FUNCTIONAL CHARACTERISATION OF THE TRANSCRIPTIONAL REGULATORS FOXC1 AND IRX3 IN ACUTE MYELOID LEUKAEMIA

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Medical and Human Sciences

2015

Timothy Dennis David Somerville

School of Medicine

List of contents

Word count: 47,087

List of figures

List of tables

List of appendices

Appendix Table 1. Genes differentially expressed in *Hoxa9/FOXC1* AMLs versus *Hoxa9/MTV* and *Hoxa9/Meis1* AMLs

Appendix Table 2. Signal-to-noise (S2N) rankings for gene set enrichment analyses.

Appendix Table 3. Non-expressed transcription factor genes additionally marked by H3K27me3 in normal human $CD34^{+}$ cells

Appendix Table 4. Absence of recurrent intergenic mutations ±1MB from the transcription start site of *FOXC1* in TCGA whole genome sequencing data for $FOXC1^{high}$ and $FOXC1^{low}$ cases

All on CD-ROM

List of abbreviations

2HG	2-hydroxygluterate
3'UTR	3' untranslated region
5hmC	5-hydroxymethlycytosine
5mC	5-methylcytosine
7-AAD	7-aminoactinomycin D
α-KG	α-ketogluterate
ac	Acetylation
AGM	Aorta-gonad-mesonephros
AID	Activation-induced cytidine deaminase
AML	Acute myeloid leukaemia
ANOVA	Analysis of variance
APL	Acute promyelocytic leukaemia
ATO	Arsenic trioxide
ATRA	All-*trans* retinoic acid
BAM	Binary sequence alignment
BFU-E	Burst-forming unit erythroid
BM	Bone marrow
BSA	Bovine serum albumin
CBF	Core binding factor
CBP	CREB binding protein
CFC	Colony-forming cell
CFU-BL	Colony-forming unit blast-like
CFU-GM	Colony-forming unit granulocyte/macrophage
CFU-M	Colony-forming unit macrophage
CFU-S	Colony-forming-units spleen
ChIP	Chromatin immunoprecipitation
CI	Confidence intervals
CML	Chronic myeloid leukaemia
CMP	Common myeloid progenitor
COMPASS	Complex proteins associated with SET
CR	Complete remission
CSC	Cancer stem cell
CXCL12	CXC chemokine ligand 12
ddH_2O	Double-distilled water
DHS	DNase hypersensitivity site
DMEM	Dulbecco's Modified Eagle's Medium

DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DOT1L	Dot1-like
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhance green fluorescent protein
Eosin	Eosinophils
EryB	Erythroblast
ES cell	Embryonic stem cell
ETP	Earliest thymic progenitors
EZH2	Enhancer of zeste
FAB	French-American-British
FACS	Fluorescent activated cell sorting
FBS	Foetal bovine serum
FDR	False discovery rate
FL	FLT3 ligand
FLT3	Fms-like tyrosine kinase receptor-3
FOX	Forkhead box
FSC	Forward scatter
G-CSF	Granulocyte-colony stimulating factor
GFP	Green fluorescent protein
GM-CSF	Granulocyte/macrophage-colony stimulating factor
GMP	Granulocyte/monocyte progenitor
GSEA	Gene set enrichment analysis
H2AK119	Histone 2A lysine 119
H2AK5	Histone 2A lysine 5
H2AK9	Histone 2A lysine 9
H2BK12	Histone 2B lysine 12
H2BK120	Histone 2B lysine 120
H2BK15	Histone 2B lysine 15
H2BK5	Histone 2B lysine 5
H3K14	Histone 3 lysine 14
H3K18	Histone 3 lysine 18
H3K23	Histone 3 lysine 23
H3K27	Histone 3 lysine 27
H3K36	Histone 3 lysine 36
H3K4	Histone 3 lysine 4

H3K56	Histone 3 lysine 56
H3K79	Histone 3 lysine 79
H3K9	Histone 3 lysine 9
H3T11	Histone 3 threonine 11
H4K12	Histone 4 lysine 12
H4K12	Histone 4 lysine 16
H4K20	Histone 3 lysine 20
H4K5	Histone 4 lysine 5
H4K8	Histone 4 lysine 8
H4K91	Histone 4 lysine 91
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HEK	Human embryonic kidney
HMT	Histone methyltransferase
HOX	Homeobox
HRP	Horseradish peroxidase
HSC	Haematopoietic stem cell
HSCT	Haematopoietic stem cell transplantation
HSLB	High salt lysis buffer
HSPC	Haematopoietic stem and progenitor cell
IFC	Integrated fluidic circuit
IL-2R	Interleukin-2 receptor
IL3	Interleukin 3
IL6	Interleukin 6
IP	Immunoprecipitation
IRES	Internal ribosomal entry site
IRX	Iroquois homeobox
ITD	Internal tandem duplication
IT-HSC	Intermediate-term haematopoietic stem cell
KD	Knockdown
KLF	Krüppel like factor
LB	Luria broth
LIC	Leukaemia-initiating cell
Lin	Lineage marker
LMPP	Lymphoid-primed multipotent progenitors
LSC	Leukaemia stem cell
LT-HSC	Long-term haematopoietic stem cell
MDS	Myelodysplastic syndrome

me1	Mono-methylation
me2	Di-methylation
me3	Tri-methylation
Mega	Megakaryocytes
MEP	Megakaryocyte/erythroid progenitor
MLL	Mixed lineage leukaemia
MLP	Immature lymphoid progenitor
Mono	Monocytes
MPN	Myeloproliferative neoplasm
MPP	Multipotent progenitor
MSCV	Murine stem cell virus
MTV	Empty vector
Neut	Neutrophils
NGS	Next generation sequencing
NK cells	Natural killer cells
NOD	Non-obese diabetic
NOG	NOD-*scidIL-2Rγ*^{null}
NSG	NOD.Cg-*Prkdc*^{scid}*-IL-2Rγt*^{m1Wjl}/SzJ
NTC	Non-targeting control
PB	Peripheral blood
PBS	Phosphate buffered saline
PcG	Polycomb group
PCGF	Polycomb group RING finger
PCR	Polymerase chain reaction
PEI	Polyethylenimine
ph	Phosphorylation
PHD	Plant homeodomain
Plat-E	Platinum-E
PML	Promyelocytic
PRC	Polycomb repressive complex
PreAmp	Pre-amplification
qPCR	Quantitative PCR
RARA	Retinoic acid receptor alpha
RNA pol II	RNA polymerase II
RNA	Ribonucleic acid
RNAseq	RNA sequencing
RPMI	Roswell Park Memorial Institute
Sca-1	Stem cell antigen-1

SCF	Stem cell factor
SCID	Severe combined immunodeficient
SD	Standard deviation
SDM3	Site directed mutagenesis #3
SDS	Sequence Detection System
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEM	Standard error of the mean
SET	su(var)3-9, enhancer of zeste, trithorax
SFEM	Serum-free expansion medium
shRNA	short hairpin RNA
SM buffer	Staining medium buffer
SNV	Single nucleotide variant
SRC	SCID-repopulating cell
SSC	Side scatter
ST-HSC	Short-term haematopoietic stem cell
TAE	Tris acetate EDTA
TALE	Three-amino acid loop extension
T-ALL	T cell acute lymphoblastic leukaemia
TCGA	The Cancer Genome Atlas Research Network
TDG	Thymine DNA glycosylase
TFIID	Transcription factor II D
TPO	Thrombopoietin
TrxG	Trithorax group
ub1	Mono-ubiquitylation
UPL	Universal probe library
UV	Ultraviolet
VAF	Variant allele frequency
WGBS	Whole-genome bisulfite sequencing
WHO	World Health Organization

Abstract

The University of Manchester

Timothy Dennis David Somerville

Degree of Doctor of Philosophy

Functional characterisation of the transcriptional regulators FOXC1 and IRX3 in acute myeloid leukaemia

September 2015

Acute myeloid leukaemia (AML) is an aggressive, clonal blood neoplasm for which current treatment strategies remain ineffective in the majority of cases. A greater understanding of the critical biological processes that are active in AML cells, but not in normal haematopoietic stem and progenitor cells (HSPCs), may lead to the identification of selective AML dependencies and the development of targeted therapies.

Through *in silico* and other analyses, *FOXC1* and *IRX3* were identified as transcriptional regulators expressed in at least 20% and 30% of human AML cases respectively. Importantly, these transcription factor genes are either not expressed or minimally expressed in normal haematopoietic populations. Functional experiments demonstrated that FOXC1 and IRX3 contribute to a block in monocyte/macrophage differentiation and enhance clonogenic potential. *FOXC1* and *IRX3* expression in AML is almost exclusively associated with expression of the HOXA/B locus and *in vivo* analyses demonstrated that both collaborate with HOXA9 to significantly enhance myeloid differentiation block in murine AML. A FOXC1-repressed gene set identified in murine leukaemia also exhibited quantitative repression in human AML in accordance with *FOXC1* expression, and *FOXC1*^{high} human AML cases exhibited reduced morphologic monocytic differentiation and inferior survival. Thus, both *FOXC1* and *IRX3* are frequently expressed to functional effect in human AML.

Continued repression of both *FOXC1* and *IRX3* in normal HSPCs is dependent upon the Polycomb Repressive Complex 2 (PRC2) as treatment of human CD34^{+} cells with distinct PRC2 inhibitors led to significant increases in their expression. However, bioinformatics analyses demonstrated that in AML there is no widespread derepression of genes normally marked by Polycomb in CD34^{+} cells. Thus the derepression of *FOXC1* and *IRX3* in AML represents a genomic locus-specific phenomenon rather than a genome-wide failure of Polycomb activity. Thus, these investigations highlight a hitherto unappreciated but frequent pathogenic mechanism in human AML: the tissue-inappropriate derepression of transcriptional regulators with unexpected functional consequences and prognostic significance.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

The following four notes on copyright and the ownership of intellectual property rights must be included as written below:

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on Presentation of Theses.

Acknowledgements

First of all I wish to thank Dr Tim Somervaille, my PhD supervisor, for inviting me to join the Leukaemia Biology group at the CRUK Manchester Institute. I can only hope that after four years of ribbing, incorrectly received emails and all round general confusion on account of our near identical names, that I have repaid the faith he placed in me when he gave me this fantastic opportunity. His support and guidance has been instrumental in these early stages of my scientific career and for this I am eternally grateful. I would also like to thank Dr Georges Lacaud, my PhD advisor, for helpful discussions relating to my project and career progression.

I count myself extremely lucky to have joined a lab that, although has evolved significantly during my PhD studies, has continued to remain full of helpful, smart and interesting people, all of whom have contributed to making my time at the institute highly enjoyable. From the early days I would like to pay special recognition to Dr James Lynch for his help, support and patience when I first joined the lab. I would also like to thank Dr Xu Huang for additional advice and guidance. Of course, I must thank Dr Gary Spencer, the cornerstone of the LB lab, for answering not only all of my questions, but everyone else's also (every lab needs a Gary!).

I also feel incredibly lucky to have worked alongside Dr "Jaeger" Dan Wiseman, Emma Williams, Filippo Ciceri, Julian Jude and Bill Harris, my fellow PhD students, all of whom I now consider as close friends. Thank you for sharing all of the ups-and-downs of life in a research lab (and the Red Lion) with me and for the support and encouragement along the way. I would also like to thank Dr Alba Maquies-Diaz, Dr Isabel Romero and Dr Gauri Deb, who have all joined the lab more recently and have proven great additions, both academically and socially. I wish you all great success in your scientific careers.

The work within this thesis would not have been possible without the excellent facilities and members of staff within the CRUK Manchester Institute, in particular the Biological Research Unit and the Flow Cytometry and Molecular Biology core facilities. Thank you to everyone that has contributed to this work.

Last, but by no means least, I would like to thank my family for their ever-present support.

Statement I: Scientific content

All scientific endeavours require collaborative efforts. Throughout this thesis there are experiments which have been performed with the assistance of others. In particular, I have had substantial support with the bioinformatics aspects of this thesis. Wherever assistance has been received this is clearly documented.

Statement II: Publication

A significant proportion of the work presented in this thesis has recently been accepted for publication, referenced below:

Somerville, T. D. D., Wiseman, D. H., Spencer, G. J., Huang, X., Lynch, J. T., Leong, H. S., Williams, E. L., Cheesman, E. and Somervaille, T. C. P. (2015). Frequent derepression of the mesenchymal transcription factor gene FOXC1 in acute myeloid leukaemia. Cancer Cell *28*, 1-14.

Chapter 1. Introduction

1.1. Haematopoiesis

Haematopoiesis is the formation and development of blood cells. In mammals, blood production begins during the early stages of embryonic development and arises in two waves, termed “primitive” and “definitive” (Orkin and Zon, 2008). Primitive haematopoiesis occurs as a transient wave in the mammalian yolk sac, with the primary function of producing erythroid cells to support the rapidly-growing embryo. Definitive haematopoiesis is first observed in the aorta-gonad-mesonephros (AGM) region and proceeds sequentially at multiple, distinct anatomical sites which include the placenta, the foetal liver and the bone marrow (BM), the latter being the primary site of definitive haematopoiesis in the adult (Orkin and Zon, 2008). As mature blood cells are predominantly short-lived, they must be constantly replenished throughout life; the adult human produces approximately 10^{11}-10^{12} new blood cells every day (Doulatov et al., 2012). To maintain this remarkable regenerative capacity, haematopoiesis is organised as a cellular hierarchy, sustained at its apex by a rare population of quiescent haematopoietic stem cells (HSCs).

HSCs are defined by their unique ability to self-renew, ensuring the stem cell pool is maintained, and their potential to differentiate into all of the terminally differentiated cells of the blood system (e.g. B cells, T cells, granulocytes and monocytes). In the adult BM, HSCs are maintained in a specialised perivascular niche, where they are localised close to blood vessels and a diverse group of cell types including endothelial cells and mesenchymal progenitor cells (Morrison and Scadden, 2014). These niche cells produce key factors, such as stem cell factor (SCF) and CXC chemokine ligand 12 (CXCL12), which are critical for the maintenance of HSC quiescence and their localisation within the BM. Human HSCs have been estimated to divide only once or twice a year (Shepherd et al., 2004; Catlin et al., 2011), but when they do, the differentiating daughter cells give rise to a series of highly proliferative progenitor cell intermediates which expand rapidly, become progressively lineage-restricted and ultimately assume the identity of a mature blood cell (Doulatov et al., 2012). The extraordinary capacity of HSCs to regenerate the entire blood system underpins the success of BM transplantation in the treatment of conditions that cause BM failure, such as autoimmune disorders and a variety of haematological malignancies (Weissman, 2000).

1.1.1. Assaying haematopoiesis

The first experiments to functionally validate the existence of a cell with HSC-like properties, i.e. self-renewal and multilineage differentiation potential, were performed by Till and McCulloch in the 1960s (Till and McCulloch, 1961; Becker et al., 1963; Siminovitch et al., 1963; Wu et al., 1968). When mouse BM cells were intravenously injected into irradiated recipient mice, a small proportion were observed to give rise to macroscopically visible colonies in the host spleen 9 to 14 days later. The number of colonies formed was shown to be linearly related to the number of cells injected and the unidentified cells capable of forming these colonies were termed CFU-S (for colony-forming-units spleen). The spleen colonies contained a variety of distinct, terminally differentiated myeloid cells and a subset of cells that were capable of forming CFU-S upon transplantation into secondary

recipients (Siminovitch et al., 1963). This indicated that CFU-S contained multilineage differentiation potential and self-renewal properties. Later experiments would show that CFU-S are derived from an even more primitive cell that also had the potential to generate lymphoid cells (Wu et al., 1968). Importantly, the cells within each spleen colony were clonal in origin (Becker et al., 1963), providing the first indication that the haematopoietic system is derived from a multipotent stem cell rather than a series of stem cells with lineage-restricted potential.

Another key observation from these clonal *in vivo* repopulation assays was the high degree of variability displayed by CFU-S in the number and types of progenitor and mature cells they produced (Siminovitch et al., 1963). This suggested that individual primitive haematopoietic cells vary in their self-renewal capacity and led to the concept that each HSC has a certain, albeit stochastic, probability of self-renewal (Till et al., 1964). This highlighted that investigations into the various stages of haematopoiesis, as well as the factors that govern HSC self-renewal, required quantitative and functional assays in which extrinsic factors could be well controlled. Moreover, the nature of early clonal *in vivo* repopulation assays restricted their application to murine haematopoiesis; whether similar principles governed human haematopoiesis remained to be determined. This prompted the development of a large number of surrogate *in vitro* assays that helped to unpick the various intermediate stages of maturation that occur between HSCs and terminally differentiated cells (Ogawa et al., 1983; Eaves, 2015). Although robust surrogates for assaying multipotent cells, the relationship between the cells capable of initiating long-term, multilineage cultures and the HSC cannot be fully determined by these assays. Operationally, HSCs can only be identified and measured by their ability to durably repopulate the haematopoietic system of a conditioned recipient following transplantation. It is for this reason that much of our understanding of haematopoiesis is generated from mouse studies. However, the development of increasingly immunocompromised mice for use in xenotransplantation, as well as robust and sophisticated *in vitro* clonal assays, has enabled significant advances to be made in defining human haematopoiesis (Dick, 2008; Doulatov et al., 2012).

The discovery of the severe combined immunodeficient (SCID) mouse, which lacks B and T cells, allowed for the first xenotransplantation experiments to be performed using primary human haematopoietic cells (Bosma et al., 1983; Doulatov et al., 2012). Formal proof that human HSCs were capable of engrafting and proliferating in immunodeficient mice was obtained when human BM cells were shown to confer long-term lymphoid- and myeloid-engraftment in conditioned recipients (Kamel-Reid and Dick, 1988; Lapidot et al., 1992). These so-called SCID-repopulating cells (SRCs) fulfilled the criteria of HSCs and hence an *in vivo* assay for the functional quantification of human HSCs was established. More robust xenograft models followed. The Scid mutation was backcrossed onto the non-obese diabetic (NOD) mouse, which exhibits defective macrophage function, generating the NOD/SCID mouse (Shultz et al., 1995; Takenaka et al., 2007). Further manipulation of the NOD/SCID mouse, through either deletion (NSG) or truncation of the IL-2R common γ chain (NOG) resulted in mice that additionally were depleted of NK cells (Ito et al., 2002; Shultz et al., 2005). The SRC xenotransplantation assay quickly emerged as the “gold

standard" surrogate assay for human HSCs, allowing for their quantification and functional characterisation following prospective isolation (Dick and Lapidot, 2005).

1.1.2. Isolating the HSC

Isolating the HSC has been a major goal of haematopoietic research for many years. In order to be defined as a stem cell, a cell must demonstrate durable self-renewal, at a clonal or single cell level, coupled with the ability to differentiate into all of the cell types that comprise the tissue. As HSCs are extremely rare within the transplantable BM (Wang et al., 1997), they must be purified from bulk differentiating cells before their function can be interrogated. Almost 30 years ago, murine HSCs became the first tissue-specific stem cells to be prospectively isolated (Spangrude et al., 1988; Ikuta and Weissman, 1992). This was achieved by fluorescent activated cell sorting (FACS) on the basis of antigens expressed on the cell surface. HSC properties were subsequently demonstrated by long-term (>3 months) multilineage engraftment upon transplantation into lethally irradiated mice. Cells positive for c-Kit (KIT) and stem cell antigen-1 (Sca-1) but negative for markers of the mature blood lineages (KSL population) were found to be highly enriched for HSCs (Spangrude et al., 1988; Ikuta and Weissman, 1992). Within the KSL subset, cells with the unique capacity for long-term multilineage reconstitution were found to reside in the $CD34^-$ fraction (Osawa et al., 1996). A similar degree of long-term reconstitution has also been found among $CD150^+CD48^-$ cells (Kiel et al., 2005). Heterogeneity has been demonstrated within HSC populations based on the length of time they are capable of sustaining multilineage haematopoiesis. Whereas long-term HSCs (LT-HSCs) are capable of maintaining haematopoiesis indefinitely, intermediate-term (IT-HSCs) and short-term HSCs (ST-HSCs) are capable generating all blood lineages but only exhibit transient engraftment for 6-8 months and 4-6 weeks, respectively (Benveniste et al., 2010).

The identification of human cells that fulfil the functional definition of HSCs poses a greater challenge than it does for mouse cells, which can be assessed in syngeneic transplantation experiments. CD34 was the first surface antigen shown to enrich for human HSCs (Civin et al., 1984), demonstrated most convincingly as CD34-enriched peripheral blood (PB) and BM cells are able to reconstitute haematopoiesis following autologous and allogeneic transplantations (Berenson et al., 1991; Dunbar et al., 1995; Link et al., 1996). Human HSCs are further enriched in the $CD34^+CD90^+Lin^-$ population, in which cells demonstrated myeloid and lymphoid differentiation potential *in vitro* and through an *in vivo* assay whereby human cells were first injected into human foetal bones which were subsequently transplanted into SCID mice (Baum et al., 1992). Using NOD/SCID mice, human HSCs were found to be exclusively contained within the $CD34^+CD38^-$ fraction (Bhatia et al., 1997). Although studies have found SRCs within the $CD34^-$ (Bhatia et al., 1998; Danet et al., 2002) and $CD38^+$ fractions (Hogan et al., 2002), the vast majority of human HSCs are $CD34^+$. In contrast to mouse HSCs, human HSCs express FLT3 (Sitnicka et al., 2003) but not CD150 (Larochelle et al., 2011). As well as CD38, CD45RA is expressed on more differentiated progeny and therefore negatively enriches for human HSCs (Lansdorp et al., 1990). Thus, the $CD34^+CD38^-CD90^+CD45RA^-Lin^-$ fraction has emerged as a highly enriched HSC

population (Seita and Weissman, 2010). However, HSC activity within this population can be even more greatly enriched by selecting the $CD49f^+$ fraction, with single cells demonstrating self-renewal and multilineage differentiation output for 30 weeks in NSG mice (Notta et al., 2011). Interestingly, a recent study demonstrated that the rare $CD34^-CD38^-CD93^+Lin^-$ fraction, found within human cord blood, BM and mobilised PB, represents a highly quiescent and immature HSC population that can give rise to $CD34^+CD38^-$ cells *in vivo* (Anjos-Afonso et al., 2013). These cells were shown to be more primitive than the $CD34^+CD38^-Lin^-$ fraction as they were more competent in serial transplantation assays and could give rise to $CD34^+CD38^-Lin^-$ cells with high self-renewal capacity. Thus, as with murine HSCs, heterogeneity exists within the human HSC population; $CD34^-CD38^-CD93^+Lin^-$ HSCs are highly quiescent and sit above the $CD34^+CD38^-CD90^+CD49f^+CD45RA^-Lin^-$ HSCs in the human haematopoietic hierarchy (Figure 1).

1.1.3. The haematopoietic hierarchy

Results from mouse transplantation studies led to the proposal of the classical model of haematopoiesis (Reya et al., 2001; Figure 1). In this model, HSCs progressively lose their self-renewal capacity and ultimately take the first step towards lineage commitment when they generate a cell that has lost its ability to self-renew. This cell retains multilineage differentiation potential and is termed the multipotent progenitor (MPP) (Morrison et al., 1997; Adolfsson et al 2001). The earliest commitment decision occurs downstream of the MPP, generating the common myeloid progenitor (CMP) and immature lymphoid progenitors (MLPs), which respectively form the myeloid and lymphoid branches of the blood system (Kondo et al., 1997; Akashi et al., 2000). Lymphoid cells consist of B, T and natural killer (NK) cells, which carry out adaptive and innate responses of the immune system. The myeloid lineage comprises a variety of functionally distinct cells including granulocytes (eosinophils, neutrophils, basophils and mast cells), monocytes/macrophages, erythrocytes and megakaryocytes. The CMP and the CLP subsequently undergo further commitment steps. The CMP gives rise to GMPs and MEPs, which become committed to produce granulocytes and monocytes or megakaryocytes and erythrocytes, respectively. On the other hand, the MLPs give rise to B cell precursors and also the earliest thymic progenitors (ETPs), which are committed to the T and NK cell lineages (Doulatov et al., 2012).

Although attractive, the picture painted by the classical model, in which restricted progenitors arise in an orderly fashion from a prototypical HSC, is an oversimplified view of haematopoiesis. HSCs may be more accurately described as groups of cells with varying developmental capacities and lymphoid-/myeloid-biased differentiation potentials (Copley et al., 2012). Indeed, lymphoid-primed multipotent progenitors (LMPPs) have been identified mediating transient myeloid and lymphoid engraftment, but with a clear bias towards the lymphoid lineage (Adolfsson et al., 2005; Månsson et al., 2007). Moreover, our knowledge of haematopoiesis is largely based on results obtained from transplantation assays using irradiated recipient mice. Recent studies using genetic-labelling strategies to track blood cell production under homeostatic conditions have yielded some surprising

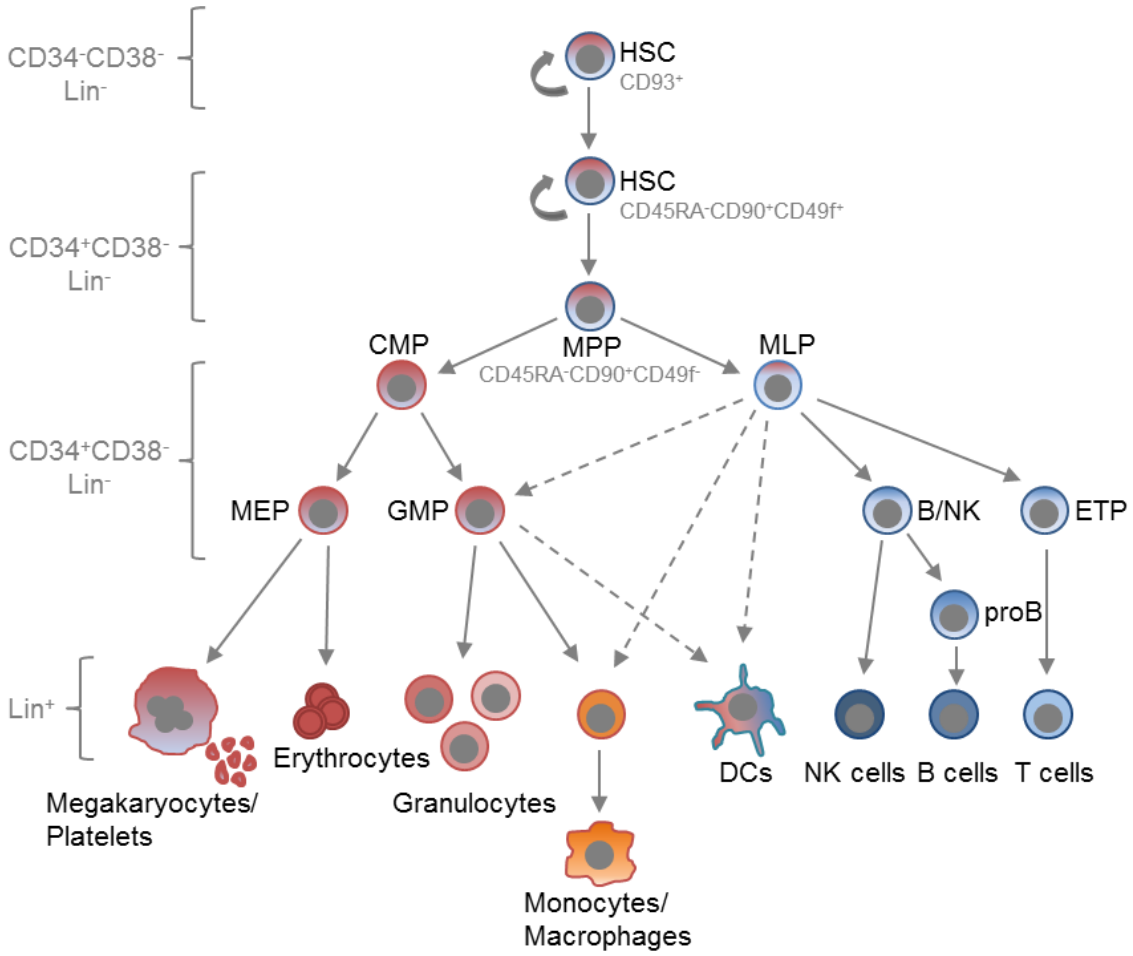


Figure 1. The human haematopoietic hierarchy.
Schematic representation of lineage commitment and differentiation during normal human haematopoiesis. Straight arrows depict differentiation and lineage relationships inferred from mouse transplantation studies. Broken arrows represent less well-defined relationships. Curved arrows indicate self-renewal. Selected cell surface phenotypes are shown which define distinct cell populations. HSC, haematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; MLP, immature lymphoid progenitor; MEP, megakaryocyte/erythroid progenitor; GMP, granulocyte/monocyte progenitor; B/NK, B/NK cell progenitor; ETP, earliest thymic progenitor; proB, B cell progenitor; DCs, dendritic cells; NK, natural killer cells; Lin, lineage markers. Image adapted from Doulatov et al. (2012).

insights (Sun et al., 2014; Busch et al., 2015). These studies suggest that ST-HSCs/MPPs, exhibiting an early myeloid/lymphoid lineage bias, are the main contributors to haematopoiesis under unperturbed conditions, with LT-HSCs making a very limited contribution. This is in contrast to transplantation assays, in which LT-HSCs are the main contributors to repopulating haematopoiesis. Additionally, results from transplantation experiments have suggested that long-term repopulation of the haematopoietic system, following its ablation by radiation, is mainly achieved by a small number of LT-HSCs. In contrast, the studies by Busch et al. (2015) and Sun et al. (2014) suggest that the composition of blood is highly polyclonal, i.e. a large number of different short-term stem and progenitor cells are recruited to drive steady-state haematopoiesis. The discrepancies between native haematopoiesis and the haematopoiesis occurring following myeloablation are likely due to the stressful conditions of the transplantation assay (Goyal and Zandstra, 2015). Nevertheless, the capacity to purify increasingly homogenous populations reflective of the various stages of haematopoietic maturation, coupled with the ability to assay their

functional potential *in vivo*, has yielded important insights into the molecular control of lineage commitment and self-renewal during haematopoiesis. Moreover, enriching and understanding the function of the cells capable of providing life-long blood production is of enormous importance for understanding BM transplantation in humans.

1.1.4. Molecular control of haematopoiesis

Maintenance of the haematopoietic hierarchy is critical for the appropriate and sustained function of the haematopoietic system; self-renewal must be maintained within the HSC compartment whereas maturing cells must lose the ability to self-renewal as they differentiate towards a specific lineage. These reciprocal processes are controlled at the molecular level by nuclear proteins called transcription factors which, together with epigenetic modifiers, orchestrate the regulation of competing genetic programmes (Rosenbauer and Tenen, 2007; Orkin and Zon, 2008). Many transcription factors exhibit lineage-specific expression patterns during haematopoiesis; for example, master regulators of myeloid development include *PU.1*, *CEBPA*, *GFI1* and *IRF8* (Rosenbauer and Tenen, 2007; Orkin and Zon, 2008). The interplay among these transcription factors and their relative expression levels during critical stages of myeloid development are key determinants of cell fate decisions (Orkin, 2000). For example, *PU.1* and *Gfi1* have been shown to directly antagonise each other's activity in the resolution of monocytic versus neutrophilic differentiation (Dahl et al., 2007). In support of this, *Gfi1*-null mice, as well as being deficient in normal neutrophils, fail to repress monocyte/macrophage gene expression programmes resulting in an expansion of this cellular compartment with age (Hock et al., 2003). Thus, these key regulators are endowed with the dual capability of promoting their own lineage differentiation whilst simultaneously repressing differentiation towards other lineages, ensuring lineage choices are resolved and reinforced.

