wang, J., Romeo, R., Bouchez, L.C., Parker, A.E., Sutton, S.E., Walker, J.R., Flaveny, C.A., Perdew, G.H., Denison, M.S., et al. (2010). Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. Science *329*, 1345–1348.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bomken, S., Buechler, L., Rehe, K., Ponthan, F., Elder, A., Blair, H., Bacon, C.M., Vormoor, J., and Heidenreich, O. (2013). Lentiviral marking of patient-derived acute lymphoblastic leukaemic cells allows in vivo tracking of disease progression. Leukemia *27*, 1792.

Bonifer, C., and Bowen, D.T. (2010). Epigenetic mechanisms regulating normal and malignant haematopoiesis: new therapeutic targets for clinical medicine. Expert Rev. Mol. Med. *12*, e6.

Bonnet, D., and Dick, J.E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat. Med. *3*, 730–737.

Bonora, G., Plath, K., and Denholtz, M. (2014). A mechanistic link between gene regulation and genome architecture in mammalian development. Curr. Opin. Genet. Dev. *27*, 92–101.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Boyer, S.W., Schroeder, A. V, Smith-Berdan, S., and Forsberg, E.C. (2011). All hematopoietic cells develop from hematopoietic stem cells through Flk2/Flt3-positive progenitor cells. Cell Stem Cell *9*, 64–73.

Breig, O., Bras, S., Martinez Soria, N., Osman, D., Heidenreich, O., Haenlin, M., and Waltzer, L. (2014). Pontin is a critical regulator for AML1-ETO-induced leukemia. Leukemia *28*, 1–9.

Bröske, A.-M., Vockentanz, L., Kharazi, S., Huska, M.R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R., et al. (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat. Genet. *41*, 1207–1215.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena Histone Acetyltransferase A: A Homolog to Yeast Gcn5p Linking Histone Acetylation to Gene Activation. Cell *84*, 843–851.

de Bruijn, M.F., Speck, N. a, Peeters, M.C., and Dzierzak, E. (2000). Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. EMBO J. *19*, 2465–2474.

De Bruijn, M.F.T.R., Ma, X., Robin, C., Ottersbach, K., Sanchez, M., and Dzierzak, E. (2002). Haematopoietic Stem Cells Localize to the Endothelial Cell Layer in the Midgestation Mouse Aorta. Immunity *16*, 673–683.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. *2015*, 21.29.1-21.29.9.

Bullinger, L., Rücker, F.G., Kurz, S., Du, J., Scholl, C., Sander, S., Corbacioglu, A., Lottaz, C., Krauter, J., Fröhling, S., et al. (2007). Gene-expression profiling identifies distinct subclasses of core binding factor acute myeloid leukemia. Blood *110*, 1291–1300.

Burda, P., Curik, N., Kokavec, J., Basova, P., Mikulenkova, D., Skoultchi, A.I., Zavadil, J., and Stopka, T. (2009). PU.1 activation relieves GATA-1-mediated repression of Cebpa and Cbfb during leukemia differentiation. Mol. Cancer Res. *7*, 1693–1703.

Burel, S.A., Harakawa, N., Zhou, L., Pabst, T., Tenen, D.G., and Zhang, D.E. (2001). Dichotomy of AML1-ETO functions: growth arrest versus block of differentiation. Mol. Cell. Biol. *21*, 5577–5590.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Buza-Vidas, N., Woll, P., Hultquist, A., Duarte, S., Lutteropp, M., Bouriez-Jones, T., Ferry, H., Luc, S., and Jacobsen, S.E.W. (2011). FLT3 expression initiates in fully multipotent mouse hematopoietic progenitor cells. Blood *118*, 1544–1548.

Byrd, J.C., Mrózek, K., Dodge, R.K., Carroll, A.J., Edwards, C.G., Arthur, D.C., Pettenati, M.J., Patil, S.R., Rao, K.W., Watson, M.S., et al. (2002). Pretreatment cytogenetic abnormalities are predictive of

induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). Blood *100*, 4325–4336.

Cabezas-Wallscheid, N., Eichwald, V., de Graaf, J., Löwer, M., Lehr, H.-A., Kreft, A., Eshkind, L., Hildebrandt, A., Abassi, Y., Heck, R., et al. (2013). Instruction of haematopoietic lineage choices, evolution of transcriptional landscapes and cancer stem cell hierarchies derived from an AML1-ETO mouse model. EMBO Mol. Med. *5*, 1804–1820.

Cai, Z., de Bruijn, M., Ma, X., Dortland, B., Luteijn, T., Downing, R.J., and Dzierzak, E. (2000). Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. Immunity *13*, 423–431.

Caligiuri, M.A., Schichman, S.A., Strout, M.P., Mrózek, K., Baer, M.R., Frankel, S.R., Barcos, M., Herzig, G.P., Croce, C.M., and Bloomfield, C.D. (1994). Molecular rearrangement of the ALL-1 gene in acute myeloid leukemia without cytogenetic evidence of 11q23 chromosomal translocations. Cancer Res. *54*, 370–373.

Caligiuri, M.A., Strout, M.P., Lawrence, D., Arthur, D.C., Baer, M.R., Yu, F., Knuutila, S., Mrózek, K., Oberkircher, A.R., Marcucci, G., et al. (1998). Rearrangement of ALL1 (MLL) in acute myeloid leukemia with normal cytogenetics. Cancer Res. *58*, 55–59.

Capron, C., Lécluse, Y., Kaushik, A.L., Foudi, A., Lacout, C., Sekkai, D., Godin, I., Albagli, O., Poullion, I., Svinartchouk, F., et al. (2006). The SCL relative LYL-1 is required for fetal and adult hematopoietic stem cell function and B-cell differentiation. Blood *107*, 4678–4686.

Cazzola, M., Della Porta, M.G., and Malcovati, L. (2013). The genetic basis of myelodysplasia and its clinical relevance. Blood *122*, 4021–4034.

Chadwick, K., Wang, L., Li, L., Menendez, P., Murdoch, B., Rouleau, A., and Bhatia, M. (2003). Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells. Blood *102*, 906–915.

Chanda, B., Ditadi, A., Iscove, N.N., and Keller, G. (2013). Retinoic Acid Signaling Is Essential for Embryonic Hematopoietic Stem Cell Development. Cell *155*, 215–227.

Chen, J.D., and Evans, R.M. (1995). A transcriptional co-repressor that interacts with nuclear hormone receptors. Nature *377*, 454–457.

Chen, L.T., and Weiss, L.P. (1975). The development of vertebral bone marrow of human fetuses. Blood *46*, 389–408.

Chen, T., and Dent, S.Y.R. (2014). Chromatin modifiers and remodellers: regulators of cellular differentiation. Nat. Rev. Genet. *15*, 93–106.

Chen, D., Ma, H., Hong, H., Koh, S.S., Huang, S.M., Schurter, B.T., Aswad, D.W., Stallcup, M.R., Wong, J., Tempst, P., et al. (1999). Regulation of transcription by a protein methyltransferase. Science *284*, 2174–2177.

Chen, J., Kao, Y.-R., Sun, D., Todorova, T.I., Reynolds, D., Narayanagari, S.-R., Montagna, C., Will, B., Verma, A., and Steidl, U. (2019). Myelodysplastic syndrome progression to acute myeloid leukemia at the stem cell level. Nat. Med. *25*, 103–110.

Chen, M.J., Yokomizo, T., Zeigler, B.M., Dzierzak, E., and Speck, N. a (2009). Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. Nature *457*, 887–891.

Chen, M.J., Li, Y., De Obaldia, M.E., Yang, Q., Yzaguirre, A.D., Yamada-Inagawa, T., Vink, C.S., Bhandoola, A., Dzierzak, E., and Speck, N. a. (2011). Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells. Cell Stem Cell *9*, 541–552.

Chen, Y., Bates, D.L., Dey, R., Chen, P.-H., Machado, A.C.D., Laird-Offringa, I.A., Rohs, R., and Chen, L. (2012). DNA Binding by GATA Transcription Factor Suggests Mechanisms of DNA Looping and Long-Range Gene Regulation. Cell Rep. *2*, 1197–1206.

Chen-Wichmann, L., Shvartsman, M., Preiss, C., Hockings, C., Windisch, R., Redondo Monte, E.,

Leubolt, G., Spiekermann, K., Lausen, J., Brendel, C., et al. (2019). Compatibility of RUNX1/ETO fusion protein modules driving CD34+ human progenitor cell expansion. Oncogene *38*, 261–272.

Cheng, C.K., Li, L., Cheng, S.H., Lau, K.M., Chan, N.P.H., Wong, R.S.M., Shing, M.M.K., Li, C.K., and Ng, M.H.L. (2008). Transcriptional repression of the RUNX3/AML2 gene by the t(8;21) and inv(16) fusion proteins in acute myeloid leukemia. Blood *112*, 3391–3402.

Cheng, C.K., Li, L., Cheng, S.H., Ng, K., Chan, N.P.H., Ip, R.K.L., Wong, R.S.M., Shing, M.M.K., Li, C.K., and Ng, M.H.L. (2011). Secreted-frizzled related protein 1 is a transcriptional repression target of the t(8;21) fusion protein in acute myeloid leukemia. Blood *118*, 6638–6648.

Cheng, K., Sportoletti, P., Ito, K., Clohessy, J.G., Teruya-Feldstein, J., Kutok, J.L., and Pandolfi, P.P. (2010). The cytoplasmic NPM mutant induces myeloproliferation in a transgenic mouse model. Blood *115*, 3341–3345.

Cheng, T., Rodrigues, N., Shen, H., Yang, Y., Dombkowski, D., Sykes, M., and Scadden, D.T. (2000). Hematopoietic Stem Cell Quiescence Maintained by p21cip1/waf1. Science (80-.). *287*, 1804–1808.

Choi, Y., Elagib, K.E., Delehanty, L.L., and Goldfarb, A.N. (2006). Erythroid Inhibition by the Leukemic Fusion AML1-ETO Is Associated with Impaired Acetylation of the Major Erythroid Transcription Factor GATA-1. Cancer Res. *66*, 2990–2996.

Chou, F.-S., Wunderlich, M., Griesinger, A., and Mulloy, J.C. (2011a). N-Ras(G12D) induces features of stepwise transformation in preleukemic human umbilical cord blood cultures expressing the AML1-ETO fusion gene. Blood *117*, 2237–2240.

Chou, F.-S., Griesinger, A., Wunderlich, M., Lin, S., Link, K.A., Shrestha, M., Goyama, S., Mizukawa, B., Shen, S., Marcucci, G., et al. (2012). The thrombopoietin/MPL/Bcl-xL pathway is essential for survival and self-renewal in human preleukemia induced by AML1-ETO. Blood *120*, 709–719.

Chou, W.-C., Chou, S.-C., Liu, C.-Y., Chen, C.-Y., Hou, H.-A., Kuo, Y.-Y., Lee, M.-C., Ko, B.-S., Tang, J.-L., Yao, M., et al. (2011b). TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. Blood *118*, 3803–3810.

Clarke, R.L., Yzaguirre, A.D., Yashiro-Ohtani, Y., Bondue, A., Blanpain, C., Pear, W.S., Speck, N. a, and Keller, G. (2013). The expression of Sox17 identifies and regulates haemogenic endothelium. Nat. Cell Biol. *15*, 502–510.

Clements, A., Poux, A.N., Lo, W.-S., Pillus, L., Berger, S.L., and Marmorstein, R. (2003). Structural Basis for Histone and Phosphohistone Binding by the GCN5 Histone Acetyltransferase. Mol. Cell *12*, 461–473.

Cockerill, P.N. (2011). Structure and function of active chromatin and DNase I hypersensitive sites. FEBS J. *278*, 2182–2210.

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., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet. *48*, 1193–1203.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Core, L.J., Waterfall, J.J., Gilchrist, D.A., Fargo, D.C., Kwak, H., Adelman, K., and Lis, J.T. (2012). Defining the Status of RNA Polymerase at Promoters. Cell Rep. *2*, 1025–1035.

Cosgrove, M.S., Boeke, J.D., and Wolberger, C. (2004). Regulated nucleosome mobility and the histone code. Nat. Struct. Mol. Biol. *11*, 1037–1043.

Cozzio, A., Passegué, E., Ayton, P.M., Karsunky, H., Cleary, M.L., and Weissman, I.L. (2003). Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. Genes Dev. *17*, 3029–3035.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., and Carey, B.W. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Crisan, M., and Dzierzak, E. (2016). The many faces of hematopoietic stem cell heterogeneity. Development *143*, 4571–4581.

Cumano, A., Dieterlen-Lievre, F., and Godin, I. (1996). Lymphoid potential, probed before circulation in mouse, is restricted to caudal intraembryonic splanchnopleura. Cell *86*, 907–916.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). Drosophila Enhancer of Zeste/ESC Complexes Have a Histone H3 Methyltransferase Activity that Marks Chromosomal Polycomb Sites. Cell *111*, 185–196.

D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., and Baetge, E.E. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat. Biotechnol. *23*, 1534–1541.

D'Souza, S.L., Elefanty, A.G., Keller, G., and Dc, W. (2005). SCL / Tal-1 is essential for hematopoietic commitment of the hemangioblast but not for its development SCL / Tal-1 is essential for hematopoietic commitment of the hemangioblast but not for its development. Blood *105*, 3862–3870.

Daga, A., Tighe, J., and Calabi, F. (1992). Leukaemia/Drosophila homology. Nature *356*, 484–484.

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

Davis, J.N., McGhee, L., and Meyers, S. (2003). The ETO (MTG8) gene family. Gene *303*, 1–10.

Davis, R.P., Ng, E.S., Costa, M., Mossman, A.K., Sourris, K., Elefanty, A.G., and Stanley, E.G. (2008). Targeting a GFP reporter gene to the MIXL1 locus of human embryonic stem cells identifies human primitive streak-like cells and enables isolation of primitive hematopoietic precursors. Blood *111*, 1876–1884.

Dawson, M.A., and Kouzarides, T. (2012). Cancer Epigenetics: From Mechanism to Therapy. Cell *150*, 12–27.

Dekker, J., and Misteli, T. (2015). Long-Range Chromatin Interactions. Cold Spring Harb. Perspect. Biol. *7*, a019356.

Deschamps, J., and van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. Development *132*, 2931–2942.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., Zhou, M.-M., and Zhou, M.-M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491–496.

Di-Poï, N., Koch, U., Radtke, F., and Duboule, D. (2010). Additive and global functions of HoxA cluster genes in mesoderm derivatives. Dev. Biol. *341*, 488–498.

Ditadi, A., and Sturgeon, C.M. (2016). Directed differentiation of definitive hemogenic endothelium and hematopoietic progenitors from human pluripotent stem cells. Methods *101*, 65–72.

Ditadi, A., Sturgeon, C.M., and Keller, G. (2016). A view of human haematopoietic development from the Petri dish. Nat. Rev. Mol. Cell Biol. *18*.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Döhner, H., Weisdorf, D.J., and Bloomfield, C.D. (2015). Acute Myeloid Leukemia. N. Engl. J. Med. *373*, 1136–1152.

Dou, D.R., Calvanese, V., Sierra, M.I., Nguyen, A.T., Minasian, A., Saarikoski, P., Sasidharan, R., Ramirez, C.M., Zack, J.A., Crooks, G.M., et al. (2016). Medial HOXA genes demarcate haematopoietic stem cell fate during human development. Nat. Cell Biol. *18*, 595–606.

Doulatov, S., Notta, F., Eppert, K., Nguyen, L.T., Ohashi, P.S., and 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*, 585–593.

Dowdy, C.R., Xie, R., Frederick, D., Hussain, S., Zaidi, S.K., Vradii, D., Javed, A., Li, X., Jones, S.N., Lian, J.B., et al. (2010). Definitive hematopoiesis requires Runx1 C-terminal-mediated subnuclear

targeting and transactivation. Hum. Mol. Genet. *19*, 1048–1057.

Drew, E., Merzaban, J.S., Seo, W., Ziltener, H.J., and McNagny, K.M. (2005). CD34 and CD43 Inhibit Mast Cell Adhesion and Are Required for Optimal Mast Cell Reconstitution. Immunity *22*, 43–57.

Drissen, R., Thongjuea, S., Theilgaard-Mönch, K., and Nerlov, C. (2019). Identification of two distinct pathways of human myelopoiesis. Sci. Immunol. *4*, eaau7148.

Duffy, J.B., and Gergen, J.P. (1991). The Drosophila segmentation gene runt acts as a position-specific numerator element necessary for the uniform expression of the sex-determining gene Sex-lethal. Genes Dev. *5*, 2176–2187.