Blood is an extraordinary tissue. From a single HSC arises multiple cell types with diverse functions encoded by distinct transcriptional programmes. A phenomenal number of these cells must be produced daily in order to meet the relentless demands of constant immune challenge and tissue oxygen consumption. Moreover, this extraordinary regenerative capacity must be maintained throughout our lifetime. The hierarchical arrangement of haematopoiesis, coupled with the tightly regulated molecular control of self-renewal and cell fate decisions, ensures faithful blood production for most of us during our lives. However, with the significant demands placed on haematopoietic cells, the accumulation of genomic alterations, such as somatic mutations and chromosomal translocations, is inevitable. Indeed, an increased mutation burden in human $CD34^+$ haematopoietic stem and progenitor cells (HSPCs) is observed with age (Welch et al., 2012). If such alterations impact on the function of key regulators of haematopoietic development, then many types of blood disorders can ensue. Indeed, the majority of transcription factor genes encoding critical regulators of haematopoiesis were identified through their intimate association with haematological disorders. For example, mutations in *GFI1* are found in severe congenital neutropenia (Klein, 2011), whereas mutations in *CEBPA* and *RUNX1* are found in familial myelodysplasia and acute myeloid leukaemia (AML) (Owen et al, 2008). Many haematopoietic

transcription factors are recurrently somatically mutated, translocated or ectopically expressed in AML (Tenen, 2003; Rosenbauer and Tenen, 2007), highlighting their significance in this aggressive blood disorder.

1.2. Acute myeloid leukaemia

Acute myeloid leukaemia (AML) is a heterogeneous group of blood neoplasms that is characterised by clonal expansion of poorly differentiated myeloid cells, termed blasts. AML blasts take over the bone marrow, severely disrupting normal haematopoiesis in the process. Thus, patients display clinical features of bone marrow failure such as increased risk of infection, symptoms of anaemia and unusual bleeding and bruising (Kumar, 2011). AML blasts may transit into the blood, elevating the white cell count, and may also infiltrate tissues such as central nervous system and the gums. AML is rapidly fatal if left untreated.

Whilst AML is a rare disease, with an incidence of approximately 4 cases per 100,000, it is responsible for approximately one third of all leukaemia types combined and more than 40% of all leukaemia-associated deaths (Cancer Research UK, 2015; Siegel et al., 2013). Although AML can affect individuals of all ages, the incidence rates rise sharply with age; the median age at diagnosis is approximately 70 years and around three quarters of newly diagnosed patients are aged 60 years or above (Derolf et al; 2009; Cancer Research UK, 2015; National Cancer Institute, 2015). Although environmental risk factors for AML include exposure to ionising radiation, benzene and cytotoxic chemotherapy (Austin, 1988; Wong, 1995; Nakanishi et al., 1999; Godley and Larson, 2008; Deschler and Lübbert, 2006), the majority of cases arise *de novo*; the exceptions being those evolving from a precursor myelodysplastic syndrome (MDS) or myeloproliferative neoplasm (MPN).

1.2.1. Treatment of AML

The standard treatment paradigm for most types of AML has seen little change for over 40 years, continuing to rely upon the use of conventional cytotoxic agents (Crowther et al., 1970). This form of therapy can be divided into two phases; induction chemotherapy, which attempts to induce complete remission (CR), and post-remission consolidation therapy, which focuses on prolonging the CR and minimising the risk of relapse (Tallman et al., 2005). Although regimens and protocols vary in different countries, the standard therapy for AML is generally considered to be one to two courses of an anthracycline (e.g. daunorubicin or idarubicin) in combination with cytarabine, followed by one or two courses of, for example, high dose cytarabine, should CR be achieved (Mayer et al., 1994; Estey, 2014). In some patients with high-risk disease, haematopoietic stem cell transplantation (HSCT) is carried out following remission. However, this decision is made on a patient-by-patient basis depending on individual fitness, donor availability and predicted response to therapy, as treatment-related death rates can be between 10-25% (Tallman et al., 2005; Cornelissen et al., 2012). Thus, the majority of AML patients face a lengthy and intensive course of chemotherapy with a substantial and often life-threatening toxicity profile. Furthermore, these

agents carry significant risk of effects later in life, such as the development of secondary malignancies.

The major exception to the standard AML treatment regimen is acute promyelocytic leukaemia (APL), which represents the most mature and curable AML subtype (Sanz et al., 2009). APL is associated with chromosomal translocations that result in the fusion of the retinoic acid receptor alpha (*RARA*) gene on chromosome 17 with one of five different partner genes, by far the most common being the promyelocytic (*PML*) gene on chromosome 15 (Grignani et al., 2000). The combination of the vitamin A derivative all-*trans* retinoic acid (ATRA) with chemotherapy emerged as the standard of care for APL as it was shown to cure ~70% of patients (Tallman et al., 2002). ATRA degrades the PML-RARA protein, ultimately relieving transcriptional repression and promoting differentiation (de Thè and Chen, 2010). Thus, APL was the first AML subtype to be treated with an agent that targets a specific genetic abnormality. In addition, arsenic trioxide (ATO) also degrades PML-RARA (de Thè and Chen, 2010) and early studies showed that it is an effective agent in the treatment of APL relapse (Douer and Tallman, 2005). Moreover, recent studies have demonstrated that the combination of ATRA plus ATO was not inferior and possibly superior to the combination of ATRA plus chemotherapy (Lo-Coco et al., 2013). Hence, ATRA plus ATO with no requirement for chemotherapy may become the new standard of care for low-risk APL patients (Coombs et al., 2015).

The successes in the treatment of APL aside, only modest improvements in AML patient survival have been observed since the early 1970s and this is primarily the result of better supportive care (Derolf et al; 2009; Othus et al., 2014). Although up to 80% of adults aged younger than 60 years achieve CR, most of these patients will ultimately relapse and die from disease within five years (Mandelli et al., 2009; Burnett et al., 2011a). The outcome is much bleaker for older patients; cure rates for those aged over 60 years are at best 10-20%, with a median survival of less than one year (Döhner et al., 2010; Wheatley et al., 2009). The dismal response to therapy observed in the elderly remains poorly understood, although it is confounded by factors such as the inability of elderly individuals to tolerate intensive induction and post-remission therapy, as well as AML in this group of individuals being more likely to be associated with poor-risk cytogenetics (Appelbaum et al., 2006; Jurcic, 2008). Recently, the DNA-hypomethylating agents azacitidine (Fenaux et al., 2010) and decitabine (Kantarjian et al., 2012) were shown to provide some benefit to a subset of older patients with AML, although overall survival rates are still very poor. As AML occurs much more frequently with age, treating this population of patients represents a major challenge in the development of drug therapy for AML.

1.2.2. Classification of AML

The principal feature of AML is a highly variable response to therapy. Therefore, classifying AML into distinct subgroups, in order to identify those patients most at risk and to inform on treatment strategies, is of huge clinical and biological relevance. In 1976 the first major strides towards a

uniform system for the classification and nomenclature of the acute leukaemias was undertaken by seven French, American and British (FAB) haematologists (Bennett et al., 1976). Based primarily on analysis of cellular morphology, as well as immunophenotype and cytochemistry, this FAB classification system recognises eight major subtypes of AML (denoted M0 to M7) according to the degree of maturation and direction of differentiation of the predominant blast population (Bennett et al., 1976; Bennett et al., 1985a; Bennett et al., 1985b; Bennett et al., 1991) (Table 1). Although this system provided a consistent morphological and cytochemical framework for classifying AML, it offered minimal prognostic information and had little impact on treatment decisions; the notable exceptions being FAB M3, which exclusively denotes the APL subtype, and M4Eo which is strongly associated with a t(8;21) and therefore confers a good prognosis.

FAB classification			
Subtype	**Description**	**Direction of differentiation**	**Defining features**
M0	Undifferentiated acute myeloblastic leukaemia	Myeloid (not lymphoid)	Minimal differentiation. Cannot be diagnosed solely by morphology. Immunological techniques required.
M1	Acute myeloblastic leukaemia without maturation	Granulocytic	Predominantly myeloblasts, evidence of granulocytic differentiation uncommon
M2	Acute myeloblastic leukaemia with maturation	Granulocytic	Myeloblasts with other immature granulocytes
M3	Acute promyelocytic leukaemia	Granulocytic	Majority abnormal promyelocytes with a characteristic pattern of heavy granulation
M4	Acute myelomonocytic leukaemia	Granulocytic/Monocytic	Both granulocytic and monocytic differentiation present
M4Eo	Acute myelomonocytic leukaemia with eosinophilia	Granulocytic/Monocytic	As M4 but with increased and abnormal bone marrow eosinophils
M5a	Acute monocytic leukaemia without maturation	Monocytic	Almost total replacement of the marrow with poorly differentiated monocytic leukaemia cells
M5b	Acute monocytic leukaemia with maturation	Monocytic	Almost total replacement of the marrow with differentiated monocytic leukaemia cells
M6	Acute erythroid leukaemia	Erythroblastic	Erythropoietic component usually exceeds 50% of nucleated cells in the bone marrow
M7	Acute megakaryoblastic leukaemia	Megakaryocytic	Immature megakaryocytes

Table 1. FAB classification of acute myeloid leukaemia. See text for references.

Shortly after the introduction of the FAB system, it became increasingly apparent that many AML cases are associated with recurring genetic abnormalities, including abnormal cytogenetics and gene mutations, the biological and clinical relevance of which could not be appreciated using the FAB criteria alone. Thus, in 2002 the World Health Organization (WHO) proposed a new system for classifying AML based on cellular morphology, cytochemistry and immunophenotype, as well as incorporating genetic and clinical features (Vardiman et al., 2002). Using these more holistic criteria, the 2008 revision of the WHO classification system identifies seven major subgroups of AML (Vardiman et al., 2009; Table 2).

WHO classification
AML with recurrent genetic abnormalities - AML with t(8;21)(q22;q22); RUNX1-RUNX1T1 - AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); CBFB-MYH11 - APL with t(15;17)(q22;q12); PML-RARA - AML with t(9;11)(p22;q23); MLL-MLLT3 - AML with t(6;9)(p23;q34); DEK-NUP214 - AML with inv(3)(q21q26.2) or t(3;3)(q21;q26.2); RPN1-EVI1 - AML (megakaryoblastic) with t(1;22)(p13;q13); RBM15-MKL1 - *Provisional entity: AML with mutated NPM1* - *Provisional entity: AML with mutated CEBPA*
Secondary AML - AML with myelodysplasia-related changes - Therapy-related myeloid neoplasms
AML, not otherwise specified - AML with minimal differentiation - AML without maturation - AML with maturation - Acute myelomonocytic leukemia - Acute monoblastic/monocytic leukemia - Acute erythroid leukemia - Pure erythroid leukemia - Erythroleukemia, erythroid/myeloid - Acute megakaryoblastic leukemia - Acute basophilic leukemia - Acute panmyelosis with myelofibrosis
Myeloid sarcoma
Myeloid proliferations related to Down syndrome
Blastic plasmacytoid dendritic cell neoplasm

Table 2. WHO classification of acute myeloid leukaemia.
According to the 2008 revised version of the WHO classification system (Vardiman et al., 2009).

The first subgroup, “AML with recurrent genetic abnormalities”, includes AML cases with recurring chromosomal rearrangements and gene mutations which influence the morphologic and clinical characteristics of AML in such a way that enables them to be defined as homogenous, biologically relevant and mutually exclusive entities. Such a clear distinction cannot be achieved for some of the most commonly recurring and prognostically significant mutations in AML, such as *NPM1*, *FLT3* and *CEBPA*, as they are not associated with any consistent morphologic and clinical features. Nevertheless, the relevance of these genes is clearly recognised; the assessment of their

mutational status is recommended in all cases of cytogenetically normal AML at diagnosis and 'AML with mutated *NPM1'* and 'AML with mutated *CEBPA'* represent provisional entities in the 2008 WHO classification system (Vardiman et al., 2009).

The WHO subgroup "AML, not otherwise specified" considers those cases that do not fulfil the characteristics of the other subgroups and mainly relies upon the FAB classification system to categorise the clinical and biological significance of the disease. At the time of the last revision, this entity accounted for approximately 30% of all AML cases. This figure will inevitably decline as the results of massive next generation sequencing efforts begin to segregate cytogenetically normal AML into more defined genetic subgroups (Patel et al., 2012; Cancer Genome Atlas Research Network, 2013).

1.2.3. Molecular heterogeneity of AML

Gross cytogenetic abnormalities, including chromosomal translocations, inversions and deletions, are found in approximately 55% of AML patients (Grimwade et al., 1998; Mrózek et al., 2001). The clinical importance of pre-treatment karyotyping for risk stratification in AML has been recognised for many years (Arthur et al., 1989; Fenaux et al., 1989; Schiffer et al., 1989). Patients can be stratified according to their cytogenetic profile into favourable, intermediate and adverse prognosis groups (Grimwade et al., 1998; Slovak et al., 2000; Grimwade et al., 2010). Nevertheless, even within these cytogenetically defined AML subsets, there remains enormous biological and clinical heterogeneity. For example, although up to 50% of AML patients present with a normal karyotype and would therefore be placed into the intermediate risk category, their response to therapy and overall survival is highly variable (Mrózek et al., 2007). However, in the post-genome era, many recurrently mutated or aberrantly expressed genes have been identified in AML, shedding light on the molecular pathogenesis and the clinical heterogeneity of this disease.

The first genes to be recognised as recurrently mutated in AML were those with known or suspected roles in normal and malignant haematopoiesis and were identified through targeted sequencing studies. These included *FLT3*, *RUNX1*, *CEBPA* and *NPM1* (Mrózek et al., 2007). More recent targeted sequencing studies identified recurrent mutations in *TET2* (Abdel-Wahab et al., 2009; Delhommeau et al., 2009). However, the advent of next generation sequencing (NGS) technology dramatically accelerated the discovery of many more nonsynonymous somatic coding mutations in AML (Riva et al., 2012). The most comprehensive NGS study to date, which analysed 200 *de novo* AML genomes, reported 23 significantly recurrent mutations (Table 3) and almost 240 genes that were found mutated in more than one patient sample (Cancer Genome Atlas Research Network, 2013). Moreover, these recurrently mutated genes can be organised into nine categories based on their biological function and putative role in AML pathogenesis (Table 4). Thus, this study not only provides a comprehensive framework for a greater understanding of the molecular classification of AML patients, but also a strong foundation for unpicking the genetic rules that govern the pathogenesis of this disease.

Gene	TCGA (2013)	Patel et al. (2012)
***FLT3* (ITD, TKD)**	27% (18, 9)	37% (30, 7)
NPM1	27%	29%
DNMT3A	26%	23%
IDH2	10%	8%
IDH1	10%	7%
RUNX1	10%	5%
TET2	9%	8%
TP53	8%	2%
NRAS	8%	10%
CEBPA	6%	9%
WT1	6%	8%
PTPN11	4%	n/a
KIT	4%	6%
U2AF1	4%	n/a
KRAS	4%	2%
SMC1A	4%	n/a
SMC3	4%	n/a
PHF6	3%	3%
STAG2	3%	n/a
RAD21	3%	n/a
FAM5C	3%	n/a
EZH2	3%	0%
HNRNPK	1%	n/a

Table 3. Recurrently mutated genes in AML.
Presented are 23 recurrently mutated genes and their mutation frequency as identified through whole-genome or -exome sequencing (Cancer Genome Atlas Research Network, 2013; n=200). As a comparison, their mutation frequencies are also shown from a large scale targeted sequencing study (Patel et al., 2012; n=398).

1.2.4. Pathogenesis of AML

More than 10 years prior to the completion of the TCGA study a “two-hit” model was proposed for AML pathogenesis (Gilliland and Griffin, 2002). This model posited that leukaemias arise from the stepwise acquisition and collaboration of mutations in one of two general classes; Class I mutations, which activate signal transduction pathways, thus conferring a survival and/or proliferative advantage to the cell (e.g. *FLT3*, *RAS*, *KIT*), and Class II mutations involving transcriptional regulators which impair differentiation (e.g. *CEBPA*, *RUNX1*, *PML-RARA*). Key to this model is the notion that the cell must acquire both class of mutation in order to become completely transformed. Results from animal models formed the cornerstone of this hypothesis (Perry and Attar, 2014). For example, an internal tandem duplication (ITD) of the *FLT3* gene is not sufficient to cause AML in mice, although it does cause a promptly lethal myeloproliferative disease

Biological category	Cases with at least 1 mutation	Example genes mutated/transclocated
Activated signalling	118 (59%)	*FLT3* (27%), *KIT* (4%)
DNA methylation	87 (43.5%)	*DNMT3A* (25%), *TET1* (1%), *TET2* (9%), *IDH1* (10%), *IDH2* (10%)
Chromatin modifiers	61 (30.5%)	*MLL*-fusions (5%), *NUP98-NSD1* (1%), *EZH2* (2%), *UTX* (2%), *ASXL1* (3%)
NPM1	54 (27%)	-
Myeloid transcription factors	44 (22%)	*RUNX1* (9%), *CEBPA* (7%)
Transcription factor fusions	36 (18%)	*PML-RARA* (8%), *MYHLL-CBFB* (5%), *RUNX1-RUNX1T1* (3.5%), *PICALM-MLLT10* (1%)
Tumour suppressors	31 (15.5%)	*TP53* (7%), *WT1* (6%), *PHF6* (3%)
Spliceosome	27 (14%)	*U2AF1* (4%), *PRPF8* (1%), *SF3B1* (0.5%), *SRSF2* (0.5%)
Cohesin complex	26 (13%)	*STAG2* (3%), *SMC1A* (4%), *SMC3* (4%), *RAD21* (2%)
All categories	**199 (99.5%)**	-

Table 4. Organisation of AML mutated genes into biologically related categories.
Number of mutated cases are presented along with their percentage mutation frequencies in brackets. Selected genes in each category are shown as examples. From Cancer Research Genome Atlas Network (2013).

(Kelly et al., 2002a); only when it is combined with other genetic lesions is AML induced (Kelly et al., 2002b). A myeloproliferative disease is also induced using a knock-in *FLT3*-ITD mouse model, with mice succumbing to disease with a median latency of 10 months (Li et al., 2008). Moreover, *in vivo* models developed to investigate the biological function of the recurrent chromosomal structural variants t(15;17), t(8;21) and inv(16), which respectively generate the *PML-RARA*, *RUNX1-RUNX1T1* and *CBF-MYH11* gene fusions, also indicated a multistep process in AML pathogenesis; they were shown to be insufficient to induce leukaemia in isolation (Yang et al., 2002), or did so with long latencies and incomplete penetrance (Moreno-Miralles et al., 2005; Castilla et al., 2004).

The fundamental aspects of the two-hit model still hold true today. The pathogenesis of AML is unquestionably a multistep process in which mutations altering distinct biological pathways are acquired over time and ultimately lead to the development of acute leukaemia. However, the functional compartmentalisation of the recurrently mutated genes in AML clearly highlights that this model is an oversimplification of AML pathogenesis (Cancer Genome Atlas Research Network, 2013). For example, according to the two-hit model, a mutation in a gene encoding a protein with a direct role in signalling would be predicted to be a ubiquitous feature of AML. However, in the TCGA study around 40% of AML cases were devoid of such a mutation. This discrepancy is likely due to the fact that this model was proposed in an era when the frequency of mutations and their impact on a variety of biological processes in AML was largely unknown. For example, the

deregulation of the epigenetic machinery has emerged as a major driving force of AML pathogenesis in recent years, with many of the recurrently mutated genes in AML being directly or indirectly involved in the epigenetic regulation of gene expression. Moreover, the identification of mutations involving components of the cohesin complex (Thol et al., 2014; Kon et al., 2013) and the splicing machinery (Cancer Genome Atlas Research Network, 2013) has revealed new pathways and a complex interplay of genomic events on the road to leukaemic transformation.

1.2.5. Clonal architecture of AML and disease drivers

As well as providing an understanding of the repertoire of mutations which contribute to AML pathogenesis, NGS studies have also generated important insights into the clonal architecture and the evolution of this disease. By performing whole genome sequencing (WGS) and subsequent analysis of the variant allele frequency (VAF), it has been shown that AML exhibits an oligoclonal anatomy, with up to four genetically related but distinct clones identified per tumour sample (Ding et al, 2012; Cancer Genome Atlas Research Network, 2013). All cells within the tumour population contain shared somatic mutations with a common ancestor cell (the "founding clone"), but additional mutations are present in subpopulations of cells that define tumour "subclones". Amongst the recurrently mutated "driver" genes in AML, a number are almost always present in the founding clone, and hence the totality of the leukaemic clone (e.g. *DNMT3A*), whereas others are almost exclusive to subclones (e.g. *FLT3*). Thus, mutations such as *DNMT3A* are considered to be "initiator" mutations as, although they are insufficient to cause acute leukaemia in isolation (Challen et al., 2012), they predispose the cell to the future acquisition of selected "cooperating" mutations that result in transformation to acute leukaemia (Corces-Zimmerman et al., 2014; Shlush et al., 2014). This clonal heterogeneity has vast biological and clinical relevance in AML as it has been shown to evolve during disease progression from MDS (Walter et al., 2012) and under the selective pressure of chemotherapy to drive relapse (Ding et al., 2012). Thus, the development of curative therapies for AML requires the eradication of not only the founding clone, but also any coexisting subclones.

Although the mutation frequency is much lower in AML in comparison to most solid tumours (Kandoth et al., 2013), AML genomes nevertheless commonly contain hundreds of mutations which are present almost ubiquitously in the cells of the tumour population (Ley et al., 2008; Ding et al., 2012). However, most of these will be random, background mutations that existed in the normal haematopoietic cell prior to the acquisition of the initiating mutation and subsequent clonal expansion (Welch et al., 2012). Hence, it is critical to identify which mutations are driving the clonal outgrowth and disease progression, and which are "passenger" mutations that merely reflect the mutational history of the cell. Although some driver mutations are easily identified as they are highly recurrently mutated, bioinformatics approaches can be applied to identify less prevalent drivers (Mazzarella et al., 2014). However, the "gold standard" for determining a driver mutation is the demonstration of a functional role in leukaemogenesis using animal models (Perry and Attar, 2014).

As well as being mutated, a gene may also function as a driver through its deregulated expression. A good example of such a gene in AML is *HOXA9*. This gene is only rarely mutated in AML, via the t(7;11) translocation which results in the formation of the *NUP98-HOXA9* gene fusion (Nakamura et al., 1996; Chou et al., 2009); however, it is frequently overexpressed as a consequence of a number of genetic and epigenetic alterations (Abramovich and Humphries, 2005). Results obtained from retroviral overexpression and BM transplantation studies have shown that *HOXA9* is capable of initiating acute leukaemia in mice, the latency of which is dramatically accelerated by the co-expression of its cofactor *MEIS1* (Kroon et al., 1998; Thorsteinsdottir et al., 2001; Thorsteinsdottir et al., 2002). Moreover, *HOXA9* and other HOX genes have been shown to be essential for the transformation and maintenance of MLL leukaemias (Ayton and Cleary, 2003; Faber et al., 2009). Thus, the overexpression of *HOXA9* is sufficient for leukaemia initiation and essential for the maintenance of certain AML subtypes and is therefore a *bona fide* driver of the disease.

1.3. Epigenetics, haematopoiesis and AML

The term epigenetics refers to a stable, somatically heritable trait that occurs independent of changes in the primary DNA sequence (Berger et al., 2009). Epigenetic regulation confers precise control over gene expression, enabling cells with identical genetic information to display distinct transcriptional outputs, thus providing a mechanism for cellular diversity. Importantly, the phenomenon of epigenetic memory ensures that cells "remember" this identity through successive rounds of cell division. The epigenetic control of gene expression is achieved through modifications of chromatin, which ultimately dictate which genes can be accessed and regulated by cellular machinery (Sharma et al., 2010). Chromatin provides the scaffold for the packaging of our entire genome. The basic functional unit of chromatin is the nucleosome, which contains 147 base pairs of DNA wrapped around an octamer of the four core histones (H2A, H2B, H3 and H4), separated by linker DNA associated with histone H1 (Kouzarides, 2007; Dawson and Kouzarides, 2012). Generally, chromatin can be divided into regions of heterochromatin, which is highly condensed and primarily contains inactive genes, and euchromatin, which is relatively open and contains most of the active genes. As well as being an efficient packaging mechanism for genetic information, this conformation allows for a large number of dynamic and highly regulated chromatin modifications. These include rather dramatic changes, such as the positional sliding of nucleosomes or the replacement of core histones with so-called histone variants, or more localised changes, such as those occurring on the N-terminal histone tails or directly to the DNA itself. Epigenetic signals or "marks" are laid down by epigenetic "writers", interpreted by "readers" and removed by "erasers", generating a "histone code" that confers upon the cell a great degree of plasticity in the regulation of its genetic blueprint (Strahl and Allis, 2000; Gardner et al., 2011).

The sum of the information encoded by these epigenetic modifications, often referred to as the "epigenome", impacts heavily on all DNA-based processes including transcription, damage repair and replication (You and Jones, 2012; Dawson and Kouzarides, 2012). Thus, genetic alterations and deregulated expression of chromatin modifiers is frequently observed in cancer and is particularly prominent in haematopoietic malignancies, including AML (Kandoth et al., 2013;

Cancer Genome Atlas Research Network, 2013; Table 4). To understand how these proteins contribute to malignancy, we must first appreciate how they function physiologically to regulate these key cellular processes. The best characterised chromatin alterations involve the covalent modifications of DNA (e.g. methylation) and histones.

1.3.1. DNA methylation

In mammals, DNA methylation primarily occurs at the 5-carbon on cytosine residues present in CpG dinucleotides to generate 5-methylcytosine (5mC). This methylation mark is not easily removed and provides a stable mechanism for silencing gene expression. CpG dinucleotides are unevenly distributed throughout the genome and this allows for a basic bimodal methylation pattern to be established early in development (Cedar and Bergman, 2009). As such, the majority of the genome contains sparsely distributed and heavily methylated CpG dinucleotides, whereas CpG-rich regions of DNA, known as “CpG islands”, remain largely constitutively unmethylated during development and differentiation and are found at around 60% of gene promoters (Wang and Leung, 2004; Jones, 2012). The exceptions include the CpG island regions surrounding the promoters of imprinted genes and those affected during X chromosome inactivation, which undergo extensive methylation during development to impose long-term transcriptional silencing (Suzuki and Bird, 2008).

There are three enzymatically active DNA methyltransferases (DNMTs) in the mammalian genome. DNMT1 is known as the “maintenance” DNA methyltransferase as it recognises hemimethylated DNA produced during replication and subsequently methylates newly synthesised CpG dinucleotides on the opposite strand, thus ensuring that methylation patterns are maintained through cell division. DNMT3A and DNMT3B act independently of replication and, although they too can recognise hemimethylated DNA, they function primarily at unmethylated regions to establish new DNA methylation patterns, hence are referred to as *de novo* DNA methyltransferases (Kim et al., 2002; Jones and Liang, 2009). The basic mechanism through which DNA methylation causes transcriptional silencing is by imposing chromatin compaction, which it can achieve through the direct or indirect recruitment of histone modifying proteins (Cedar and Bergman, 2009; Jones, 2012). This prevents transcriptional activation by denying transcription factors access to their target binding sites on DNA. This form of gene silencing was initially believed to be irreversible (Ooi and Bestor, 2008), however the process of active DNA demethylation is now widely accepted (Branco et al., 2012; Wu and Zhang, 2014). This involves the stepwise actions of the TET methylcytosine dioxygenases and the enzymes activation-induced cytidine deaminase (AID) and thymine DNA glycosylase (TDG) (Branco et al., 2012). TET proteins catalyse the conversion of 5mC to 5-hydroxymethlycytosine (5hmC), which undergoes further oxidation steps prior to removal via the base excision repair pathway.

1.3.2. Histone modifications

At least 16 distinct classes of histone modification have been described to date (including acetylation, methylation, phosphorylation and ubiquitylation) and they exert major influence over a number of DNA-templated processes such as transcription, repair and replication (Kouzarides, 2007; Tan et al., 2011). With respect to transcription, histone modifications at specific residues characterise genomic regulatory regions and may be associated with activation or repression of gene expression. Histones can be modified on many sites, predominantly localised to their N-terminal tails which protrude out from the globular nucleosome core. Each modification may impose an alternative functional output, introducing great complexity to this form epigenetic regulation. An extra level of complexity arises in the context of histone methylation; lysine residues may be mono-, di- or trimethylated, whereas arginines may be mono- or dimethylated, with each layer conferring functional significance. These modification patterns are far from static and represent a dynamic and complex landscape that evolves in a cell context-dependent fashion (Kouzarides, 2007). This is coordinated by the counteracting epigenetic writers and erasers, such as the histone methyltransferases (HMTs) and demethylases with respect to methylation, and the histone acetyltransferases (HATs) and deacetylases (HDACs) with respect to acetylation. These marks are interpreted by epigenetic readers, such as the bromodomain-containing proteins that recognise lysine acetylation, which recruit protein machinery that can further modify the local and global chromatin architecture.

In terms of transcription, histone modifications can be divided into those that correlate with gene activation and those that correlate with repression (Kouzarides, 2007; Bannister and Kouzarides, 2011). For example, histone acetylation, which occurs at lysine residues, is invariably associated with active transcription, whereas histone ubiquitylation, which also occurs on lysine residues, is associated with repression. Again, the situation regarding histone methylation is more complex as both lysine and arginine residues can be methylated to varying extent and may be associated with activation or repression of transcription. For example, the trimethylation of H3 at lysine 27 (H3K27me3) or lysine 9 (H3K9me3) are generally repressive marks, whereas trimethylation of the same histone at lysine 4 (H3K4me3) or lysine 36 (H3K36me3) are marks of active transcription. Although these trends usually accurately reflect the transcriptional activity of a given gene, it is important to note that the distribution of these marks can impact on their ability to act as markers of activation or repression. For example, methylation of H3K9 and H3K36 can have both negative and positive effects on transcription depending on whether they are located at the promoter or coding regions of the gene, respectively (Vakoc et al., 2005; Kouzarides, 2007). Moreover, certain histone modifications mark functional genomic regions. For example, monomethylated H3K4 (H3K4me1) and acetylated H3K27 (H3K27ac) are found at active enhancers, whereas H3K36me3 can be found on the gene bodies of active genes, where it may facilitate transcriptional elongation (Baylin and Jones, 2011; Guenther et al., 2007). Additionally, marks of active and repressive transcription can be found localised to the same gene. This is particularly prominent in embryonic (ES) and somatic stem cells which possess “bivalent domains” containing coexisting active (H3K4me3) and

repressive (H3K27me3) marks on the promoters of developmentally important and lineage restricted genes (Bernstein et al., 2006; Azuara et al., 2006), which is discussed further below.

Certain histone modifications can directly alter the compaction of chromatin, thus controlling accessibility to the genome. For example, the addition of a negatively charged acetyl group to the lysine side chain neutralises the lysine's positive charge, weakening its interaction with DNA. This results in chromatin decompaction, facilitating access of cellular machinery and transcriptional activation (Bannister and Kouzarides, 2011). However, it remains an open question as to whether histone modification marks are drivers of transcriptional regulation or simply correlative of the transcriptional state (Henikoff and Shilatifard, 2011; Gardner et al., 2011; Simon and Kingston, 2013). On the one hand, it can be argued that genes are not regulated by chromatin modifications *per se*, but rather it is the components of the canonical transcriptional machinery that recruit activating or repressive complexes to genomic loci that are the true modulators of transcriptional activity. Thus, in the regulation of transcriptional output, the deposition of the histone marks may be considered a secondary event to the remodelling of the chromatin architecture into a more or less permissive state. However, prior to the recruitment of the activating or repressive complexes, there is clearly an element of targeting required in order to determine which genes have been selected for activation or silencing; studies have suggested that histone modifications are a prerequisite for this. For example, the transcription factor complex TFIID contains two subunits that recognise and directly bind to histone modifications; the TAF1 subunit contains a double bromodomain through which it binds diacetylated H4 (Jacobson et al., 2000), whereas the TAF3 subunit harbours a PHD finger domain allowing it to selectively bind H3K4me3 (Vermeulen et al., 2007). Loss of the latter chromatin mark results in reduced association of TFIID with certain promoters and impaired transcriptional activity (Vermeulen et al., 2007; Gardner et al., 2011). This implies that the establishment of histone modifications is, in some cases, a requirement for the recruitment of transcription factors and subsequent transcriptional modulation. Whether histone modification marks function solely as a consequence of, or as a prerequisite for, the recruitment of the canonical transcriptional machinery is very much a "chicken or egg?" question (Gardner et al., 2011); it is likely that both arguments remain true in a gene and cell context-dependent manner.

Each histone mark can significantly influence the deposition, reading and erasing of other histone modifications, resulting in a large degree of "histone crosstalk" that tightly controls the regulation of key cellular processes (Lee et al., 2010). Extensive crosstalk also occurs between the DNA methylation machinery and histone modifying enzymes (Cedar and Bergman, 2009). For example, methyl binding proteins recognise sites of DNA methylation and recruit HDACs to the site, resulting in local deacetylation of histones and chromatin compaction (Gardner et al., 2011). Additionally, the HMT EZH2 can interact with DNMT3A and DNMT3B and may be required for their recruitment to, and stable silencing of, EZH2-target genes (Viré et al., 2006). This bidirectional crosstalk is important for the re-establishment of chromatin modification patterns following cell division, as well as the dynamic regulation of gene expression during lineage-restricted differentiation (Gardner et al., 2011). The epigenetic mechanisms that regulate lineage commitment, as well as a range of

other key cellular processes, involve the actions of the repressive Polycomb group (PcG) and activating Trithorax group (TrxG) protein complexes.

1.3.3. Polycomb and Trithorax group complexes

PcG and TrxG genes were first identified in *Drosophila* as regulators of anterior-posterior body patterning through the respective repression and activation of the Homeobox (HOX) genes (Lewis, 1978; Simon, 1995; Kennison, 1995). They act antagonistically to regulate chromatin dynamics and transcriptional output, thus determining cellular identity during the development of this organism (Schwartz and Pirrotta, 2007). Not only is this mechanism of cell fate control highly conserved, mammalian homologues of these genes are now known to play key roles in a broad range of biological processes including stem cell maintenance, differentiation, cell cycle control, DNA damage repair and tumourigenesis (Mills, 2010; Surface et al., 2010; Sauvageau and Sauvageau, 2010).

1.3.3.1. Polycomb repressive complexes

Members of the PcG protein family are structurally diverse and assemble into repressive chromatin-modifying complexes composed of multiple subunits. Although their composition is variable and cell context-dependent, two main complexes have been identified in mammals and are termed Polycomb repressive complex 1 (PRC1) and 2 (PRC2) (Sauvageau and Sauvageau, 2010). The core of PRC2 consists of one of four EED isoforms, SUZ12 and the HMTs EZH1 or EZH2. PRC2 induces transcriptional repression of target genes via the activity of EZH1/2, which contain a methyltransferase SET domain and catalyse the methylation of H3K27, performing each of the three methyl transfers to generate H3K27me3. EED and SUZ12 do not possess HMT activity but are essential for the binding of the complex to the nucleosome and the catalytic activity of EZH1/2 (Cao and Zhang, 2004a; Cao and Zhang, 2004b; Nekrasov et al., 2005). Other PRC2-associated proteins, which may help with complex recruitment or modulation of its enzymatic activity, include RBBP4/7, PCL1/2, JARID2, EPC1/2 and ASXL1 (Sauvageau and Sauvageau, 2010; Simon and Kingston, 2013).

The PRC1 family of complexes displays a greater degree of subunit variability (Simon and Kingston, 2013). At their core all PRC1 complexes contain the RING finger domain-containing proteins RING1A/B and one subunit of the PCGF proteins, such as BMI1. RING1 proteins are E3 ubiquitin ligases which catalyse the monoubiquitylation of lysine 119 on histone 2A (H2AK119ub1), a mark associated with gene repression. The mechanism(s) by which PRC1 promotes gene repression are not completely understood, but have been suggested to involve: the induction of chromatin compaction and subsequent blocking of the transcription factor machinery from binding target promoters; the impeding of RNA polymerase II (RNA pol II) initiation and elongation; and the inhibition of transcriptional activation by blocking the association of the Mediator complex with genomic loci (Simon and Kingston, 2013; Mills, 2010). PRC1 complexes that contain a CBX

isoform are able to recognise the H3K27me3 repressive mark via the chromodomain of these proteins. Thus, a linear model of gene silencing has been suggested whereby a gene is first marked for silencing by PRC2 and subsequent recruitment of PRC1 to the locus reinforces transcriptional repression (Simon and Kingston, 2009). However, PRC1 complexes exist that do not contain the CBX subunit and therefore must be targeted to chromatin through different mechanisms, which may be independent of PRC2 (Simon and Kingston, 2013).

1.3.3.2. Activating Trithorax complexes

Akin to PcG proteins, TrxG proteins assemble into multi-subunit complexes to execute their chromatin modifying functions. Although their composition is not very well characterised, TrxG protein complexes can be divided into two major classes based on their molecular function; the SET domain-containing histone modifiers and the ATP-dependent nucleosome remodellers (Schuettengruber et al., 2011).

The histone modifiers contain HMTs that establish histone marks associated with active transcription such as H3K4me3, a hallmark of active genes. The first HMT identified to specifically catalyse the mono-, di- and trimethylation of H3K4 was the yeast protein SET domain containing 1 (Set1), found within a protein complex called COMPASS (complex proteins associated with SET) (Schuettengruber et al., 2011). Six COMPASS-related complexes have been identified in mammals, which are able to methylate H3K4 through their conserved SET domain. MLL1, whose gene was originally discovered through its role in human leukaemia as part of the 11q23 translocation (Tkachuk et al., 1992), was the first COMPASS-like component to be identified. Both MLL1- and MLL2-containing COMPASS-like complexes include the subunit Menin, which is required for the localisation of the complex to chromatin (Yokoyama and Cleary, 2008). These complexes are required to establish H3K4me3 marks at HOX genes, and hence regulate their transcriptional activation (Wu et al., 2008). Additionally, MLL1 has been shown to associate with MOF, a H4K16 HAT, linking histone methylation and acetylation in the process of gene activation (Dou et al., 2005). MLL3- and MLL4-containing complexes also contain the H3K27 demethylase UTX as a subunit, suggesting that these TrxG complexes function to directly relieve PRC2-mediated gene repression in addition to their roles in sustaining active transcription at target genes (Schuettengruber et al., 2011). Interestingly, the Drosophila TrxG protein ASH1 has been shown to associate with the HAT CREB binding protein (CBP) and facilitate acetylation of H3K27 (Schwartz et al., 2010). Thus, competition for the modification of H3K27 may be another mechanism through which TrxG complexes antagonise gene repression imposed by PRC2.

In contrast to the histone modifiers, the ATP-dependent chromatin remodelling complexes regulate transcription by altering the overall chromatin architecture, through the induction of nucleosome sliding and eviction, resulting in chromatin decompaction. Different families of these TrxG protein complexes have been identified, based on the homology of their ATPase subunit, and include the SWI/SNF nucleosome remodelling complexes. In mammals, SWI/SNF complexes contain the BRM

and BRG1 proteins, which harbour bromodomains, enabling the complex to be recruited to genomic loci marked by histone acetylation (Schuettengruber et al., 2011).

As already alluded to, PcG and TrxG complexes can impose their respective repressive and activating functions independently of their intrinsic enzymatic activities. They are able to recruit, either directly or indirectly, other chromatin modifiers to target genomic sites. For example, PcG complexes are able to recruit HDACs, histone demethylases and DNMTs to facilitate gene repression (Mills, 2010). In contrast, TrxG complexes are able to recruit HATs (e.g. MOZ) and histone demethylases (e.g. UTX) to assist with transcriptional activation. This is consistent with the extensive, bidirectional crosstalk that occurs between the histone and DNA modifying machinery to confer dynamic regulation of gene expression. This dynamic behaviour is essential for the maintenance HSC self-renewal as well as the appropriate commitment to a specific lineage during differentiation.

1.3.4. Epigenetic regulation of self-renewal and lineage commitment

During differentiation from stem cell compartments, genes regulating self-renewal must become repressed, whereas those controlling commitment to a specific lineage must be activated in the appropriate cellular and temporal context. Studies of embryonic stem (ES) cell differentiation have provided key insights into how these processes are molecularly controlled.

1.3.4.1. Lessons from ES cells

ES cells have the ability to self-renew, whilst maintaining the potential to differentiate into all of the cell types of the adult, a feature known as pluripotency. For appropriate development, ES cells must respond rapidly to developmental cues and initiate lineage commitment in the correct temporal and cellular context. Genome-wide chromatin immunoprecipitation (ChIP) studies in mammalian ES cells have revealed that PcG complexes bind to and repress many genes which are enriched for transcription regulators of highly conserved developmental pathways, such as the *Hox*, *Sox*, *Fox*, *Pax*, *Irx* and *Pou* families (Boyer et al., 2006; Lee et al., 2006; Spivakov and Fisher, 2007). Accordingly, interaction of PcG proteins with chromatin at these loci is associated with increased levels of the repressive H3K27me3 mark. However, further studies revealed that many of these loci are simultaneously marked with active histone modifications, such as H3K4me3 (Bernstein et al., 2006; Ku et al., 2008; Stock et al., 2007; Azuara et al., 2006). These “bivalent” loci remain transcriptionally repressed in ES cells and appear to be in a “poised” state ready for activation, as indicated by the presence of phosphorylated RNA pol II (Stock et al., 2007). These bivalent domains are usually resolved to monovalent modifications denoting complete repression (H3K27me3) or activation (H3K4me3) upon differentiation. Thus, it has been proposed that these bivalent domains restrain lineage-specific gene expression in early stem cells whilst maintaining a poised state for rapid transcriptional activation to drive appropriate lineage commitment in response to developmental cues.

1.3.4.2. Self-renewal and lineage commitment in haematopoietic cells

Interestingly, a large number of bivalent domains have also been identified in haematopoietic stem and progenitor cells (HSPCs), suggesting that this "lineage priming" may also regulate lineage commitment in normal haematopoiesis (Cui et al., 2009; Weishaupt et al., 2010; Oguro et al., 2010; Adli et al., 2010). Cui et al. (2009) examined genome-wide distributions of chromatin modifications in human HSPCs and differentiated erythrocyte precursors and showed that haematopoietic differentiation is associated with dramatic changes in histone modifications at critical regions, such as the HOX genes. Down-regulation of HOX genes was associated with reorganisation of H3K4me3 and H3K27me3 marks at these loci (Cui et al., 2009). Master regulators of haematopoietic lineages, such as the B cell regulator *Pax5*, show high levels of both H3K27me3 and H3K4me3 in HSPCs, whereas master developmental regulators not involved in the haematopoietic lineage commitment, such as *Pax3* and *Sox3*, were enriched only for the repressive H3K27me3 mark (Adli et al., 2010). Thus, haematopoietic lineage genes appear to be specifically primed in HSCs for subsequent activation during differentiation (Cedar and Bergman, 2011). Furthermore, loss of the PRC1 complex component *Bmi1* in HSCs results in premature derepression of the bivalently-marked genes *Ebf1* and *Pax5*, resulting in accelerated lymphoid differentiation and a reduction in HSPC numbers (Oguro et al., 2010). This is consistent with results showing that BMI1 plays a crucial role in the maintenance of self-renewal and proliferation of HSCs (Park et al., 2003; Lessard and Sauvageau, 2003). In addition, over-expression of the critical PRC2 member EZH2 enhances the repopulating potential of HSCs and prevents their exhaustion (Kamminga et al., 2006). In support of a requirement for the PRC2 complex in the maintenance of HSCs, conditional knock out of *Eed*, which results in ablation of PRC2 function, results in loss of adult HSCs primarily through apoptosis (Xie et al., 2014). Taken together, these studies suggest that combinatorial effects of chromatin modifications at bivalent domains in HSPCs provide epigenetic memory for the regulation of self-renewal and lineage-specific genetic programmes.

DNA methylation also plays a critical role in the regulation of adult haematopoiesis. Analysis of genome-wide changes in DNA methylation patterns has shown that many genes that are methylated in HSPCs undergo selective demethylation in a lineage-specific manner (Ji et al., 2010). On the contrary, genes that play key roles in HSC self-renewal and proliferation, such as *Meis1* (Abramovich and Humphries, 2005; Argiropoulos et al., 2007), remain unmethylated in HSPCs and undergo progressive hypermethylation and transcriptional silencing as differentiation progresses. Although the mechanisms through which gene-specific *de novo* DNA methylation occurs are largely unknown, many of these changes occur at Polycomb-target genes and may involve the recruitment of DNMTs by PcG protein complexes (Viré et al., 2006). Thus, stable silencing mechanisms through DNA methylation may serve to strengthen lineage commitment during haematopoiesis. Indeed, DNA methylation has been shown to have a direct role in regulating HSC self-renewal and lymphoid versus myeloid cell fate decisions (Bröske et al., 2009; Trowbridge et al., 2009). HSCs with diminished *Dnmt1* expression not only displayed impaired self-renewal ability but also myeloid lineage skewing *in vivo*. This suggests that DNMT1 is required to maintain the methylation and silencing of myeloid specific regulators during normal

haematopoiesis, and their premature demethylation and inappropriate activation may force cells into the myeloid lineage pathway. Furthermore, DNMT3A has be shown to be required for the silencing of HSC multipotency genes, which is essential for their appropriate differentiation (Challen et al., 2012). Conditional knock out of *Dnmt3a* in HSCs progressively impairs their differentiation over serial transplantation, concomitant with an increase in HSC numbers in the bone marrow. Interestingly, differential methylation patterns were observed at distinct loci, such as hypermethylation at CpG islands, and this was associated with silencing of differentiation factors as well as activation of genes regulating HSC self-renewal (Cahllen et al., 2012). Moreover, TET proteins, which are essential for active DNA demethylation as they convert 5mC to 5hmC, play important roles during normal haematopoiesis. Mouse genetic studies have revealed that *Tet2*-defiency causes enhanced self-renewal of HSCs and the development of myeloid malignancies (Ko et al., 2011; Li et al., 2011; Moran-Crusio et al., 2011). Therefore, during normal haematopoiesis it is evident that, as well as histone modifications, DNA methylation patterns confer dynamic and highly regulated control over self-renewal and lineage commitment decisions.

1.3.5. Epigenetic deregulation in AML

For a cell to become malignant, an initiating event that enhances its fitness must first be acquired, stably encoded and subsequently inherited. Additional oncogenic events can then be accumulated in emerging clonal lineages, ultimately resulting in complete transformation. As genetic mutations are ideal vehicles for persistent phenotypic change, cancer has traditionally been viewed as a group of diseases based principally on genetics. However, as discussed, epigenetic events are also heritable and exert powerful effects on cellular phenotype. Although both genetic and epigenetic changes can alter transcriptional output, the stable yet plastic nature of the epigenome confers upon the cell a greater ability to adapt to selection pressures (such as anti-cancer therapies) during the evolution of the malignant clone. Thus, it is becoming increasingly appreciated that genetic and epigenetic alterations can influence each other and work cooperatively to drive cancer initiation, maintenance and progression (You and Jones, 2012; Shen and Laird, 2013). Epigenetic dysfunction, caused either directly by genetic mutations or indirectly by alterations in transcriptional networks, has emerged as a major driving force in AML (Mazzarella et al., 2014).

During haematopoiesis, multiple terminally differentiated cells with highly distinct transcriptional programmes arise from a single multipotent HSC. This is a precisely coordinated, dynamic process which relies heavily on the flexibility of epigenetic control mechanisms. This reliance on chromatin-modifying machinery is reflected in the fact that around one third of all recurrent driver mutations in AML involve genes encoding such proteins, much higher than in many solid tumours (Cancer Genome Atlas Research Network, 2013; Kandoth et al., 2013). Many of the recurrent chromosomal translocations in AML involve histone modifiers, the best studied of which are those involving the *MLL* gene. MLL is mutated by translocation with more than 60 different partners, as well as by partial tandem duplication, in 5-10% of adult leukaemias (Krivtsov and Armstrong, 2007). As discussed above, the *MLL* gene is a human homologue of the *Drosophila* trithorax, and therefore

contains a SET domain conferring methyltransferase activity for the methylation of H3K4 at target gene promoters and enhancers. Key targets that are positively regulated by MLL include the HOX cluster genes and *MEIS1*, which encode transcription factors essential for normal haematopoiesis and stem cell self-renewal (Argiropoulos et al., 2007). A common feature of MLL fusion leukaemia is the aberrant activation of these MLL target genes. However, although MLL fusions retain the N-terminal domains required for its targeting to chromatin, they all result in the loss of its C-terminal SET domain (Krivtsov and Armstrong, 2007). Thus, the primary mechanism of pathogenesis involves the aberrant recruitment of epigenetic factors, through association with the fusion partner, to MLL target genes. For example, a number of MLL fusion partners interact with DOT1L, a methyltransferase that catalyses the methylation of H3K79, a mark of transcriptional elongation (Vakoc et al., 2005). Thus, the aberrant recruitment of DOT1L substitutes H3K4me3 for H3K79me3 at target loci, such as *HOXA9* (Okada et al., 2005; Krivtsov et al., 2008). H3K4 methylation at target promoters is still required for transcriptional activation and leukaemic function however, and is provided by the ever-present wild type MLL allele (Thiel et al., 2010). Importantly, deletion or chemical inhibition of DOT1L in MLL fusion leukaemias has been shown to reduce H3K79 methylation at MLL target genes, selectively kill leukaemia cells and enhance survival (Daigle et al., 2011; Bernt et al., 2011), implicating DOT1L as a therapeutic target in MLL AML.

The role of MLL in AML represents a paradigm for how genetic and epigenetic alterations collaborate to alter transcriptional output in leukaemogenesis. A number of other recurrent translocations in AML result in the aberrant recruitment of epigenetic machinery to genes important for the regulation of self-renewal and differentiation during haematopoiesis. For example, the t(5;11) results in the fusion of *NUP98*, whose translated product binds to the HATs p300 and CBP (Kasper et al., 1999), with *NSD1*, which encodes a H3K36 methyltransferase. Again, generation of this fusion protein results in the upregulation of *HOXA9* and *MEIS1* (Wang et al., 2007). Additionally, ectopic recruitment of chromatin modifiers is essential for the functions of the transcription factor fusions observed in AML. Mutations affecting the core binding factor (CBF) complex, a key regulator of definitive haematopoiesis, are a frequent occurrence in AML (Speck and Gilliland, 2002). The most common chromosomal abnormalities in CBF leukaemias are t(8;21) and inv(16) which generate the RUNX1-RUNX1T1 and MYH11-CBF fusion proteins, respectively. RUNX1-RUNX1T1 and MYH11-CBF recruit repressive factors, such as HDACs and DNMTs, to CBF target gene promoters resulting in their transcriptional silencing and a block in differentiation (Uribesalgo and Di Croce, 2011). A similar mechanism is observed in t(15;17) APL, whereby the PML-RARA fusion protein induces extensive histone deacetylation and DNA hypermethylation at RARA gene promoters through the inappropriate recruitment of corepressors and DNMTs, resulting in their transcriptional silencing and a late myeloid differentiation block (Di Croce et al., 2002).

1.3.5.1. Polycomb deregulation in AML

Deregulation of Polycomb proteins is observed in many types of cancer, including the haematopoietic neoplasms (Radulović et al., 2013). This is not surprising given that many of the genes marked and transcriptionally suppressed by Polycomb in HSPCs encode key regulators of self-renewal and differentiation, as discussed above. Both overexpression and loss of function mutations in *EZH2* have been observed in AML and MDS, indicating that it can act as both a tumour suppressor and an oncogene (Lund et al., 2014). Thus, strict Polycomb regulation must be enforced as, depending on factors such as cell context and gene dosage, either loss of Polycomb activity or excessive Polycomb-mediated repression can lead to the inappropriate expression of oncogenes or the silencing of tumour suppressors, respectively (Mills, 2010; Sauvageau and Sauvageau, 2010). Mutations and copy number alterations in PRC2 (e.g. *EZH2*, *SUZ12* and *EED*) and PRC1 (e.g. CBX5 and CBX7) components are collectively observed in approximately 8% of de novo AML cases (Cancer Genome Atlas Research Network, 2013). The observation that around 3% of adult de novo AML cases exhibit EZH2 mutations suggest that loss of H3K27me3 may contribute to AML pathogenesis. This has been shown to be the case in MDS/MPN as a result of both *EZH2* and *ASXL1* mutations (Ernst et al., 2010; Abdel-Wahab et al., 2012). Interestingly, the H3K27 demethylase KDM6A (also known as UTX) is mutated and inactivated in around 2% of AML cases (Cancer Genome Atlas Research Network, 2013), further highlighting the complexity of balancing Polycomb function in this disease.

Polycomb proteins may also be involved in leukaemogenesis in the absence of any evidence of underlying genetic alteration. For example, overexpression of the gene encoding the PRC1 component BMI1 is commonly observed in AML (Chowdhury et al., 2007); BMI1 has been shown to be required for the self-renewal of both normal and malignant haematopoietic cells (Park et al., 2003; Lessard and Sauvageau, 2003). Moreover, the PRC2 complex is aberrantly recruited to RARA-target genes by PML-RARA and knockdown of the PRC2 component SUZ12 is sufficient to disrupt this aberrant interaction and induce granulocytic differentiation in APL (Villa et al., 2007).

1.3.5.2. Aberrant DNA methylation in AML

Global hypomethylation of DNA was the first epigenetic aberration to be observed in cancer and is a common feature of malignant cells (Feinberg and Vogelstein, 1983). However, it is now appreciated that, along with genome-wide hypomethylation, there is site-specific hypermethylation of 5-10% of the normally unmethylated CpG island promoters in the cancer epigenome (Dawson and Kouzarides, 2012). Indeed, aberrant DNA methylation patterns are also observed in AML versus normal BM HSPCs (Figueroa et al., 2010a; Bullinger et al., 2010). Moreover, distinct DNA methylation patterns were shown to be associated with well-defined genetic alterations and transcriptional outputs (Figueroa et al., 2010a; Bullinger et al., 2010; Akalin et al., 2012). Highlighting the significance of aberrant DNA methylation in the pathogenesis of AML is the observation that around 44% of AML patients display mutations in DNA methylation-related genes

(Table 4). These include the DNMTs, most significantly DNMT3A (Ley et al., 2010), and the α-KG-dependent 'DNA demethylase' TET proteins, mostly TET2 (Delhommeau et al., 2009).

Mutations in the genes encoding the metabolic enzymes IDH1 and IDH2 also significantly impact the process of DNA methylation and are frequently observed in AML (Mardis et al., 2009; Figueroa et al., 2010b; Ward et al., 2010; Marcucci et al., 2010; Paschka et al., 2010). Wild-type IDH1/2 catalyse the decarboxylation of isocitrate to α-ketoglutarate (α-KG). In contrast, the mutant enzymes produce 2-hydroxyglutorate (2HG), which is a competitive inhibitor of enzymes that are dependent on α-KG for their function, including TET2 (Rose et al., 2011; Figueroa et al., 2010b). Thus, suppression of TET2 function and DNA hypermethylation appears to be a convergent mechanism for transformation in IDH1/2 and TET2 mutant AML. Indeed, these mutations are mutually exclusive in AML and are both characterised by a hypermethylated genome (Figueroa et al., 2010b). Another class of enzymes dependent on α-KG for their function are the Jumonji family of histone lysine demethylases, providing further evidence that IDH1/2 mutations contribute to AML pathogenesis via epigenetic deregulation (Fathi and Abdel-Wahab, 2012). Interestingly, mutations in DNMT3A, IDH1/2 and TET2 occur early in the evolution of AML (Kosmider et al., 2010; Shlush et al., 2014; Welch et al., 2012; Ding et al., 2012), suggesting that the establishment of an aberrant DNA methylation pattern is an initiating event in AML pathogenesis. Recently, a potent inhibitor of the IDH2 mutant protein (AG221) has been developed and initial reports from a phase I clinical trial are promising: AG221 is well tolerated, induces sustained reduction of plasma 2HG levels as well as marked differentiation of myeloblasts that was associated with objective responses observed in six out of ten patients examined (Stein et al., 2014).

1.4. The cancer stem cell model and leukaemia stem cells

The stem cell model of cancer posits that the growth and progression of many cancers may be driven by a subpopulation of cancer stem cells (CSCs), creating a hierarchical organisation that is analogous to the architecture of normal tissues. CSCs sit at the apex of the hierarchy as they exhibit long-term proliferative potential and can undergo epigenetic changes, akin to the differentiation of normal cells, giving rise to less tumourigenic cells that form the bulk of the tumour (Reya et al., 2001; Shackleton et al., 2009; Pattabiraman and Weinberg, 2014). As such, there is phenotypic and functional heterogeneity among tumour cells. However, not all cancers appear to follow this stem cell model (Magee et al., 2012). Indeed, intratumour heterogeneity may arise among common tumourigenic cells through stochastic genetic or epigenetic changes as predicted by the clonal evolution model (Nowell, 1976; Shackleton et al., 2009). It is important to recognise however, that the CSC model and the concept of clonal evolution are not mutually exclusive in cancers that are hierarchically organised; clonal evolution can still occur in CSCs under selective pressures, such as those imposed by signalling from the microenvironment or anti-cancer treatments (Barabė et al., 2007; Shackleton et al., 2009). The CSC model has sparked great interest in the field of cancer research as it has important clinical implications; if only a subpopulation of cells within the tumour have the potential to contribute to disease progression, then these are the cells that must be targeted with therapy. It should be appreciated that the CSC

model only addresses which cells have the 'potential' to initiate disease in an experimental setting and does not necessarily address the behaviour of these cells within the patient (Shackleton et al., 2009). This argument is academic however, as rational approaches to therapy must be focused on eliminating all of the cells within the patient that have the potential to contribute to disease.

1.4.1. The hierarchical organisation of AML

As discussed above, AML is a clonal neoplastic disorder characterised by the presence of rapidly proliferating and functionally immature myeloid blasts. The cells of the leukaemic clone(s) are phenotypically and functionally heterogeneous and, analogous to normal haematopoiesis, organised into a hierarchy that is sustained by a sub-population of cells with long-term proliferative potential, often termed leukaemia stem cells (LSCs; Figure 2). The hierarchical organisation of AML has been recognised for many years (Kreso and Dick, 2014). For example, early *in vitro* studies showed that only a small percentage of AML blasts were capable of forming colonies when plated in semi-solid media (Dicke et al., 1976; Griffin and Löwenberg, 1986). Moreover, these colony-forming cells (CFCs) were shown to be immunophenotypically distinct from the bulk blast population (Sabbath et al., 1985). The hypothesis that the leukaemic clone is composed of phenotypically and functionally distinct cells acquired further support from seminal xenotransplantation studies that showed only a small population of AML blasts were capable of engrafting SCID and NOD/SCID mice (Lapidot et al., 1994; Bonnet and Dick, 1997). As these cells are able to re-initiate leukaemic haematopoiesis upon transplantation, although not necessarily capable of inducing disease, they were termed SCID leukaemia-initiating cells (LICs) and display clonal long-term repopulation and self-renewal capacity indicative of AML cells with stem cell properties. Hence, xenotransplantation assays are used as a surrogate for assaying AML LSC potential. Although these LICs were initially found to be rare, with estimated frequencies of 1 in 10^4 to 1 in 10^7 cells within the blast population, they could be enriched by selecting the immunophenotypically immature $CD34^+CD38^-$ fraction and were absent in the more mature $CD34^+CD38^+$ fraction (Bonnet and Dick, 1997). Thus, it was concluded that AML LSCs were analogous to normal HSCs, i.e. they possessed self-renewal capabilities, were rare and phenotypically immature.