Dunne, J., Cullmann, C., Ritter, M., Soria, N.M., Drescher, B., Debernardi, S., Skoulakis, S., Hartmann, O., Krause, M., Krauter, J., et al. (2006). siRNA-mediated AML1/MTG8 depletion affects differentiation and proliferation-associated gene expression in t(8;21)-positive cell lines and primary AML blasts. Oncogene *25*, 6067–6078.

Eilken, H.M., Nishikawa, S.-I., and Schroeder, T. (2009). Continuous single-cell imaging of blood generation from haemogenic endothelium. Nature *457*, 896–900.

Elagib, K.E., Racke, F.K., Mogass, M., Khetawat, R., Delehanty, L.L., and Goldfarb, A.N. (2003). RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. Blood *101*, 4333–4341.

Elias, S., Yamin, R., Golomb, L., Tsukerman, P., Stanietsky-Kaynan, N., Ben-Yehuda, D., and Mandelboim, O. (2014). Immune evasion by oncogenic proteins of acute myeloid leukemia. Blood *123*, 1535–1543.

Emmrich, S., Katsman-Kuipers, J.E., Henke, K., Khatib, M.E., Jammal, R., Engeland, F., Dasci, F., Zwaan, C.M., den Boer, M.L., Verboon, L., et al. (2014). miR-9 is a tumor suppressor in pediatric AML with t(8;21). Leukemia *28*, 1022–1032.

Eppert, K., Takenaka, K., Lechman, E.R., Waldron, L., Nilsson, B., van Galen, P., Metzeler, K.H., Poeppl, A., Ling, V., Beyene, J., et al. (2011). Stem cell gene expression programs influence clinical outcome in human leukemia. Nat. Med. *17*, 1086–1093.

Erickson, P., Gao, J., Chang, K.S., Look, T., Whisenant, E., Raimondi, S., Lasher, R., Trujillo, J., Rowley, J., and 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*, 1825–1831.

Erickson, P., Dessev, G., Lasher, R., Philips, G., Robinson, M., and Drabkin, H. (1996). ETO and AML1 phosphoproteins are expressed in CD34+ hematopoietic progenitors: implications for t(8;21) leukemogenesis and monitoring residual disease. Blood *88*.

Escribano-Díaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J.T.F., Tkáč, J., Cook, M.A., Rosebrock, A.P., Munro, M., Canny, M.D., et al. (2013). A Cell Cycle-Dependent Regulatory Circuit Composed of 53BP1-RIF1 and BRCA1-CtIP Controls DNA Repair Pathway Choice. Mol. Cell *49*, 872–883.

Esposito, M.T., Zhao, L., Fung, T.K., Rane, J.K., Wilson, A., Martin, N., Gil, J., Leung, A.Y., Ashworth, A., and Eric So, C.W. (2015). Synthetic lethal targeting of oncogenic transcription factors in acute leukemia by PARP inhibitors. Nat. Med. *21*, 1481–1490.

Falini, B., Bolli, N., Shan, J., Martelli, M.P., Liso, A., Pucciarini, A., Bigerna, B., Pasqualucci, L., Mannucci, R., Rosati, R., et al. (2006). Both carboxy-terminus NES motif and mutated tryptophan(s) are crucial for aberrant nuclear export of nucleophosmin leukemic mutants in NPMc+ AML. Blood *107*, 4514–4523.

Falini, B., Nicoletti, I., Martelli, M.F., and Mecucci, C. (2007). Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. Blood *109*, 874–885.

Faure, A.J., Schmidt, D., Watt, S., Schwalie, P.C., Wilson, M.D., Xu, H., Ramsay, R.G., Odom, D.T., and Flicek, P. (2012). Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-

regulatory modules. Genome Res. *22*, 2163–2175.

Fazi, F., Racanicchi, S., Zardo, G., Starnes, L.M., Mancini, M., Travaglini, L., Diverio, D., Ammatuna, E., Cimino, G., Lo-Coco, F., et al. (2007). Epigenetic Silencing of the Myelopoiesis Regulator microRNA-223 by the AML1/ETO Oncoprotein. Cancer Cell *12*, 457–466.

Feinstein, P.G., Kornfeld, K., Hogness, D.S., and Mann, R.S. (1995). Identification of homeotic target genes in Drosophila melanogaster including nervy, a proto-oncogene homologue. Genetics *140*, 573–586.

Fishburn, J., Mohibullah, N., and Hahn, S. (2005). Function of a Eukaryotic Transcription Activator during the Transcription Cycle. Mol. Cell *18*, 369–378.

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., et al. (2016). The leukemia-associated RUNX1/ETO oncoprotein confers a mutator phenotype. Leukemia *30*, 251–254.

Frankenberger, M., Hofer, T.P.J., Marei, A., Dayyani, F., Schewe, S., Strasser, C., Aldraihim, A., Stanzel, F., Lang, R., Hoffmann, R., et al. (2012). Transcript profiling of CD16-positive monocytes reveals a unique molecular fingerprint. Eur. J. Immunol. *42*, 957–974.

Fu, L., Huang, W., Jing, Y., Jiang, M., Zhao, Y., Shi, J., Huang, S., Xue, X., Zhang, Q., Tang, J., et al. (2014). AML1-ETO triggers epigenetic activation of early growth response gene I, inducing apoptosis in t(8;21) acute myeloid leukemia. FEBS J. *281*, 1123–1131.

Fuchs, G., Voichek, Y., Benjamin, S., Gilad, S., Amit, I., and Oren, M. (2014). 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. Genome Biol. *15*, 1–10.

Fujita, Y., Nishimura, M., Taniwaki, M., Abe, T., and Okuda, T. (2001). Identification of an Alternatively Spliced Form of the Mouse AML1/RUNX1 Gene Transcript AML1c and Its Expression in Early Hematopoietic Development. Biochem. Biophys. Res. Commun. *281*, 1248–1255.

Fujiwara, Y., Browne, C.P., Cunniff, K., Goff, S.C., and Orkin, S.H. (1996). Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. Proc. Natl. Acad. Sci. U. S. A. *93*, 12355–12358.

Gaidzik, V., and Döhner, K. (2008). Prognostic implications of gene mutations in acute myeloid leukemia with normal cytogenetics. Semin. Oncol. *35*, 346–355.

Galy, A., Travis, M., Cen, D., and Chen, B. (1995). Human T, B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset. Immunity *3*, 459–473.

Gao, J., Erickson, P., Gardiner, K., Le Beau, M.M., Diaz, M.O., Patterson, D., Rowley, J.D., and Drabkin, H.A. (1991). Isolation of a yeast artificial chromosome spanning the 8;21 translocation breakpoint t(8;21)(q22;q22.3) in acute myelogenous leukemia. Proc. Natl. Acad. Sci. U. S. A. *88*, 4882–4886.

Gardini, A., Cesaroni, M., Luzi, L., Okumura, A.J., Biggs, J.R., Minardi, S.P., Venturini, E., Zhang, D.-E., Pelicci, P.G., and Alcalay, M. (2008). AML1/ETO Oncoprotein Is Directed to AML1 Binding Regions and Co-Localizes with AML1 and HEB on Its Targets. PLoS Genet. *4*, e1000275.

Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J.M., Wagner, E.J., Lazar, M.A., Shilatifard, A., and Shiekhattar, R. (2014). Integrator Regulates Transcriptional Initiation and Pause Release following Activation. Mol. Cell *56*, 128–139.

Gelmetti, V., Zhang, J., Fanelli, M., Minucci, S., Pelicci, P.G., and Lazar, M.A. (1998). Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO. Mol. Cell. Biol. *18*, 7185–7191.

Di Genua, C., Norfo, R., Rodriguez-Meira, A., Wen, W.X., Drissen, R., Booth, C.A.G., Povinelli, B., Repapi, E., Gray, N., Carrelha, J., et al. (2019). Cell-intrinsic depletion of Aml1-ETO-expressing pre-leukemic hematopoietic stem cells by K-Ras activating mutation. Haematologica haematol.2018.205351.

Gershenzon, N.I., and Ioshikhes, I.P. (2005). Synergy of human Pol II core promoter elements

revealed by statistical sequence analysis. Bioinformatics *21*, 1295–1300.

Gertow, K., Hirst, C.E., Yu, Q.C., Ng, E.S., Pereira, L.A., Davis, R.P., Stanley, E.G., and Elefanty, A.G. (2013). WNT3A Promotes Hematopoietic or Mesenchymal Differentiation from hESCs Depending on the Time of Exposure. Stem Cell Reports *1*, 53–65.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells. Cell *153*, 1149–1163.

Gilchrist, D.A., Santos, G. Dos, Fargo, D.C., Xie, B., Gao, Y., Li, L., and Adelman, K. (2010). Pausing of RNA Polymerase II Disrupts DNA-Specified Nucleosome Organization to Enable Precise Gene Regulation. Cell *143*, 540–551.

Gilliland, D.G., and Tallman, M.S. (2002). Focus on acute leukemias. Cancer Cell *1*, 417–420.

Gilmour, J., Assi, S. a, Jaegle, U., Kulu, D., van de Werken, H., Clarke, D., Westhead, D.R., Philipsen, S., and Bonifer, C. (2014). A crucial role for the ubiquitously expressed transcription factor Sp1 at early stages of hematopoietic specification. Development *141*, 2391–2401.

Goardon, N., Marchi, E., Atzberger, A., Quek, L., Schuh, A., Soneji, S., Woll, P., Mead, A., Alford, K.A., Rout, R., et al. (2011). Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. Cancer Cell *19*, 138–152.

Godin, I., Dieterlen-Lièvre, F., and Cumano, a (1995). Emergence of multipotent hemopoietic cells in the yolk sac and paraaortic splanchnopleura in mouse embryos, beginning at 8.5 days postcoitus. Proc. Natl. Acad. Sci. U. S. A. *92*, 773–777.

Gomes, A.M., Kurochkin, I., Chang, B., Daniel, M., Law, K., Satija, N., Lachmann, A., Wang, Z., Ferreira, L., Ma'ayan, A., et al. (2018). Cooperative Transcription Factor Induction Mediates Hemogenic Reprogramming. Cell Rep. *25*, 2821–2835.e7.

Goode, D.K., Obier, N., Vijayabaskar, M.S., Lie-A-Ling, M., Lilly, A.J., Hannah, R., Lichtinger, M., Batta, K., Florkowska, M., Patel, R., et al. (2016). Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. Dev. Cell.

Görgens, A., Radtke, S., Möllmann, M., Cross, M., Dürig, J., Horn, P.A., and Giebel, B. (2013a). Revision of the human hematopoietic tree: granulocyte subtypes derive from distinct hematopoietic lineages. Cell Rep. *3*, 1539–1552.

Görgens, A., Radtke, S., Horn, P.A., and Giebel, B. (2013b). New relationships of human hematopoietic lineages facilitate detection of multipotent hematopoietic stem and progenitor cells. Cell Cycle *12*, 3478–3482.

Goyama, S., Schibler, J., Cunningham, L., Zhang, Y., Rao, Y., Nishimoto, N., Nakagawa, M., Olsson, A., Wunderlich, M., Link, K.A., et al. (2013). Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells. J. Clin. Invest. *123*, 3876–3888.

Goyama, S., Huang, G., Kurokawa, M., and Mulloy, J.C. (2015). Posttranslational modifications of RUNX1 as potential anticancer targets. Oncogene *34*, 3483–3492.

Goyama, S., Schibler, J., Gasilina, A., Shrestha, M., Lin, S., Link, K.A., Chen, J., Whitman, S.P., Bloomfield, C.D., Nicolet, D., et al. (2016). UBASH3B/Sts-1-CBL axis regulates myeloid proliferation in human preleukemia induced by AML1-ETO. Leukemia *30*, 728–739.

Grass, J.A., Boyer, M.E., Pal, S., Wu, J., Weiss, M.J., and Bresnick, E.H. (2003). GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. Proc. Natl. Acad. Sci. *100*, 8811–8816.

Gratzner, H.G., and Leif, R.C. (1981). An immunofluorescence method for monitoring DNA synthesis by flow cytometry. Cytometry *1*, 385–389.

Grimwade, D., Hills, R.K., Moorman, a V, Walker, H., Chatters, S., Goldstone, a H., Wheatley, K., Harrison, C.J., and Burnett, a K. (2010). Refinement of cytogenetic classification in acute myeloid leukaemia: Determination of prognostic significance of rarer recurring chromosomal abnormalities amongst 5,876 younger adult patients treated in the UK Medical Research Council trials. Br. J.

Haematol. *Conference*, 17.

Growney, J.D., Shigematsu, H., Li, Z., Lee, B.H., Adelsperger, J., Rowan, R., Curley, D.P., Kutok, J.L., Akashi, K., Williams, I.R., et al. (2005). Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. Blood *106*, 494–505.

Grünberg, S., and Hahn, S. (2013). Structural insights into transcription initiation by RNA polymerase II. Trends Biochem. Sci. *38*, 603–611.

Guan, D., Zhang, W., Zhang, W., Liu, G., and Belmonte, J.C.I. (2013). Switching cell fate, ncRNAs coming to play. Cell Death Dis. *4*, e464-6.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. Cell *130*, 77–88.

Guo, C., Hu, Q., Yan, C., and Zhang, J. (2009). Multivalent binding of the ETO corepressor to E proteins facilitates dual repression controls targeting chromatin and the basal transcription machinery. Mol. Cell. Biol. *29*, 2644–2657.

de Guzman, C.G., Warren, A.J., Zhang, Z., Gartland, L., Erickson, P., Drabkin, H., Hiebert, S.W., and Klug, C.A. (2002). Hematopoietic stem cell expansion and distinct myeloid developmental abnormalities in a murine model of the AML1-ETO translocation. Mol. Cell. Biol. *22*, 5506–5517.

Haddad, R., Guimiot, F., Six, E., Jourquin, F., Setterblad, N., Kahn, E., Yagello, M., Schiffer, C., Andre-Schmutz, I., Cavazzana-Calvo, M., et al. (2006). Dynamics of thymus-colonizing cells during human development. Immunity *24*, 217–230.

Haferlach, C., Dicker, F., Herholz, H., Schnittger, S., Kern, W., and Haferlach, T. (2008). Mutations of the TP53 gene in acute myeloid leukemia are strongly associated with a complex aberrant karyotype. Leukemia *22*, 1539–1541.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nat. Struct. Mol. Biol. *11*, 394–403.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W.H., Ye, C., Ping, J.L.H., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nat. Genet. *43*, 630–638.

Hao, Q.L., Zhu, J., Price, M.A., Payne, K.J., Barsky, L.W., and Crooks, G.M. (2001). Identification of a novel, human multilymphoid progenitor in cord blood. Blood *97*, 3683–3690.

Harris, M.B., Mostecki, J., and Rothman, P.B. (2005). Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. J. Biol. Chem. *280*, 13114–13121.

Hatlen, M. a., Wang, L., and Nimer, S.D. (2012). AML1-ETO driven acute leukemia: insights into pathogenesis and potential therapeutic approaches. Front. Med. *6*, 248–262.

He, N., Chan, C.K., Sobhian, B., Chou, S., Xue, Y., Liu, M., and Alber, T. (2011a). Human Polymerase-Associated Factor complex (PAFc) connects the Super Elongation Complex (SEC) to RNA polymerase II on chromatin. Proc. Natl. Acad. Sci. U. S. A. *108*, E636-45.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011b). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303–1307.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.

Henikoff, S., and Smith, M.M. (2015). Histone Variants and Epigenetics. Cold Spring Harb. Perspect. Biol. *7*, a019364.

Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. Genes Dev. *32*, 26–41.

Heo, K., Kim, H., Choi, S.H., Choi, J., Kim, K., Gu, J., Lieber, M.R., Yang, A.S., and An, W. (2008). FACT-Mediated Exchange of Histone Variant H2AX Regulated by Phosphorylation of H2AX and ADP-Ribosylation of Spt16. Mol. Cell *30*, 86–97.

Higuchi, M., O'Brien, D., Kumaravelu, P., Lenny, N., Yeoh, E.J., and 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.

Hildebrand, D., Tiefenbach, J., Heinzel, T., Grez, M., and Maurer, A.B. (2001). Multiple Regions of ETO Cooperate in Transcriptional Repression. J. Biol. Chem. *276*, 9889–9895.

Hodge, D., Coghill, E., Keys, J., Maguire, T., Hartmann, B., McDowall, A., Weiss, M., Grimmond, S., and Perkins, A. (2006). A global role for EKLF in definitive and primitive erythropoiesis. Blood *107*, 3359–3370.