1.4.2. The heterogeneity of AML LSCs

Contrary to the initial reports, LSCs have been detected in several different phenotypic compartments, varying with respect to their CD34 and/or CD38 expression (Taussig et al., 2008; Sarry et al., 2011; Eppert et al., 2011). Moreover, continuous refinement of the *in vivo* model, such as the generation of more profoundly immunodeficient mice (Shultz et al., 2005; Sanchez et al., 2009; Taussig et al., 2008), transplanting neonates as opposed to adults (Ishikawa et al., 2007) and the use of humanised mice (Wunderlich et al., 2010), has resulted in enhanced levels of engraftment and increased LSC frequencies detected. Thus AML LSCs can display immunopheno-

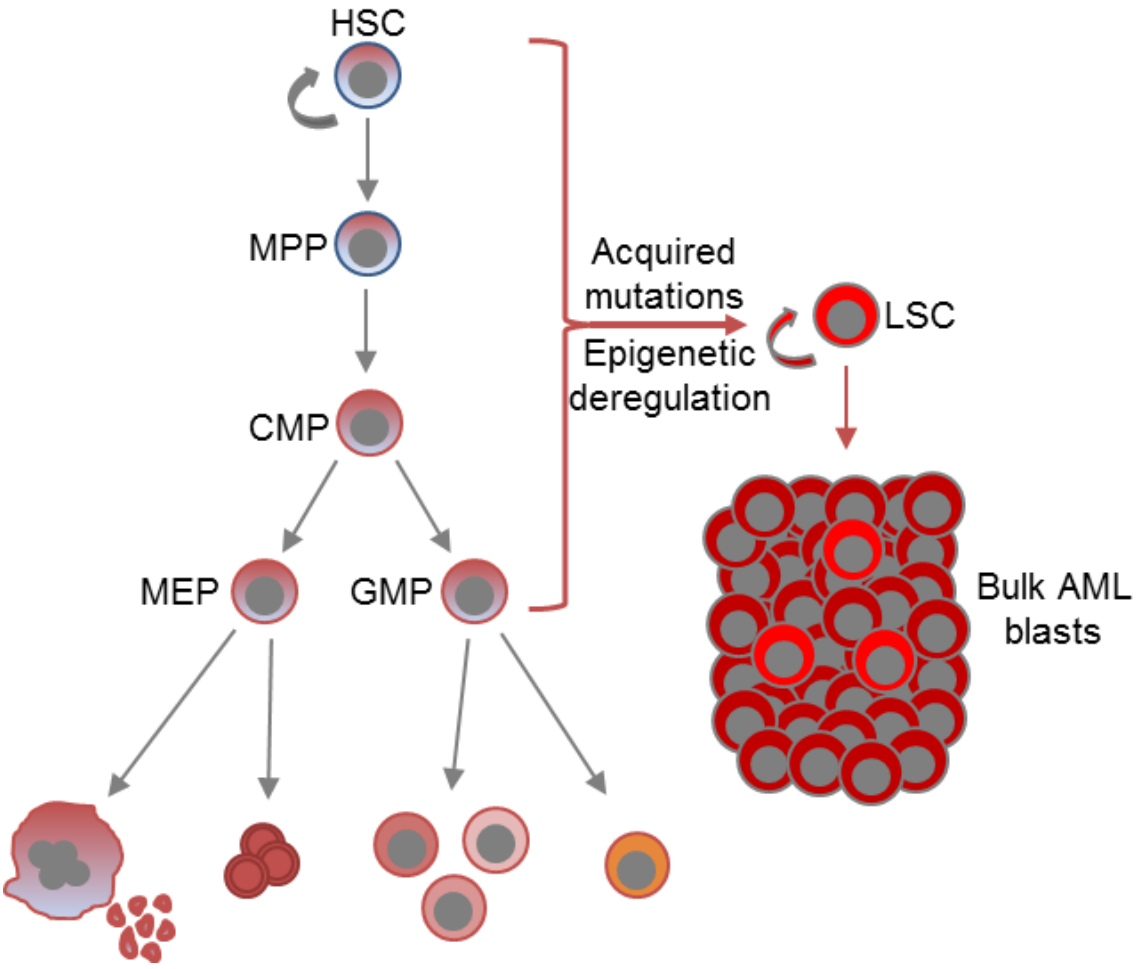


Figure 2. Hierarchical organisation of AML.
Schematic representation of AML. Normal haematopoietic stem a progenitor cells accumulate genetic mutations and epigenetic changes which ultimately lead to a myeloid differentiation block, generating highly proliferative AML blasts. The AML blast population is sustained by leukaemia stem cells (LSCs) which have the ability to self-renew. LSC, leukaemia stem cell; HSC, haematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; MEP, megakaryocyte/erythroid progenitor; GMP, granulocyte/monocyte progenitor.

types indicative of cells downstream of the HSC, i.e. committed progenitor cells, and may be more common within the leukaemic clone than initially suspected. Support for these observations comes from studies employing syngeneic mouse models of AML to investigate the nature of LSCs. For example, in murine models of MLL-translocated AML, the vast majority of LSCs express mature myeloid markers, such as CD11b and Gr1, placing them phenotypically downstream of the HSC (Somervaille and Cleary, 2006; Somervaille et al., 2009). Moreover, limiting dilution analyses suggested frequencies of LSCs approaching 1% of the bulk AML population, significantly higher than frequencies reported in xenotransplantation assays (Somervaille and Cleary, 2006; Krivtsov et al., 2006). Further experiments demonstrated that AML CFCs were enriched for LSCs and could therefore be used as a surrogate measure of LSC frequency. In this case, the LSC frequency was shown to be as high as 30% in MLL-AF9 AML (Somervaille and Cleary, 2006). Thus, transplantation experiments may significantly underreport the true frequency of AML LSCs. This is likely due to the barriers imposed by the transplantation assay, such as homing to the niche

(Somervaille and Cleary, 2006), which are much higher in xenotransplantation experiments and may be responsible for the very low frequency of LSCs reported in these assays.

The observations that AML LSCs can express mature cell markers suggests that the stem cell property of self-renewal can become associated with the characteristics of non-self-renewing progenitor cells during leukaemic transformation. This raises the concept of the 'cell of origin' in AML. This describes the cell in which the initial leukaemogenic event occurs, giving rise to the founding clone, and is not synonymous with the cell of transformation. The hierarchical organisation of the haematopoietic system has evolved to reduce its mutation burden. A small number of quiescent, self-renewing HSCs persist throughout life and give rise to highly proliferative, short-lived progenitor cells that expand rapidly and generate the functionally mature cells of the blood system. As such, long-lived HSCs are more likely than their transient progenitors to sequentially accumulate genomic alterations that ultimately result in transformation (Welch et al., 2012). Additionally, by definition, a LSC must exhibit the potential to self-renew. As the self-renewal machinery is already activated within the HSC, it can be inferred that fewer mutations may be required to transform a HSC to a LSC (Reya et al., 2001). Thus, the HSC is the most likely cell of origin in AML. Indeed, the presence of pre-leukaemic HSCs has been recently demonstrated (Shlush et al., 2014). However, experimental evidence has shown that genetic lesions targeted to non-self-renewing downstream progenitor cells can result in leukaemic transformation (Cozzio et al., 2003; Krivtsov et al., 2006; Wang et al., 2010a). In these cases, the mutations ectopically activate self-renewal mechanisms in the progenitor cells, such as the reactivation of the Wnt/β-catenin pathway or self-renewal-associated genes such as *Hoxa9* and *Meis1* (Krivtsov et al., 2006; Wang et al., 2010a). Furthermore, oncogenes that are capable of transforming committed progenitors can couple the properties of these cells, such as cell cycle status, with the aberrantly acquired ability to self-renew, hence generating LSCs that are actively proliferating and not quiescent (Somervaille et al., 2009). Therefore, both the type of genetic/epigenetic lesion, as well as the cellular context in which this lesion occurs, contributes to the biological heterogeneity of AML LSCs (Wiseman et al., 2014).

1.4.3. Clinical significance of LSCs

If it is assumed that all AML clones are sustained by a population of LSCs, then their eradication will be necessary and perhaps sufficient to cure this disease. As discussed above, the majority of AML patients that achieve complete remission (CR) following cytotoxic chemotherapy will subsequently relapse and die (Burnett et al., 2011a). Although increasing the dose of chemotherapy treatment significantly increases the CR rates in AML patients, no differences are observed in overall survival (Burnett et al., 2009; Löwenberg et al., 2009). This suggests that current therapies for AML effectively target the bulk tumour population, but in most cases residual LSCs are able to survive chemotherapy and re-establish the disease. Indeed, studies spanning the last decade have implicated AML LSCs in chemotherapy resistance and highlighted their clinical relevance. For example, a high frequency of LSCs at diagnosis was shown to correlate with high frequency of residual disease in CR following chemotherapy and this was associated with a poor

outcome (van Rhenen et al., 2005). Likewise, residual disease detected following chemotherapy has been shown to be enriched in leukaemic $CD34^+CD38^-$ cells and their abundance in CR correlated with relapse (Gerber et al., 2012). Moreover, the ability to detect AML LSCs in xenotransplantation assays has been shown to be associated with a poor outcome in AML patients (Pearce et al., 2006). More recently, stem cell related gene signatures generated from AML LSCs, immunophenotypically and/or functionally defined, have been found to be significant independent predicators of patient survival (Gentles et al., 2010; Eppert et al., 2011).

The contribution of chemotherapy-resistant LSCs to disease relapse has also been investigated using mouse models of disease. In AML-patient-derived xenografts, LSCs have been shown to be resistant to chemotherapeutic treatment and responsible for driving relapse (Ishikawa et al., 2007). AML LSCs were shown to home to the endosteal surface of the BM where they were largely non-cycling, suggesting the mechanism for chemotherapeutic resistance was their quiescent nature. Indeed, exposure of AML LSCs to granulocyte-colony stimulating factor (G-CSF) promoted cell cycle re-entry and, when combined with chemotherapy, reduced disease initiation upon secondary transplantation (Saito et al., 2010). The largely quiescent nature of AML LSCs has also been reported elsewhere (Guzman et al., 2001; Guan et al., 2003). However, mouse models of MLL-translocated AML demonstrate that LSCs can be both cycling (Somervaille et al., 2009) and resistant to chemotherapy (Zuber et al., 2009). Thus, although quiescence may be one important mechanism through which LSCs evade cytotoxic chemotherapy, reflective of their heterogeneous nature (Wiseman et al., 2014), other mechanisms of resistance may also exist (Vidal et al., 2014).

Perhaps the best support for the role of LSCs in drug resistance and sustaining disease is observed in chronic myeloid leukaemia (CML). Akin to AML, CML is a clonal haematological disorder (Fialkow et al., 1967; Fialkow et al., 1977) sustained by a rare population of LSCs (Bedi et al., 1993; Huntly and Gilliland, 2005). It is characterised by the t(9;22) which generates the *BCR-ABL* fusion gene encoding a constitutively active tyrosine kinase (Sawyers, 1999). Like normal HSCs, CML LSCs are quiescent (Holyoake et al., 1999) and capable of regenerating multilineage haematopoiesis (Wang et al., 1998). However, their differentiated progeny have a proliferative advantage that enables *BCR-ABL*$^+$ clones to displace normal haematopoiesis. CML was the first neoplasm associated with a diagnostic molecular alteration to be treated with a targeted therapy when the tyrosine kinase inhibitor imatinib mesylate emerged as a front line treatment (Druker et al., 2001; O'Brien et al. 2003). Imatinib treatment reverses the proliferative advantage of the CML progenitors over normal haematopoietic cells, eliminating the bulk of the disease and inducing complete remission. Consistent with the successful targeting of most of the CML cells, the majority of patients exhibit dramatic reductions of *BCR-ABL* transcript levels (Hughes et al., 2003). However, disease recurrence is normally observed upon withdrawal from imatinib treatment, even in patients deemed to be in complete molecular remission (Cortes et al., 2004; Rousselot et al., 2007). This suggests that, unlike their differentiated and proliferative progeny, CML LSCs are not sensitive to imatinib treatment and are therefore able to re-establish the disease after chronic drug treatment. Indeed, the presence of *BCR-ABL* can be detected in $CD34^+$ cells residing in the BM of CML patients in prolonged remission following imatinib treatment (Bhatia et al., 2003; Chu et al.,

2011). Early *in vitro* studies suggested that quiescent CML LSCs are insensitive to imatinib-induced apoptosis (Graham et al., 2002; Holtz et al., 2005) and their contribution to imatinib resistance has been subsequently demonstrated in a number of mouse models of CML (Zhang et al., 2010; Li et al., 2012; Goff et al., 2013). Thus, studies in both AML and CML demonstrate the critical role played by LSCs in the resistance to current therapies and highlight the importance of their selected targeting in the search for curative strategies for these diseases.

1.4.4. Targeting AML LSCs

As discussed above, LSCs share many properties with normal HSCs. This creates the significant challenge of identifying selective targets for the eradication of LSCs without incurring substantial toxicities arising from the elimination of normal HSCs. Despite these challenges, a number of LSC targets have been identified and some targeted-therapies are currently under investigation in clinical trials (Pollyea et al., 2014). Distinct cell surface markers have been identified that are preferentially expressed on LSCs versus HSCs and are therefore potential targets for monoclonal antibody therapies. These include CLL-1 (van Rhenen et al., 2007), CD44 (Jin et al., 2006), CD123 (Jin et al., 2009), CD96 (Hosen et al., 2007), CD47 (Majeti et al., 2009) and HAVCR2 (Kikushige et al., 2010). CD33 is expressed on AML LSCs as well as normal HSCs (Sabbath et al., 1985; Taussig et al., 2005) and can be targeted with gemtuzumab ozogamicin, which has recently been reported to modestly increase cure rates when combined with standard chemotherapy in some AML subtypes (Burnett et al., 2011b; Castaigne et al., 2012). Other strategies currently under investigation involve the inhibition of components of signalling pathways that LSCs have been found to be preferentially dependent upon, including NFκB (Guzman et al., 2002; Guzman et al., 2005) and the Src family kinase member HCK (Saito et al., 2010; Saito et al., 2013). Despite the increasing number of new agents available for the treatment of human leukaemias, no single agent has been validated for the effective targeting and eradication of the LSC. Until such therapies are discovered and developed, AML is more than likely to remain a highly incurable disease.

1.5. Project aims

A review of the literature outlined above highlights a number of key points. Firstly, more effective treatment strategies for AML are clearly required. The standard treatment for most types of AML has not changed since the 1970s and it is intensive, toxic, often life-threatening and ineffective in most cases. Secondly, there is strong evidence to suggest that LSCs are of significant biological and clinical relevance in AML and that their eradication is required to cure this disease. Therefore, a deeper understanding of biological processes that are active in AML LSCs is likely to guide the development of novel therapeutic strategies. Thirdly, transcriptional regulators play essential roles in both normal and leukaemic haematopoiesis. Although a number of haematopoietic transcription factor genes have been found to be mutated or ectopically expressed in AML, those that are expressed uniquely in AML are entirely unknown. Given the widespread epigenetic deregulation in

AML, it is likely that a number of unappreciated transcriptional regulators may be inappropriately expressed and functionally relevant in the pathogenesis of AML.

Reflective of the remarkable advances in sequencing technologies that have emerged in recent years, there has rightly been intense focus on the role mutations play in initiating and accelerating malignant transformation. Nevertheless, this focus may have resulted in the cancer research community overlooking other important and distinct oncogenic mechanisms. The work presented herein reports the discovery and functional validation of one such mechanism: the tissue-inappropriate derepression of transcriptional regulators to functional effect.

The aims of the following studies are to:

1. Identify genes encoding transcriptional regulators that are significantly expressed in AML LSCs but not normal HSPCs.
2. Characterise the functional contribution of candidate novel AML transcription factors in the pathogenesis of AML.
3. Investigate the potential mechanisms of ‘derepression’ of these transcription factor genes.

Chapter 2. Materials and Methods

2.1. Cell culture techniques

2.1.1. Cell culture conditions

All cell cultures were maintained at 5% CO_2, 37°C in a humidified atmosphere in a Leec research incubator (Leec - www.leec.co.uk).

2.1.2. Cell culture media

D10: 450mL Dulbecco's Modified Eagle's Medium (DMEM) medium 1x (Sigma Aldrich, St Louis, MO) supplemented with 2mM L-Glutamine (Life Technologies, Warrington, UK) and 50mL (10%) foetal bovine serum (FBS) (Sigma Aldrich).

R10/20: 450/400mL Roswell Park Memorial Institute (RPMI) 1640 medium 1x (Sigma Aldrich) supplemented with 2mM L-Glutamine (Life Technologies) and 50/100mL (10%/20%) FBS (Sigma Aldrich).

R20 X5: 375mL RPMI 1640 medium 1x (Sigma Aldrich) supplemented with 2mM L-Glutamine (Life Technologies), 100mL (20%) FBS (Sigma Aldrich) and 25mL (5%) X63 conditioned medium. X63 cells were a gift from Fritz Melchers and constitutively secrete interleukin 3 (IL3) (Karasuyama and Melchers, 1988).

StemSpan®: StemSpan® serum-free expansion medium (SFEM) is supplemented with bovine serum albumin (BSA), recombinant human insulin, iron-saturated human transferrin and 2-mercaptoethanol and has been optimised for the expansion of haematopoietic progenitors in the presence of cytokines (Stem Cell Technologies, Vancouver, BC).

2.1.3. Adherent cell lines

Generally, adherent cells were maintained in 75cm^2 or 225cm^2 tissue culture flasks (Corning - www.corning.com) in 20 or 40mL D10 medium respectively. Cultures were split 1:10 by trypsinisation every 2-3 days, or when the monolayer of adherent cells neared 100% confluence. This was achieved by washing the cells once in sterile PBS followed by addition of 2 or 4mL Trypsin-EDTA (1x) solution (Sigma Aldrich). Flasks were then incubated at 37°C, 5% CO_2 until cells became detached, typically within 2-3 minutes. Before replating, cells were washed once in cell culture medium to remove trypsin.

2.1.3.1. HEK 293FT cells

The 293FT cell line is a human embryonic kidney (HEK) line that combines the fast growing ability of the 293F cell line with the high titre lentiviral capacity of the 293T cell line. Constitutive expression of SV40 large T antigen permits episomal replication of lentiviral plasmids containing the SV40 origin of replication (Life Technologies).

2.1.3.2. Platinum-E cells

The Platinum-E (Plat-E) cell line is based on 293T but contains ecotropic viral structural genes for the production of functional retroviral particles capable of infecting murine cells. Expression of these structural genes is driven by the powerful EF1α promoter, resulting in high titre retroviral production (Morita et al., 2000). Plat-E cells were a gift from Toshio Kitamura, University of Tokyo.

2.1.4. Suspension cells

2.1.4.1. Human leukaemia cell lines

THP1, HL60 and MonoMac-1 cells were from DMSZ (Braunschweig, Germany). MV(4;11), U937, Fujioka and NB4 cells were gifts from Dr. Vaskar Saha (Children's Cancer Group, Manchester Cancer Research Centre). K562 cells were a gift from Prof. Caroline Dive (Clinical and Experimental Pharmacology Group, CRUK Manchester Institute). Kasumi-1 cells were a gift from Dr. Constanze Bonifer (School of Cancer Sciences, University of Birmingham). All cell lines were cultured as recommended by DMSZ and were authenticated by short tandem repeat DNA profiling in-house by the Cancer Research UK Manchester Institute Molecular Biology Core Facility.

2.1.4.2. Murine MLL-AF9 AML cells

Experimentally initiated MLL-AF9 AML cells were generated as previously described (Somervaille and Cleary, 2006). Murine HSPCs were retrovirally infected with the *MLL-AF9* fusion oncogene, expressed from MSCV-IRES-GFP vector, and were subsequently transplanted into syngeneic C57BL/6 mice. AML cells were harvested from the bone marrow (BM) and spleens of sick mice and were cryopreserved prior to further manipulation. These cells were cultured in R20 X5 medium and split 1:10 every 2 days with fresh medium to maintain a cell density of less than $5x10^5$/mL.

2.1.4.3. Murine CD117$^+$ normal haematopoietic stem and progenitor cells

Cells were harvested from the long bones of four to eight week-old mice by crushing with a mortar and pestle. Next CD117$^+$ BM stem and progenitor cells were recovered using immunomagnetic

beads and an AutoMACS Pro device (Miltenyi Biotec, Bergisch Gladbach, Germany). $CD117^+$ cells were cultured at a density of $2x10^5$-$1x10^6$/mL in R20 medium supplemented with 20ng/mL SCF, 10ng/mL IL6, 10ng/ml GM-CSF and 10ng/ml IL3 (all from Peprotech, London, UK).

2.1.4.4. Human $CD34^+$ normal haematopoietic stem and progenitor cells

Primary human normal cells were isolated from peripheral blood stem cell samples using anti-CD34 immunomagnetic beads (clone QBEND10) and the POSSELD program of an AutoMACS Pro device (both from Miltenyi Biotec) according to the manufacturer's instructions. For liquid culture, cells were placed in StemSpan® (Stem Cell Technologies) at a density of $2.5x10^5$/mL supplemented with 100ng/mL each of FL, IL6, TPO and SCF (all from Peprotech).

2.1.5. Cryopreservation of Cells

In order to maintain sustainable long-term stocks of cell lines, aliquots of early passage cells were frozen down and stored in liquid nitrogen. For this, $1\text{-}2x10^6$ cells were suspended in freeze mix (90% FBS, 10% DMSO), transferred to cryovials (Nunc - www.thermofisher.com) and placed in a Mr Frosty™ container (Nalgene - www.nalgenelabware.com) at -80°C. After 24 hours the vials were transferred to liquid nitrogen for long-term storage.

2.1.6. Cell thawing

2.1.6.1. Cell lines

Cryovials were removed from liquid nitrogen and rapidly thawed in a 37°C water bath. Cells were then transferred to a suitable centrifuge tube and made up to 10mL with the desired (and pre-warmed) culture medium and centrifuged for 4 minutes at 290xg. The pellet was resuspended in desired tissue culture medium, plated in a culture flask and incubated in standard conditions.

2.1.6.2. Primary cells

Recovery of primary cell samples was facilitated by thawing cells directly into PBS supplemented with 20% human serum albumin, 4µg/mL DNaseI, 2.5mM $MgCl_2$ and 16.4mM trisodium citrate ("DAMP solution"). Cells were removed from liquid nitrogen and immediately thawed in a 37°C water bath. Cells were transferred to a 15 mL sterile falcon tube and recovered by drop-wise addition of 10mL of DAMP solution over a 20 minute period. This step was performed at room temperature to enhance the activity of the DNaseI, with constant agitation to prevent clumping of the cells. The cells were centrifuged at 200xg for 10 minutes and after discarding the supernatant, the pellet was washed twice in DAMP solution, resuspended in StemSpan® and cultured as

described above in section 2.1.4.4. '*Human CD34$^+$ normal haematopoietic stem and progenitor cells*'.

2.1.7. Cell counting and Trypan Blue exclusion

For all experiments the viability of cells at the start of culture was confirmed by Trypan blue dye exclusion (Life Technologies) and accurate cell counts were achieved by haemocytometer cell counting, for which a Neubauer BS.748 haemocytometer chamber was used (Hawksley - www.hawksley.co.uk).

2.2. Human tissue and ethical approvals

Use of human tissue was in compliance with the ethical and legal framework of the United Kingdom's Human Tissue Act, 2004. Normal CD34$^+$ HSPCs surplus to requirements were from patients undergoing autologous transplantation for lymphoma or myeloma. Cells were mobilised using chemotherapy and G-CSF as described (Lee et al., 2005). Their use was authorised by the Salford and Trafford Research Ethics Committee (08/H1004/114) and, for samples collected since 2006, following the written informed consent of donors. Normal human BM was collected with informed consent from healthy adult male donors, with the ethical approval of the Yorkshire Independent Research Ethics Committee (D1330N00007). Primary human AML samples were from Manchester Cancer Research Centre's Tissue Biobank (instituted with approval of the South Manchester Research Ethics Committee). Their use was authorised following ethical review by the Tissue Biobank's scientific sub-committee, and with the informed consent of donors. Primary AML cells are referred to in this report by their Biobank number to maintain patient confidentiality.

2.3. Mice and transplantation experiments

Experiments using mice were approved by the Cancer Research UK Manchester Institute's Animal Ethics Committee and performed under a project license (70/7766) issued by the United Kingdom Home Office, in keeping with the Home Office Animal Scientific Procedures Act, 1986. C57BL/6 (CD45.2$^+$) mice were purchased from Harlan (Shardlow, UK). B6.SJL-*Ptprca Pepcb*/BoyJ (CD45.1$^+$) mice were purchased from Jackson Laboratories (Bar Harbor, ME) and bred in-house.

2.3.1. Murine transplantation experiments

Prior to retroviral transduction, BM cells were harvested from the long bones of four to eight week-old CD45.1$^+$ mice, as described in section 2.1.4.3. '*Murine CD117$^+$ normal haematopoietic stem and progenitor cells*', and CD117$^+$ cells were incubated overnight in R20 with added growth factors: 20ng/mL SCF, 10ng/mL IL6, 10ng/mL GM-CSF and 10ng/mL IL3 (all from Peprotech, London, UK) to promote cell cycle entry. Next, cells were spinoculated, as described in section 2.4.2.3 '*Murine*

CD117^{+}*normal haematopoietic stem and progenitor cells*'. Following spinoculation, cells were incubated overnight in R20 with growth factors to allow for expression of antibiotic resistance in transduced cells prior to the addition of puromycin (3µg/mL; Sigma-Aldrich) and, when required, neomycin (1.5mg/mL; Sigma-Aldrich). Following 96 hours of antibiotic selection $1x10^6$ viable cells (as determined by Trypan blue dye exclusion) were injected into the tail veins of lethally irradiated (800 cGy) recipient CD45.2^{+} mice. Mice receiving doubly transduced cells received in addition $2x10^5$ CD45.2^{+} unfractionated BM cells. Tail bleeds were performed every four weeks over a 16 week period and blood engraftment levels and lineage contributions determined by flow cytometry analysis of leukocyte populations following ammonium chloride lysis of erythrocytes. When transplanted mice exhibited signs of ill health they were euthanised and leukaemia cells from BM and spleen cryopreserved for later use. For secondary transplantations, cryopreserved BM samples were thawed rapidly and $1x10^5$ viable cells were injected into the tail veins of sub-lethally irradiated (450 cGy) recipient CD45.2^{+} mice.

2.4. Manufacture of lentiviral and retroviral particles and infection of mammalian cells

2.4.1. Polyethylenimine transfection into packaging cells

Polyethylenimine (PEI) 25kD linear transfection reagent (Polysciences, Warrington, PA) was prepared in ddH_2O to a concentration of 1mg/mL. PEI condenses DNA into positively charged particles to facilitate uptake into target cells. The day prior to transfection, 293FT or Plat-E cells, for the manufacture of lentiviruses or retroviruses respectively, were plated out into 10cm dishes (BD Biosciences – www.bdbiosciences.com) at a density of $4.5x10^6$ cells per dish in 9mL D10. Next day, the cells were typically at ~90% confluence. For the transfection, 21ug PEI was firstly diluted in room temperature, serum-free DMEM to a total volume of 500uL per 10cm dish. For the manufacture of lentiviral particles, plasmids containing viral structural genes were combined with lentiviral expression plasmids and diluted as follows:

Plasmid DNA for the manufacture of lentiviral particles	
shRNA or expression construct in lentiviral vector	4µg
pCMVΔ8.91 (containing *gag/pol* genes necessary for active lentivirus)	2µg
pMDG.2 (containing *env* genes for pseudotyping)	1µg
Serum-free DMEM	to 500µL

Table 5. Plasmid DNA for the manufacture of lentiviral particles by 293FT cells.

For the manufacture of retroviral particles, Plat-E cells have been developed to produce functional retroviral particles without addition of further viral plasmids (Morita et al., 2000). Expression of *gag-pol* and *env* are each driven by an EF1α promoter within two pMX-IRES-EGFP endogenously expressed plasmids. MSCV retroviral expression plasmids were diluted as follows:

Plasmid DNA for the manufacture of retroviral particles	
Gene of interest in retroviral vector (i.e. MSCV)	7µg
Serum-free DMEM	to 500µL

Table 6. Plasmid DNA for the manufacture of retroviral particles by Plat-E cells.

Equal volumes of the diluted PEI and the diluted plasmid DNA constructs were then combined, gently mixed by pipetting and left to incubate for 20-30 minutes at room temperature to allow sufficient formation of DNA-PEI complexes. The mixture was then added drop wise to the respective dish of 293FT or Plat-E cells. After rocking the plates to ensure an even distribution, they were incubated overnight. The following day the medium was replaced with 10mL of fresh, pre-warmed D10 medium per dish prior to further overnight incubation and subsequent harvest of viral particle-containing supernatants. Depending upon cell viability, viral particles were again harvested after a further overnight incubation. All viral supernatants were filtered through a 0.45µm polyethersulfone filter prior to use (Corning). Retroviral supernatants were used immediately or kept at 4°C for no more than 2-3 days prior to use. Lentiviral supernatants were either used immediately or stored long-term at -80°C. All virus-containing medium was routinely decontaminated in 2% Trigene/Distel laboratory disinfectant (Starlab, Milton Keynes, UK) for 24 hours prior to disposal.

2.4.2. Viral infection of target cells and selection of transduced cells

Prior to target cell infection, all viral supernatants were supplemented with the cationic polymer Polybrene® (Millipore, Billerica, MA) at a final concentration of 8µg/mL to increase transduction efficiency. Polybrene® aids the infection process by neutralising the negative charge repulsion between viral particles and cell membranes.

2.4.2.1. Human cell lines

For lentiviral infection of human cell lines $1x10^6$ cells were resuspended in 2mL viral supernatant in a 12-well tissue culture plate (Corning) and centrifuged for 30 minutes at 900xg and 37°C (a process called 'spinoculation' or 'spinfection'). After centrifugation, cells were incubated at standard conditions overnight prior to the addition of 2mL of fresh pre-warmed R10 medium the following morning to reduce Polybrene® toxicity. For human cell lines infected with shRNA constructs, 24

hours following spinoculation 3µg/mL puromycin (Sigma-Aldrich) was added for 48 hours to select for successfully transduced cells prior to further manipulation. For human cell lines infected pLentiGS-EGFP constructs, GFP^+ cells were sorted using a FACS Aria II® or an Influx® flow cytometer (both from BD Biosciences) as described in section 2.13.3. '*FACS analysis and sorting*'.

2.4.2.2. Primary $CD34^+$ cells

For lentiviral infection of primary $CD34^+$ cells, recovered cells were incubated overnight in StemSpan™ supplemented with growth factors (FL, IL6, TPO and SCF), as described above in section 2.1.4.3. '*Human $CD34^+$ normal haematopoietic stem and progenitor cells*'. Next day, $2.5x10^5$ cells per well were cultured overnight in a CELLSTAR® 24-well suspension culture plate (Sigma-Aldrich) with 0.5mL lentiviral supernatant supplemented with Polybrene® and the same added growth factors. An additional 0.5mL StemSpan™ plus growth factors was added the following morning. 24 hours following infection 2µg/mL puromycin was added and cells were then incubated for 48 hours to select for transduced cells prior to further manipulation.

2.4.2.3. Murine $CD117^+$ normal haematopoietic stem and progenitor cells

For retroviral infection of primary murine HSPCs, $CD117^+$ BM cells were incubated overnight in R20 with growth factors (SCF, IL6, GM-CSF, IL3), as described in section 2.1.4.3. '*Murine $CD117^+$ normal haematopoietic stem and progenitor cells*', to promote cell cycle entry. $1.5\text{-}2.5x10^5$ cells per well were then spinoculated (30 minutes, 900xg, 37ºC) with 1mL retroviral supernatant in the presence of these cytokines and Polybrene® in 24-well plates (Corning). For co-transduction experiments cells were spinoculated sequentially with fresh retroviral supernatant. Following spinoculation, cells were incubated overnight in R20 with growth factors to allow for expression of antibiotic resistance in transduced cells prior to the addition of puromycin (3µg/mL) and, when required, neomycin (1.5mg/mL; Sigma-Aldrich). Cells were then incubated for 48 hours to select for transduced cells prior to further manipulation.

2.4.2.4. $Hoxa9^+$ cells

Murine HSPCs successfully transduced in pairwise combination with retroviral vectors expressing Hoxa9, under neomycin selection, plus FOXC1 (*Hoxa9*/*FOXC1*) or an empty vector control (*Hoxa9*/MTV), both under puromycin selection, were subsequently infected with a lentiviral *KLF4*-expressing vector using the same infection protocol as for normal murine HSPCs. 6µg/mL blasticidin (Sigma-Aldrich) for 5 days was used to select for *KLF4*-expressing cells.

2.5. RNA extraction and quantitative PCR (qPCR)

2.5.1. RNA extraction

RNA extraction was performed using the RNeasy® Plus Micro kit (for $5x10^5$ cells or less) or Mini kit (for greater than $5x10^5$ cells) and QIAshredder spin columns (Qiagen - www.qiagen.com). Cells were washed twice in PBS and lysed by vortexing in 350µL of RLT lysis buffer supplemented with 1% β-mercaptoethanol. Cell lysate was subsequently passed through a QIAshredder spin column for homogenisation and homogenised lysates were then passed through a gDNA eliminator spin column to remove genomic DNA contamination. 350µL of 70% ethanol (VWR international - www.vwr.co.uk) was added prior to loading the sample onto a MinElute spin column. Following several washes of the column and a 5 minute high speed spin to remove residual ethanol from the column, RNA bound to the column was eluted with RNase-free water. RNA yield was quantified through spectrophotometric analysis using a Nanodrop 1000® spectrophotometer.

2.5.2. cDNA production

For reverse transcription, between 1µg and 100ng of extracted RNA from each cell population was diluted in 10µL of nuclease-free water and combined in a 200µL thin-walled PCR tube (VWR International - www.vwr.co.uk) with 10µL of a reverse transcriptase 'mastermix' (High Capacity Reverse Transcription kit; Applied Biosystems - www.appliedbiosystems.com), made up as follows:

Reverse Transcriptase 'mastermix'	
Components	**Volume**
10x RT buffer	2µL
25x dNTPs (100mM)	0.8µL
10x RT Random Primers	2µL
MultiScribe™ Reverse Transcriptase (50U/µL)	1µL
RNase Inhibitor (20U/µL)	1µL
Nuclease-free H_2O	3.2µL
Total	**10µL**

Table 7. Reverse transcriptase mastermix for cDNA production.

Samples were then transferred to a Bio-Rad DNA Engine Dyad Thermal Cycler® (Bio-Rad - www.bio-rad.com) and treated as follows:

Reverse transcription reaction		
Step	**Temperature (°C)**	**Duration (minutes)**
Incubation	25	10
Extension	37	120
Reaction inactivation	85	5

Table 8. Thermal cycling conditions for the reverse transcription reaction.

The cDNA generated was diluted with nuclease-free water to an appropriate concentration (typically 10ng/µL) and either used immediately as template in a qPCR reaction or stored at -20°C for future use.

2.5.3. qPCR assay and data analysis

All qPCR assays were performed in MicroAmp® optical 384-well reaction plates (Applied Biosystems) and analysed using an Applied Biosystems 7900HT Sequence Detection System (SDS)**.** Reactions were performed in triplicate and typically involved 1-25ng cDNA per well and included primers for the analysis of β-Actin (*ACTB*) as a housekeeping gene loading control. No template controls were also included for detection of non-specific signal. qPCR analyses involved either 20x TaqMan® primer/probe assays (Life Technologies) or separate TaqMan® probes from the Universal Probe Library (UPL) (Roche - www.roche.com) in conjunction with specific oligonucleotides (IDT - www.idtdna.com) which offered extensive transcriptome coverage. qPCR 384-well plate set-up for both systems was as follows:

qPCR mix for 20x TaqMan® assays	
Components	**Volume (µL) for 1x 384 well**
2x TaqMan® Fast Universal PCR Mastermix	5
20x TaqMan® assay	0.5
cDNA	1-25ng
Deionised water	up to 10µL

Table 9. qPCR mix for 20x TaqMan® assays.

qPCR mix for specific oligonucleotides and UPL probes	
Components	**Volume (µL) for 1x 384 well**
2x TaqMan® Fast Universal PCR Mastermix	5
Forward primer (10µM)	0.25
Reverse primer (10µM)	0.25
Gene-specific probe (10µM)	0.25
cDNA	1-25ng
Deionised water	up to 10µL

Table 10. qPCR mix for specific oligonucleotides and UPL probes.

Reactions within the 7900HT qPCR machine were performed as follows:

qPCR reaction			
Step	**Temperature (ºC)**	**Duration (seconds)**	**Number of cycles**
AmpliTaq® DNA polymerase activation	95	20	1
Denaturation	95	1	40
Annealing/ Extension	60	20	

Table 11. Thermal cycling conditions for the qPCR reaction the 7900HT system.

Applied Biosystems SDS 2.1 software, provided with the 7900HT system, was used to convert the raw fluorescence data from the qPCR plate reader to CT values for each individual reaction. The CT value corresponds to the point at which the amplification cycle reaches the exponential phase, determined empirically by setting a specific threshold. Normalising for housekeeping gene expression generates the ΔCT value, which can be further normalised to baseline samples (such as a non-targeting or empty vector control) to generate the ΔΔCT allowing relative quantitation of mRNA transcript levels.

2.5.4. Single cell qPCR analysis

Single cell qPCR analysis was performed on FACS-purified BM AML blasts ($CD45^+CD34^+CD117^+$) from Biobank patient 259. Single cell capture, cell lysis, reverse transcription and pre-amplification (PreAmp) was performed using the Fluidigm® C1™ system and a small (5-10µm) C1 integrated fluidic circuit (IFC), according to the manufacturer's instructions. Briefly, FACS-purified AML blasts

were resuspended at a concentration of $8x10^5$/mL in SM buffer and 60µL of the cell suspension was combined with 40µL of C1 suspension reagent. 5µL of this 100µL cell suspension mix, containing approximately 2000 cells, was then pipetted into the appropriate cell inlet of a primed C1 IFC. The IFC was then loaded onto the C1 system and cells were drawn through the chip and captured in cell capture sites by microfluidic technology within the chip. Successfully captured cells (72 in total) were confirmed by microscopy and the locations of their capture sites were recorded. The IFC was then prepared for the lysis, reverse transcription and PreAmp steps, which were all performed on the C1 system. A PreAmp step is required to enrich for loci of interest whilst maintaining relative abundance between loci and was achieved by preparing a 180nM pooled TaqMan primer mix containing the assays for the genes of interest. The following thermal cycling conditions were used for the reverse transcription and PreAmp steps on the C1 system:

Reverse Transcription		
Step	**Temperature (ºC)**	**Time (minutes)**
Incubation	25	10
Extension	42	60
Reaction inactivation	85	5

Table 12. Thermal cycling conditions for the reverse transcription reaction on the C1.

Pre-amplification			
Step	**Temperature (ºC)**	**Time**	**Cycles**
Enzyme activation/ RT inactivation	95	10 min	1
Denaturation	95	15 sec	18
Annealing/extension	60	4 min	
Hold	4	∞	∞

Table 13. Thermal cycling conditions for the pre-amplification reaction on the C1.

The cDNA amplicons were then harvested and gene expression analysis performed with a 96.96 Dynamic Array® IFC using the Biomark™HD system. The PCR thermal protocol on the Biomark™HD system was as follows:

qPCR reaction			
Step	**Temperature (ºC)**	**Duration (seconds)**	**Number of cycles**
Thermal mix	70	2400	1
	60	30	1
DNA polymerase activation	98	60	1
Denaturation	97	5	35
Annealing/ Extension	60	20	

Table 14. Thermal cycling conditions for the qPCR reaction the Biomark™HD system.

Fluidigm® Real-Time PCR Analysis software was used for data analysis.

2.5.5. qPCR primers and probes

The following TaqMan® primer/probe assays (Life Technologies) were used for qPCR detection in this study:

TaqMan® primer/probe assays used for qPCR	
Gene	**Assay ID**
FOXC1	Hs00559473_s1
IRX5	Hs04334749_m1
KLF4	Hs00358836_m1
RUNX1	Hs01021971_m1
CEBPA	Hs00269972_s1

Table 15. List of the TaqMan® primer/probe assays used for qPCR.

Listed overleaf are the separate primer and probes from the Universal Probe Library (UPL) system (Roche) used for qPCR detection of each gene in this study.

UPL primers and probes used for qPCR			
Gene	**Species**	**Primer**	**Probe #**
ACTB	Human	Fwd: ATTGGCAATGAGCGGTTC Rev: GGATGCCACAGGACTCCAT	11
FOXC1	Human	Fwd: TGAACGGGAATAGTAGCTGTCA Rev: GGACGTGCGGTACAGAGAC	11
HOXA9	Human	Fwd: AAAACAATGCTGAGAATGAGAGC Rev: TATAGGGGCACCGCTTTTT	3
IRX3	Human	Fwd: AAAAGTTACTCAAGACAGCTT Rev: GGATGAGGAGAGAGCCGATA	57
MYB	Human	Fwd: TGCTCCTAATGTCAACCGAGA Rev: AGCTGCATGTGTGGTTCTGT	56
RPS5	Human	Fwd: CTCAGGCTGTGTTCTCAGGAT Rev: TCCACTTCCCAAAGAGCTTG	53
18S	Human	Fwd: GCAATTATTCCCCATGAACG Rev: GGGACTTAATCAACGCAAGC	48
Actb	Mouse	Fwd: TGACAGGATGCAGAAGGAGA Rev: CGCTCAGGAGGAGCAATG	106
Foxc1	Mouse	Fwd: GCTTTCCTGCTCATTCGTCTT Rev: AAATATCTTACAGGTGAGAGG	34
Hoxa9	Mouse	Fwd: TCCCTGACTGACTATGCTTGTG Rev: GTTGGCAGCCGGGTTATT	25
Meis1	Mouse	Fwd: GACGCTTTAAAGAGAGATAAAGATGC Rev: CATTTCTCAAAAATCAGTGCTAAGA	103
Mafb	Mouse	Fwd: GCAGGTATAAACGCGTCCAG Rev: TGAATGAGCTGCGTCTTCTC	80
Klf4	Mouse	Fwd: GCTCCTCTACAGCCGAGAATC Rev: ATGTCCGCCAGGTTGAAG	10
Myc	Mouse	Fwd: CCTAGTGCTGCATGAGGAGA Rev: TCCACAGACACCACATCAATT	77

Table 16. List of the UPL primers and probes used for qPCR.

2.6. Western blotting

2.6.1. Cell lysis

Cells to be lysed were first counted, pelleted by centrifugation and resuspended twice in ice cold PBS in order to wash away media and any debris from cell culture. Cells were lysed in ice-cold high salt lysis buffer (HSLB; 45mM HEPES (pH 7.5), 400mM NaCl, 1mM EDTA, 10% Glycerol, 0.5% NP40, 6.25mM NaF, 20mM β-glycerophosphate, 1mM DTT, 20mM sodium butyrate and 1x Protease Inhibitor cocktail (Roche, Burgess Hill, UK)), typically at concentrations of $1x10^6$ cells in 100uL of lysis buffer. Pelleted cells were resuspended in HSLB by pipetting and vortexed at the highest setting for 5 seconds prior to rotation at 4ºC for 15 minutes. Samples were then centrifuged at 20,000xg, 4ºC for 15 minutes to pellet cell debris and the supernatant was collected. Lysates were stored at -80ºC.

2.6.2. Gel electrophoresis

Proteins were separated by SDS-PAGE. Equal amounts of lysate were diluted in ddH_2O containing 10x NuPAGE® sample reducing agent and 4x NuPAGE® lithium dodecyl sulphate (LDS) sample loading buffer (both from Life Technologies). Samples were then incubated at 95ºC for 10 minutes in order to ensure complete unfolding of the protein secondary structure. Lysates were then loaded into pre-cast NuPAGE® 4-12% Bis-Tris acrylamide gels in a gel tank filled with 1x MOPS® running buffer (50mM MOPS, 50mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.7) to ensure electric conduction. Gels, tanks and MOPS® running buffer were all from Life Technologies. For molecular weight estimation, 5µL of PageRuler Plus prestained protein ladder (ThermoFisher) was run together with the samples. Empty wells were filled with 4x NuPAGE® LDS loading buffer diluted in ddH_2O to ensure an even run of the samples. Gels were electrophoresed for at 150-180 volts for approximately 1 hour, until the lowest molecular weight marker was close to the bottom of the gel.

2.6.3. Nitrocellulose membrane transfer

Following electrophoresis the pre-cast gel was transferred to a nitrocellulose membrane (Whatman Protram® - www.ge.com). Transfer was performed at 4°C at 70 volts for 1 hour 15 minutes in a semi-dry transfer tank (Bio-Rad) filled with transfer buffer. Transfer buffer was prepared by diluting 50mL of transfer buffer 10x solution (30g Tris and 143g Glycine made up to 1L with deionised water) and 100mL of methanol (Fisher Scientific) with deionised water to a final volume of 500 mL. Following completion of transfer the nitrocellulose membrane was retrieved from the apparatus and stained with Ponceau Red (Sigma Aldrich) in order to confirm equal loading of the samples and successful transfer to the nitrocellulose membrane.

2.6.4. Nitrocellulose membrane incubation

Following Ponceau Red staining, membranes were rinsed with tap water and cut with a sterile scalpel to isolate proteins of the appropriate molecular weight for subsequent staining. Ponceau Red was washed away with 1x PBS-Tween (prepared from a 20x stock solution consisting of 560g NaCl, 14g KCl, 100.8g Na_2HPO_4, 16.8g KH_2PO_4, 70mL Tween20 diluted in deionised water to a final volume of 10L), prior to blocking with 5% skimmed milk in 1x PBS-Tween for 30 minutes at room temperature to reduce non-specific binding. Residual milk was washed away with 1x PBS-Tween and primary antibody incubation was performed on rollers at 4ºC overnight. Primary and secondary antibodies were either diluted in 5% milk in 1x PBS-T or 'magic mix' consisting of 5% BSA (Sigma Aldrich) and 2% Western Blocking reagent (Roche) in 1x PBS-Tween. After 3x 10 minutes washes with 1x PBS-Tween, membranes were incubated with secondary Horseradish peroxidase (HRP)-linked secondary antibodies (GE Healthcare - www.gelifesciences.com) on rollers for 30 minutes at room temperature. After 5 further 10 minute washes with 1x PBS-Tween, membranes were incubated with either ECL (enhanced chemiluminescence; GE Healthcare) or

Supersignal (Pierce, Rockford, IL, USA) and the signal generated by the HRP-conjugated immune complexes was exposed using a high performance chemiluminescence film (Amersham™ Hyperfilm - www.ge.com) and an X-ray cassette and detected using a Curix 60 film processor (AGFA - www.agfahealhcare.com) in a dark room.

Antibodies used for western blotting were as follows:

Primary antibodies for western blotting			
Antibody	**Company (catalogue #)**	**Dilution**	**Diluent**
FOXC1	Santa Cruz (sc-21396)	1:1000	5% milk
FOXC1	Cell Signaling (8758)	1:1000	Magic mix
ACTB	Millipore (MAB1501R)	1:10,000	Magic mix
LSD1	Cell Signaling (2184)	1:1000	Magic mix
IRX3	abcam (ab25703)	1:1000	Magic mix
HOXA9	Millipore (07-178)	1:1000	Magic mix

Table 17. List of primary antibodies used for western blotting.

2.7. Subcellular fractionation

In order to detect IRX3 protein expression in murine MLL-AF9 AML cells, nuclear proteins were isolated from $2x10^6$ cells using a subcellular protein fractionation kit (ThermoFisher; product #78840) according to the manufacturer's instructions. IRX3 protein was detected as described in section 2.6. '*Western blotting*'.

2.8. Immunoprecipitation followed by western blotting

Immunoprecipitation (IP) experiments were performed using lysate generated from HEK 293FT cells exogenously expressing FLAG-tagged-HOXA9 or MYC-tagged-FOXC1 in pairwise combination or with an empty vector control. HEK 293FT cells were plated and transfected as described in section 2.4.1. '*Polyethylenimine transfection into packaging cells*', with the following exceptions; 10µg of total plasmid DNA was combined with 30µg PEI and transfections were performed for 6 hours instead of overnight. 48 hours following transfection, cells were washed twice with ice-cold PBS and lysed in 1mL of ice-cold TNN buffer (50mM Tris-Cl (pH7.5), 100mM NaCl, 5mM EDTA, 0.5% Nonidet p40) supplemented with 6.25mM NaF, 20mM β-glycerophosphate, 1mM DTT and 1µL Benzonase® nuclease (Sigma Aldrich), by rotation at 40rpm, 4ºC for 15 minutes. Samples were then centrifuged at 20,000xg, 4ºC for 15 minutes to pellet cell

debris and 10µL of the supernatant per sample was taken for input control, the rest was used for the IP. For the IP, Protein G Sepharose® Fast Flow beads (Sigma Aldrich; 20µL per sample) were washed 3 times in TNN buffer before resuspending in 100µL TNN buffer with 5µL of the appropriate antibody (either anti-FLAG (Sigma, cat# F1804) or anti-MYC-tag (Cell Signaling, cat# 2276) or isotype control. The beads were incubated with antibodies for a minimum of 2 hours at 4ºC with constant rotation. Following this incubation, beads were centrifuged at 1700xg at 4ºC for 1 minute, combined with the prepared lysates and rotated at 40rpm, 4ºC overnight. The following morning the antibody-bound beads were centrifuged (1700xg at 4ºC for 1 minute) and washed 4 times in 1mL TNN buffer before resuspending in 20µL elution buffer (10x NuPAGE® sample reducing agent, 4x NuPAGE® LDS). Proteins bound to the beads were eluted by heating the samples for 10 minutes at 70ºC and the beads were subsequently removed by centrifuging the sample through a 0.45µm Spin-X® centrifuge tube filter within a 2mL DNase/RNase-free polypropylene tube (costar®, Corning). Immunoprecipitated and co-immunoprecipitated proteins were assayed by western blotting as described in section 2.6. '*Western blotting*'.

2.9. Chromatin immunoprecipitation followed by qPCR

In order to interrogate the chromatin marks present in AML cell lines in which *FOXC1* is either highly expressed (THP1 cells) or not expressed (K562 cells) chromatin immunoprecipitation (ChIP) followed by qPCR analysis was performed. ChIP was performed on $5x10^6$ cells per IP using a HighCell# ChIP kit (Diagenode - www.diagenode.com) for the H3K27me3 and H3K4me3 chromatin marks, according to the manufacturer's instructions. Briefly, 1µg of antibody or IgG control was bound to Protein-A magnetic beads by rotation for 2-4 hours at 4ºC. $5x10^6$ THP1 or K562 cells per IP were washed twice in PBS and cross-linked with 1% formaldehyde for 8 minutes before the reaction was stopped by the addition of 125mM glycine. Cells were then centrifuged (300xg, 4ºC, 5 minutes), washed twice with ice-cold PBS supplemented with 20mM sodium butyrate and subsequently lysed with L1 and L2 lysis buffer. Lysed cells were next resuspended in 200µL shearing buffer S1 containing protease inhibitor within 1.5mL Bioruptor™ Microtubes (Diagenode) and the released crosslinked DNA-protein complexes sheared using a Bioruptor™ sonicator (UCD-200; Diagenode); lysates were sonicated for 15 cycles of 30 seconds 'ON' and 30 seconds 'OFF' on the highest setting. Following shearing, samples were centrifuged for 10 minutes at 10,000xg and 4ºC to remove residual cellular debris and combined with the antibody-bound Protein-A magnetic beads and rotated overnight at 4ºC; 10% of the supernatant containing sheared chromatin was kept and used as the input sample. Following extensive washing, DNA-protein complexes were de-crosslinked with proteinase K and DNA eluted from the beads with DNA isolation buffer by two successive 15 minutes incubations at 55ºC and 100ºC on a heat block. Eluted DNA was transferred to new microcentrifuge tubes and used immediately for qPCR analysis or stored at -20ºC.

Antibodies used for ChIP were as follows:

ChIP antibodies		
Antibody	**Company (catalogue #)**	**Conc. (µg)**
H3K27me3	Diagenode (C15410069)	1
H3K4me3	Diagenode (C15310003)	1
Rabbit IgG	Diagenode (K01541008)	1

Table 18. Antibodies used for ChIP.

qPCR primers and probes used to assess chromatin occupancy around the FOXC1 promoter were as follows:

UPL primers and probes used for qPCR		
Primer (distance from *FOXC1* TSS*)	**Primer sequence**	**Probe #**
-962	Fwd: AATGAGAGCGAGCCAGCA Rev: GTGCAGACCCTTCCTTCG	81
-541	Fwd: GGTGACGGATGCTCAAAAGT Rev: CACCTTTTCTCGGTCTCTCG	28
-519	Fwd: AGAAGTTTTCCCAATGCTTCC Rev: CGTCACCTTTTCTCGGTCTC	28
+231	Fwd: AAGCCGCCCTATAGCTACATC Rev: TTCAGGGTGATCTTCTTGTCC	9

Table 19. UPL primers and probes used for ChIP-qPCR of the *FOXC1* promoter region.
*TSS = transcription start site

2.10. Colony Forming Unit (CFU) Assays

For clonogenic assays of human $CD34^+$ cells, cells were cultured in methylcellulose medium (H4320, Stem Cell Technologies) at a starting density of 10^4/mL with 10ng/mL SCF, 10ng/mL IL6, 50ng/mL GM-CSF, 20ng/mL IL3, 10ng/mL IL11, 50ng/mL TPO, 10ng/mL FL, (all from Peprotech), 50ng/mL G-CSF (Chugai, London, UK) and 4U/mL EPO (Janssen Cilag, Beerse, Belgium). When required, puromycin was added at a concentration of 2µg/mL. Colonies were enumerated after 14 days. For clonogenic assays of human AML cell lines, a similar approach was followed although cytokines were not added, starting cell densities were lower ($2\text{-}5x10^3$/mL), puromycin was used at 3µg/mL and colonies were enumerated after 7-10 days. Clonogenic assays of retrovirally transduced murine $CD117^+$ BM cells were performed in methylcellulose medium (M3231, Stem Cell Technologies) containing 20ng/mL SCF, 10ng/mL IL6, 10ng/mL GM-CSF and 10ng/mL IL3 (Peprotech) with puromycin (3µg/mL) and neomycin (1.5mg/mL) as required. Starting culture

densities were $2x10^3$-$5x10^4$/mL and colonies were enumerated after 4-5 days. For serial replating, the cells were washed with PBS and resuspended as single cell suspensions; starting culture densities were $2x10^3$-$2x10^4$/mL and the same growth factors and selection antibodies as previous rounds were maintained.

2.11. Cytospin analyses

2-$5x10^4$ cells were resuspended in 150µL PBS and, through centrifugation at 60x*g* for 5 minutes, were spun onto a microscope glass slide and left to air dry. Cells were fixed by incubation in methanol for 10 minutes followed by May-Grunwald (Sigma; diluted 1:1 with Sorenson's Buffer (33.3mM KH_2PO_4, 64.75mM Na_2HPO_4, pH 6.8)) staining for 20 minutes and subsequent staining with Giemsa (Sigma; 10x diluted with Sorenson's Buffer) for 30 minutes. Finally, stained cells were washed under running tap water and left in Sorenson's buffer for five minutes prior to one final brief wash with tap water. Slides were left to air dry before cells were permanently mounted with a coverslip and DPX neutral mounting media (VWR, Radnor, PA). Images were obtained using a Leica SCN400 histology scanner (Leica, Solms, Germany) and analysed using SlidePath Gateway software v1.0 (Leica).

2.12. Routine microscopy

Cell counts and assessment of viability were performed with an Axiovert 40 CFL microscope (Zeiss - www.zeis.com). Enumeration of colonies in semi-solid media was performed with an Olympus CK2 inverted microscope (Olympus - www.olympus.com) and images of colonies were acquired using an Axiocam camera (Zeiss).

2.13. Flow cytometry analysis

2.13.1. Staining medium buffer

Staining Medium (SM) buffer was used as cell suspension media in all FACS experiments, for incubation with antibodies for immunophenotypic analysis as well as FACS purification. SM buffer consisted of 479mL phenol red free RPMI 1640 medium (Sigma Aldrich) supplemented with 15mL (3%) FBS and 1mM EDTA (Fisher Scientific).

2.13.2. Red Cell Lysis

To lyse red cells from mouse peripheral blood samples prior to FACS analysis or purification, cells were re-suspended in 5x sample volume of red cell lysis buffer and incubated on ice for 10 minutes followed by dilution and 2 wash steps with 5mL SM buffer. Following centrifugation, cells were re-

suspended SM buffer containing the appropriate cocktail of antibodies. Red cell lysis buffer consisted of 8.3g NH_4Cl, 1.0g $KHCO_3$ (both from Sigma Aldrich) and 1.8mL 500mM EDTA solution made up to 1 litre with distilled water before being passed through sterile, non-pyrogenic 500mL filter system containing a 0.22µm polyethersulfone filter (Corning).

2.13.3. FACS analysis and sorting

Flow cytometry analyses were performed using a LSR Model II flow cytometer (BD Biosciences, Oxford, UK). Cell sorting experiments were performed using either an Influx® or a FACS Aria II® flow cytometer (both from BD Biosciences) with assistance from members of the Imaging and Cytometry Core Facility at the Cancer Research UK Manchester Institute. Antibodies were diluted and incubated with cells in 100µL SM buffer on ice for 15 minutes in the dark. GFP^+ cells were resuspended in SM buffer and immediately sorted. Staining of cells for flow cytometry analyses and sorting was performed using the following conjugated antibodies obtained from eBioscience (Hatfield, UK):

FACS antibodies used for analysis and sorting				
Target species	**Antigen**	**Fluorochrome**	**Clone**	**Dilution**
Human	CD117	APC	104D2	1:200
	CD11b	PE	ICRF44	1:200
	CD14	FITC	61D3	1:200
Mouse	CD45.1	APC	A20	1:200
	CD45.2	PerCP-Cy5.5	104	1:200
	CD117	APC	2B8	1:200
	CD11b	PE	M1/70	1:200
	Gr1	PE-Cy7	R86-BC5	1:200
	F4/80	eFluor®450	BM8	1:100
	B220	PE	RA3-6B2	1:200
	CD19	PE-Cy7	eBio1D3	1:200
	TCRβ	APC-eFluor®780	H57-597	1:100
	CD8a	PE	53-6.7	1:500
	CD4	eFluor®450	GK1.5	1:200

Table 20. List of antibodies used for FACS analysis and sorting.

2.13.4. Apoptosis analysis

Apoptosis was assessed using a BD Pharmingen APC-Annexin V kit (Oxford, UK). Cells were stained for Annexin V and 7-aminoactinomycin (7-AAD), diluted according to the manufacturer's instructions and incubated with cells in 100µL 1x Binding Buffer (provided with the kit) for 15 minutes in the dark. Apoptotic cells were evaluated by flow cytometry. Annexin V binds with high affinity to phosphatidylserine expressed on the surface of early apoptotic cells, whereas late stage apoptotic cells are positive for 7-AAD due to loss of plasma membrane integrity.

2.13.5. Cell cycle analysis

Cells were stained with propidium iodide and their DNA content assessed by flow cytometry. 1-2x10^5 cells were first washed with PBS then carefully resuspended in an ice-cold solution of 70% methanol in PBS and fixed at -20ºC overnight. Following fixation, cells were washed twice with PBS and resuspended in a propidium iodide staining solution (20µg/mL propidium iodide (Sigma Aldrich) and 0.5mg RNase (Sigma Aldrich) in PBS) and incubated at 37ºC for one hour. Immediately following incubation, cells were analysed on a LSR Model II flow cytometer (BD Biosciences) whilst in the PI staining solution. RNase was prepared by resuspending RNase powder (Sigma Aldrich) in a 10mM Tris (pH 7.5), 15mM NaCl solution to a concentration of 10mg/mL. This solution was then heated to 100ºC for 15 minutes to eliminate any contaminating DNase prior to sterile filtration.

2.13.6. Flow sorting of normal BM populations

Normal human BM populations were isolated by Dr. Xu Huang as described in detail elsewhere (Huang et al., 2014). From the same samples at the same time eosinophils were isolated from the $CD34^-$ population as side $scatter^{hi}CD14^-CD15^-CD16^-$ cells. To purify megakaryocytes, $CD61^+$ cells were selected from 5x10^7 fresh nucleated BM cells following ammonium chloride red cell lysis of erythrocytes, using anti-CD61 microbeads (clone Y2/51) (Miltenyi Biotec) and the POSSEL program of an AutoMACS Pro device (Miltenyi Biotec). Selected cells were stained with the following antibody cocktail (each at 1/50 dilution, except where stated): anti CD61 APC (Y2/51) (Dako, Glostrup, Denmark), anti-CD14 Pacific Blue (TUK4) (Invitrogen, Paisley, UK), anti-CD15 biotin (HI98) (eBioscience, San Diego, CA) & streptavidin-conjugated Pacific Orange (Invitrogen) (used at 1/200 dilution) and anti-CD16 PECy7 (3G8) (BD Biosciences, Franklin Lakes, NJ). Megakaryocytes were sorted as forward $scatter^{hi}CD61^+CD14^-CD15^-CD16^-$ cells. The purity of sorted populations was confirmed by morphological analysis of May-Grunwald-Giemsa stained cytospin preparations.

2.14. Molecular methods

2.14.1. Polymerase Chain Reaction (PCR)

All PCR reactions were prepared on ice in 200µL thin-walled PCR tubes (VWR International) using the following typical reagents:

PCR mix		
Components	**Amount**	**Final concentration**
5x HF buffer	10µL	1x
10mM dNTPs	1µL	200µM
Forward primer (10µM)	2.5µL	0.5µM
Reverse primer (10µM)	2.5µL	0.5µM
Template DNA	10ng	0.2ng/µL
DMSO	1.5µL	3%
Phusion® DNA polymerase	0.5uL	0.02 U/µL
Deionised water	up to 50µL	-

Table 21. Typical PCR mix components with Phusion® DNA polymerase.

Amplification of DNA fragments was performed using the proof-reading Phusion® DNA polymerase (New England Biolabs (NEB) - www.neb.com), as its 3'-5' exonuclease activity guarantees higher fidelity (mutation rate: $4.4x10^{-7}$) using the supplied HF buffer. For amplification of GC-rich regions, or for when PCR reactions required further optimisation, the supplied GC buffer was used.

The PCR tubes were then transferred to a Bio-Rad DNA Engine Dyad Thermal Cycler® (Bio-Rad) and programmed as follows:

PCR reaction			
Step	**Temperature (ºC)**	**Duration (seconds)**	**Number of cycles**
Initial denaturation	98	30	1
Denaturation	98	10	34
Annealing	Mean T_m of primers ±3ºC	45	
Extension	72	30/Kb	
Final extension	72	600	1
Hold	4	∞	∞

Table 22. Thermal cycling conditions for PCR with Phusion® DNA polymerase.

PCR products were immediately visualised on a 1% agarose gel as described below in section 2.14.3. '*Agarose gel electrophoresis*'.