Hoebeke, I., De Smedt, M., Stolz, F., Pike-Overzet, K., Staal, F.J.T., Plum, J., and Leclercq, G. (2007). T-, B- and NK-lymphoid, but not myeloid cells arise from human CD34(+)CD38(-)CD7(+) common lymphoid progenitors expressing lymphoid-specific genes. Leukemia *21*, 311–319.

Hoogenkamp, M., Krysinska, H., Ingram, R., Huang, G., Barlow, R., Clarke, D., Ebralidze, A., Zhang, P., Tagoh, H., Cockerill, P.N., et al. (2007). The Pu.1 locus is differentially regulated at the level of chromatin structure and noncoding transcription by alternate mechanisms at distinct developmental stages of hematopoiesis. Mol. Cell. Biol. *27*, 7425–7438.

Hoogenkamp, M., Lichtinger, M., Krysinska, H., Lancrin, C., Clarke, D., Williamson, A., Mazzarella, L., Ingram, R., Jorgensen, H., Fisher, A., et al. (2009). Early chromatin unfolding by RUNX1: A molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program. Blood *114*, 299–309.

Hotchkiss, R.D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J. Biol. Chem. *175*, 315–332.

Hsieh, C.L. (1999). In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b. Mol. Cell. Biol. *19*, 8211–8218.

Huang, G., Shigesada, K., Ito, K., Wee, H., Yokomizo, T., and Ito, Y. (2001). Dimerization with PEBP2beta protects RUNX1/AML1 from ubiquitin-proteasome-mediated degradation. EMBO J. *20*, 723–733.

Huang, G., Zhang, P., Hirai, H., Elf, S., Yan, X., Chen, Z., Koschmieder, S., Okuno, Y., Dayaram, T., Growney, J.D., et al. (2008). PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. Nat. Genet. *40*, 51–60.

Ichikawa, M., Asai, T., Saito, T., Seo, S., Yamazaki, I., Yamagata, T., Mitani, K., Chiba, S., Ogawa, S., Kurokawa, M., et al. (2004). AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. Nat. Med. *10*, 299–304.

Ichikawa, M., Goyama, S., Asai, T., Kawazu, M., Nakagawa, M., Takeshita, M., Chiba, S., Ogawa, S., and Kurokawa, M. (2008). AML1/Runx1 negatively regulates quiescent hematopoietic stem cells in adult hematopoiesis. J. Immunol. *180*, 4402–4408.

Ikonomi, P., Rivera, C.E., Riordan, M., Washington, G., Schechter, A.N., and Noguchi, C.T. (2000). Overexpression of GATA-2 inhibits erythroid and promotes megakaryocyte differentiation. Exp. Hematol. *28*, 1423–1431.

Imai, Y., Kurokawa, M., Tanaka, K., Friedman, A.D., Ogawa, S., Mitani, K., Yazaki, Y., and Hirai, H. (1998). TLE, the human homolog of groucho, interacts with AML1 and acts as a repressor of AML1-induced transactivation. Biochem. Biophys. Res. Commun. *252*, 582–589.

Imai, Y., Kurokawa, M., Yamaguchi, Y., Izutsu, K., Nitta, E., Mitani, K., Satake, M., Noda, T., Ito, Y., and Hirai, H. (2004). The Corepressor mSin3A Regulates Phosphorylation-Induced Activation, Intranuclear Location, and Stability of AML1. Mol. Cell. Biol. *24*, 1033–1043.

Ivanovs, A., Rybtsov, S., Welch, L., Anderson, R.A., Turner, M.L., and Medvinsky, A. (2011). Highly

potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. J. Exp. Med. *208*, 2417–2427.

Ivanovs, A., Rybtsov, S., Anderson, R.A., Turner, M.L., and Medvinsky, A. (2014). Identification of the Niche and Phenotype of the First Human Hematopoietic Stem Cells. Stem Cell Reports *2*, 449–456.

Ivanovs, A., Rybtsov, S., Ng, E.S., Stanley, E.G., Elefanty, A.G., and Medvinsky, A. (2017). Human haematopoietic stem cell development: from the embryo to the dish. Development *144*, 2323–2337.

Iwamoto, M., Björklund, T., Lundberg, C., Kirik, D., and Wandless, T.J. (2010). A general chemical method to regulate protein stability in the mammalian central nervous system. Chem. Biol. *17*, 981–988.

Jaffredo, T., Gautier, R., Eichmann, A., and Dieterlen-Lièvre, F. (1998). Intraaortic Hemopoietic Cells are Derived from Endothelial Cells During Ontogeny. Development *125*, 4575–4583.

Jaffredo, T., Bollerot, K., Sugiyama, D., Gautier, R., and Drevon, C. (2005). Tracing the hemangioblast during embryogenesis: Developmental relationships between endothelial and hematopoietic cells. Int. J. Dev. Biol. *49*, 269–277.

Ji, H., Ehrlich, L.I.R., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K., et al. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature *467*, 338–342.

Jiang, W., Shi, Y., Zhao, D., Chen, S., Yong, J., Zhang, J., Qing, T., Sun, X., Zhang, P., Ding, M., et al. (2007). In vitro derivation of functional insulin-producing cells from human embryonic stem cells. Cell Res. *17*, 333–344.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 11–13.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife *3*, 1–25.

Kania, M.A., Bonner, A.S., Duffy, J.B., and Gergen, J.P. (1990). The Drosophila segmentation gene runt encodes a novel nuclear regulatory protein that is also expressed in the developing nervous system. Genes Dev. *4*, 1701–1713.

Kanno, T., Kanno, Y., Chen, L.F., Ogawa, E., Kim, W.Y., and Ito, Y. (1998). Intrinsic transcriptional activation-inhibition domains of the polyomavirus enhancer binding protein 2/core binding factor alpha subunit revealed in the presence of the beta subunit. Mol. Cell. Biol. *18*, 2444–2454.

Kao, T., Labonne, T., Niclis, J.C., Chaurasia, R., Lokmic, Z., Qian, E., Bruveris, F.F., Howden, S.E., Motazedian, A., Schiesser, J. V., et al. (2016). GAPTrap: A Simple Expression System for Pluripotent Stem Cells and Their Derivatives. Stem Cell Reports *7*, 518–526.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, 493D–496.

Kaufman, D.S., Hanson, E.T., Lewis, R.L., Auerbach, R., and Thomson, J.A. (2001). Hematopoietic colony-forming cells derived from human embryonic stem cells. Proc. Natl. Acad. Sci. U. S. A. *98*, 10716–10721.

Kauts, M.-L., Vink, C.S., and Dzierzak, E. (2016). Hematopoietic (stem) cell development - how divergent are the roads taken? FEBS Lett. *590*, 3975–3986.

Kawamoto, H., Ohmura, K., Fujimoto, S., and Katsura, Y. (1999). Emergence of T cell progenitors without B cell or myeloid differentiation potential at the earliest stage of hematopoiesis in the murine fetal liver. J. Immunol. *162*, 2725–2731.

Kawamoto, H., Ikawa, T., Masuda, K., Wada, H., and Katsura, Y. (2010). A map for lineage restriction of progenitors during hematopoiesis: the essence of the myeloid-based model. Immunol. Rev. *238*, 23–36.

Kayser, S., Schlenk, R.F., Londono, M.C., Breitenbuecher, F., Wittke, K., Du, J., Groner, S., Späth, D., Krauter, J., Ganser, A., et al. (2009). Insertion of FLT3 internal tandem duplication in the tyrosine

kinase domain-1 is associated with resistance to chemotherapy and inferior outcome. Blood *114*, 2386–2392.

Keller, G.M. (1995). In vitro differentiation of embryonic stem cells. Curr. Opin. Cell Biol. *7*, 862–869.

Keller, G., Kennedy, M., Papayannopoulou, T., and Wiles, M. V (1993). Hematopoietic commitment during embryonic stem cell differentiation in culture. Mol. Cell. Biol. *13*, 473–486.

Kelly, L.M., and Gilliland, D.G. (2002). Genetics of myeloid leukemias. Annu. Rev. Genomics Hum. Genet. *3*, 179–198.

Kelly, L.M., Liu, Q., Kutok, J.L., Williams, I.R., Boulton, C.L., and Gilliland, D.G. (2002). FLT3 internal tandem duplication mutations associated with human acute myeloid leukemias induce myeloproliferative disease in a murine bone marrow transplant model. Blood *99*, 310–318.

Kennedy, M., Awong, G., Sturgeon, C.M., Ditadi, A., LaMotte-Mohs, R., Zúñiga-Pflücker, J.C., and Keller, G. (2012). T Lymphocyte Potential Marks the Emergence of Definitive Hematopoietic Progenitors in Human Pluripotent Stem Cell Differentiation Cultures. Cell Rep. *2*, 1722–1735.

Kim, A.D., Stachura, D.L., and Traver, D. (2014a). Cell signaling pathways involved in hematopoietic stem cell specification. Exp. Cell Res. *329*, 227–233.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Kim, I., Saunders, T.L., and Morrison, S.J. (2007). Sox17 Dependence Distinguishes the Transcriptional Regulation of Fetal from Adult Hematopoietic Stem Cells. Cell *130*, 470–483.

Kim, J., Guermah, M., and Roeder, R.G. (2010). The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS. Cell *140*, 491–503.

Kim, W., Klarmann, K.D., Keller, J.R., Yu, C., Smith, K.A., Mueller, B.U., Narravula, S., Torbett, B.E., Orkin, S.H., and Tenen, D.G. (2014b). Gfi-1 regulates the erythroid transcription factor network through Id2 repression in murine hematopoietic progenitor cells. Blood *124*, 1586–1596.

Kimelman, D. (2006). Mesoderm induction: from caps to chips. Nat. Rev. Genet. *7*, 360–372.

Kimura, H., and Shiota, K. (2003). Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. J. Biol. Chem. *278*, 4806–4812.

Kissa, K., and Herbomel, P. (2010). Blood stem cells emerge from aortic endothelium by a novel type of cell transition. Nature *464*, 112–115.

Kitabayashi, I., Yokoyama, A., Shimizu, K., and Ohki, M. (1998). Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. EMBO J. *17*, 2994–3004.

Klampfer, L., Zhang, J., Zelenetz, A.O., Uchida, H., and Nimer, S.D. (1996). The AML1/ETO fusion protein activates transcription of BCL-2. Proc. Natl. Acad. Sci. U. S. A. *93*, 14059–14064.

Koenderman, L., Buurman, W., and Daha, M.R. (2014). The innate immune response. Immunol. Lett. *162*, 95–102.

Kondo, M., Weissman, I.L., and Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell *91*, 661–672.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868–871.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. Cell *98*, 285–294.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Kozu, T., Fukuyama, T., Yamami, T., Akagi, K., and Kaneko, Y. (2005). MYND-less splice variants of AML1-MTG8 (RUNX1-CBFA2T1) are expressed in leukemia with t(8;21). Genes, Chromosom. Cancer *43*, 45–53.

Krämer, O.H., Müller, S., Buchwald, M., Reichardt, S., and Heinzel, T. (2008). Mechanism for ubiquitylation of the leukemia fusion proteins AML1-ETO and PML-RARα. FASEB J. *22*, 1369–1379.

Krauth, M.-T., Eder, C., Alpermann, T., Bacher, U., Nadarajah, N., Kern, W., Haferlach, C., Haferlach, T., and Schnittger, S. (2014). High number of additional genetic lesions in acute myeloid leukemia with t(8;21)/RUNX1-RUNX1T1: frequency and impact on clinical outcome. Leukemia *28*, 1449–1458.

Krejci, O., Wunderlich, M., Geiger, H., Chou, F.S., Schleimer, D., Jansen, M., Andreassen, P.R., and Mulloy, J.C. (2008). P53 signaling in response to increased DNA damage sensitizes AML1-ETO cells to stress-induced death. Blood *111*, 2190–2199.

Krüger, I., Vollmer, M., Simmons, D., Elsässer, H.P., Philipsen, S., and Suske, G. (2007). Sp1/Sp3 compound heterozygous mice are not viable: Impaired erythropoiesis and severe placental defects. Dev. Dyn. *236*, 2235–2244.

Kruse, E. a, Loughran, S.J., Baldwin, T.M., Josefsson, E.C., Ellis, S., Watson, D.K., Nurden, P., Metcalf, D., Hilton, D.J., Alexander, W.S., et al. (2009). Dual requirement for the ETS transcription factors Fli-1 and Erg in hematopoietic stem cells and the megakaryocyte lineage. Proc. Natl. Acad. Sci. U. S. A. *106*, 13814–13819.

Kuchenbauer, F., Feuring-Buske, M., and Buske, C. (2005). AML1-ETO needs a partner: new insights into the pathogenesis of t(8;21) leukemia. Cell Cycle *4*, 1716–1718.

Kuchenbauer, F., Schnittger, S., Look, T., Gilliland, G., Tenen, D., Haferlach, T., Hiddemann, W., Buske, C., and Schoch, C. (2006). Identification of additional cytogenetic and molecular genetic abnormalities in acute myeloid leukaemia with t(8;21)/AML1-ETO. Br. J. Haematol. *134*, 616–619.

Kurokawa, M., Tanaka, T., Tanaka, K., Ogawa, S., Mitani, K., Yazaki, Y., and Hirai, H. (1996). Overexpression of the AML1 proto-oncoprotein in NIH3T3 cells leads to neoplastic transformation depending on the DNA-binding and transactivational potencies. Oncogene *12*, 883–892.

Kurotaki, D., Osato, N., Nishiyama, A., Yamamoto, M., Ban, T., Sato, H., Nakabayashi, J., Umehara, M., Miyake, N., Matsumoto, N., et al. (2013). Essential role of the IRF8-KLF4 transcription factor cascade in murine monocyte differentiation. Blood *121*, 1839–1849.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. Science (80-.). *339*, 950–953.

Kwok, C., Zeisig, B.B., Qiu, J., Dong, S., and So, C.W.E. (2009). Transforming activity of AML1-ETO is independent of CBFbeta and ETO interaction but requires formation of homo-oligomeric complexes. Proc. Natl. Acad. Sci. U. S. A. *106*, 2853–2858.

de la Serna, I.L., Ohkawa, Y., Berkes, C.A., Bergstrom, D.A., Dacwag, C.S., Tapscott, S.J., and Imbalzano, A.N. (2005). MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex. Mol. Cell. Biol. *25*, 3997–4009.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Lam, K., and Zhang, D.-E. (2012). RUNX1 and RUNX1-ETO: roles in hematopoiesis and leukemogenesis. Front. Biosci. (Landmark Ed. *17*, 1120–1139.

Lam, E.Y.N., Hall, C.J., Crosier, P.S., Crosier, K.E., and Flores, M.V. (2010). Live imaging of Runx1 expression in the dorsal aorta tracks the emergence of blood progenitors from endothelial cells. Blood *116*, 909–914.

Lamandé, S.R., Yuan, Y., Gresshoff, I.L., Rowley, L., Belluoccio, D., Kaluarachchi, K., Little, C.B., Botzenhart, E., Zerres, K., Amor, D.J., et al. (2011). Mutations in TRPV4 cause an inherited arthropathy of hands and feet. Nat. Genet. *43*, 1142–1146.

Lamartina, S., Sporeno, E., Fattori, E., and Toniatti, C. (2000). Characteristics of the adeno-associated virus preintegration site in human chromosome 19: open chromatin conformation and transcription-competent environment. J. Virol. *74*, 7671–7677.

Lancrin, C., Sroczynska, P., Stephenson, C., Allen, T., Kouskoff, V., and Lacaud, G. (2009). The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. Nature

457, 892–895.

Lancrin, C., Mazan, M., Stefanska, M., Patel, R., Lichtinger, M., Costa, G., Vargel, Ö., Wilson, N.K., Möröy, T., Bonifer, C., et al. (2012). GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. Blood *120*, 314–322.

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics *24*, 719–720.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M.A., and Dick, J.E. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature *367*, 645–648.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. Genomics *13*, 1095–1107.

Lausen, J., Cho, S., Liu, S., and Werner, M.H. (2004). The nuclear receptor co-repressor (N-CoR) utilizes repression domains I and III for interaction and co-repression with ETO. J. Biol. Chem. *279*, 49281–49288.

Lawson, N.D., Scheer, N., Pham, V.N., Kim, C.H., Chitnis, A.B., Campos-Ortega, J.A., and Weinstein, B.M. (2001). Notch signaling is required for arterial-venous differentiation during embryonic vascular development. Development *128*, 3675–3683.

Lebert-Ghali, C.-E., Fournier, M., Dickson, G.J., Thompson, A., Sauvageau, G., and Bijl, J.J. (2010). HoxA cluster is haploinsufficient for activity of hematopoietic stem and progenitor cells. Exp. Hematol. *38*, 1074–1086.e5.