2.14.2. Endonuclease restriction enzyme digestion

Endonuclease restriction enzyme digestion was performed either to obtain DNA fragments with compatible ends for ligation or during diagnostic digest screens for plasmids generated by a ligation reaction. All restriction enzymes in this study were from NEB. Typically, either 1µg or 5µg of vector DNA was respectively digested in a 20µL or 50µL reaction for 1 or 3 hours on a 37˚C heating block. All reactions contained the appropriate buffers for each enzyme as well as BSA, as recommended by NEB. Following restriction digestion, DNA was run on a 1% agarose gel in order to visualise the excised fragment for further purification or for screening the results of a ligation reaction.

2.14.3. Agarose gel electrophoresis

PCR products and digested DNA fragments were resolved by electrophoresis on 1% agarose gels, prepared by dissolving 1g of ultra-pure agarose powder (Life Technologies) in 100mL of 1x TAE buffer in a microwave oven. 1x TAE buffer was prepared in deionised water from a 50x stock solution containing 54.4g of $NaAc.3H_2O$, 96.9g of Tris, 7.44g of EDTA and 40mL of acetic acid in a final volume of 1L of deionised water. The agarose solution was then cooled by contact with fresh running tap water before adding 5µL of ethidium bromide (Sigma-Aldrich). After gentle mixing, the agarose solution was poured into a sealed gel rig with a spacer comb which determined the size and number of lanes present on the gel. Once set, gels were transferred to an electrophoresis tank containing enough 1x TAE to immerse the gel fully. 10x DNA loading buffer, consisting of bromophenol blue 0.25% and glycerol 60% in double distilled water, was added to each DNA

sample (1x final concentration) before they were loaded into the wells. One lane was reserved for a DNA marker ladder (1 kb plus, Life Technologies) in order to determine resolved fragment sizes. The gels were electrophoresed for 30-45 minutes at 120 volts and DNA bands were visualised under ultraviolet (UV) light using the Geneflash Gel documentation system (Syngene - www.syngene.com).

2.14.4. DNA fragment gel extraction

Bands containing the desired DNA fragments were excised using a sterile scalpel following visualisation using an AppliGene PRB-073 UV transilluminator (www.qbiogene.com). The DNA fragments were then purified using a MinElute Gel Extraction Kit (Qiagen) as per the manufacturer's instructions. Briefly, the gel slice containing the DNA fragments was dissolved in 3x volume (w/v) of buffer QG by incubation on a 50ºC heat block and gentle vortexing. One gel volume of isopropanol was then added and, following mixing by inversion, the sample was centrifuged (11,300xg for 1 minute) through a MinElute spin column. Following washing steps with buffer QG and buffer PE (containing isopropanol) the DNA was eluted with 10-20µL of elution buffer. Purified DNA fragments were then immediately used for ligation.

2.14.5. Ligation reaction

T4 DNA ligase (NEB) was used for the ligation of the compatible ends of gel-purified DNA fragments and linearised vectors. DNA ligase catalyses the formation of a phosphodiester bond between 5'-phosphate and 3'-hydroxyl termini in double-stranded DNA. Briefly, ligation reactions were performed using 100ng of vector DNA with the quantity of insert DNA determined by a 1:3 molar vector:insert ratio (http://www.insilico.uni-duesseldorf.de/Lig_Input.html). 20µL reactions were performed with 10x T4 DNA ligase buffer (NEB) and 1µL of enzyme for 1 hour or overnight at 16ºC. As controls, mock ligations containing only insert or only vector were also performed.

2.14.6. Transformation of competent cells

Competent cells are specific strains of *Escherichia coli* treated with rubidium chloride in order to increase their ability to uptake exogenous plasmid DNA. For long-term storage competent cells were kept at -80ºC in 200µL aliquots. DH5α cells (Life Technologies) were commonly used for transformation with MSCV-based retroviral expression vectors. Lentiviral expression vectors were instead transformed into Stbl3 competent cells (Life Technologies) in which the genes mediating homologous recombination between lentiviral long terminal repeats (LTR) have been removed in order to reduce unwanted recombination and provide higher yield from both mini- and maxi-preps (see section 2.14.7. '*Plasmid preparation techniques*').

For transformation, 3µL of ligated product was carefully added to competent cells that were thawed on ice and the mixture was stirred gently. The competent cells and plasmid DNA were kept on ice for 30 minutes to allow the association of plasmid DNA with bacterial membranes. The cells were subsequently heat-shocked for 45 seconds in a 42ºC water bath to induce uptake of the plasmid DNA, followed by recovery on ice for 2 minutes. 500µL of LB medium was then added to each tube and cells were incubated at 37ºC for 1 hour in an SM30 orbital incubator (Edmund Büler GmbH - www.edmund-buehler.de). Finally, cells were resuspended in 100µL of LB medium, spread on LB agar plates containing the appropriate antibiotics and incubated overnight at 37ºC. The next day sterile tips were used to pick cells from single colonies which were transferred to antibiotic-containing LB medium for expansion prior to plasmid DNA extraction. Each colony formed on the plate derives from proliferation of a single transformed antibiotic-resistant bacterial cell.

2.14.7. Plasmid preparation techniques

Following the successful ligation of a DNA fragment and the appropriate vector, small scale (mini-prep) or larger scale (maxi-prep) plasmid preparations were performed as described below.

2.14.7.1. Mini-prep

Small-scale plasmid DNA preparations allowed for the presence of the correct insert and its orientation to be verified by performing a diagnostic endonuclease restriction digest and/or DNA sequencing analysis. For mini-prep, single colonies were picked from LB agar plates using sterile pipette tips and placed in 6mL of Terrific Broth (see section 2.15.1. '*Bacterial culture medium*' for details) containing the appropriate antibiotic (typically 1% ampicillin (Sigma-Aldrich)) in a 14mL sterile Falcon tube (Scientific laboratory Supplies (SLS) - www.scientificlabs.eu) and incubated overnight (12-16 hours) at 37ºC in an SM30 orbital incubator (Edmund Büler GmbH). 1mL of the bacterial cell culture was removed and stored temporarily at 4ºC in a 1.5mL Eppendorf tube (these cells were later used for maxi-prep starter cultures following confirmation of successful uptake of the appropriate insert in the correct orientation); the remaining bacterial cells were then pelleted by centrifugation at 2350xg for 5 minutes and media was aspirated. Mini-prep was performed using a commercially available kit (Nucleospin® Plasmid; Macherey-Nagel - www.mn-net.com) according to manufacturer's instructions.

Briefly, bacterial cells were resuspended by pipetting and vortexing in 250µL of A1 resuspension buffer supplemented with RNaseA (100mg/mL stock solution; supplied with the kit). An equal volume of A2 lysis buffer was then added and the contents were mixed by inversion before a 5 minute room temperature incubation step to allow lysis of the bacterial cells. Protein and carbohydrates released were then precipitated by adding 300µL of A3 buffer and pelleting the debris by centrifugation at 17,900xg for 10 minutes. Supernatant was then loaded onto a Nucleospin® column by centrifugation at 17,900xg for 1 minute. Sequential washes with 500µL of

AW wash buffer and then 700µL of A4 buffer were subsequently performed, also by centrifugation at 17,900xg for 1 minute. Finally, residual ethanol was removed from the column by centrifugation at 17,900xg for 10 minutes, prior to elution of purified plasmid DNA from the column with 50µL of EB buffer and centrifugation at 17,900xg for 1 minute. Eluted plasmid DNA was collected into clean 1.5mL microcentrifuge tube and stored at -20ºC.

2.14.7.2. Maxi-prep

Maxi-preps were performed to produce large stocks of DNA for use in long-term or large-scale experimental procedures. Starter cultures were initially prepared from the appropriate 1mL bacteria cell culture that had been confirmed by mini-prep and subsequent DNA analysis to contain the desired insert in the correct orientation. Starter cultures were prepared by adding the 1mL of bacteria cells to 4mL of Terrific Broth containing 1% ampicillin in a 14mL sterile Falcon tube (SLS) and incubating for 8 hours at 37ºC in an SM30 orbital incubator (Edmund Büler GmbH). Cells were then transferred to 300mL of Terrific Broth supplemented with 1% ampicillin in a 1L conical flask (Pyrex - www.pyrex.com) to continue culture overnight (12-16 hours) at 37ºC in an SM30 orbital incubator. The following day, maxi-prep was performed using a commercially-available kit (Nucleobond® Xtra Maxi; Macherey-Nagel) according to the manufacturer's instructions.

Briefly, bacterial cells were pelleted in sterile buckets by centrifugation at 3,345xg for 15 minutes and resuspended in 24mL of RES resuspension buffer supplemented with RNAseA (100mg of powder supplied with the kit dissolved in 1L of RES buffer). Complete resuspension was performed by vortexing before 24mL of LYS lysis buffer was added. Contents were then mixed by inversion prior to incubation at room temperature for 5 minutes to allow for lysis of bacterial membranes. The lysis reaction was then stopped and proteins and carbohydrates precipitated by adding 24mL of NEU neutralisation buffer with mixing by inversion. The mixture was then poured directly into a EQU solution-primed Nucleobond® column containing filter paper. After one additional wash using 25mL of EQU solution the paper filter was discarded and each column was washed with 15mL of WASH solution. DNA was eluted from the column with 12mL of ELU elution buffer and collected in 50mL Falcon tubes. To concentrate the DNA, 10.5mL isopropanol was added to the eluted DNA and mixed by vortexing. Finally, precipitated DNA was pelleted by centrifugation at 3,345xg for 45 minutes at 4ºC. The pellet was then washed with 70% ethanol and centrifuged again at 3,345xg for 10 minutes. The ethanol was removed by aspiration and the DNA air-dried at room temperature for 5-10 minutes before resuspension in 500µL of sterile, deionised water. Plasmid DNA was then transferred to a 1.5mL microcentrifuge tube (Eppendorf) and kept at -20ºC.

2.14.8. Measurement of nucleic acid concentrations

The concentration of plasmid DNA preperations and purified gel fragments was determined with a Nanodrop 1000® spectrophotometer and analysed using the Nanodrop software (Thermo Scientific).

2.14.9. DNA sequencing and analysis

DNA sequencing was performed by members of the Molecular Biology Core Facility at the Cancer Research UK Manchester Institute. DNA samples to be sequenced were prepared by combining 330ng of template DNA with 2.5µM of the appropriate sequencing primer in a final volume of 12µL within a 0.5mL microcentrifuge tube. Sequencing was carried out using a BigDye® Terminator v3.1 Cycle Sequencing kit (Life Technologies), with the following cycling parameters:

DNA sequencing			
Step	**Temperature (ºC)**	**Time (seconds)**	**Cycles**
Initial denaturation	96	60	1
Denaturation	96	10	40
Annealing	50	5	
Extension	60	120	

Table 23. Thermal cycling conditions for DNA sequencing.

Samples were then kept at 4°C before purification using an Agencourt® CleanSEQ® kit (Beckman Coulter - www.beckmancoutler.com) prior to loading into the ABI3130 capillary system genetic analyser (Applied Biosystems). Sequence reads were visualised using Chromas software (version 1.43; Technelysium - www.technelysium.com.au).

2.14.9.1. DNA sequencing primers

DNA sequencing primers used in this study are listed below:

DNA sequencing primers	
pOTB7-*IRX3*	T7: TAATACGACTCACTATAGGG SP6: TATTTAGGTGACACTATAG
pMSCV(P2)	5': CCCTTGAACCTCCTCGTTCGACC 3': GAGACGTGCTACTTCCATTTGTC
pLentiGS	5': CCTCAGACAGTCGTTCAAAGT 3': GTTAGGGGCGGGACTATGG
FOXC1 sequence verification primers	5' 456: GACCGTGGACCCGGACTCCTA 5' 756: GCGGTGCCCAAGATCGAGAGC
IRX3 sequence verification primers	5' 431: GCTGGGATACCAATACATCCG 5' 527: TGCCGGAGCCTCGGAGCTGAA 5' 1212: GTTCCCGGCTTGGACCAACC

Table 24. List of DNA sequencing primers.

2.15. Bacterial culture methods

2.15.1. Bacterial culture medium

Sterile, autoclaved Luria Broth (LB) medium was obtained from in-house laboratory services (CRUK Manchester Institute) containing 10g/L tryptone, 5g/L yeast extract and 0.5g/L NaCl. LB medium was used routinely to recover competent cells after the heat shock process. Sterile, autoclaved Terrific Broth medium was also obtained from in-house laboratory services and contained 11.8g/L tryptone, 23.6g/L yeast extract, 9.4g/L K_2HPO_4, 2.2g/L KH_2PO_4 and 0.4% glycerol. As Terrific Broth contains a higher content of nutrients compared to LB medium it was routinely used for both mini- and maxi-preps to generate higher yields of plasmid DNA.

2.15.2. Bacterial culture agar plates

LB-agar plates were either obtained from in-house laboratory services or made individually by melting autoclaved LB containing 1.5% agar in a microwave oven until all agar was dissolved. Upon cooling to below 55°C, 100µg/mL ampicillin or 12.5µg/mL chloramphenicol was added and mixed thoroughly. The resultant mixture was then poured out into 10cm culture dishes (Falcon www.bdbiosciences.com) and allowed to set.

2.15.3. Bacterial growth conditions

Bacteria on LB agar plates were kept at 37ºC in a dedicated incubator (Haraeus Instruments - www.thermoscientific.com). Bacteria in LB or Terrific Broth in 1L conical flasks (Pyrex - www.pyrex.com) or 14mL sterile Falcon tubes (SLS) were kept at constant agitation at 37ºC in an SM30 orbital incubator (Edmund Büler GmbH).

2.16. Retroviral and lentiviral expression vectors

2.16.1. Retroviral and lentiviral expression vectors containing *IRX3*

2.16.1.1. pLentiGS-EGFP-IRX3

To generate a lentiviral expression vector for *IRX3*, *IRX3* cDNA (NM_024336.2) was firstly excised from pOTB7-*IRX3* (purchased from the Plasmid ID Repository, Harvard Medical School) through triple digest with the *EcoRI*, *XhoI* and *ClaI* restriction enzymes and, using the *EcoRI* and *XhoI* restriction sites, sub-cloned into the pLentiGS-EGFP lentiviral expression vector. The pLentiGS-EGFP vector was made by Dr. Gary Spencer, as previously described (Huang et al., 2014). Successful subcloning and confirmation of the *IRX3* cDNA sequence was determined by DNA sequencing.

2.16.1.2. pMSCV(P2)-puro-IRX3-MYC-tag

To generate a retroviral expression vector for *IRX3*, the coding sequence for *IRX3* was PCR amplified from pLentiGS-EGFP-*IRX3* using primers designed to incorporate respectively the *EcoRI* and *XhoI* restriction sites, complete with a MYC-tag at the 3' end immediately 5' to the *XhoI* restriction site. The following primers (5'-3') were used:

- Forward: cacgaattcaccatgtccttcccccagctggga
- Reverse: cacctcgagtcacagatcctcttctgagatgagttttgttcacccgaaccggatgaggagagagccgataaga

NB: Restriction sites in red, Kozak sequence underlined, MYC-tag in blue, GSG-linker in green.

Gel purified, amplified DNA fragments were subcloned directly into the *EcoRI* and *XhoI* sites of pMSCV(P2)-puro, described elsewhere (Somervaille et al., 2009).

2.16.2. Retroviral and lentiviral expression vectors containing FOXC1

2.16.2.1. pMSCV(P2)-puro-FOXC1

To generate a retroviral expression vector for *FOXC1*, *FOXC1* cDNA (NM_001453) was excised from pcDNA3.1-*FOXC1* (a gift from Jane Sowden) using the *EcoRI* and *XhoI* restriction sites and sub-cloned into the *EcoRI* and *XhoI* sites of pMSCV(P2)-puro.

2.16.2.2. pMSCV(P2)-puro-FOXC1-MYC-tag

To generate MYC-tagged-*FOXC1*, the coding sequence for *FOXC1* was PCR amplified from pMSCV(P2)-puro-*FOXC1* using primers designed to incorporate respectively the *EcoRI* and *XhoI* restriction sites, complete with a MYC-tag at the 3' end immediately 5' to the *XhoI* restriction site. The following primers (5'-3') were used:

- Forward gaattcaccatgcaggcgcgctactc
- Reverse: cacctcgagtcacagatcctcttctgagatgagtttttgttcacccgaaccaaacttgctacagtcgtagac

NB: Restriction sites in red, Kozak sequence underlined, MYC-tag in blue, GSG-linker in green.

2.16.2.3. pLentiGS-EGFP-FOXC1 SDM3

To generate a lentiviral vector for expression of *FOXC1* cDNA resistant to knockdown using *FOXC1* KD3, site directed mutagenesis (SDM) was performed using overlap extension PCR followed by *DpnI* digestion with pMSCV(P2)-puro-*FOXC1* as a template. The following primers (5'-3') were used:

- Forward: cag tct ctg tac cgc acg tcc ggt gcc ttt gtt tat gat tgc agc aag ttt tga cac acc ctc
- Reverse: tt gag ggt gtg tca aaa ctt gct gca atc ata aac aaa ggc acc gga cgt gcg gta cag aga ctg

NB: 7 single point silent mutations are indicated in red font.

For the overlapping extension PCR, separate PCR reactions were first set up for the mutated forward and reverse primers in thin-walled PCR tubes (VWR International) as indicated below:

PCR components for site directed mutagenesis	
Components	**Volume (µL)**
5x GC buffer	10
dNTPs (10mM)	2
Forward/Reverse primer (10µM)	2.5
Template (50ng/µL)	5
DMSO	1.5
Deionised H_2O	to 50µL
Phusion® DNA polymerase	1

Table 25. PCR components for overlap extension PCR.

The following thermal cycling conditions were then used for the first PCR reaction:

Overlap extension: PCR reaction 1			
Step	**Temperature (ºC)**	**Duration (seconds)**	**Number of cycles**
Initial denaturation	98	180	1
Denaturation	98	20	18
Annealing	55	30	
Extension	72	60	1
Hold	4	∞	∞

Table 26. Thermal cycling conditions for overlap extension PCR reaction 1.

Following the reaction, 25µL of the each reaction product, amplified using either the forward or reverse primer, were combined in a PCR tube and an additional 1µL Phusion® DNA polymerase added. The following thermal cycling conditions were then used for the second PCR reaction:

Site directed mutagenesis: PCR reaction 2			
Step	**Temperature (ºC)**	**Duration (seconds)**	**Number of cycles**
Initial denaturation	98	180	1
Denaturation	98	20	10
Annealing	65	30	
Extension	72	60	1
Hold	4	∞	∞

Table 27. Thermal cycling conditions for overlap extension PCR reaction 2.

The resulting reaction products were digested with *Dpn1* for 1 hour at 37ºC before transformation of Stbl3 competent cells and growth overnight on LB-ampicillin plates. Colonies were then picked, mini-prepped and the site-specific mutations validated by DNA sequencing. Mutated *FOXC1* cDNA was then sub-cloned into the *EcoRI* and *XhoI* sites of pLentiGS-EGFP.

2.16.3. Retroviral and lentiviral expression vectors containing KLF4, Hoxa9 and Meis1

KLF4 cDNA (NM_004235) (a gift from Georges Lacaud) was excised from pRRL_PPT_SF_*KLF4* using the *EcoRI* and *XhoI* restriction sites and sub-cloned into the *EcoRI* and *XhoI* sites of pLentiGS-blasticidin (Huang et al., 2014).

A puromycin resistant FLAG-*Hoxa9* construct was generated by excising FLAG-*Hoxa9* cDNA from pMSCVneo-*Hoxa9* using EcoR1 and Xho1 restriction sites followed by subcloning into the *EcoRI* and *XhoI* sites of pMSCV(P2)-puro. Retroviral constructs encoding FLAG-*Hoxa9* (pMSCVneo-*Hoxa9*) and HA-*Meis1* (pMSCVpuro-*Meis1*) have been described previously (Schnabel et al., 2000; Wang et al., 2010b).

2.16.4. shRNA lentiviral constructs

shRNA constructs, designed by the RNAi Consortium (TRC) at the Broad Institute (www.broadinstitute.org), targeting human *FOXC1*, *IRX3* and *MYB*, as well as murine *Foxc1* and *Irx3*, were obtained as glycerol stocks from Sigma-Aldrich in either pLKO.1_puro (TRC version 1) or pLKO.2_puro (TRC version 2) backbones. A non-targeting shRNA construct (SHC002) with at least four base pair mismatches from known human and murine genes was used as a non-targeting control (NTC). shRNA constructs used in this study are listed below:

shRNA constructs						
Gene target	**Species**	**Construct #**	**Target sequence (sense strand)**	**Target region**	**TRC ID**	**TRC version**
FOXC1	Human	KD1	GTCACAGAGGATCGGCTTGAA	CDS	TRCN0000235691	2
		KD2	ACTCTCCAGTGAACGGGAATA	CDS	TRCN0000235692	2
		KD3	GAGCTTTCGTCTACGACTGTA	CDS	TRCN0000235693	2
IRX3	Human	KD1	GAGATCGATTTGGAGAACTTA	CDS	TRCN0000016898	1
		KD2	AGTGCCTTGGAAGTGGAGAAA	CDS	TRCN0000016899	1
		KD3	GCGCCTCAAGAAGGAGAATAA	CDS	TRCN0000016900	1
		KD4	CCCTGGTCTTATCGGCTCTCT	CDS	TRCN0000016901	1
		KD5	CAACGTGCTCTCGTCCGTGTA	CDS	TRCN0000016902	1
		KD6	GTTTGTTTGTCCGGTTGATTT	3'UTR	TRCN0000436197	2
MYB	Human	KD2	GCTATCAAGAACCACTGGAAT	CDS	TRCN0000288601	2
		KD3	GCTCCTAATGTCAACCGAGAA	CDS	TRCN0000040060	1
		KD4	CCCTCTCATCTAGTAGAAGAT	CDS	TRCN0000306971	2
		KD5	GCATCAGAAGATGAAGACAAT	CDS	TRCN0000040062	1
		KD6	AACAGAATGGAACAGATGAC	CDS	TRCN0000009853	1
		KD7	CAGATGACATCTTCCAGTCAA	CDS	TRCN0000009854	1
		KD8	TAAGTGTATGGTCTCAGAACT	3'UTR	TRCN0000010388	1
Foxc1	Mouse	KD1	CGTCTATGACTGTAGCAAATT	CDS	TRCN0000085448	1
		KD2	TGGGAATAGTAGCTGTCAGAT	CDS	TRCN0000085449	1
		KD3	TAGCAAATTCTGACCCTATTC	CDS	TRCN0000321412	2
Irx3	Mouse	KD1	CGCTCTGTATAGTGACGAGTT	3'UTR	TRCN0000096099	1
		KD2	CCCTATCCAATGTGCTTTCAT	CDS	TRCN0000096100	1
		KD3	AGAGTGGAACAGATCGCTGTA	CDS	TRCN0000096101	1

Table 28. List of lentiviral shRNA constructs.
All lentiviral shRNA constructs were tested in the appropriate AML cells (either human AML cell lines or murine MLL-AF9 AML cells). However, for clarity only constructs for which robust and consistent KD could be demonstrated (at the mRNA and protein level) are presented in the results sections of this thesis.

2.17. Bioinformatics and statistical analyses

2.17.1. Exon array analyses

2.17.1.1. RNA extraction from murine leukaemias and microarray setup

For analysis of murine leukaemia samples, BM from sick mice (three animals per cohort) was thawed, and donor derived ($CD45.1^+$) viable $CD117^+Gr1^+$ leukaemia cells were flow sorted. RNA was immediately extracted as described in section 2.5.1 '*RNA extraction*' and its quality confirmed using an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA). Next, cDNA amplification of 40ng RNA was performed using the Ovation Pico WTA System V2 (NuGEN, San Carlos, CA), according to manufacturer's instructions. The cDNA was then fragmented, biotin-labelled and hybridised to GeneChip Mouse Exon 1.0 ST Arrays (Affymetrix, Santa Clara, CA) using an Encore Biotin Module kit (NuGEN), according to manufacturer's instructions. Product quality was confirmed at each step with the Agilent Bioanalyzer. Hybridized arrays were washed using the GeneChip Fluidics Station 450 (Affymetrix) using protocol FS450_001 and scanned using a GeneChip Scanner 3000 7G (Affymetrix). All procedures from the RNA extraction to exon array hybridisation were performed by John Weightman in the Molecular Biology Core Facility at the Cancer Research UK Manchester Institute.

2.17.1.2. Normal primary human neutrophils and monocytes

For analysis of normal primary human neutrophils and monocytes (from four separate individuals), fresh BM was sorted by Dr. Xu Huang with the assistance of members of the Imaging and Cytometry Core Facility at the Cancer Research UK Manchester Institute, as previously described (Huang et al., 2014). RNA was extracted and processed as above but with the following exceptions: cDNA amplification of RNA was with the Ovation Pico WTA System (NuGEN), a WT-Ovation Exon Module kit (NuGEN) was used to generate sense transcript cDNA and GeneChip Human Exon 1.0 ST Arrays were used (Affymetrix). Procedures from the RNA extraction to exon array hybridisation were performed by members of the Molecular Biology Core Facility at the Cancer Research UK Manchester Institute.

2.17.1.3. Exon array data analysis

Exon array data were processed using an R/Bioconductor package, as previously described (Okoniewski and Miller, 2008). Background adjustment, quantile normalisation and summarisation of probe-level intensity were performed using the robust multi-array average algorithm (Irizarry et al., 2003). Probesets annotated in Ensembl murine genome build 66, or for human array analysis in human genome build 54, as "exonic" and not "multitarget", with their expression values, were extracted in CSV file format. The CSV file was used to calculate "gene level summaries", which

represent the mean expression of all exonic probesets mapping to each Ensembl annotated protein coding gene. Mapping of murine genes to human homologs and gene set enrichment analyses were performed as described in detail elsewhere (Subramanian et al., 2005; Harris et al., 2012). To facilitate downstream analyses such as GSEA, the AML dataset from Wouters et al. (2009) was collapsed to gene symbols using the Max_probe mode of the CollapseDataset function of GSEA software v2.1.14; this was necessary because the dataset features multiple probesets mapping to a single gene. Gene Ontology (GO) analyses were performed using the DAVID bioinformatics online tool (http://david.abcc.ncifcrf.gov/). Processing and analysis of all exon array data was performed with the assistance of Dr. Tim Somervaille.

2.17.2. Analysis of gene expression in human AML

2.17.2.1. Comparison of AML LSCs versus normal HSPCs

For comparative analyses of normal versus leukaemic stem and progenitor cells, processed datasets were downloaded from public databases: (i) ArrayExpress E-TABM-978 (Goardon et al., 2011), (ii) Gene Expression Omnibus GSE24395 (Kikushige et al., 2010) and (iii) CIBEX accession ID CBX111 (Saito et al., 2010). Expression values for known or candidate transcription factors were extracted using the VLOOKUP function in Microsoft Excel. Transcription factors genes differentially expressed in AML stem and progenitor cells versus normal HSPCs were identified using an unpaired t-test (assuming unequal variance) ($p<0.005$) and were ranked according to mean fold change in expression. Cluster analyses were performed and heat maps generated using dChIP (build date November 5, 2010) (Li and Wong, 2001). These comparative analyses were performed with the assistance of Dr. Tim Somervaille.

2.17.2.2. Analysis of human AML datasets

For analyses of expression patterns of individual genes or gene sets in human AML, datasets were analysed in Microsoft Excel (Gene Expression Omnibus GSE14468; Wouters et al., 2009), or accessed through the cBioPortal (Cerami et al., 2012; Cancer Genome Atlas Research Network, 2013). Analysis of human AML datasets was performed with the assistance of Dr. Tim Somervaille.

2.17.3. ENCODE consortium data

ENCODE consortium data was analysed and images generated using the WashU Epigenome Browser (Zhou et al., 2011). Non-expressed transcription factor genes additionally marked by trimethyl-H3K27 in normal human $CD34^+$ cells were identified by manually inspecting RNA sequencing and H3K27me3 ChIP-sequencing tracks respectively. Lack of expression was determined by RNA sequencing with a maximum peak read frequency of 10 across the gene locus. Genes marked by H3K27me3 were determined as having a minimum peak read frequency of 20

across the gene locus. Assistance with the analysis of ENCODE consortium data was received from Dr. Tim Somervaille.

2.17.4. Intergenic mutation analysis using whole genome sequencing data

TCGA AML cases with high *FOXC1* expression (log_2 expression value >9.2; case numbers 2964, 2966, 2971, 2984, 2990 and 2993) or low *FOXC1* expression (log_2 expression value <4; case numbers 2977, 2988, 2994, 2998, 3000, 3007) were identified using cBioPortal. Whole genome sequencing data from these cases and matched normal controls was downloaded from TCGA as part of dbGAP authorized project #8623. The binary sequence alignment (BAM) files contained reads aligned to GRCh37-lite with BWA (v0.5.9). VarScan (Koboldt et al., 2012) (version 2.3.8, parameters: min-var-freq=0.05, strand-filter=1, min-coverage-normal=8, min-coverage-tumour=8) was used to detect somatic mutations in genomic DNA regions 1MB upstream and downstream of the transcription start site of *FOXC1* (chr6:1610681) (Appendix Table 4). A custom R script was used to compare the frequency of occurrence of each these somatic mutations across samples. This analysis was performed with the assistance of Dr. Hui Sun Leong.

2.17.5. Additional statistical analyses

Survival curves were generated using Prism software v6.0 (GraphPad Software, La Jolla, CA). One-way ANOVA with Fisher's LSD post-hoc analyses were also performed using Prism software v6.0. Microsoft Excel or Prism software v6.0 were used to perform t-tests. Multivariate analysis of independent predictors for overall survival was performed by Cox proportional hazards regression, generating hazard ratios and 95% confidence intervals for each co-variate included in the model (age; gender; cytogenetic risk; mutation status for *FLT3/NPM1*, *CEBPA*, *IDH1*, *IDH2*, *KRAS*, *NRAS*; and *FOXC1* expression level). Analyses were performed using SPSS for Mac (v22; IBM, Hampshire, UK) with the assistance of Dr. Daniel Wiseman.

Chapter 3. Frequent derepression of the mesenchymal transcription factor gene *FOXC1* in acute myeloid leukaemia

3.1. Introduction

Acute myeloid leukaemia (AML) is a hierarchically-organised, clonal neoplastic disorder sustained by a sub-population of cells with long-term proliferative potential, often termed leukaemia stem cells (LSCs) (Wiseman et al., 2014). In recent years there has been a concerted effort to understand the genetic, epigenetic and transcriptional differences between AML cells and their normal cellular counterparts, with the longer term aim of developing therapeutic strategies that selectively target leukaemia cells. One significant approach has been to prospectively isolate AML cells with immature progenitor immunophenotypes (immunophenotypic LSC) and compare their transcriptional profiles with normal haematopoietic stem and progenitor cells (HSPCs) (Saito et al., 2010; Kikushige et al., 2010; Goardon et al., 2011). These studies have highlighted the cell surface receptors CD25, CD32 and HAVCR2 (also known as TIM3) as candidate therapeutic targets in AML. In addition, such comparative datasets provide a rich resource for further exploration of biological processes active in human AML. Of particular interest is the set of transcription factors expressed by leukaemic cells, in view of their essential roles in regulation of development, cell fate specification, cellular signalling and the cell cycle.