Lechman, E.R., Gentner, B., van Galen, P., Giustacchini, A., Saini, M., Boccalatte, F.E., Hiramatsu, H., Restuccia, U., Bachi, A., Voisin, V., et al. (2012). Attenuation of miR-126 Activity Expands HSC In Vivo without Exhaustion. Cell Stem Cell *11*, 799–811.

Lechman, E.R., Gentner, B., Ng, S.W.K., Schoof, E.M., van Galen, P., Kennedy, J.A., Nucera, S., Ciceri, F., Kaufmann, K.B., Takayama, N., et al. (2016). miR-126 Regulates Distinct Self-Renewal Outcomes in Normal and Malignant Hematopoietic Stem Cells. Cancer Cell *29*, 214–228.

Leddin, M., Perrod, C., Hoogenkamp, M., Ghani, S., Assi, S., Heinz, S., Wilson, N.K., Follows, G., Schönheit, J., Vockentanz, L., et al. (2011). Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells. Blood *117*, 2827–2838.

Lee, T.I., and Young, R.A. (2000). Transcription of Eukaryotic Protein-Coding Genes. Annu. Rev. Genet. *34*, 77–137.

Lee, D.Y., Hayes, J.J., Pruss, D., and Wolffe, A.P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell *72*, 73–84.

van Leeuwen, F., Gafken, P.R., and Gottschling, D.E. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell *109*, 745–756.

Legrier, M.-E., Bièche, I., Gaston, J., Beurdeley, A., Yvonnet, V., Déas, O., Thuleau, A., Château-Joubert, S., Servely, J.-L., Vacher, S., et al. (2016). Activation of IFN/STAT1 signalling predicts response to chemotherapy in oestrogen receptor-negative breast cancer. Br. J. Cancer *114*, 177–187.

Lekstrom-Himes, J. a (2001). The role of C/EBP(epsilon) in the terminal stages of granulocyte differentiation. Stem Cells *19*, 125–133.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat. Rev. Genet. *13*, 233–245.

Levanon, D., Goldstein, R.E., Bernstein, Y., Tang, H., Goldenberg, D., Stifani, S., Paroush, Z., and Groner, Y. (1998). Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors. Proc. Natl. Acad. Sci. U. S. A. *95*, 11590–11595.

Levantini, E., Lee, S., Radomska, H.S., Hetherington, C.J., Alberich-Jorda, M., Amabile, G., Zhang, P., Gonzalez, D. a, Zhang, J., Basseres, D.S., et al. (2011). RUNX1 regulates the CD34 gene in haematopoietic stem cells by mediating interactions with a distal regulatory element. EMBO J. *30*, 4059–4070.

Li, J., and Gilmour, D.S. (2013). Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. EMBO J. *32*, 1829–1841.

Li, Y., Gao, L., Luo, X., Wang, L., Gao, X., Wang, W., Sun, J., Dou, L., Li, J., Xu, C., et al. (2013). Epigenetic silencing of microRNA-193a contributes to leukemogenesis in t(8;21) acute myeloid leukemia by activating the PTEN/PI3K signal pathway. Blood *121*, 499–509.

Li, Y., Wang, H., Wang, X., Jin, W., Tan, Y., Fang, H., Chen, S., Chen, Z., and Wang, K. (2016). Genome-wide studies identify a novel interplay between AML1 and AML1/ETO in t(8;21) acute myeloid leukemia. Blood *127*, 233–242.

Li, Z., Chen, P., Su, R., Li, Y., Hu, C., Wang, Y., Arnovitz, S., He, M., Gurbuxani, S., Zuo, Z., et al. (2015). Overexpression and knockout of miR-126 both promote leukemogenesis. Blood *126*, 2005–2015.

Lichtinger, M., Hoogenkamp, M., Krysinska, H., Ingram, R., and Bonifer, C. (2010). Chromatin regulation by RUNX1. Blood Cells, Mol. Dis. *44*, 287–290.

Lichtinger, M., Ingram, R., Hannah, R., Müller, D., Clarke, D., Assi, S. a, Lie-A-Ling, M., Noailles, L., Vijayabaskar, M.S., Wu, M., et al. (2012). RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. EMBO J. *31*, 4318–4333.

Liddiard, K., Hills, R., Burnett, A.K., Darley, R.L., and Tonks, A. (2010). OGG1 is a novel prognostic indicator in acute myeloid leukaemia. Oncogene *29*, 2005–2012.

Lin, S., Mulloy, J.C., and Goyama, S. (2017a). RUNX1-ETO Leukemia. In RUNX Proteins in Development and Cancer, Advances in Experimental Medicine and Biology VOL 952, (Springer Nature Singapore Pte Ltd.), pp. 151–173.

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., et al. (2017b). A FOXO1-induced oncogenic network defines the AML1-ETO preleukemic program. Blood *130*, 1213–1222.

Lin, X., Taube, R., Fujinaga, K., and Peterlin, B.M. (2002). P-TEFb containing cyclin K and Cdk9 can activate transcription via RNA. J. Biol. Chem. *277*, 16873–16878.

Lindsley, R.C., Mar, B.G., Mazzola, E., Grauman, P. V, Shareef, S., Allen, S.L., Pigneux, A., Wetzler, M., Stuart, R.K., Erba, H.P., et al. (2015). Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. Blood *125*, 1367–1376.

Linggi, B., Müller-Tidow, C., van de Locht, L., Hu, M., Nip, J., Serve, H., Berdel, W.E., van der Reijden, B., Quelle, D.E., Rowley, J.D., et al. (2002). The t(8;21) fusion protein, AML1–ETO, specifically represses the transcription of the p14ARF tumor suppressor in acute myeloid leukemia. Nat. Med. *8*, 743–750.

Link, K.A., Lin, S., Shrestha, M., Bowman, M., Wunderlich, M., Bloomfield, C.D., Huang, G., and Mulloy, J.C. (2016). Supraphysiologic levels of the AML1-ETO isoform AE9a are essential for transformation. Proc. Natl. Acad. Sci. U. S. A. *113*, 9075–9080.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Liu, F., Walmsley, M., Rodaway, A., and Patient, R. (2008). Fli1 Acts at the Top of the Transcriptional Network Driving Blood and Endothelial Development. Curr. Biol. *18*, 1234–1240.

Liu, S., Shen, T., Huynh, L., Klisovic, M.I., Rush, L.J., Ford, J.L., Yu, J., Becknell, B., Li, Y., Liu, C., et al. (2005). Interplay of RUNX1/MTG8 and DNA Methyltransferase 1 in Acute Myeloid Leukemia. Cancer Res. *65*, 1277–1284.

Liu, X., Zhang, Q., Zhang, D.-E., Zhou, C., Xing, H., Tian, Z., Rao, Q., Wang, M., and Wang, J. (2009).

Overexpression of an isoform of AML1 in acute leukemia and its potential role in leukemogenesis. Leukemia *23*, 739–745.

Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H., and Wong, J. (2013). UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nat. Commun. *4*, 1563.

Liu, Y., Cheney, M.D., Gaudet, J.J., Chruszcz, M., Lukasik, S.M., Sugiyama, D., Lary, J., Cole, J., Dauter, Z., Minor, W., et al. (2006). The tetramer structure of the Nervy homology two domain, NHR2, is critical for AML1/ETO's activity. Cancer Cell *9*, 249–260.

Lo, M.-C., Peterson, L.F., Yan, M., Cong, X., Jin, F., Shia, W.-J., Matsuura, S., Ahn, E.-Y., Komeno, Y., Ly, M., et al. (2012). Combined gene expression and DNA occupancy profiling identifies potential therapeutic targets of t(8;21) AML. Blood *120*, 1473–1484.

Loke, J., Assi, S.A., Imperato, M.R., Ptasinska, A., Cauchy, P., Grabovska, Y., Soria, N.M., Raghavan, M., Delwel, H.R., Cockerill, P.N., et al. (2017). RUNX1-ETO and RUNX1-EVI1 Differentially Reprogram the Chromatin Landscape in t(8;21) and t(3;21) AML. Cell Rep. *19*, 1654–1668.

Lombardo, A., Cesana, D., Genovese, P., Di Stefano, B., Provasi, E., Colombo, D.F., Neri, M., Magnani, Z., Cantore, A., Lo Riso, P., et al. (2011). Site-specific integration and tailoring of cassette design for sustainable gene transfer. Nat. Methods *8*, 861–869.

Loughran, S.J., Kruse, E. a, Hacking, D.F., de Graaf, C. a, Hyland, C.D., Willson, T. a, Henley, K.J., Ellis, S., Voss, A.K., Metcalf, D., et al. (2008). The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. Nat. Immunol. *9*, 810–819.

Luckett, W.P. (1978). Origin and differentiation of the yolk sac and extraembryonic mesoderm in presomite human and rhesus monkey embryos. Am. J. Anat. *152*, 59–97.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251–260.

Lugthart, S., Gröschel, S., Beverloo, H.B., Kayser, S., Valk, P.J.M., van Zelderen-Bhola, S.L., Jan Ossenkoppele, G., Vellenga, E., van den Berg-de Ruiter, E., Schanz, U., et al. (2010). Clinical, molecular, and prognostic significance of WHO type inv(3)(q21q26.2)/t(3;3)(q21;q26.2) and various other 3q abnormalities in acute myeloid leukemia. J. Clin. Oncol. *28*, 3890–3898.

Lutterbach, B., Westendorf, J.J., Linggi, B., Patten, A., Moniwa, M., Davie, J.R., Huynh, K.D., Bardwell, V.J., Lavinsky, R.M., Rosenfeld, M.G., et al. (1998a). ETO, a target of t(8;21) in acute leukemia, interacts with the N-CoR and mSin3 corepressors. Mol. Cell. Biol. *18*, 7176–7184.

Lutterbach, B., Sun, D., Schuetz, J., and Hiebert, S.W. (1998b). The MYND motif is required for repression of basal transcription from the multidrug resistance 1 promoter by the t(8;21) fusion protein. Mol. Cell. Biol. *18*, 3604–3611.

Madan, V., Han, L., Hattori, N., Teoh, W.W., Mayakonda, A., Sun, Q.-Y., Ding, L.-W., Binte Mohd Nordin, H., Lim, S.L., Shyamsunder, P., et al. (2018). ASXL2 regulates hematopoiesis in mice and its deficiency promotes myeloid expansion. Haematologica haematol.2018.189928.

Maiques-Diaz, A., Chou, F.S., Wunderlich, M., Gómez-López, G., Jacinto, F. V., Rodriguez-Perales, S., Larrayoz, M.J., Calasanz, M.J., Mulloy, J.C., Cigudosa, J.C., et al. (2012). Chromatin modifications induced by the AML1-ETO fusion protein reversibly silence its genomic targets through AML1 and Sp1 binding motifs. Leukemia *26*, 1329–1337.

Mandoli, A., Singh, A.A., Prange, K.H.M., Tijchon, E., Oerlemans, M., Dirks, R., Ter Huurne, M., Wierenga, A.T.J., Janssen-Megens, E.M., Berentsen, K., et al. (2016). The Hematopoietic Transcription Factors RUNX1 and ERG Prevent AML1-ETO Oncogene Overexpression and Onset of the Apoptosis Program in t(8;21) AMLs. Cell Rep. *17*, 2087–2100.

Månsson, R., Hultquist, A., Luc, S., Yang, L., Anderson, K., Kharazi, S., Al-Hashmi, S., Liuba, K., Thorén, L., Adolfsson, J., et al. (2007). Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. Immunity *26*, 407–419.

Manz, M.G., Miyamoto, T., Akashi, K., and Weissman, I.L. (2002). Prospective isolation of human

clonogenic common myeloid progenitors. Proc. Natl. Acad. Sci. U. S. A. *99*, 11872–11877.

Marcucci, G., Maharry, K., Wu, Y.-Z., Radmacher, M.D., Mrózek, K., Margeson, D., Holland, K.B., Whitman, S.P., Becker, H., Schwind, S., et al. (2010). IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. J. Clin. Oncol. *28*, 2348–2355.

Marcucci, G., Haferlach, T., and Döhner, H. (2011). Molecular genetics of adult acute myeloid leukemia: prognostic and therapeutic implications. J. Clin. Oncol. *29*, 475–486.

Marcucci, G., Metzeler, K.H., Schwind, S., Becker, H., Maharry, K., Mrózek, K., Radmacher, M.D., Kohlschmidt, J., Nicolet, D., Whitman, S.P., et al. (2012). Age-related prognostic impact of different types of DNMT3A mutations in adults with primary cytogenetically normal acute myeloid leukemia. J. Clin. Oncol. *30*, 742–750.

Marinić, M., Aktas, T., Ruf, S., and Spitz, F. (2013). An Integrated Holo-Enhancer Unit Defines Tissue and Gene Specificity of the Fgf8 Regulatory Landscape. Dev. Cell *24*, 530–542.

Marneth, A.E., Botezatu, L., Hönes, J.M., Israël, J.C.L., Schütte, J., Vassen, L., Lams, R.F., Bergevoet, S.M., Groothuis, L., Mandoli, A., et al. (2018). GFI1 is required for RUNX1/ETO positive acute myeloid leukemia. Haematologica *103*, e395–e399.

Marshall, C.J., Moore, R.L., Thorogood, P., Brickell, P.M., Kinnon, C., and Thrasher, A.J. (1999). Detailed characterization of the human aorta-gonad-mesonephros region reveals morphological polarity resembling a hematopoietic stromal layer. Dev. Dyn. *215*, 139–147.

Marshall, C.J., Kinnon, C., and Thrasher, A.J. (2000). Polarized expression of bone morphogenetic protein-4 in the human aorta-gonad-mesonephros region. Blood *96*.

Marshall, C.J., Sinclair, J.C., Thrasher, A.J., and Kinnon, C. (2007). Bone morphogenetic protein 4 modulates c-Kit expression and differentiation potential in murine embryonic aorta-gonad-mesonephros haematopoiesis in vitro. Br. J. Haematol. *139*, 321–330.

Martens, J.H. a., Mandoli, a., Simmer, F., Wierenga, B.-J., Saeed, S., Singh, a. a., Altucci, L., Vellenga, E., and 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.

Martinez, N., Drescher, B., Riehle, H., Cullmann, C., Vornlocher, H.-P., Ganser, A., Heil, G., Nordheim, A., Krauter, J., and Heidenreich, O. (2004). The oncogenic fusion protein RUNX1-CBFA2T1 supports proliferation and inhibits senescence in t(8;21)-positive leukaemic cells. BMC Cancer *4*, 44.

Martinez-Soria, N., McKenzie, L., Draper, J., Ptasinska, A., Issa, H., Potluri, S., Blair, H.J., Pickin, A., Isa, A., Chin, P.S., et al. (2019). The Oncogenic Transcription Factor RUNX1/ETO Corrupts Cell Cycle Regulation to Drive Leukemic Transformation. Cancer Cell *35*, 705.

Martinez Soria, N., Tussiwand, R., Ziegler, P., Manz, M.G., and Heidenreich, O. (2009). Transient depletion of RUNX1/RUNX1T1 by RNA interference delays tumour formation in vivo. Leuk. Off. J. Leuk. Soc. Am. Leuk. Res. Fund, U.K *23*, 188–190.

Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional Regulatory Elements in the Human Genome. Annu. Rev. Genomics Hum. Genet. *7*, 29–59.

Mathas, S., Hinz, M., Anagnostopoulos, I., Krappmann, D., Lietz, A., Jundt, F., Bommert, K., Mechta-Grigoriou, F., Stein, H., Dörken, B., et al. (2002). Aberrantly expressed c-Jun and JunB are a hallmark of Hodgkin lymphoma cells, stimulate proliferation and synergize with NF-kappaB. EMBO J. *21*, 4104–4113.

Matsuura, S., Yan, M., Lo, M.-C., Ahn, E.-Y., Weng, S., Dangoor, D., Matin, M., Higashi, T., Feng, G.-S., and Zhang, D.-E. (2012). Negative effects of GM-CSF signaling in a murine model of t(8;21)-induced leukemia. Blood *119*, 3155–3163.

McGinnis, W., and Krumlauf, R. (1992). Homeobox genes and axial patterning. Cell *68*, 283–302.

McGrath, K.E., Frame, J.M., Fegan, K.H., Bowen, J.R., Conway, S.J., Catherman, S.C., Kingsley,

P.D., Koniski, A.D., and Palis, J. (2015). Distinct sources of hematopoietic progenitors emerge before HSCs and provide functional blood cells in the mammalian embryo. Cell Rep. *11*, 1892–1904.

McKercher, S.R., Torbett, B.E., Anderson, K.L., Henkel, G.W., Vestal, D.J., Baribault, H., Klemsz, M., Feeney, a J., Wu, G.E., Paige, C.J., et al. (1996). Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. EMBO J. *15*, 5647–5658.

Medvinsky, A., and Dzierzak, E. (1996). Definitive hematopoiesis is autonomously initiated by the AGM region. Cell *86*, 897–906.

Medvinsky, A., Rybtsov, S., and Taoudi, S. (2011). Embryonic origin of the adult hematopoietic system: advances and questions. Development *138*, 1017–1031.

Medvinsky, A.L., Samoylina, N.L., Müller, A.M., and Dzierzak, E.A. (1993). An early pre-liver intraembryonic source of CFU-S in the developing mouse. Nature *364*, 64–67.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell *58*, 499–507.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–771.

Melgar, K., Walker, M.M., Jones, L.M., Bolanos, L.C., Hueneman, K., Wunderlich, M., Jiang, J.-K., Wilson, K.M., Zhang, X., Sutter, P., et al. (2019). Overcoming adaptive therapy resistance in AML by targeting immune response pathways. Sci. Transl. Med. *11*, eaaw8828.

Merika, M., and Thanos, D. (2001). Enhanceosomes. Curr. Opin. Genet. Dev. *11*, 205–208.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic Plasticity of Chromatin Proteins in Pluripotent Embryonic Stem Cells. Dev. Cell *10*, 105–116.

Mestas, J., and Hughes, C.C.W. (2004). Of mice and not men: differences between mouse and human immunology. J. Immunol. *172*, 2731–2738.

Metzeler, K.H., Becker, H., Maharry, K., Radmacher, M.D., Kohlschmidt, J., Mrózek, K., Nicolet, D., Whitman, S.P., Wu, Y.-Z., Schwind, S., et al. (2011). ASXL1 mutations identify a high-risk subgroup of older patients with primary cytogenetically normal AML within the ELN Favorable genetic category. Blood *118*, 6920–6929.

Meyer, C., Schneider, B., Jakob, S., Strehl, S., Attarbaschi, A., Schnittger, S., Schoch, C., Jansen, M.W.J.C., van Dongen, J.J.M., den Boer, M.L., et al. (2006). The MLL recombinome of acute leukemias. Leukemia *20*, 777–784.

Meyers, S., Downing, J.R., and Hiebert, S.W. (1993). Identification of AML-1 and the (8;21) translocation protein (AML-1/ETO) as sequence-specific DNA-binding proteins: the runt homology domain is required for DNA binding and protein-protein interactions. Mol. Cell. Biol. *13*, 6336–6345.

Meyers, S., Lenny, N., and Hiebert, S.W. (1995). The t(8;21) fusion protein interferes with AML-1B-dependent transcriptional activation. Mol. Cell. Biol. *15*, 1974–1982.

Micol, J.-B., Duployez, N., Boissel, N., Petit, A., Geffroy, S., Nibourel, O., Lacombe, C., Lapillonne, H., Etancelin, P., Figeac, M., et al. (2014). Frequent ASXL2 mutations in acute myeloid leukemia patients with t(8;21)/RUNX1-RUNX1T1 chromosomal translocations. Blood *124*, 1445–1449.

Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat. Genet. *47*, 598–606.

Migliaccio, G., Migliaccio, A.R., Petti, S., Mavilio, F., Russo, G., Lazzaro, D., Testa, U., Marinucci, M., and Peschle, C. (1986). Human embryonic hemopoiesis. Kinetics of progenitors and precursors underlying the yolk sac----liver transition. J. Clin. Invest. *78*, 51–60.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and

lineage-committed cells. Nature *448*, 553–562.

Miller, M.E., Rosten, P., Lemieux, M.E., Lai, C., and Humphries, R.K. (2016). Meis1 Is Required for Adult Mouse Erythropoiesis, Megakaryopoiesis and Hematopoietic Stem Cell Expansion. PLoS One *11*.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., and Hess, J.L. (2002). MLL Targets SET Domain Methyltransferase Activity to Hox Gene Promoters. Mol. Cell *10*, 1107–1117.

Min, I.M., Pietramaggiori, G., Kim, F.S., Passegué, E., Stevenson, K.E., and Wagers, A.J. (2008). The Transcription Factor EGR1 Controls Both the Proliferation and Localization of Hematopoietic Stem Cells. Cell Stem Cell *2*, 380–391.

Minucci, S., Maccarana, M., Cioce, M., De Luca, P., Gelmetti, V., Segalla, S., Di Croce, L., Giavara, S., Matteucci, C., Gobbi, A., et al. (2000). Oligomerization of RAR and AML1 Transcription Factors as a Novel Mechanism of Oncogenic Activation. Mol. Cell *5*, 811–820.

Miyamoto, T., Weissman, I.L., and Akashi, K. (2000). AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. Proc. Natl. Acad. Sci. U. S. A. *97*, 7521–7526.

Miyoshi, H., Shimizu, K., Kozu, T., Maseki, N., Kaneko, Y., and Ohki, M. (1991). t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. Proc. Natl. Acad. Sci. U. S. A. *88*, 10431–10434.

Miyoshi, H., Kozu, T., Shimizu, K., Enomoto, K., Maseki, N., Kaneko, Y., Kamada, N., and Ohki, M. (1993). The t(8;21) translocation in acute myeloid leukemia results in production of an AML1-MTG8 fusion transcript. EMBO J. *12*, 2715–2721.

Miyoshi, H., Ohira, M., Shimizu, K., Mitani, K., Hirai, H., Imai, T., Yokoyama, K., Soceda, E., and Ohkl, M. (1995). Alternative splicing and genomic structure of the *AML1* gene involved in acute myeloid leukemia. Nucleic Acids Res. *23*, 2762–2769.

Mizzen, C.A., Yang, X.-J., Kokubo, T., Brownell, J.E., Bannister, A.J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S.L., Kouzarides, T., et al. (1996). The TAFII250 Subunit of TFIID Has Histone Acetyltransferase Activity. Cell *87*, 1261–1270.

Molloy, E.L., Adams, A., Moore, J.B., Masterson, J.C., Madrigal-Estebas, L., Mahon, B.P., and O'Dea, S. (2008). BMP4 induces an epithelial–mesenchymal transition-like response in adult airway epithelial cells. Growth Factors *26*, 12–22.

Morris, J.R., Petrov, D.A., Lee, A.M., and Wu, C.-T. (2004). Enhancer choice in cis and in trans in Drosophila melanogaster: role of the promoter. Genetics *167*, 1739–1747.

Morrison, S., Wandycz, A., Hemmati, H., Wright, D., and Weissman, I. (1997). Identification of a lineage of multipotent hematopoietic progenitors. Development *124*, 1929–1939.

Mrózek, K., Marcucci, G., Paschka, P., Whitman, S.P., and Bloomfield, C.D. (2007). Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: are we ready for a prognostically prioritized molecular classification? Blood *109*, 431–448.

Mucenski, M.L., McLain, K., Kier, a B., Swerdlow, S.H., Schreiner, C.M., Miller, T. a, Pietryga, D.W., Scott, W.J., and Potter, S.S. (1991). A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. Cell *65*, 677–689.

Müller, A.M., Medvinsky, A., Strouboulis, J., Grosveld, F., and Dzierzakt, E. (1994). Development of hematopoietic stem cell activity in the mouse embryo. Immunity *1*, 291–301.

Müller-Tidow, C., Steffen, B., Cauvet, T., Tickenbrock, L., Ji, P., Diederichs, S., Sargin, B., Köhler, G., Stelljes, M., Puccetti, E., et al. (2004). Translocation products in acute myeloid leukemia activate the Wnt signaling pathway in hematopoietic cells. Mol. Cell. Biol. *24*, 2890–2904.

Mulloy, J.C., Cammenga, J., MacKenzie, K.L., Berguido, F.J., Moore, M.A., and Nimer, S.D. (2002). The AML1-ETO fusion protein promotes the expansion of human hematopoietic stem cells. Blood *99*, 15–23.

Mulloy, J.C., Cammenga, J., Berguido, F.J., Wu, K., Zhou, P., Comenzo, R.L., Jhanwar, S., Moore, M. a. S., and Nimer, S.D. (2003). Maintaining the self-renewal and differentiation potential of human CD34+ hematopoietic cells using a single genetic element. Blood *102*, 4369.

Mulloy, J.C., Jankovic, V., Wunderlich, M., Delwel, R., Cammenga, J., Krejci, O., Zhao, H., Valk, P.J.M., Lowenberg, B., and Nimer, S.D. (2005). AML1-ETO fusion protein up-regulates TRKA mRNA expression in human CD34+ cells, allowing nerve growth factor-induced expansion. Proc. Natl. Acad. Sci. *102*, 4016–4021.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat. Genet. *39*, 1507–1511.

Nakagawa, M., Shimabe, M., Watanabe-Okochi, N., Arai, S., Yoshimi, A., Shinohara, A., Nishimoto, N., Kataoka, K., Sato, T., Kumano, K., et al. (2011). AML1/RUNX1 functions as a cytoplasmic attenuator of NF-κB signaling in the repression of myeloid tumors. Blood *118*, 6626–6637.

Nakajima-Takagi, Y., Osawa, M., Oshima, M., Takagi, H., Miyagi, S., Endoh, M., Endo, T. a, Takayama, N., Eto, K., Toyoda, T., et al. (2013). Role of SOX17 in hematopoietic development from human embryonic stem cells. Blood *121*, 447–458.

Nan, X., Ng, H.-H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386–389.

Navarro-Montero, O., Ayllon, V., Lamolda, M., López-Onieva, L., Montes, R., Bueno, C., Ng, E., Guerrero-Carreno, X., Romero, T., Romero-Moya, D., et al. (2017). *RUNX1c* Regulates Hematopoietic Differentiation of Human Pluripotent Stem Cells Possibly in Cooperation with Proinflammatory Signaling. Stem Cells *35*, 2253–2266.

Ng, A.P., Loughran, S.J., Metcalf, D., Hyland, C.D., de Graaf, C.A., Hu, Y., Smyth, G.K., Hilton, D.J., Kile, B.T., and Alexander, W.S. (2011). Erg is required for self-renewal of hematopoietic stem cells during stress hematopoiesis in mice. Blood *118*, 2454–2461.

Ng, C.E.L., Yokomizo, T., Yamashita, N., Cirovic, B., Jin, H., Wen, Z., Ito, Y., and Osato, M. (2010). A Runx1 Intronic Enhancer Marks Hemogenic Endothelial Cells and Hematopoietic Stem Cells. Stem Cells *28*, 1869–1881.

Ng, E.S., Davis, R.P., Azzola, L., Stanley, E.G., and Elefanty, A.G. (2005). Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust, reproducible hematopoietic differentiation. Blood *106*, 1601–1603.

Ng, E.S., Davis, R., Stanley, E.G., and Elefanty, A.G. (2008). A protocol describing the use of a recombinant protein-based, animal product-free medium (APEL) for human embryonic stem cell differentiation as spin embryoid bodies. Nat. Protoc. *3*, 768–776.

Ng, E.S., Azzola, L., Bruveris, F.F., Calvanese, V., Phipson, B., Vlahos, K., Hirst, C., Jokubaitis, V.J., Yu, Q.C., Maksimovic, J., et al. (2016). Differentiation of human embryonic stem cells to HOXA+ hemogenic vasculature that resembles the aorta-gonad-mesonephros. Nat. Biotechnol. *34*, 1168–1179.

Nick, H.J., Kim, H.-G., Chang, C.-W., Harris, K.W., Reddy, V., and Klug, C.A. (2012). Distinct classes of c-Kit-activating mutations differ in their ability to promote RUNX1-ETO-associated acute myeloid leukemia. Blood *119*, 1522–1531.

Nisson, P.E., Watkins, P.C., and Sacchi, N. (1992). Transcriptionally active chimeric gene derived from the fusion of the AML1 gene and a novel gene on chromosome 8 in t(8;21) leukemic cells. Cancer Genet. Cytogenet. *63*, 81–88.

Nobuhisa, I., Osawa, M., Uemura, M., Kishikawa, Y., Anani, M., Harada, K., Takagi, H., Saito, K., Kanai-Azuma, M., Kanai, Y., et al. (2014). Sox17-mediated maintenance of fetal intra-aortic hematopoietic cell clusters. Mol. Cell. Biol. *34*, 1976–1990.

North, T., Gu, T.L., Stacy, T., Wang, Q., Howard, L., Binder, M., Marín-Padilla, M., and Speck, N. a (1999). Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. Development *126*,

2563–2575.

North, T.E., De Bruijn, M.F.T.R., Stacy, T., Talebian, L., Lind, E., Robin, C., Binder, M., Dzierzak, E., and Speck, N. a. (2002). Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. Immunity *16*, 661–672.

North, T.E., Stacy, T., Matheny, C.J., Speck, N.A., and De Bruijn, M.F.T.R. (2004). Runx1 Is Expressed in Adult Mouse Hematopoietic Stem Cells and Differentiating Myeloid and Lymphoid Cells, But Not in Maturing Erythroid Cells. Stem Cells *22*, 158–168.

Nostro, M.C., Cheng, X., Keller, G.M., and Gadue, P. (2008). Wnt, Activin, and BMP Signaling Regulate Distinct Stages in the Developmental Pathway from Embryonic Stem Cells to Blood. Cell Stem Cell *2*, 60–71.

Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O.I., Wilson, G., Kaufmann, K.B., McLeod, J., Laurenti, E., Dunant, C.F., et al. (2015). Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science aab2116.

Nottingham, W.T., Jarratt, A., Burgess, M., Speck, C.L., Cheng, J., Prabhakar, S., Rubin, E.M., Li, P., Sloane-stanley, J., Kong-a-san, J., et al. (2007). Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata / Ets / SCL-regulated enhancer. Regulation *110*, 4188–4197.

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., et al. (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., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745.

Oberlin, E., Tavian, M., Blazsek, I., and Péault, B. (2002). Blood-forming potential of vascular endothelium in the human embryo. Development *129*, 4147–4157.

Obier, N., Cauchy, P., Assi, S.A., Gilmour, J., Lie-A-Ling, M., Lichtinger, M., Hoogenkamp, M., Noailles, L., Cockerill, P.N., Lacaud, G., et al. (2016). Cooperative binding of AP-1 and TEAD4 modulates the balance between vascular smooth muscle and hemogenic cell fate. Development *143*, 4324–4340.

Odaka, Y., Mally, A., Elliott, L.T., and Meyers, S. (2000). Nuclear import and subnuclear localization of the proto-oncoprotein ETO (MTG8). Oncogene *19*, 3584–3597.

Ogata, T., Kozuka, T., and Kanda, T. (2003). Identification of an insulator in AAVS1, a preferred region for integration of adeno-associated virus DNA. J. Virol. *77*, 9000–9007.

Ogawa, M., Nishikawa, S., Yoshinaga, K., Hayashi, S., Kunisada, T., Nakao, J., Kina, T., Sudo, T., Kodama, H., and Nishikawa, S. (1993). Expression and function of c-Kit in fetal hemopoietic progenitor cells: transition from the early c-Kit-independent to the late c-Kit-dependent wave of hemopoiesis in the murine embryo. Development *117*, 1089–1098.

Ogryzko, V. V, Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The Transcriptional Coactivators p300 and CBP Are Histone Acetyltransferases. Cell *87*, 953–959.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell *99*, 247–257.

Okuda, T., Van Deursen, J., Hiebert, S.W., Grosveld, G., and 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.

Okuda, T., Cai, Z., Yang, S., Lenny, N., Lyu, C.J., van Deursen, J.M., Harada, H., and 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*, 3134–3143.

Okumura, A.J., Peterson, L.F., Okumura, F., Boyapati, A., and Zhang, D.-E. (2008). t(8;21)(q22;q22) Fusion proteins preferentially bind to duplicated AML1/RUNX1 DNA-binding sequences to differentially

regulate gene expression. Blood *112*, 1392–1401.

Org, T., Duan, D., Ferrari, R., Montel-hagen, A., Handel, B. Van, Marc, A., Sasidharan, R., Rubbi, L., Fujiwara, Y., Pellegrini, M., et al. (2015). Scl binds to primed enhancers in mesoderm to regulate hematopoietic and cardiac fate divergence. EMBO J. *34*, 759–777.

Orlic, D., Anderson, S., Biesecker, L.G., Sorrentino, B.P., and Bodine, D.M. (1995). Pluripotent hematopoietic stem cells contain high levels of mRNA for c-kit, GATA-2, p45 NF-E2, and c-myb and low levels or no mRNA for c-fms and the receptors for granulocyte colony-stimulating factor and interleukins 5 and 7. Proc. Natl. Acad. Sci. U. S. A. *92*, 4601–4605.