To this end, the datasets from these three studies were interrogated for transcription regulators expressed in AML LSCs but not normal HSPCs and unexpectedly the mesenchymal transcription factor gene *FOXC1* was identified as among the most highly up regulated. FOXC1 is a member of the forkhead box family of transcription factors which are critical regulators of development and differentiation. It is normally expressed in paraxial mesoderm, prechondrogenic mesenchyme, neural crest, endothelium and the developing eye and kidney. In keeping with a requirement for FOXC1 in mesenchymal differentiation, *Foxc1* null mice die perinatally with skeletal, cardiac and renal abnormalities, hydrocephalus, iris hypoplasia and open eyelids (Kume et al., 1998). Humans with inherited haploinsufficiency of *FOXC1* due to mutation or deletion exhibit the Axenfeld-Rieger syndrome which is characterised by abnormalities of the anterior segment of the eye, glaucoma, hearing loss, and skeletal, dental and cardiac malformations (Nishimura et al., 1998; Kume et al., 1998). High *FOXC1* expression is associated with poor prognosis in a range of solid malignancies including those of breast and liver (Ray et al., 2010; Xia et al., 2013). Its forced expression promotes an epithelial-to-mesenchymal transition, and enhanced proliferation, migration, invasion and drug resistance, through downstream mediators such as NF-κB, MMP7 or NEDD9 (Bloushtain-Qimron et al., 2008; Sizemore and Keri, 2012; Wang et al., 2012; Xia et al., 2013). Interestingly, *Foxc1* is highly expressed by *CXCL12*-expressing adipo-osteogenic progenitor cells in mouse bone marrow (BM) and its deletion ablates haematopoietic stem cell niches leading to substantial reductions in BM cellularity (Omatsu et al., 2014).

Given the absence of information regarding any cell autonomous role for FOXC1 in normal or leukaemic haematopoiesis, the potential functional role of this transcription factor in AML was investigated. Remarkably, it was discovered that *FOXC1*, which is not normally expressed in

haematopoiesis, is derepressed in at least 20% of cases of human AML, leading to significant functional consequences including enhanced differentiation block and inferior survival.

3.2. *FOXC1* is expressed in human AML

To identify transcriptional regulators expressed in human AML LSCs but not normal HSPCs, the expression levels of known or candidate transcription factor genes (Vaquerizas et al., 2009) were analysed in three recently published datasets (Saito et al., 2010; Kikushige et al., 2010; Goardon et al., 2011). Of those exhibiting significantly higher expression in AML LSCs versus normal HSPCs, *FOXC1* was among the most highly up regulated genes in each study when ranked by fold-change increase in expression (Figures 3A-C). Other genes exhibiting significantly higher expression in AML LSCs versus normal HSPCs included those encoding transcription factors such as CEBPA, CEBPD, IKZF1 and IRF8 which are known to be highly expressed in myeloid lineage cells (Figures 3A-C). High level *FOXC1* expression (where values for *FOXC1* were among the top 10% of protein-coding gene probesets) was observed in 48%, 50% and 33% of samples analysed in each study respectively (Figures 3A-C). Thus, in multiple independent studies *FOXC1* is among the most highly up regulated transcription factor genes in AML LSCs.

Deletion of *Foxc1* in normal murine HSCs does not affect haematopoiesis (Omatsu et al., 2014), suggesting redundancy or lack of expression. To determine whether *FOXC1* is expressed in normal human haematopoiesis, quantitative PCR (qPCR) was performed on flow sorted populations of BM HSPCs and terminally differentiated cells, including neutrophils and monocytes (Figure 3D) (Huang et al., 2014). *FOXC1* expression was either absent or detected at very low level. In contrast, *FOXC1* transcripts were detected at high level (greater than 200-fold increase over expression levels in the lowest expressing AML sample) in 5/29 (17%) bulk AML blast samples tested (Figure 3D; Table 29), and at intermediate level in 8/29 (28%) samples (20-200-fold increase over expression levels in the lowest expressing AML sample). Increased *FOXC1* transcripts in AML cells led to increased protein expression (Figure 3E) and FOXC1 protein was not detected in normal human $CD34^+$ cells (Figure 3E). *Foxc1* is highly expressed by *CXCL12*-expressing adipo-osteogenic progenitor cells in the mouse BM niche (Omatsu et al., 2014). To confirm that the high *FOXC1* expression observed in AML LSCs and bulk AML samples was in fact due to its aberrant expression in AML cells and not from contaminating BM niche cells, qPCR was also performed on single blast cells isolated from a *FOXC1*-high patient BM sample (Figure 3F). *FOXC1* transcripts were detected in 56/72 (78%) single AML blast cells at comparable levels to the haematopoietic transcription factor genes *CEBPA* and *RUNX1*, which were detected in 62/72 (86%) and 40/72 (56%) single cells respectively (Figure 3F).

These data demonstrate that the mesenchymal transcription factor FOXC1 (which is neither expressed in nor required for normal haematopoiesis) is frequently highly expressed in human AML in both the stem/progenitor and bulk blast compartments.

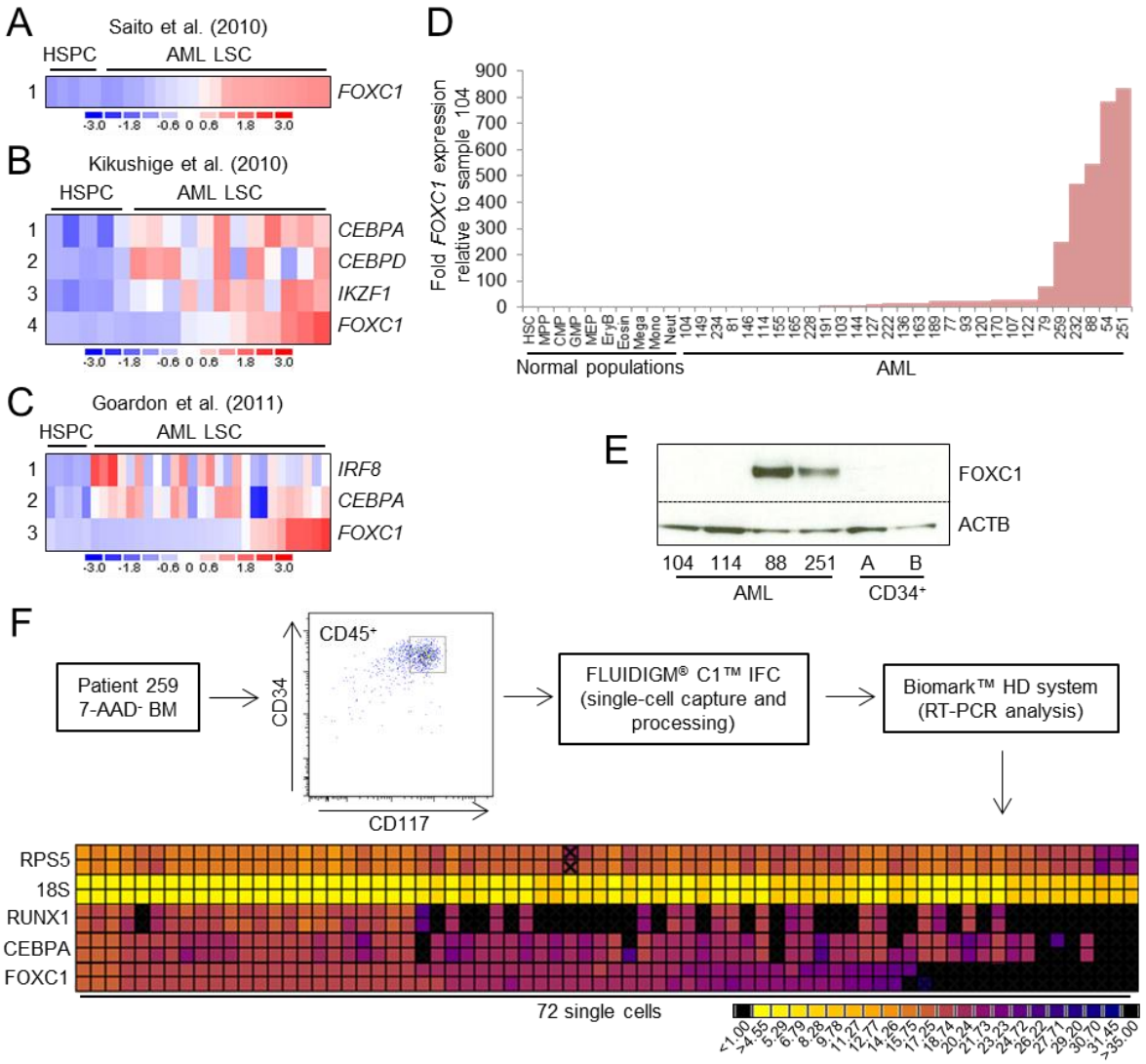


Figure 3. Expression of *FOXC1* in human AML.

(A-C) Heat maps show the most highly up regulated transcription factor genes in the indicated studies. Transcription factor genes differentially expressed in AML HSPCs versus normal adult immunophenotypic BM HSPCs were identified using an unpaired t test (with $p<0.005$) and were ranked according to the mean fold change increase in expression (ranking number shown on the left of each heat map row). Colour scale indicates the standardised expression level for each gene. The definition of HSPCs in each study was (A) $CD34^+CD38^-$ for both normal (n=5) and AML cells (n=21); (B) $CD34^+CD38^-Lin^-$ for normal cells (n=5) and $CD34^+CD38^-$ for AML cells (n=12), and (C) $CD34^+CD38^-CD90^+CD45RA^-Lin^-$ for normal cells (n=5) and GMP-like (n=21), MPP-like (n=3) or $CD34^+CD38^-CD90^-CD45RA^+$ (n=3) for AML samples. (D) Bar chart shows relative expression of *FOXC1* in bulk primary human AML samples (n=29) and prospectively sorted normal human cell populations (n=3 separate individuals per cell type; Huang et al., 2014). See also Table 29. HSC, haematopoietic stem cell ($CD34^+38^-90^+45RA^-Lin^-$); MPP, multipotent progenitor; CMP, common myeloid progenitor; GMP, granulocyte/macrophage progenitor; MEP, megakaryocyte/erythrocyte progenitor; EryB, erythroblast; Eosin, eosinophils; Mega, megakaryocytes; Mono, monocytes; Neut, neutrophils. AML sample numbers refer to the Biobank identifier. (E) Western blot shows expression of FOXC1 in the indicated AML and normal $CD34^+$ cell samples. (F) Top panel shows the flow sorting strategy for the indicated *FOXC1*^{high} AML patient sample and the procedure for single cell qPCR analysis. Heat map (bottom panel) shows the expression levels of *RUNX1*, *CEBPA* and *FOXC1* in 72 single cells. Expression levels of two the ribosomal genes *RPS5* and *18S* are shown as housekeeping controls. Colour scale indicates Ct expression levels. Assistance with single cell qPCR analysis was received from Chris Clark.

ID	Sample	Disease Status	Gender	Age	Cytogenetics
54	Blood	2nd Relapse	Male	75	47,XY,+11[1]/48,sl,+8[7]/49,sdl,+4[2]
77	BM	Diagnosis	Female	67	Normal
79	BM	Diagnosis	Female	65	47,XX,der(3)t(3;9;14;4)(q2?1;q34;q24;q21),der(4)t(3;9;14;4),del(7)(q22q32),add(8)(q24),der(9)del(9)(p2?3)t(3;9;14;4),der(14)t(3;9;14;4),+18,add(19)(q13.3)x2[9]/46,XX[1]
81	Blood	1st Relapse	Male	64	Normal
88	BM	Refractory	Female	64	46,XX,del(12)(p1?2p1?3)[18]/46,XX[2]
93	BM	Diagnosis	Male	59	Normal
103	BM	Diagnosis	Male	51	Normal
104	BM	1st Relapse	Female	31	46,XX,t(6;9;11)(p2?1;p22;q23)[6]/45,idem,der(15)t(15;17)(p11.2;q11.2),-17[4]
107	BM	1st Relapse	Female	16	45,XX,del(7)(q11.2q3?2),t(8;21)(q22;q22),der(12)t(12;18)(p11.2;q11.2),-18[10]/46,XX[1]
114	BM	Diagnosis	Male	56	Normal
120	BM	Diagnosis	Female	68	Normal
122	Blood	Diagnosis	Male	66	Normal
127	BM	Diagnosis	Female	63	46,XX,inv(16)(p13q22)[8]/46,XX[2]
136	BM	Diagnosis	Female	61	Normal
144	Blood	Diagnosis	Male	71	Normal
146	BM	Diagnosis	Male	52	46,Y,?t(X;6)(p22.1;q?25),t(15;17)(q22;q11.2)[9]/46,XY[1]
149	BM	Diagnosis	Female	49	46,XX,t(15;17)(q22;q11.2)[7]/46,sl,-6,add(16)(q12),+mar[3]/46,XX[3]
155	Blood	Diagnosis	Male	63	Normal
163	BM	Diagnosis	Male	20	45,X,-Y,t(8;21)(q22;q22)[8]/46,XY[2]
165	BM	Diagnosis	Female	48	46,XX,t(8;22)(p11;q13),del(9)(q13q32)[10]
170	BM	1st Relapse	Male	50	46,XY,add(6)(p22)[8]/46,XY[2]
189	BM	Diagnosis	Female	76	Normal
191	BM	Diagnosis	Female	38	Normal
222	BM	Diagnosis	Female	80	Normal
228	BM	Diagnosis	Male	54	46,XY,t(1;2)(q2?4;p1?3)[9]/46,XY[1]
232	Blood	Diagnosis	Female	85	Normal
234	Blood	Diagnosis	Male	84	47,XY,+13[3]/93~94,idemx2[cp3]/92~94<4n>,XXYY,+X,der(5)t(1;5)(q21;q31),+13[cp5]/46,XY[4]
251	BM	Diagnosis	Male	16	46,XY,t(6;9)(p22;q34)[9]/46,XY,der(6)t(6;9),der(9)t(6;9)del(9)(q21q34)[2]
259	BM	Diagnosis	Male	75	47,XY,+8[10]

Table 29. Karyotype of 29 Manchester Cancer Research Centre Biobank AML samples analysed for *FOXC1* expression. Samples in red exhibit high *FOXC1* expression (see Figure 3D). BM = bone marrow blasts.; sl = stemline; sdl = sideline; der = derivative; t = translocation; del = deletion; add = additional material of unknown origin; idem = denotes the stemline karyotype in a subclone; inv = inversion; mar = marker chromosome; cp = composite karyotype. Related to Figure 3. Assistance with sample curation and karyotype was received from Dr. Daniel Wiseman.

3.3. *FOXC1* expression in human AML is associated with mutations in *NPM1*, and t(6;9)

The qPCR analysis in bulk AML samples was confirmed by expression data from two recent studies. *FOXC1* was expressed at high level in 100/461 (22%) (Wouters et al., 2009) and 36/163 (22%) (Cancer Genome Atlas Research Network, 2013) of samples analysed (Figures 4A and 4B). High *FOXC1* expression was strongly associated with intermediate cytogenetic risk, normal karyotype, and the presence of an *NPM1* mutation or a t(6;9) translocation (Tables 30 and 31). High *FOXC1* expression was negatively associated with good cytogenetic risk, its associated karyotypes and the presence of double *CEBPA* mutations (Table 30). In the Wouters study there was a strong positive association of high *FOXC1* expression with the presence of FLT3 internal tandem duplications but this was not seen in the TCGA study (Tables 30 and 31). There was no association of high *FOXC1* expression with other recurring mutations in AML (Table 31).

Characteristic		*FOXC1*^{low} (n=290)	*FOXC1*^{high} (n=100)	p value
Gender		141:149 M:F	49:51 M:F	NS
Age		42 (range 15-61)	43 (range 15-60)	NS
Cytogenetic risk	Good	83 (29%)	0 (0%)	<0.0001
	Intermediate	150 (52%)	69 (69%)	0.0034
	Poor	53	25	NS
	Unknown	4	6	NS
Karyotype	Trisomy 8	12	3	NS
	11q23	6	4	NS
	5 or7(q) loss	18	8	NS
	Complex	11	1	NS
	Normal	105 (36%)	50 (50%)	0.018
	Other	35	13	NS
	t(6;9)	0 (0%)	6 (6%)	0.0003
	t(8;21)	27 (9%)	0 (0%)	0.0004
	inv(16)	30 (10%)	0 (0%)	0.0001
	t(15;17)	20 (7%)	0 (0%)	0.0031
	abn(3q)	0	1	NS
	minus 9q	5	0	NS
	t(9;22)	1	0	NS
	Unknown	20	14	NS
Gene mutations	*CEBPA* double	19 (7%)	1 (1%)	0.033
	IDH1	18	13	NS
	IDH2	20	12	NS
	FLT3 ITD	68 (23%)	43 (43%)	0.0003
	FLT3 TKD	40	7	NS
	NRAS	29	8	NS
	KRAS	3	0	NS
	NPM1	66 (23%)	55 (55%)	<0.0001
	FLT3 ITD and *NPM1*	31 (11%)	34 (34%)	<0.0001
High *MECOM* expression		16	11	NS

Table 30. Association of *FOXC1* expression with diagnostic and genetic features in AML.
Array data from 461 AML cases where extended genotyping analysis was available were downloaded and analysed (Wouters et al., 2009). Cases were divided into *FOXC1*^{low} and *FOXC1*^{high} groups based on $\log_2$ expression values of <6.1 or >7.1 respectively (probeset 213260_at). Statistical significance for the indicated comparisons was assessed using Fisher's Exact Test, except for age where an unpaired t-test was used. FAB = French-American-British. NS = not significant. Related to Figure 4. Assistance with bioinformatics analysis was received from Dr. Tim Somervaille.

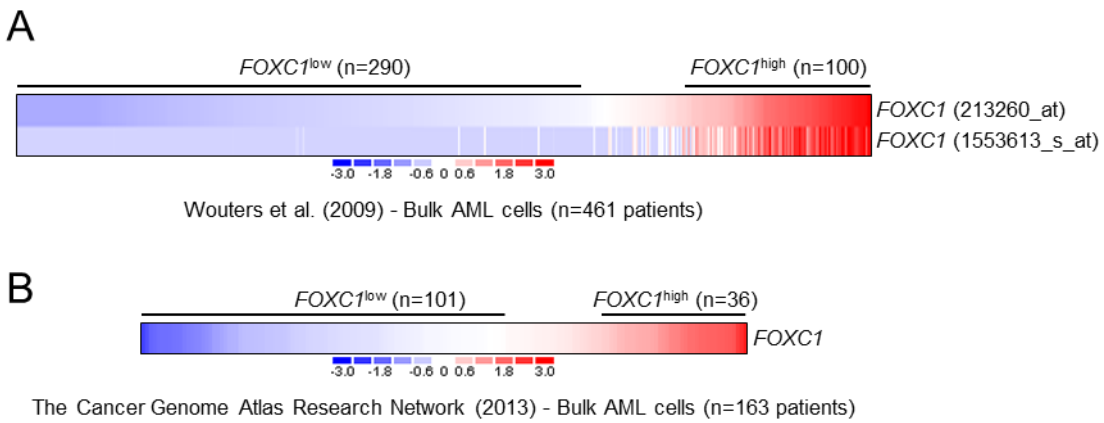


Figure 4. *FOXC1* expression in large human AML datasets.
Heat maps show (A) expression of *FOXC1* in 461 bulk AML samples (data extracted from Wouters et al., 2009). Samples were divided into *FOXC1*^{low} and *FOXC1*^{high} groups based on $\log_2$ expression values of <6.1 or >7.1 respectively (probeset 213260_at); and (B) expression of *FOXC1* in 163 bulk AML samples (data extracted from cBioPortal; Cerami et al., 2012; Cancer Genome Atlas Research Network, 2013). Samples were divided into *FOXC1*^{low} and *FOXC1*^{high} groups based on $\log_2$ expression values of ≤6.25 or >8 respectively. Colour scale indicates standardised expression level. Assistance with bioinformatics analysis was received from Dr. Tim Somervaille.

Mutated gene	*FOXC1*^{low} (n=101)	*FOXC1*^{high} (n=36)	p value
DNMT3A	20	10	NS
TET2	8	4	NS
NPM1	18 (18%)	22 (61%)	<0.0001
IDH1	10	5	NS
IDH2	9	5	NS
RUNX1	9	3	NS
TP53	7	1	NS
WT1	4	5	NS
PTPN11	6	1	NS
KIT	6	0	NS
EZH2	2	0	NS
HNRNPK	1	0	NS
U2AF1	3	1	NS
SMC1A	5	1	NS
SMC3	5	2	NS
BRINP3	3	0	NS
PHF6	3	1	NS
STAG2	3	1	NS
RAD21	1	1	NS
FLT3 ITD	17	10	NS
FLT3 D835X	8	3	NS
FLT3 ITD and *NPM1* mutation	4 (4%)	8 (22%)	0.003

Table 31. Association of *FOXC1* expression with genetic features in AML.
Expression data from 163 AML cases with exome or whole genome sequencing were analysed using cBioPortal (Cerami et al., 2012; Cancer Genome Atlas Research Network, 2013). Cases were divided into *FOXC1*^{low} and *FOXC1*^{high} groups based on $\log_2$ expression values of ≤6.25 or >8 respectively. Statistical significance for the indicated associations was assessed using Fisher's Exact Test. Related to Figure 4. Assistance with bioinformatics analysis was received from Dr. Tim Somervaille.

3.4. FOXC1 sustains clonogenic potential and differentiation block in AML cells

To investigate whether *FOXC1* expression in AML makes a functional contribution to transformation, knockdown (KD) experiments were firstly performed in human THP1 AML cells which exhibit both a t(9;11) translocation, the cytogenetic hallmark of *MLL-AF9*, and high *FOXC1* expression. *FOXC1* KD led to loss of clonogenic potential due to induction of differentiation and G1 arrest (Figure 5). Following initiation of KD there was down regulation of the stem cell marker CD117, up regulation of the myeloid differentiation marker CD11b (Figure 5D), morphological differentiation (Figure 5E), G1 arrest (Figure 5F) and apoptosis (Figure 5G).

To confirm that the observed phenotype was an on-target consequence of *FOXC1* KD, similar experiments were performed in a line constitutively expressing a *FOXC1* cDNA engineered by site-directed mutagenesis to generate KD-resistant transcripts. FOXC1 forced expression and resistance to KD was confirmed by western blot (Figure 6A). Expression of KD-resistant *FOXC1* in *FOXC1* KD cells completely prevented loss of clonogenic potential, indicating an on-target phenotype (Figure 6B).

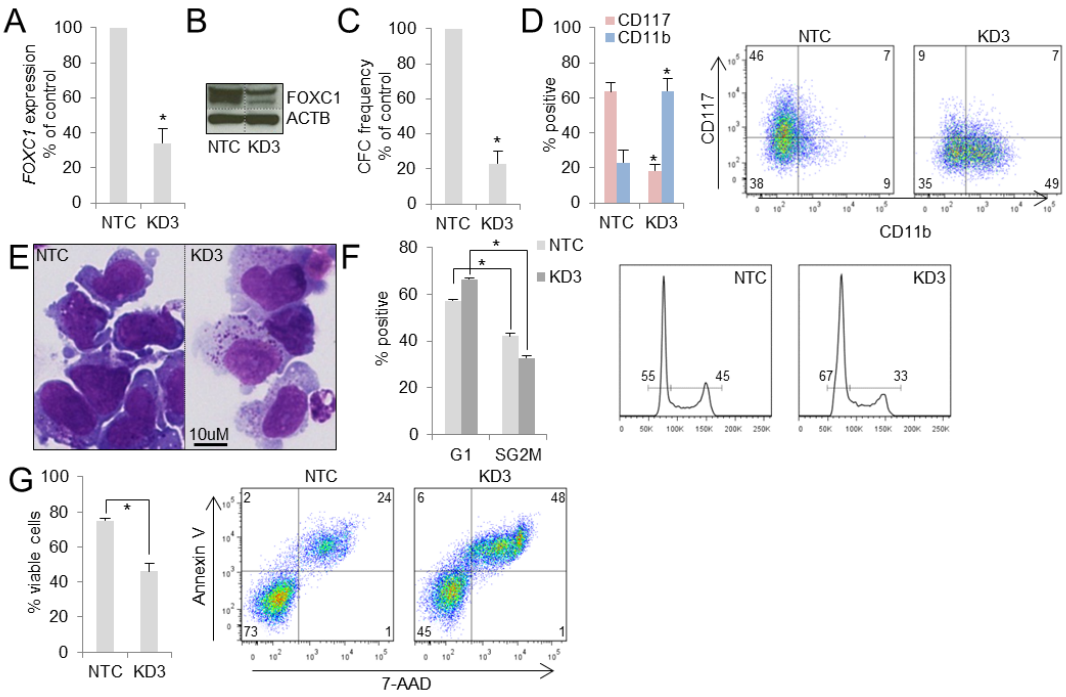


Figure 5. FOXC1 sustains the differentiation block and clonogenic potential of human AML cells. Human THP1 AML cells were infected with a lentivirus targeting *FOXC1* for KD (KD3) or a non-targeting control vector (NTC). Three KD constructs were tested, however only KD3 is shown as this construct displayed the most consistent and greatest level of KD. (A) Bar chart shows mean+SEM relative transcript expression in KD versus control cells (n=3). (B) Western blot shows expression of the indicated proteins in the indicated conditions. (C) Bar chart shows the mean+SEM colony-forming cell (CFC) frequencies of KD cells relative to control cells enumerated after ten days in semi-solid culture (n=4). (D) Bar chart (left panel) shows mean+SEM percentage of cells positive for the indicated cell surface markers, as determined by flow cytometry analysis six days following initiation of KD (n=4). Representative flow cytometry plots (right panel) are also shown. (E) Representative images of cytospins of cells from (D). (F) Bar chart (left panel) shows the percentage of viable control and KD cells in the indicated phases of the cell cycle as determined by propidium iodide staining six days following initiation of KD. Representative cell cycle profiles (right panel) are also shown. (G) Bar chart (left panel) shows mean+SEM proportion of viable cells as determined by Annexin-V/7-AAD analysis seven days following initiation of KD (n=6). Representative flow cytometry plots (right panel) are also shown. For A, C, D, F and G * indicates $p<0.05$ using an unpaired t-test.

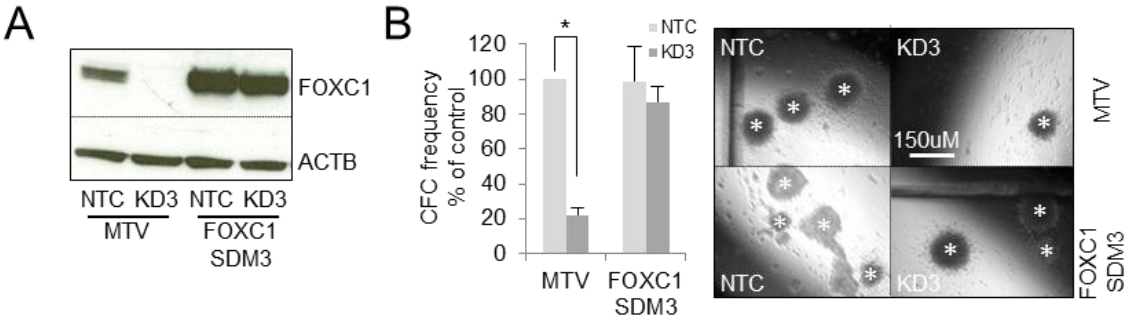


Figure 6. Rescue of clonogenic potential with FOXC1 SDM3.
Human THP1 AML cells stably expressing *FOXC1* SDM3 (for site directed mutagenesis #3) or a control retroviral vector (MTV) were infected with the same lentiviral *FOXC1* KD vector (KD3) as shown in Figure 5 or a non-targeting control vector (NTC). (A) Western blot shows expression of the indicated proteins in the indicated conditions. (B) Bar chart (left panel) shows mean+SEM CFC frequencies of THP1 AML cells expressing either *FOXC1* SDM3 or MTV in *FOXC1* KD cells relative to control cells. Colonies were enumerated after ten days in semi-solid culture (n=3). Image (right panel) shows representative colonies, each of which is marked by a white asterisk. For B, * indicates $p<0.05$ using an unpaired t-test.

KD of *FOXC1* in other *FOXC1* expressing AML cell lines representative of a variety of molecular subtypes also led to loss of clonogenic potential (Figure 7). In contrast, the clonogenic and multilineage differentiation potential of normal human CD34$^+$ cells (which do not express *FOXC1*) was unaffected following infection with the same *FOXC1* KD construct (Figure 8). *Foxc1* expression could also be detected in murine MLL-AF9 AML cells and its KD in these cells with multiple constructs likewise led to loss of clonogenic potential and induction of differentiation (Figure 9).

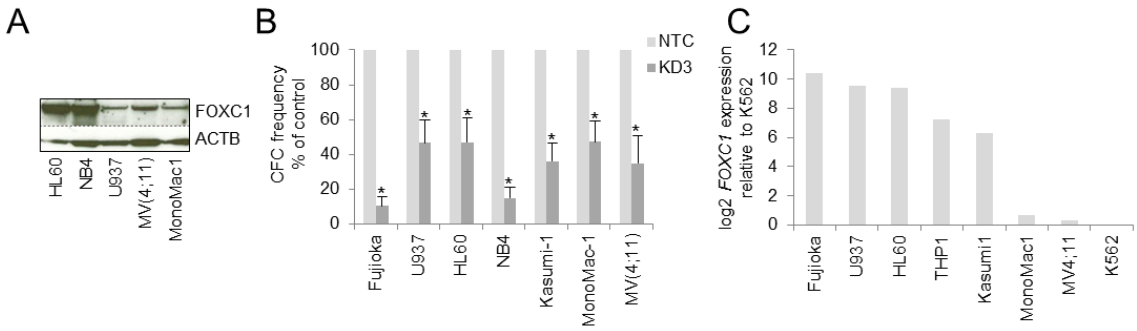


Figure 7. Consequences of *FOXC1* KD in human AML cell lines.
(A) Western blot shows expression of the indicated proteins in the indicated cell lines. (B) Bar chart shows mean+SEM colony-forming cell (CFC) frequencies of KD cells relative to control cells, enumerated after ten days in semi-solid culture (n=3). (C) Bar chart shows the relative expression levels of *FOXC1* in the indicated cell lines as determined by quantitative PCR. For B, * indicates $p<0.05$ using an unpaired t-test.