Oudet, P., Gross-Bellard, M., and Chambon, P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell *4*, 281–300.

Pabst, T., Mueller, B.U., Harakawa, N., Schoch, C., Haferlach, T., Behre, G., Hiddemann, W., Zhang, D.-E., and Tenen, D.G. (2001). AML1–ETO downregulates the granulocytic differentiation factor C/EBPα in t(8;21) myeloid leukemia. Nat. Med. *7*, 444–451.

Palis, J., Robertson, S., Kennedy, M., Wall, C., and Keller, G. (1999). Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. Development *126*, 5073–5084.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. Cell Stem Cell *1*, 299–312.

Pang, C.J., Lemsaddek, W., Alhashem, Y.N., Bondzi, C., Redmond, L.C., Ah-Son, N., Dumur, C.I., Archer, K.J., Haar, J.L., Lloyd, J. a., et al. (2012). Kruppel-Like Factor 1 (KLF1), KLF2, and Myc Control a Regulatory Network Essential for Embryonic Erythropoiesis. Mol. Cell. Biol. *32*, 2628–2644.

Park, S., Chen, W., Cierpicki, T., Tonelli, M., Cai, X., Speck, N.A., and Bushweller, J.H. (2009). Structure of the AML1-ETO eTAFH domain-HEB peptide complex and its contribution to AML1-ETO activity. Blood *113*, 3558–3567.

Paschka, P., Marcucci, G., Ruppert, A.S., Mrózek, K., Chen, H., Kittles, R.A., Vukosavljevic, T., Perrotti, D., Vardiman, J.W., Carroll, A.J., et al. (2006). Adverse prognostic significance of KIT mutations in adult acute myeloid leukemia with inv(16) and t(8;21): a Cancer and Leukemia Group B Study. J. Clin. Oncol. *24*, 3904–3911.

Patel, J.P., Gönen, M., Figueroa, M.E., Fernandez, H., Sun, Z., Racevskis, J., Van Vlierberghe, P., Dolgalev, I., Thomas, S., Aminova, O., et al. (2012). Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. N. Engl. J. Med. *366*, 1079–1089.

De Pater, E., Kaimakis, P., Vink, C.S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P.S., Camper, S. a, Speck, N., and Dzierzak, E. (2013). Gata2 is required for HSC generation and survival. J. Exp. Med. *210*, 2843–2850.

Pauler, F.M., Sloane, M.A., Huang, R., Regha, K., Koerner, M. V, Tamir, I., Sommer, A., Aszodi, A., Jenuwein, T., and Barlow, D.P. (2009). H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. Genome Res. *19*, 221–233.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295.

Peterlin, B.M., and Price, D.H. (2006). Controlling the Elongation Phase of Transcription with P-TEFb. Mol. Cell *23*, 297–305.

Peterson, L.F., and Zhang, D.-E. (2004). The 8;21 translocation in leukemogenesis. Oncogene *23*, 4255–4262.

Peterson, L.F., Boyapati, A., Ahn, E., Biggs, J.R., Okumura, A.J., Lo, M., and Yan, M. (2007). Acute myeloid leukemia with the 8q22 ; 21q22 translocation : secondary mutational events and alternative t (8 ; 21) transcripts. Blood *110*, 799–805.

Petrykowska, H.M., Vockley, C.M., and Elnitski, L. (2008). Detection and characterization of silencers

and enhancer-blockers in the greater CFTR locus. Genome Res. *18*, 1238–1246.

Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H., and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature *349*, 257–260.

Pick, M., Azzola, L., Mossman, A., Stanley, E.G., and Elefanty, A.G. (2007). Differentiation of human embryonic stem cells in serum-free medium reveals distinct roles for bone morphogenetic protein 4, vascular endothelial growth factor, stem cell factor, and fibroblast growth factor 2 in hematopoiesis. Stem Cells *25*, 2206–2214.

Pilon, A.M., Ajay, S.S., Kumar, S.A., Steiner, L. a., Cherukuri, P.F., Wincovitch, S., Anderson, S.M., Mullikin, J.C., Gallagher, P.G., Hardison, R.C., et al. (2011). Genome-wide ChIP-Seq reveals a dramatic shift in the binding of the transcription factor erythroid Kruppel-like factor during erythrocyte differentiation. Blood *118*, 139–149.

Popescu, D.-M., Botting, R.A., Stephenson, E., Green, K., Jardine, L., Calderbank, E.F., Efremova, M., Acres, M., Maunder, D., Vegh, P., et al. (2019). Decoding the development of the blood and immune systems during human fetal liver haematopoiesis. bioRxiv 654210.

Pozner, A., Lotem, J., Xiao, C., Goldenberg, D., Brenner, O., Negreanu, V., Levanon, D., and Groner, Y. (2007). Developmentally regulated promoter-switch transcriptionally controls Runx1 function during embryonic hematopoiesis. BMC Dev. Biol. *7*, 84.

Pradhan, S., Bacolla, A., Wells, R.D., and Roberts, R.J. (1999). Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. J. Biol. Chem. *274*, 33002–33010.

Privalsky, M.L. (2004). The Role of Corepressors in Transcriptional Regulation by Nuclear Hormone Receptors. Annu. Rev. Physiol. *66*, 315–360.

Ptasinska, A., Assi, S.A., Mannari, D., James, S.R., Williamson, D., Dunne, J., Hoogenkamp, M., Wu, M., Care, M., McNeill, H., et al. (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, S.A., Martinez-Soria, N., Imperato, M.R., Piper, J., Cauchy, P., Pickin, A., James, S.R., Hoogenkamp, M., Williamson, D., et al. (2014). Identification of a dynamic core transcriptional network in t(8;21) AML that regulates differentiation block and self-renewal. Cell Rep. *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., et al. (2019). RUNX1-ETO Depletion in t(8;21) AML Leads to C/EBPα- and AP-1-Mediated Alterations in Enhancer-Promoter Interaction. Cell Rep. *28*, 3022–3031.e7.

Pulikkan, J. a., Madera, D., Xue, L., Bradley, P., Landrette, S.F., Kuo, Y.-H., Abbas, S., Zhu, L.J., Valk, P., and Castilla, L.H. (2012). Thrombopoietin/MPL participates in initiating and maintaining RUNX1-ETO acute myeloid leukemia via PI3K/AKT signaling. Blood *120*, 868–879.

Qian, K., Huang, C.T.-L., Huang, C.-L., Chen, H., Blackbourn, L.W., Chen, Y., Cao, J., Yao, L., Sauvey, C., Du, Z., et al. (2014). A simple and efficient system for regulating gene expression in human pluripotent stem cells and derivatives. Stem Cells *32*, 1230–1238.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982.

Rada-iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., Mccuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc Regulates Transcriptional Pause Release. Cell *141*, 432–445.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Razin, A., and Riggs, A.D. (1980). DNA methylation and gene function. Science *210*, 604–610.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593–599.

Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. Proc. Natl. Acad. Sci. U. S. A. *99*, 6883–6888.

Regha, K., Assi, S. a., Tsoulaki, O., Gilmour, J., Lacaud, G., and Bonifer, C. (2015). Developmental-stage-dependent transcriptional response to leukaemic oncogene expression. Nat. Commun. *6*, 7203.

Reya, T., Morrison, S.J., Clarke, M.F., and Weissman, I.L. (2001). Stem cells, cancer, and cancer stem cells. Nature *414*, 105–111.

Rhoades, K.L., Hetherington, C.J., Harakawa, N., Yergeau, D. a, Zhou, L., Liu, L.Q., Little, M.T., Tenen, D.G., and Zhang, D.E. (2000). Analysis of the role of AML1-ETO in leukemogenesis, using an inducible transgenic mouse model. Blood *96*, 2108–2115.

Rieger, M.A., and Schroeder, T. (2012). Hematopoiesis. Cold Spring Harb Perspect Biol *4*, 1–18.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47–e47.

Robb, L., Lyons, I., Li, R., Hartley, L., Köntgen, F., Harvey, R.P., Metcalf, D., and Begley, C.G. (1995). Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. Proc. Natl. Acad. Sci. U. S. A. *92*, 7075–7079.

Robertson, A.L., Avagyan, S., Gansner, J.M., and Zon, L.I. (2016). Understanding the regulation of vertebrate hematopoiesis and blood disorders - big lessons from a small fish. FEBS Lett. *590*, 4016–4033.

Rosenbauer, F., Owens, B.M., Yu, L., Tumang, J.R., Steidl, U., Kutok, J.L., Clayton, L.K., Wagner, K., Scheller, M., Iwasaki, H., et al. (2006). Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU.1. Nat. Genet. *38*, 27–37.

Rossi, D.J., Bryder, D., Seita, J., Nussenzweig, A., Hoeijmakers, J., and Weissman, I.L. (2007). Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. Nature *447*, 725–729.

Rothkamm, K., Krüger, I., Thompson, L.H., and Löbrich, M. (2003). Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. Mol. Cell. Biol. *23*, 5706–5715.

Roudaia, L., Cheney, M.D., Manuylova, E., Chen, W., Morrow, M., Park, S., Lee, C.T., Kaur, P., Williams, O., Bushweller, J.H., et al. (2009). CBFbeta is critical for AML1-ETO and TEL-AML1 activity. Blood *113*, 3070–3079.

Rowley, J.D. (1973). Identificaton of a translocation with quinacrine fluorescence in a patient with acute leukemia. Ann. Génétique *16*, 109–112.

Rowley, J.D. (1984). Biological Implications of Consistent Chromosome Rearrangements in Leukemia and Lymphoma. Cancer Res. *44*, 3159–3168.

Rubin, C.M., Larson, R.A., Anastasi, J., Winter, J.N., Thangavelu, M., Vardiman, J.W., Rowley, J.D., and Le Beau, M.M. (1990). t(3;21)(q26;q22): a recurring chromosomal abnormality in therapy-related myelodysplastic syndrome and acute myeloid leukemia. Blood *76*, 2594–2598.

Ruffell, D., Mourkioti, F., Gambardella, A., Kirstetter, P., Lopez, R.G., Rosenthal, N., and Nerlov, C. (2009). A CREB-C/EBPbeta cascade induces M2 macrophage-specific gene expression and promotes muscle injury repair. Proc. Natl. Acad. Sci. U. S. A. *106*, 17475–17480.

Rybtsov, S., Sobiesiak, M., Taoudi, S., Souilhol, C., Senserrich, J., Liakhovitskaia, A., Ivanovs, A., Frampton, J., Zhao, S., and Medvinsky, A. (2011). Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. J. Exp. Med. *208*, 1305–1315.

Rybtsov, S., Batsivari, A., Bilotkach, K., Paruzina, D., Senserrich, J., Nerushev, O., and Medvinsky, A. (2014). Tracing the Origin of the HSC Hierarchy Reveals an SCF-Dependent, IL-3-Independent CD43− Embryonic Precursor. Stem Cell Reports *3*, 489–501.

Rybtsov, S., Ivanovs, A., Zhao, S., and Medvinsky, A. (2016). Concealed expansion of immature precursors underpins acute burst of adult HSC activity in foetal liver. Development *143*, 1284–1289.

Saeed, S., Logie, C., Francoijs, K.-J., Frige, G., Romanenghi, M., Nielsen, F.G., Raats, L., Shahhoseini, M., Huynen, M., Altucci, L., et al. (2012). Chromatin accessibility, p300, and histone acetylation define PML-RAR and AML1-ETO binding sites in acute myeloid leukemia. Blood *120*, 3058–3068.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 129–143.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246–3248.

Sande, S., and Privalsky, M.L. (1996). Identification of TRACs (T3 receptor-associating cofactors), a family of cofactors that associate with, and modulate the activity of, nuclear hormone receptors. Mol. Endocrinol. *10*, 813–825.

Schessl, C., Rawat, V.P.S., Cusan, M., Deshpande, A., Kohl, T.M., Rosten, P.M., Spiekermann, K., Humphries, R.K., Schnittger, S., Kern, W., et al. (2005). The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice. J. Clin. Invest. *115*, 2159–2168.

Schuldiner, M., Yanuka, O., Itskovitz-Eldor, J., Melton, D.A., and Benvenisty, N. (2000). Effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells. Proc. Natl. Acad. Sci. U. S. A. *97*, 11307–11312.

Scott, E., Simon, M., Anastasi, J., and Singh, H. (1994). Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science (80-.). *265*, 1573–1577.

Scott, E.W., Fisher, R.C., Olson, M.C., Kehrli, E.W., Simon, M.C., and Singh, H. (1997). PU.1 functions in a cell-autonomous manner to control the differentiation of multipotential lymphoid-myeloid progenitors. Immunity *6*, 437–447.

Sen, S., Block, K.F., Pasini, A., Baylin, S.B., and Easwaran, H. (2016). Genome-wide positioning of bivalent mononucleosomes. BMC Med. Genomics *9*, 60.

Setoguchi, R., Tachibana, M., Naoe, Y., Muroi, S., Akiyama, K., Tezuka, C., Okuda, T., and Taniuchi, I. (2008). Repression of the transcription factor Th-POK by Runx complexes in cytotoxic T cell development. Science (80-.). *319*, 822–825.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. Cell *119*, 941–953.

Shia, W.-J., Okumura, A.J., Yan, M., Sarkeshik, A., Lo, M.-C., Matsuura, S., Komeno, Y., Zhao, X., Nimer, S.D., Yates, J.R., et al. (2012). PRMT1 interacts with AML1-ETO to promote its transcriptional activation and progenitor cell proliferative potential. Blood *119*, 4953–4962.

Shima, T., Miyamoto, T., Kikushige, Y., Yuda, J., Tochigi, T., Yoshimoto, G., Kato, K., Takenaka, K., Iwasaki, H., Mizuno, S., et al. (2014). The ordered acquisition of Class II and Class I mutations directs formation of human t (8 ; 21) acute myelogenous leukemia stem cell. Exp. Hematol. *42*, 955–965.e5.

Shimada, H., Ichikawa, H., Nakamura, S., Katsu, R., Iwasa, M., Kitabayashi, I., and Ohki, M. (2000). Analysis of genes under the downstream control of the t(8;21) fusion protein AML1-MTG8: overexpression of the TIS11b (ERF-1, cMG1) gene induces myeloid cell proliferation in response to G-CSF. Blood *96*, 655–663.

Shivdasani, R. a (1996). The role of transcription factor NF-E2 in megakaryocyte maturation and

platelet production. Stem Cells *14*, 112–115.

Shlush, L.I., Zandi, S., Mitchell, A., Chen, W.C., Brandwein, J.M., Gupta, V., Kennedy, J.A., Schimmer, A.D., Schuh, A.C., Yee, K.W., et al. (2014). Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature *506*, 328–333.

Shultz, L.D., Brehm, M.A., Garcia-Martinez, J.V., and Greiner, D.L. (2012). Humanized mice for immune system investigation: progress, promise and challenges. Nat. Rev. Immunol. *12*, 786–798.

Simon, M.C., Pevny, L., Wiles, M. V, Keller, G., Costantini, F., and Orkin, S.H. (1992). Rescue of erythroid development in gene targeted GATA-1- mouse embryonic stem cells. Nat. Genet. *1*, 92–98.

Sinenko, S.A., Hung, T., Moroz, T., Tran, Q.-M., Sidhu, S., Cheney, M.D., Speck, N.A., and Banerjee, U. (2010). Genetic manipulation of AML1-ETO-induced expansion of hematopoietic precursors in a Drosophila model. Blood *116*, 4612–4620.

Six, E.M., Bonhomme, D., Monteiro, M., Beldjord, K., Jurkowska, M., Cordier-Garcia, C., Garrigue, A., Dal Cortivo, L., Rocha, B., Fischer, A., et al. (2007). A human postnatal lymphoid progenitor capable of circulating and seeding the thymus. J. Exp. Med. *204*, 3085–3093.

Skene, P.J., and Henikoff, S. (2013). Histone variants in pluripotency and disease. Development *140*, 2513–2524.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA Polymerase II Core Promoter. Annu. Rev. Biochem. *72*, 449–479.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220.

Smith, E., Lin, C., and Shilatifard, A. (2011). The super elongation complex (SEC) and MLL in development and disease. Genes Dev. *25*, 661–672.

Song, W.J., Sullivan, M.G., Legare, R.D., Hutchings, S., Tan, X., Kufrin, D., Ratajczak, J., Resende, I.C., Haworth, C., Hock, R., et al. (1999). Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. Nat. Genet. *23*, 166–175.

Souilhol, C., Gonneau, C., Lendinez, J.G., Batsivari, A., Rybtsov, S., Wilson, H., Morgado-Palacin, L., Hills, D., Taoudi, S., Antonchuk, J., et al. (2016). Inductive interactions mediated by interplay of asymmetric signalling underlie development of adult haematopoietic stem cells. Nat. Commun. *7*, 10784.

Souroullas, G.P., Salmon, J.M., Sablitzky, F., Curtis, D.J., and Goodell, M.A. (2009). Adult Hematopoietic Stem and Progenitor Cells Require Either Lyl1 or Scl for Survival. Cell Stem Cell *4*, 180–186.

Speck, N. a, and Gilliland, D.G. (2002). Core-binding factors in haematopoiesis and leukaemia. Nat. Rev. Cancer *2*, 502–513.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Spyropoulos, D.D., Pharr, P.N., Lavenburg, K.R., Jackers, P., Papas, T.S., Ogawa, M., and Watson, D.K. (2000). Hemorrhage, impaired hematopoiesis, and lethality in mouse embryos carrying a targeted disruption of the Fli1 transcription factor. Mol. Cell. Biol. *20*, 5643–5652.

Srinivasan, L., and Atchison, M.L. (2004). YY1 DNA binding and PcG recruitment requires CtBP. Genes Dev. *18*, 2596–2601.

Sripathy, S.P., Stevens, J., and Schultz, D.C. (2006). The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. Mol. Cell. Biol. *26*, 8623–8638.

Sroczynska, P., Lancrin, C., Kouskoff, V., and Lacaud, G. (2009). The differential activities of Runx1 promoters define milestones during embryonic hematopoiesis. Blood *114*, 5279–5289.

Staber, P.B., Zhang, P., Ye, M., Welner, R.S., Nombela-Arrieta, C., Bach, C., Kerenyi, M., Bartholdy, B. a., Zhang, H., Alberich-Jordà, M., et al. (2013). Sustained PU.1 Levels Balance Cell-Cycle Regulators to Prevent Exhaustion of Adult Hematopoietic Stem Cells. Mol. Cell *49*, 934–946.

Stalder, J., Larsen, A., Engel, J.D., Dolan, M., Groudine, M., and Weintraub, H. (1980). Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. Cell *20*, 451–460.

Steffen, B., Knop, M., Bergholz, U., Vakhrusheva, O., Rode, M., Köhler, G., Henrichs, M.-P., Bulk, E., Hehn, S., Stehling, M., et al. (2011). AML1/ETO induces self-renewal in hematopoietic progenitor cells via the Groucho-related amino-terminal AES protein. Blood *117*, 4328–4337.

Steger, D.J., Lefterova, M.I., Ying, L., Stonestrom, A.J., Schupp, M., Zhuo, D., Vakoc, A.L., Kim, J.-E., Chen, J., Lazar, M.A., et al. (2008). DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Mol. Cell. Biol. *28*, 2825–2839.

Sturgeon, C.M., Ditadi, A., Awong, G., Kennedy, M., and Keller, G. (2014). Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells. Nat. Biotechnol. *32*, 554–561.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.

Sun, X.-J., Wang, Z., Wang, L., Jiang, Y., Kost, N., Soong, T.D., Chen, W.-Y., Tang, Z., Nakadai, T., Elemento, O., et al. (2013). A stable transcription factor complex nucleated by oligomeric AML1–ETO controls leukaemogenesis. Nature *500*, 93–97.

Swiers, G., Baumann, C., O'Rourke, J., Giannoulatou, E., Taylor, S., Joshi, A., Moignard, V., Pina, C., Bee, T., Kokkaliaris, K.D., et al. (2013). Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. Nat. Commun. *4*.

Tachibana, M., Sugimoto, K., Fukushima, T., and Shinkai, Y. (2001). Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. J. Biol. Chem. *276*, 25309–25317.

Tachibana, M., Tezuka, C., Muroi, S., Nishimoto, S., Katsumoto, T., Nakajima, A., Kitabayashi, I., and Taniuchi, I. (2008). Phosphorylation of Runx1 at Ser249, Ser266, and Ser276 is dispensable for bone marrow hematopoiesis and thymocyte differentiation. Biochem. Biophys. Res. Commun. *368*, 536–542.

Tagoh, H., Ingram, R., Wilson, N., Salvagiotto, G., Warren, A.J., Clarke, D., Busslinger, M., and Bonifer, C. (2006). The mechanism of repression of the myeloid-specific c-fms gene by Pax5 during B lineage restriction. EMBO J. *25*, 1070–1080.

Takahashi, H., Parmely, T.J., Sato, S., Tomomori-sato, C., Banks, C.A.S., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-brown, S., Washburn, M.P., et al. (2011). Human Mediator Subunit MED26 Functions as a Docking Site for Transcription Elongation Factors. Cell *146*, 92–104.

Takata, M., Sasaki, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A., and Takeda, S. (1998). Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. EMBO J. *17*, 5497–5508.

Tallack, M.R., Whitington, T., Yuen, W.S., Wainwright, E.N., Keys, J.R., Gardiner, B.B., Nourbakhsh, E., Cloonan, N., Grimmond, S.M., Bailey, T.L., et al. (2010). A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. Genome Res. *20*, 1052–1063.

Tanaka, T., Tanaka, K., Ogawa, S., Kurokawa, M., Mitani, K., Nishida, J., Shibata, Y., Yazaki, Y., and Hirai, H. (1995). An acute myeloid leukemia gene, AML1, regulates hemopoietic myeloid cell differentiation and transcriptional activation antagonistically by two alternative spliced forms. EMBO J. *14*, 341–350.

Tanaka, T., Kurokawa, M., Ueki, K., Tanaka, K., Imai, Y., Mitani, K., Okazaki, K., Sagata, N., Yazaki, Y., Shibata, Y., et al. (1996). The extracellular signal-regulated kinase pathway phosphorylates AML1,

an acute myeloid leukemia gene product, and potentially regulates its transactivation ability. Mol. Cell. Biol. *16*, 3967–3979.

Tang, J.-L., Hou, H.-A., Chen, C.-Y., Liu, C.-Y., Chou, W.-C., Tseng, M.-H., Huang, C.-F., Lee, F.-Y., Liu, M.-C., Yao, M., et al. (2009). AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations. Blood *114*, 5352–5361.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., et al. (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell *163*, 1611–1627.

Taoudi, S., Gonneau, C., Moore, K., Sheridan, J.M., Blackburn, C.C., Taylor, E., and Medvinsky, A. (2008). Extensive Hematopoietic Stem Cell Generation in the AGM Region via Maturation of VE-Cadherin+CD45+ Pre-Definitive HSCs. Cell Stem Cell *3*, 99–108.

Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science *272*, 408–411.

Tavian, M., Coulombel, L., Luton, D., Clemente, H., Dieterlen-Lievre, F., and Peault, B. (1996). Aorta-associated CD34+ hematopoietic cells in the early human embryo. Blood *87*.

Tavian, M., Hallais, M.F., and Peault, B. (1999). Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo. Development *126*.

Tavian, M., Robin, C., Coulombel, L., and Péault, B. (2001). The Human Embryo, but Not Its Yolk Sac, Generates Lympho-Myeloid Stem Cells: Mapping Multipotent Hematopoietic Cell Fate in Intraembryonic Mesoderm. Immunity *15*, 487–495.

Tay, L.S., Krishnan, V., Sankar, H., Chong, Y.L., Chuang, L.S.H., Tan, T.Z., Kolinjivadi, A.M., Kappei, D., and Ito, Y. (2018). RUNX Poly(ADP-Ribosyl)ation and BLM Interaction Facilitate the Fanconi Anemia Pathway of DNA Repair. Cell Rep. *24*, 1747–1755.

Telfer, J.C., and Rothenberg, E. V. (2001). Expression and Function of a Stem Cell Promoter for the Murine CBFα2 Gene: Distinct Roles and Regulation in Natural Killer and T Cell Development. Dev. Biol. *229*, 363–382.

Thambyrajah, R., Mazan, M., Patel, R., Moignard, V., Stefanska, M., Marinopoulou, E., Li, Y., Lancrin, C., Clapes, T., Möröy, T., et al. (2015). GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. Nat. Cell Biol. *18*, 21–32.

The Cancer Genome Atlas Research Network (2013). Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. N. Engl. J. Med. *368*, 2059–2074.

Thota, S., Viny, A.D., Makishima, H., Spitzer, B., Radivoyevitch, T., Przychodzen, B., Sekeres, M.A., Levine, R.L., and Maciejewski, J.P. (2014). Genetic alterations of the cohesin complex genes in myeloid malignancies. Blood *124*, 1790–1798.

Tighe, J.E., and Calabi, F. (1995). t(8;21) breakpoints are clustered between alternatively spliced exons of MTG8. Clin. Sci. (Lond). *89*, 215–218.

Tighe, J.E., Daga, A., and Calabi, F. (1993). Translocation breakpoints are clustered on both chromosome 8 and chromosome 21 in the t(8;21) of acute myeloid leukemia. Blood *81*, 592–596.

Tober, J., Koniski, A., McGrath, K.E., Vemishetti, R., Emerson, R., De Mesy-Bentley, K.K.L., Waugh, R., and Palis, J. (2007). The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. Blood *109*, 1433–1441.

Tonks, A., Pearn, L., Musson, M., Gilkes, A., Mills, K.I., Burnett, A.K., and Darley, R.L. (2007). Transcriptional dysregulation mediated by RUNX1-RUNX1T1 in normal human progenitor cells and in acute myeloid leukaemia. Leukemia *21*, 2495–2505.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Trombly, D.J., Whitfield, T.W., Padmanabhan, S., Gordon, J.A., Lian, J.B., van Wijnen, A.J., Zaidi, S.K., Stein, J.L., and Stein, G.S. (2015a). Genome-wide co-occupancy of AML1-ETO and N-CoR defines the t(8;21) AML signature in leukemic cells. BMC Genomics *16*, 309.

Tropberger, P., and Schneider, R. (2013). Scratching the (lateral) surface of chromatin regulation by histone modifications. Nat. Struct. Mol. Biol. *20*, 657–661.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J. *13*, 3822–3831.

Tsien, R.Y. (1998). The green fluorescent protein. Annu. Rev. Biochem. *67*, 509–544.

Tsuzuki, S., Hong, D., Gupta, R., Matsuo, K., Seto, M., and Enver, T. (2007). Isoform-Specific Potentiation of Stem and Progenitor Cell Engraftment by AML1/RUNX1. PLoS Med. *4*, e172.

Ugarte, G.D., Vargas, M.F., Medina, M.A., León, P., Necuñir, D., Elorza, A.A., Gutiérrez, S.E., Moon, R.T., Loyola, A., and De Ferrari, G. V (2015). Wnt signaling induces transcription, spatial proximity and translocation of fusion gene partners in human hematopoietic cells. Blood *126*, 1785–1790.

Vairapandi, M., Balliet, A.G., Hoffman, B., and Liebermann, D.A. (2002). GADD45b and GADD45g are cdc2/cyclinB1 kinase inhibitors with a role in S and G2/M cell cycle checkpoints induced by genotoxic stress. J. Cell. Physiol. *192*, 327–338.

Vangala, R.K., Heiss-Neumann, M.S., Rangatia, J.S., Singh, S.M., Schoch, C., Tenen, D.G., Hiddemann, W., and Behre, G. (2003). The myeloid master regulator transcription factor PU.1 is inactivated by AML1-ETO in t(8;21) myeloid leukemia. Blood *101*, 270–277.

Vardiman, J.W., Thiele, J., Arber, D.A., Brunning, R.D., Borowitz, M.J., Porwit, A., Harris, N.L., Le Beau, M.M., Hellström-Lindberg, E., Tefferi, A., et al. (2009). The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood *114*, 937–951.

Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E., and Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. Genome Res. *24*, 896–905.

Vettese-Dadey, M., Grant, P.A., Hebbes, T.R., Crane- Robinson, C., Allis, C.D., and Workman, J.L. (1996). Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. EMBO J. *15*, 2508–2518.

Viale, A., De Franco, F., Orleth, A., Cambiaghi, V., Giuliani, V., Bossi, D., Ronchini, C., Ronzoni, S., Muradore, I., Monestiroli, S., et al. (2009). Cell-cycle restriction limits DNA damage and maintains self-renewal of leukaemia stem cells. Nature *457*, 51–56.

de Villiers, J., and Schaffner, W. (1981). A small segment of polyoma virus DNA enhances the expression of a cloned β-globin gene over a distance of 1400 base pairs. Nucleic Acids Res. *9*, 6251–6264.

Vinjamur, D.S., Wade, K.J., Mohamad, S.F., Haar, J.L., Sawyer, S.T., and Lloyd, J. a. (2014). Kruppel-like transcription factors KLF1 and KLF2 have unique and coordinate roles in regulating embryonic erythroid precursor maturation. Haematologica *99*, 1565–1573.

Vodyanik, M.A., Bork, J.A., Thomson, J.A., and Slukvin, I.I. (2005). Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. Blood *105*, 617–626.

Voigt, P., LeRoy, G., Drury, W.J., Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically Modified Nucleosomes. Cell *151*, 181–193.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., et al. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev. *12*, 343–356.

Wahlster, L., and Daley, G.Q. (2016). Progress towards generation of human haematopoietic stem

cells. Nat. Cell Biol. *18*, 1111–1117.

Walrad, P.B., Hang, S., Joseph, G.S., Salas, J., and Gerge, J.P. (2010). Distinct Contributions of Conserved Modules to Runt Transcription Factor Activity. Mol. Biol. Cell *21*, 2315–2316.

Wamstad, J.A., Alexander, J.M., Truty, R.M., Shrikumar, A., Li, F., Eilertson, K.E., Ding, H., Wylie, J.N., Pico, A.R., Capra, J.A., et al. (2012). Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage. Cell *151*, 206–220.

Wang, Y., and Nakayama, N. (2009). WNT and BMP signaling are both required for hematopoietic cell development from human ES cells. Stem Cell Res. *3*, 113–125.

Wang, H., Huang, Z.Q., Xia, L., Feng, Q., Erdjument-Bromage, H., Strahl, B.D., Briggs, S.D., Allis, C.D., Wong, J., Tempst, P., et al. (2001). Methylation of Histone H4 at Arginine 3 Facilitating Transcriptional Activation by Nuclear Hormone Receptor. Science (80-.). *293*, 853–857.

Wang, J., Hoshino, T., Redner, R.L., Kajigaya, S., and Liu, J.M. (1998). ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. Proc. Natl. Acad. Sci. U. S. A. *95*, 10860–10865.

Wang, L., Gural, A., Sun, X.-J., Zhao, X., Perna, F., Huang, G., Hatlen, M. a, Vu, L., Liu, F., Xu, H., et al. (2011a). The leukemogenicity of AML1-ETO is dependent on site-specific lysine acetylation. Science *333*, 765–769.

Wang, L., Man, N., Sun, X.-J., Tan, Y., García-Cao, M., Cao, M.G., Liu, F., Hatlen, M., Xu, H., Huang, G., et al. (2015). Regulation of AKT signaling by Id1 controls t(8;21) leukemia initiation and progression. Blood *126*, 640–650.

Wang, M., Wang, H., Wen, Y., Chen, X., Liu, X., Gao, J., Su, P., Xu, Y., Zhou, W., Shi, L., et al. (2018). MEIS2 regulates endothelial to hematopoietic transition of human embryonic stem cells by targeting TAL1. Stem Cell Res. Ther. *9*, 340.

Wang, Q., Stacy, T., Miller, J.D., Lewis, a F., Gu, T.L., Huang, X., Bushweller, J.H., Bories, J.C., Alt, F.W., Ryan, G., et al. (1996). The CBFbeta subunit is essential for CBFalpha2 (AML1) function in vivo. Cell *87*, 697–708.

Wang, W., Schwemmers, S., Hexner, E.O., and Pahl, H.L. (2010). AML1 is overexpressed in patients with myeloproliferative neoplasms and mediates JAK2V617F-independent overexpression of NF-E2. Blood *116*, 254–266.

Wang, Y.-Y., Zhou, G.-B., Yin, T., Chen, B., Shi, J.-Y., Liang, W.-X., Jin, X.-L., You, J.-H., Yang, G., Shen, Z.-X., et al. (2005). AML1-ETO and C-KIT mutation/overexpression in t(8;21) leukemia: implication in stepwise leukemogenesis and response to Gleevec. Proc. Natl. Acad. Sci. U. S. A. *102*, 1104–1109.

Wang, Y.-Y., Zhao, L.-J., Wu, C.-F., Liu, P., Shi, L., Liang, Y., Xiong, S.-M., Mi, J.-Q., Chen, Z., Ren, R., et al. (2011b). C-KIT mutation cooperates with full-length AML1-ETO to induce acute myeloid leukemia in mice. Proc. Natl. Acad. Sci. *108*, 2450–2455.

Warren, A.J., Colledge, W.H., Carlton, M.B.L., Evans, M.J., Smith, A.J.H., and Rabbitts, T.H. (1994). The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. Cell *78*, 45–57.

Wasylyk, B., and Chambon, P. (1979). Transcription by Eukaryotic RNA Polymerases A and B of Chromatin Assembled in vitro. Eur. J. Biochem. *98*, 317–327.

Watson, P.J., Fairall, L., and Schwabe, J.W.R. (2012). Nuclear hormone receptor co-repressors: structure and function. Mol. Cell. Endocrinol. *348*, 440–449.

Weber, C.M., Ramachandran, S., and Henikoff, S. (2014). Nucleosomes Are Context-Specific, H2A.Z-Modulated Barriers to RNA Polymerase. Mol. Cell *53*, 819–830.

Wei, Y., Liu, S., Lausen, J., Woodrell, C., Cho, S., Biris, N., Kobayashi, N., Wei, Y., Yokoyama, S., and Werner, M.H. (2007). A TAF4-homology domain from the corepressor ETO is a docking platform for positive and negative regulators of transcription. Nat. Struct. Mol. Biol. *14*, 653–661.

Weiss, M.J., Keller, G., and Orkin, S.H. (1994). Novel insights into erythroid development revealed through in vitro differentiation of GATA-1- embryonic stem cells. Genes Dev. *8*, 1184–1197.

Welner, R.S., Pelayo, R., and Kincade, P.W. (2008). Evolving views on the genealogy of B cells. Nat. Rev. Immunol. *8*, 95–106.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat. Genet. *41*, 246–250.

Westendorf, J.J., Yamamoto, C.M., Lenny, N., Downing, J.R., Selsted, M.E., and 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. Mol. Cell. Biol. *18*, 322–333.

Whitmarsh, A.J., and Davis, R.J. (1996). Transcription factor AP-1 regulation by mitogen-activated protein kinase signal transduction pathways. J. Mol. Med. (Berl). *74*, 589–607.

Wichmann, C., Quagliano-Lo Coco, I., Yildiz, Ö., Chen-Wichmann, L., Weber, H., Syzonenko, T., Döring, C., Brendel, C., Ponnusamy, K., Kinner, A., et al. (2015). Activating c-KIT mutations confer oncogenic cooperativity and rescue RUNX1/ETO-induced DNA damage and apoptosis in human primary CD34+ hematopoietic progenitors. Leukemia *29*, 279–289.

Wiemels, J.L., Xiao, Z., Buffler, P. a, Maia, A.T., Ma, X., Dicks, B.M., Martyn, T., Zhang, L., Feusner, J., Wiencke, J., et al. (2010). In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *99*, 3801–3805.

Wier, A.D., Mayekar, M.K., Héroux, A., Arndt, K.M., and Vandemark, A.P. (2013). Structural basis for Spt5-mediated recruitment of the Paf1 complex to chromatin. Proc. Natl. Acad. Sci. U. S. A. *110*, 17290–17295.

Wilkinson, R.N., Pouget, C., Gering, M., Russell, A.J., Davies, S.G., Kimelman, D., and Patient, R. (2009). Hedgehog and Bmp Polarize Hematopoietic Stem Cell Emergence in the Zebrafish Dorsal Aorta. Dev. Cell *16*, 909–916.

Wiznerowicz, M., Jakobsson, J., Szulc, J., Liao, S., Quazzola, A., Beermann, F., Aebischer, P., and Trono, D. (2007). The Kruppel-associated box repressor domain can trigger de novo promoter methylation during mouse early embryogenesis. J. Biol. Chem. *282*, 34535–34541.

Woll, P.S., Morris, J.K., Painschab, M.S., Marcus, R.K., Kohn, A.D., Biechele, T.L., Moon, R.T., and Kaufman, D.S. (2008). Wnt signaling promotes hematoendothelial cell development from human embryonic stem cells. Blood *111*, 122–131.

Wolyniec, K., Wotton, S., Kilbey, A., Jenkins, A., Terry, A., Peters, G., Stocking, C., Cameron, E., and Neil, J.C. (2009). RUNX1 and its fusion oncoprotein derivative, RUNX1-ETO, induce senescence-like growth arrest independently of replicative stress. Oncogene *28*, 2502–2512.

Xiao, Z., Greaves, M.F., Buffler, P., Smith, M.T., Segal, M.R., Dicks, B.M., Wiencke, J.K., and Wiemels, J.L. (2001). Molecular characterization of genomic AML1-ETO fusions in childhood leukemia. Leukemia *15*, 1906–1913.

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013). Epigenomic Analysis of Multilineage Differentiation of Human Embryonic Stem Cells. Cell *153*, 1134–1148.

Yamada, Y., Warren, a J., Dobson, C., Forster, a, Pannell, R., and Rabbitts, T.H. (1998). The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. Proc. Natl. Acad. Sci. U. S. A. *95*, 3890–3895.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation. Cell *97*, 41–51.

Yamaguchi, Y., Inukai, N., Narita, T., Wada, T., and Handa, H. (2002). Evidence that Negative Elongation Factor Represses Transcription Elongation through Binding to a DRB Sensitivity-Inducing Factor / RNA Polymerase II Complex and RNA. Mol. Cell. Biol. *22*, 2918–2927.

Yamanaka, R., Barlow, C., Lekstrom-Himes, J., Castilla, L.H., Liu, P.P., Eckhaus, M., Decker, T., Wynshaw-Boris, a, and Xanthopoulos, K.G. (1997). Impaired granulopoiesis, myelodysplasia, and early lethality in CCAAT/enhancer binding protein epsilon-deficient mice. Proc. Natl. Acad. Sci. U. S. A. *94*, 13187–13192.

Yan, M., Burel, S.A., Peterson, L.F., Kanbe, E., Iwasaki, H., Boyapati, A., Hines, R., Akashi, K., and Zhang, D.E. (2004). Deletion of an AML1-ETO C-terminal NcoR/SMRT-interacting region strongly induces leukemia development. Proc Natl Acad Sci U S A *101*, 17186–17191.

Yan, M., Kanbe, E., Peterson, L.F., Boyapati, A., Miao, Y., Wang, Y., Chen, I.-M., Chen, Z., Rowley, J.D., Willman, C.L., et al. (2006). A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. Nat. Med. *12*, 945–949.

Yang, G., Khalaf, W., van de Locht, L., Jansen, J.H., Gao, M., Thompson, M.A., van der Reijden, B.A., Gutmann, D.H., Delwel, R., Clapp, D.W., et al. (2005). Transcriptional repression of the Neurofibromatosis-1 tumor suppressor by the t(8;21) fusion protein. Mol. Cell. Biol. *25*, 5869–5879.

Yang, X.-J., Ogryzko, V. V., Nishikawa, J., Howard, B.H., and Nakatani, Y. (1996). A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. Nature *382*, 319–324.

Yeh, J.-R.J., Munson, K.M., Elagib, K.E., Goldfarb, A.N., Sweetser, D.A., and Peterson, R.T. (2009). Discovering chemical modifiers of oncogene-regulated hematopoietic differentiation. Nat. Chem. Biol. *5*, 236–243.

Yergeau, D.A., Hetherington, C.J., Wang, Q., Zhang, P., Sharpe, A.H., Binder, M., Marín-Padilla, M., Tenen, D.G., Speck, N.A., and Zhang, D.E. (1997). Embryonic lethality and impairment of haematopoiesis in mice heterozygous for an AML1-ETO fusion gene. Nat. Genet. *15*, 303–306.

Ying, J., Srivastava, G., Hsieh, W.-S., Gao, Z., Murray, P., Liao, S.-K., Ambinder, R., and Tao, Q. (2005). The Stress-Responsive Gene GADD45G Is a Functional Tumor Suppressor, with Its Response to Environmental Stresses Frequently Disrupted Epigenetically in Multiple Tumors. Clin. Cancer Res. *11*, 6442–6449.

Yokomizo, T., Ogawa, M., Osato, M., Kanno, T., Yoshida, H., Fujimoto, T., Fraser, S., Nishikawa, S., Okada, H., Satake, M., et al. (2001). Requirement of Runx1/AML1/PEBP2alfaB for the generation of haematopoietic cells from endothelial cells. Genes Cells *6*, 13–23.

Yu, P., Pan, G., Yu, J., and Thomson, J.A. (2011). FGF2 Sustains NANOG and Switches the Outcome of BMP4-Induced Human Embryonic Stem Cell Differentiation. Cell Stem Cell *8*, 326–334.

Yuan, Y., Zhou, L., Miyamoto, T., Iwasaki, H., Harakawa, N., Hetherington, C.J., Burel, S. a, Lagasse, E., Weissman, I.L., Akashi, K., et al. (2001). AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations. Proc. Natl. Acad. Sci. U. S. A. *98*, 10398–10403.

Zaidi, S.K., Dowdy, C.R., Wijnen, A.J. Van, Lian, J.B., Stein, J.L., Croce, C.M., and Stein, G.S. (2009). Altered Runx1 Subnuclear Targeting Enhances Myeloid Cell Proliferation and Blocks Differentiation by activating a miR-24/ MKP-7/MAP Kinase Network. Cancer Res. *69*, 8249–8255.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: Establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zeng, C., McNeil, S., Pockwinse, S., Nickerson, J., Shopland, L., Lawrence, J.B., Penman, S., Hiebert, S., Lian, J.B., van Wijnen, A.J., et al. (1998). Intranuclear targeting of AML/CBFalpha regulatory factors to nuclear matrix-associated transcriptional domains. Proc. Natl. Acad. Sci. U. S. A. *95*, 1585–1589.

Zeng, H., Yücel, R., Kosan, C., Klein-Hitpass, L., and Möröy, T. (2004). Transcription factor Gfi1 regulates self-renewal and engraftment of hematopoietic stem cells. EMBO J. *23*, 4116–4125.

Zhang, M.Q. (1998). Identification of human gene core promoters in silico. Genome Res. *8*, 319–326.

Zhang, D.E., Zhang, P., Wang, N.D., Hetherington, C.J., Darlington, G.J., and Tenen, D.G. (1997). Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. Proc. Natl. Acad. Sci. U. S. A. *94*, 569–574.

Zhang, H., Alberich-Jorda, M., Amabile, G., Yang, H., Staber, P.B., Di Ruscio, A., Diruscio, A., Welner, R.S., Ebralidze, A., Zhang, J., et al. (2013a). Sox4 is a key oncogenic target in C/EBPα mutant acute myeloid leukemia. Cancer Cell *24*, 575–588.

Zhang, H., Ye, M., Welner, R.S., and Tenen, D.G. (2014). Sox4 Is Required for the Formation and Maintenance of Multipotent Progenitors. Blood *124*.

Zhang, J., Hug, B.A., Huang, E.Y., Chen, C.W., Gelmetti, V., Maccarana, M., Minucci, S., Pelicci, P.G., and Lazar, M.A. (2001). Oligomerization of ETO is obligatory for corepressor interaction. Mol. Cell. Biol. *21*, 156–163.

Zhang, J., Kalkum, M., Yamamura, S., Chait, B.T., and Roeder, R.G. (2004a). E Protein Silencing by the Leukemogenic AML1-ETO Fusion Protein. Science (80-.). *305*, 1286–1289.

Zhang, J., Wu, Z., Savin, A., Yang, M., Hsu, Y.-H.R., Jantuan, E., Bacani, J.T.C., and Ingham, R.J. (2018). The c-Jun and JunB transcription factors facilitate the transit of classical Hodgkin lymphoma tumour cells through G1. Sci. Rep. *8*, 16019.

Zhang, K., Tang, H., Huang, L., Blankenship, J.W., Jones, P.R., Xiang, F., Yau, P.M., and Burlingame, A.L. (2002a). Identification of Acetylation and Methylation Sites of Histone H3 from Chicken Erythrocytes by High-Accuracy Matrix-Assisted Laser Desorption Ionization–Time-of-Flight, Matrix-Assisted Laser Desorption Ionization–Postsource Decay, and Nanoelectrospray Ionization Tandem Mass Spectrometry. Anal. Biochem. *306*, 259–269.

Zhang, L., Fried, F.B., Guo, H., and Friedman, A.D. (2008a). Cyclin-dependent kinase phosphorylation of RUNX1/AML1 on 3 sites increases transactivation potency and stimulates cell proliferation. Blood *111*, 1193–1200.

Zhang, P., Iwasaki-Arai, J., Iwasaki, H., Fenyus, M.L., Dayaram, T., Owens, B.M., Shigematsu, H., Levantini, E., Huettner, C.S., Lekstrom-Himes, J. a., et al. (2004b). Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBPα. Immunity *21*, 853–863.

Zhang, Y., Strissel, P., Strick, R., Chen, J., Nucifora, G., Le Beau, M.M., Larson, R.A., and Rowley, J.D. (2002b). Genomic DNA breakpoints in AML1/RUNX1 and ETO cluster with topoisomerase II DNA cleavage and DNase I hypersensitive sites in t(8;21) leukemia. Proc. Natl. Acad. Sci. U. S. A. *99*, 3070–3075.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008b). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, Y., Wang, J., Wheat, J., Chen, X., Jin, S., Sadrzadeh, H., Fathi, A.T., Peterson, R.T., Kung, A.L., Sweetser, D.A., et al. (2013b). AML1-ETO mediates hematopoietic self-renewal and leukemogenesis through a COX/β-catenin signaling pathway. Blood *121*, 4906–4916.

Zhao, S., Zhang, Y., Sha, K., Tang, Q., Yang, X., Yu, C., Liu, Z., Sun, W., Cai, L., Xu, C., et al. (2014). KRAS (G12D) Cooperates with AML1/ETO to Initiate a Mouse Model Mimicking Human Acute Myeloid Leukemia. Cell. Physiol. Biochem. *33*, 78–87.

Zhao, X., Jankovic, V., Gural, A., Huang, G., Pardanani, A., Menendez, S., Zhang, J., Dunne, R., Xiao, A., Erdjument-Bromage, H., et al. (2008). Methylation of RUNX1 by PRMT1 abrogates SIN3A binding and potentiates its transcriptional activity. Genes Dev. *22*, 640–653.

Zhao, X.D., Han, X., Chew, J.L., Liu, J., Chiu, K.P., Choo, A., Orlov, Y.L., Sung, W., Shahab, A., Kuznetsov, V.A., et al. (2007). Whole-Genome Mapping of Histone H3Lys4 and 27 Trimethylations Reveals Distinct Genomic Compartments in Human Embryonic Stem Cells. Cell Stem Celll *1*, 286–298.

Zhao, Y., Li, X., Zhao, W., Wang, J., Yu, J., Wan, Z., Gao, K., Yi, G., Wang, X., Fan, B., et al. (2019). Single-cell transcriptomic landscape of nucleated cells in umbilical cord blood. Gigascience *8*.

Zheng, X., Beissert, T., Kukoc-Zivojnov, N., Puccetti, E., Altschmied, J., Strolz, C., Boehrer, S., Gul, H., Schneider, O., Ottmann, O.G., et al. (2004). gamma-Catenin contributes to leukemogenesis induced by AML-associated translocation products by increasing the self-renewal ofvery primitive

progenitor cells. Blood *93*, 3167–3215.

Zhou, F., Li, X., Wang, W., Zhu, P., Zhou, J., He, W., Ding, M., Xiong, F., Zheng, X., Li, Z., et al. (2016). Tracing haematopoietic stem cell formation at single-cell resolution. Nature *533*.

Zhou, F., Liu, Y., Rohde, C., Pauli, C., Gerloff, D., Köhn, M., Misiak, D., Bäumer, N., Cui, C., Göllner, S., et al. (2017). AML1-ETO requires enhanced C/D box snoRNA/RNP formation to induce self-renewal and leukaemia. Nat. Cell Biol. *19*, 844–855.

Zhou, Q., Li, T., and Price, D.H. (2012). RNA Polymerase II Elongation Control. Annu. Rev. Biochem. *81*, 119–143.

Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., Jager, P.L. De, et al. (2013). Genome-wide Chromatin State Transitions Associated with Developmental and Environmental Cues. Cell *152*, 642–654.

Zovein, A.C., Hofmann, J.J., Lynch, M., French, W.J., Turlo, K. a., Yang, Y., Becker, M.S., Zanetta, L., Dejana, E., Gasson, J.C., et al. (2008). Fate Tracing Reveals the Endothelial Origin of Hematopoietic Stem Cells. Cell Stem Cell *3*, 625–636.