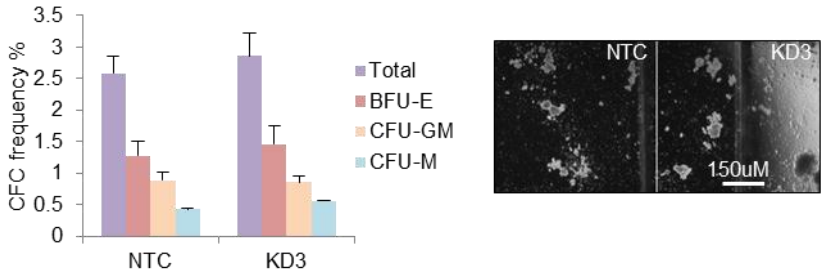


Figure 8. Consequences of *FOXC1* KD in normal human CD34$^+$ HSPCs.
Normal human CD34$^+$ HSPCs were infected with the *FOXC1* KD3 vector (KD3) or a non-targeting control (NTC). Bar chart (left panel) shows mean+SEM total and types of CFCs (n=3 separate individuals). Colonies were enumerated after 14 days. Image (right panel) shows representative colonies. BFU-E, burst-forming unit erythroid; CFU-GM, colony-forming unit granulocyte/macrophage; CFU-M, colony-forming unit macrophage.

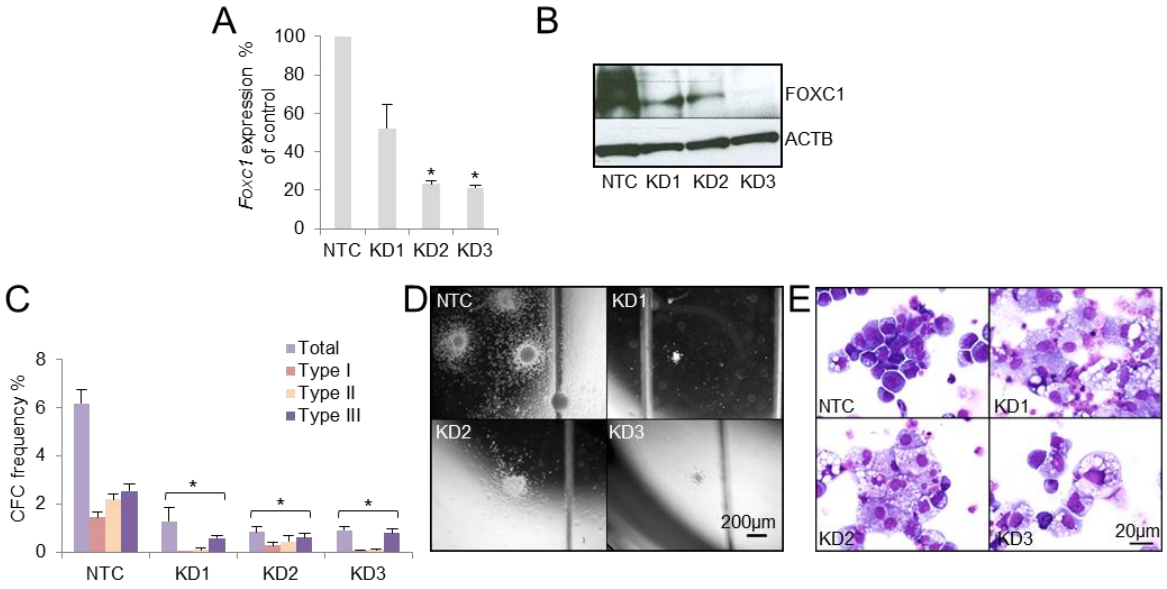


Figure 9. FOXC1 sustains the differentiation block and clonogenic potential of murine MLL-AF9 AML cells. Murine MLL-AF9 AML cells were infected with lentiviral vectors targeting *Foxc1* for KD (KD1-3) or a non-targeting control vector (NTC). (A) Bar chart shows mean+SEM relative transcript expression in KD versus control cells (n=3). (B) Western blot shows expression of the indicated proteins in the indicated conditions. (C) Bar chart shows the mean+SEM colony-forming cell (CFC) frequencies enumerated after five to seven days in semi-solid medium (n=3). Type I colonies contain poorly differentiate myeloblasts, Type II colonies contain a mixed population of blasts and differentiating myeloid cells, Type III colonies contain terminally differentiated myeloid cells. (D&E) Images show representative colonies (D) and cytospins (E) from (C). *indicates $p<0.05$ by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of each KD sample versus NTC.

3.5. FOXC1 transiently impairs myeloid differentiation in normal HSPCs

To investigate the consequences of forced *FOXC1* expression in normal HSPCs, murine KIT$^+$ (from this point on referred to as CD117$^+$) BM cells were transduced with retroviral vectors (Figure 10A). In serial replating experiments using successfully transduced cells, a significant but transient myeloid differentiation block and enhanced proliferation was observed in *FOXC1* expressing cells (FOXC1$^+$ cells). In the second round FOXC1$^+$ cells generated approximately five times as many colonies as did control cells (Figure 10B), with a substantially higher number of tightly packed blast-like colonies and a lower proportion of fully mature macrophage colonies (Figure 10C). In keeping with these observations, FOXC1$^+$ cell populations at the end of the first round (used to set up second round cultures) contained a higher proportion of myeloblasts versus mature neutrophils or macrophages (Figure 10D) and a higher proportion of cells in the SG_2M phase of the cell cycle (Figure 10E). Immunophenotypic analysis demonstrated lower expression of the mature myeloid differentiation marker Gr1 and the monocyte/macrophage marker F4/80 in FOXC1$^+$ cells by comparison with control cells (Figure 10F). Despite the significant second round differences, the consequences of *FOXC1* over expression were only transient because there was no significant difference in the clonogenic potential of third round cells, with FOXC1$^+$ cells at the end of that round displaying features of terminal differentiation (Figure 10G).

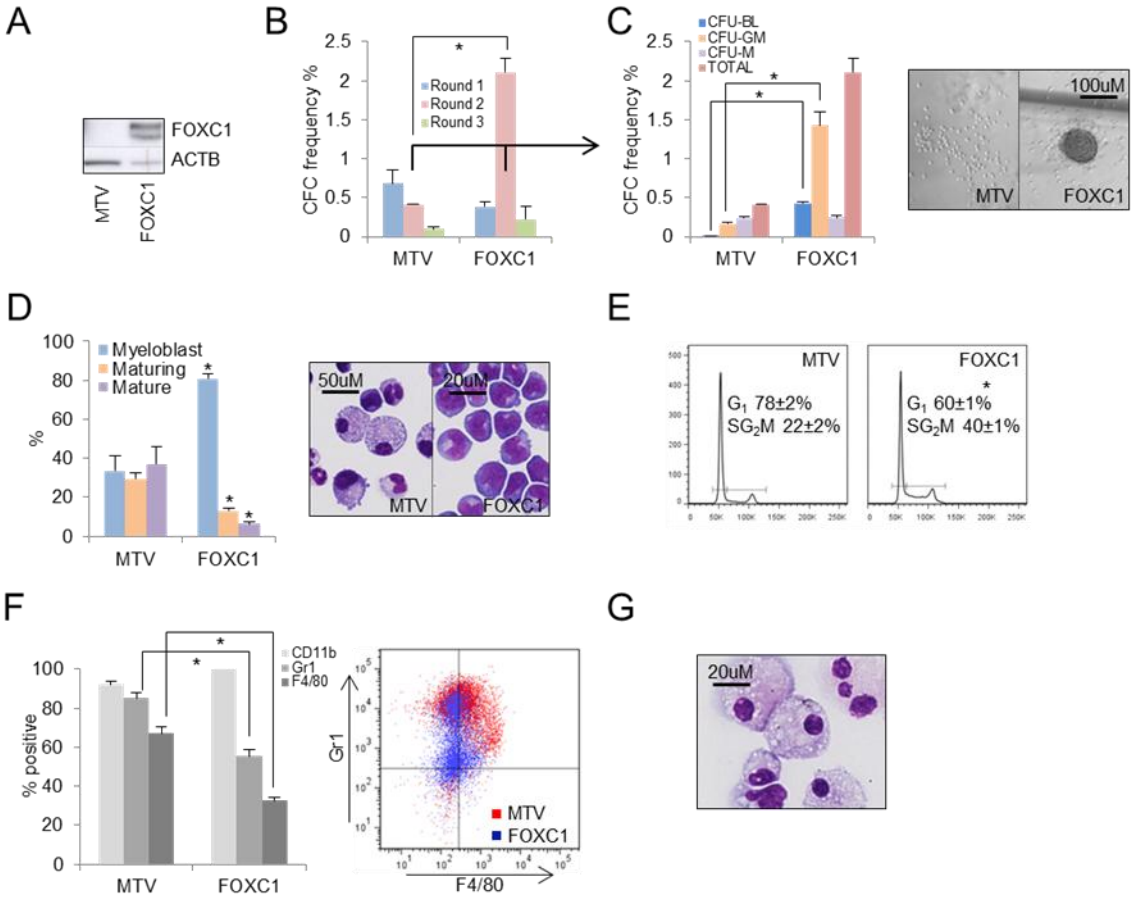


Figure 10. FOXC1 transiently impairs myeloid differentiation of normal HSPCs *in vitro*.
Murine CD117^{+} BM cells were transduced with *FOXC1*-expressing or control (MTV) retroviral vectors and serially replated *in vitro*. (A) Western blot shows FOXC1 expression in CD117^{+} BM cells 48 hours following drug selection and 72 hours post-spinoculation. (B) Bar chart shows mean+SEM CFC frequencies at the end of each round of serial replating (n=3). *indicates p<0.0001 for the indicated comparison by one way ANOVA with Fisher's least significant difference *post hoc* test. (C) Bar chart (left panel) shows mean+SEM types of colonies formed in the second round of culture (n=3). Image (right panel) shows representative colonies. CFU-BL, blast-like colony-forming unit; CFU-GM, granulocyte/macrophage colony-forming unit; CFU-M, macrophage colony-forming unit. * indicates p<0.005 for the indicated comparisons by an unpaired t test. (D) Bar chart (left panel) shows mean+SEM percentage of the indicated cell types in cytospin preparations from the end of round one (n=3). Representative images (right panel) are shown. * indicates p<0.05 for FOXC1^{+} cells versus the respective cell type infected with control vector by an unpaired t test. (E) Representative cell cycle profiles of viable control and FOXC1^{+} cells from the end of the second round of serial replating (n=3). * indicates p<0.001 by an unpaired t test. (F) Bar chart (left panel) shows mean+SEM percentage of cells positive for the indicated cell surface markers following 11 days of liquid culture (n=3). A representative flow cytometry plot (right panel) is shown. * indicates p<0.01 for the indicated comparisons (unpaired t-test). (G) Representative cytospin image of FOXC1^{+} cells at the end of round three.

To investigate the *in vivo* consequences of *FOXC1* expression in HSPCs, transplantation experiments were performed. Murine CD117$^+$ HSPCs were infected with retroviral vectors expressing *FOXC1*, *Meis1* or *Hoxa9*, or an empty vector, and transplanted into irradiated congenic recipients. *Meis1* and *Hoxa9* were chosen as comparators because forced expression of *Meis1* has no effect on BM chimerism whereas forced expression of *Hoxa9* is sufficient to cause HSC expansion followed by long latency AML (Thorsteinsdottir et al., 2001; Thorsteinsdottir et al., 2002). Assessment of the extent of donor:recipient chimerism in the blood of transplanted mice over 16 weeks demonstrated, as expected, no significant difference between those receiving MEIS1$^+$ cells versus MTV cells (Figure 11A). Also as expected, mice receiving HOXA9$^+$ cells exhibited significantly higher levels of donor:recipient chimerism (Figure 11A). By contrast, at all time points recipients of FOXC1$^+$ cells exhibited significantly lower donor:recipient chimerism in blood (the BM was not sampled) (Figure 11A). Consideration of the lineage composition of engrafted cells four weeks following transplant demonstrated donor-derived multilineage engraftment in all cohorts (Figure 11B and 11F). There were proportionately similar levels of myeloid, B-lineage and T-lineage engraftment between MEIS1$^+$, HOXA9$^+$ and MTV recipients, but significantly higher levels proportionately of myeloid engraftment, and lower levels of B-lymphoid engraftment, among FOXC1$^+$ recipients (Figures 11B and 11F). At the later time points (8, 12 and 16 weeks) HOXA9$^+$ recipients showed progressive expansion of the myeloid compartment and a proportionate reduction in T-lineage engraftment compared to MEIS1$^+$ and MTV recipients (Figures 11C-E). FOXC1$^+$ recipients maintained myeloid skewing of donor derived cells. Thus, by comparison with MEIS1$^+$ and MTV control cells, expression of FOXC1 in HSPCs reduces donor:recipient chimerism in blood and skews differentiation towards the myeloid lineage and away from the B-cell lineage (Figure 11G).

In keeping with progressive expansion of the myeloid compartment, and previous reports (Thorsteinsdottir et al., 2001), mice transplanted with HOXA9$^+$ cells succumbed to AML with a median latency of 103 days (Figure 11H), whereas at the termination of the experiment 200 days post-transplant none of the mice from any other cohort had died of a donor-derived haematological neoplasm or exhibited any features thereof in BM analyses (data not shown). The three mice that did die succumbed to recipient origin T-cell leukaemia (MTV recipient) or recipient origin T-cell lymphoma (MEIS1$^+$ and FOXC1$^+$ recipients). Taken together these data indicate that expression of *FOXC1* in HSPCs, while not overtly leukaemogenic, is sufficient to induce a transient myeloid differentiation block *in vitro* and to skew differentiation towards the myeloid lineage *in vivo*.

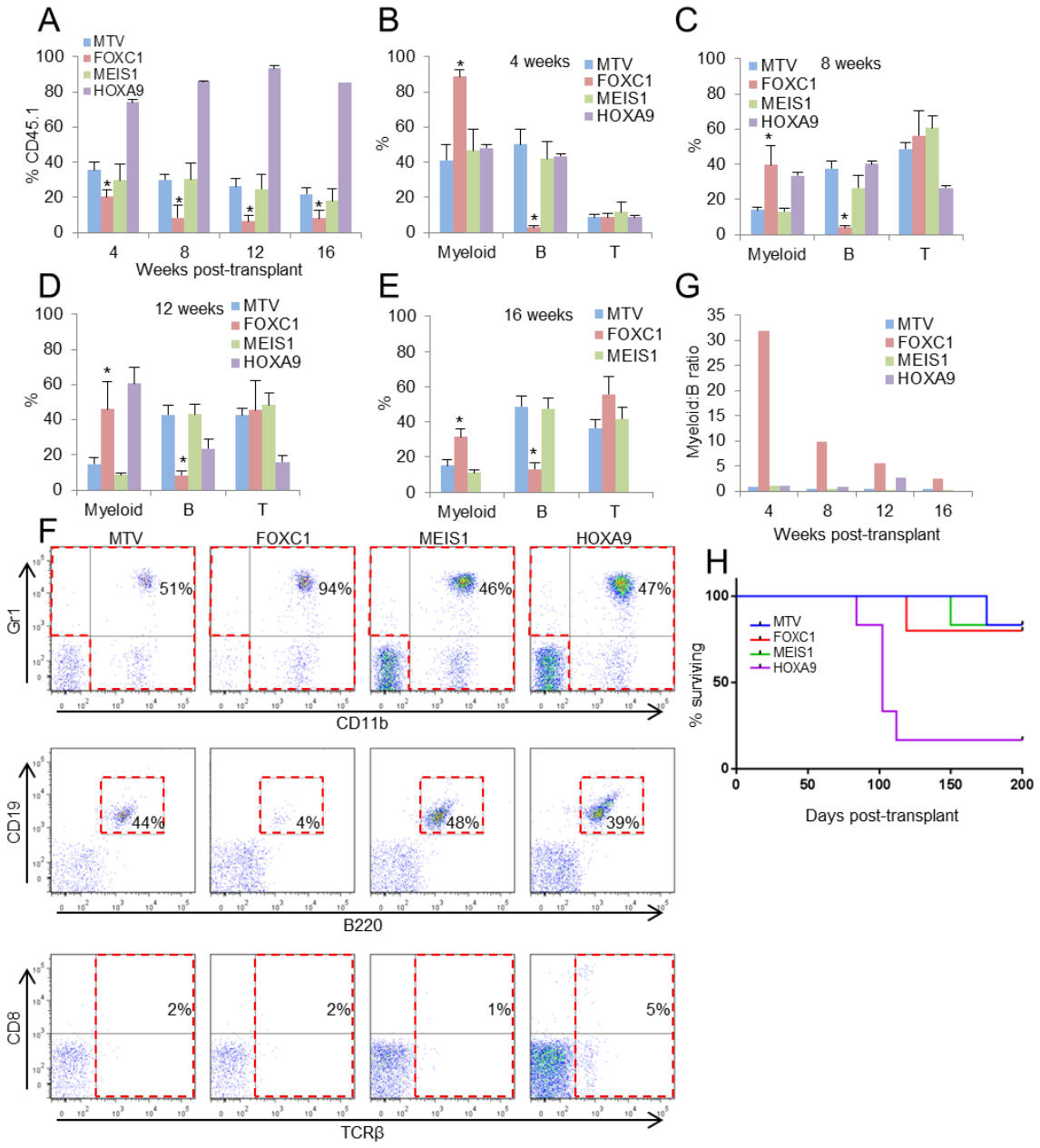


Figure 11. FOXC1 skews differentiation towards the myeloid lineage *in vivo.*

Murine CD45.1$^+$ CD117$^+$ BM cells were infected with the indicated retroviral vectors; 10^6 drug resistant cells were transplanted into CD45.2$^+$ irradiated congenic recipients 96 hours following spinoculation. (A) Bar chart shows mean+SEM percentage donor-derived CD45.1$^+$ cells in blood at the indicated times post-transplantation. * indicates $p<0.05$ for comparison of FOXC1$^+$ recipients versus all others, and at each time point, by one way ANOVA with Fisher's least significant difference *post hoc* test. (B-E) Bar charts show mean+SEM percentage contribution of donor-derived cells to the indicated lineages (myeloid-lineage, Gr-1$^+$ and/or CD11b$^+$; B-lineage, CD19$^+$B220$^+$; T-lineage, T-cell receptor β$^+$) at the indicated times post-transplantation. * indicates $p<0.05$ for comparison of FOXC1$^+$ recipients versus all others by one way ANOVA with Fisher's least significant difference *post hoc* test. (F) Representative flow cytometry plots indicate the gating strategy for identifying the linage contributions of donor-derived CD45.1$^+$ cells in blood. Images shown are from four weeks post-transplantation. (G) Bar chart shows the mean myeloid:B-lineage ratio of donor-derived CD45.1$^+$ cells in blood at the indicated times post-transplantation. (H) Survival curve of mice transplanted with cells infected with the indicated vectors (n=6-7 mice per cohort).

3.6. *FOXC1* expression in human AML is associated with high HOX gene expression

Given that expression of *FOXC1* alone is insufficient to induce AML, it was next investigated whether it might collaborate with other factors to promote leukaemogenesis. To identify transcription regulator genes whose expression was associated with that of *FOXC1* in human AML, *FOXC1*^{high} AMLs were compared with *FOXC1*^{low} AMLs (Table 30; Wouters et al., 2009) and *HOXA9*, *HOXA5* and *HOXB3* were found to be the most highly up regulated genes in the *FOXC1*^{high} group (Figure 12A). Notably, high level expression of HOX genes is a shared characteristic feature of both NPM1-mutated AML and those with a t(6;9) translocation, potentially explaining the particular association of high *FOXC1* expression with these molecular subtypes. Quantitative PCR (Figure 12B) and analysis of published microarray datasets (Figures 12C and 12D) confirmed the strong association of high *FOXC1* expression with high *HOXA9* expression. By qPCR, all five samples from our cohort with high *FOXC1* expression also exhibited high *HOXA9* expression, and 43% of high *HOXA9* expressing AML samples exhibited high *FOXC1* expression (Figure 12B). In the case of the microarray studies, 95/100 (95%) and 34/36 (94%) respectively with high *FOXC1* expression also exhibited high *HOXA9* expression (Figures 12C and 12D), and 95/320 (30%) and 34/115 (30%) respectively of high *HOXA9* expressing AML samples exhibited high *FOXC1* expression (Figures 12C and 12D). Of the total of seven *FOXC1*^{high}, *HOXA9*^{low} cases from both microarray studies, three exhibited high *HOXB2*, *HOXB3* or *HOXB4* expression, leaving just 4/129 (3%) of *FOXC1*^{high} cases across both studies lacking HOX gene expression.

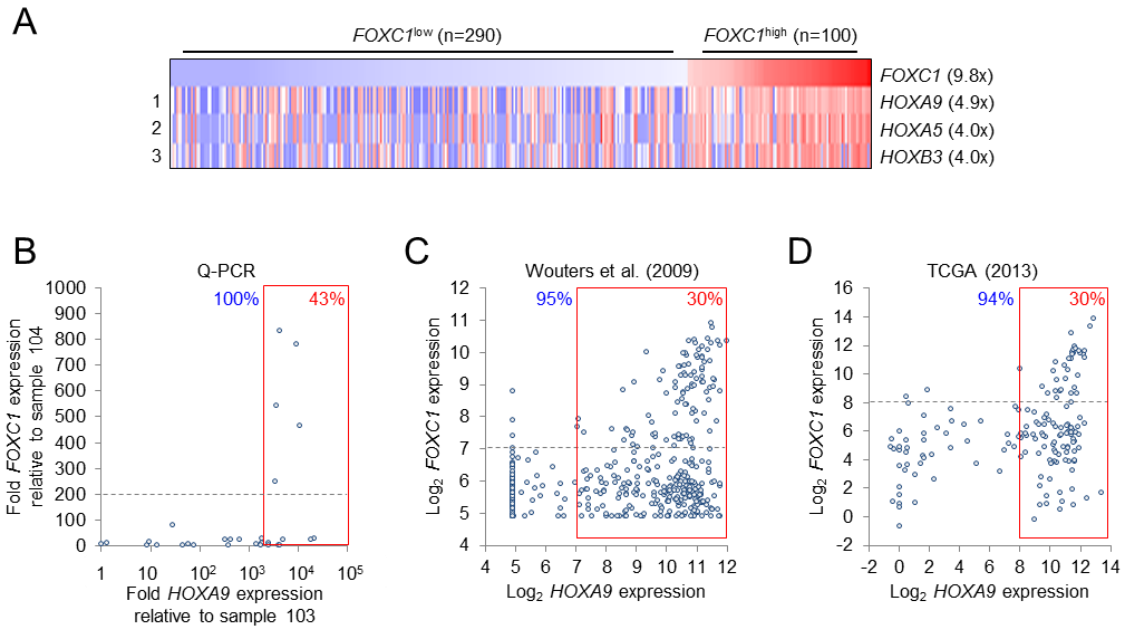
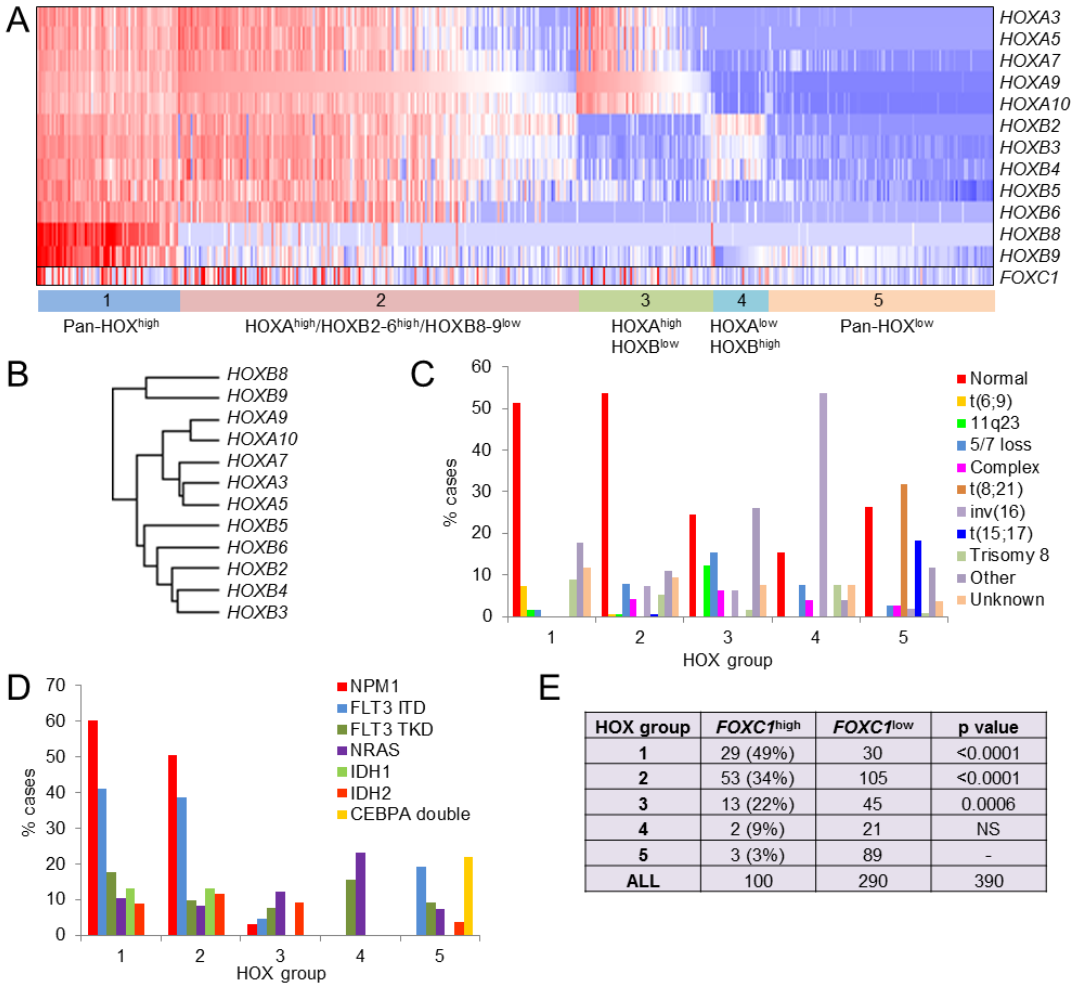


Figure 12. Association of *FOXC1* expression with *HOXA9* in human AML.
(A) Heat map shows comparison of *FOXC1*^{high} versus *FOXC1*^{low} AML cases from Wouters et al. (2009) indicating the most differentially expressed transcription factor genes, and in brackets their mean fold change. (B-D) Scatter plots show the expression of *FOXC1* versus *HOXA9* in primary AML patient samples as determined by (B) quantitative PCR (n=29) or (C&D) array expression values from the indicated studies. Percentages in blue text indicate proportion of *FOXC1*^{high} samples exhibiting high *HOXA9* expression. Percentages in red text indicate the proportion of *HOXA9*^{high} samples (in the red box) additionally exhibiting high *FOXC1* expression (above the dotted grey line). Assistance with bioinformatics analysis was received from Dr. Tim Somervaille.

To examine the association of *FOXC1* expression with HOX expression in more detail, 461 AML cases (Wouters et al., 2009) were grouped into five categories according to their pattern of HOX gene expression (Figure 13A). Twelve HOX genes were expressed at very high level in AML (where at least 1% of 461 samples exhibited probeset values in the top 10% of protein coding genes, i.e. $>\log_2$ 8.3) and unsupervised clustering of their expression patterns demonstrated three major clusters: *HOXA*, *HOXB2-6* and *HOXB8/9* (Figure 13B). The karyotypes and mutation spectra of the five groups were as expected (Figures 13C and 13D) (Valk et al., 2004). The *FOXC1*^{high} cases were most strongly associated with HOX group 1 (pan-HOX^{high}) and to a lesser extent with HOX groups 2 (HOXA^{high}/HOXB2-6^{high}/HOXB8-9^{low}) and 3 (HOXA^{high}/HOXB^{low}) (Figure 13E). This analysis indicates that *FOXC1* expression in AML is strongly associated with high level expression across the *HOXA* and *HOXB* locus, and not just *HOXA9*. The observation that approximately 30% of *HOXA9*-expressing human AML samples express *FOXC1* in a tissue-inappropriate manner provides a strong basis to suggest that FOXC1 may collaborate with HOX family transcription factors to enhance leukaemogenesis.

3.7. FOXC1 collaborates with HOXA9 to enhance clonogenic potential, differentiation block and accelerate onset of symptomatic leukaemia

To investigate this question further, murine CD117^{+} HSPCs were infected in pairwise combinations with retroviral vectors expressing *Hoxa9*, *FOXC1* or *Meis1*, or a control vector (to generate *Hoxa9/*MTV, *Hoxa9*/*FOXC1* and *Hoxa9*/*Meis1* cells, respectively) and their clonogenic potential was assessed in serial replating assays. MEIS1 is an established HOXA9 cofactor (Kroon et al., 1998). As expected, *Hoxa9* expression induced sustained clonogenic potential of BM HSPCs in serial replating assays, an effect which was enhanced by co-expression of *Meis1* (Figure 14A). Interestingly, the co-expression of *FOXC1* with *Hoxa9* significantly enhanced the clonogenic potential of BM HSPCs versus cells overexpressing *Hoxa9* alone (Figure 14A). *Hoxa9/FOXC1* co-expression also significantly enhanced the differentiation block by comparison with cells expressing *Hoxa9* alone: by the fourth round of culture *Hoxa9*/*FOXC1* cells formed approximately four times more Type I colonies (which contain poorly differentiated myeloid cells) (Figures 14A-C), displayed a more extensive morphologic differentiation block in cytospin preparations (Figure 14D), and expressed significantly lower levels of both the mature myeloid marker Gr1 and the monocyte/macrophage marker F4/80 (Figures 14E and 14F). Experiments performed in HEK 293FT cells show that FOXC1 and HOXA9 co-immunoprecipitate, suggesting that the interaction between the proteins may underlie this enhanced differentiation block *in vitro* (Figure 14G and 14H), although further experiments are required to confirm this hypothesis.



HOX group	*FOXC1*high	*FOXC1*low	p value
1	29 (49%)	30	<0.0001
2	53 (34%)	105	<0.0001
3	13 (22%)	45	0.0006
4	2 (9%)	21	NS
5	3 (3%)	89	-
ALL	100	290	390

Figure 13. Association of *FOXC1* expression with HOX gene expression in human AML. Cluster analysis of *HOXA* and *HOXB* gene expression in 461 cases of AML from Wouters et al. (2009). To focus on HOX genes which are highly expressed in a significant sub-fraction of AML samples, only genes where at least 1% of samples exhibited expression values in the top 10% of array expression values (i.e. >$\log_2$ 8.3) were retained. Where multiple probesets for the same gene were present, the probeset with the greatest range of values was selected. (A) Heat map illustrates five distinct patterns of HOX gene expression in AML, based on (B) an unsupervised cluster analysis of the 12 informative probesets. (C) Association of karyotype with HOX gene cluster. (D) Association of mutation status with HOX gene cluster. (E) Association of *FOXC1*^{high} cases with HOX gene cluster. Statistical significance for the distribution of *FOXC1*^{high} versus *FOXC1*^{low} cases in each HOX cluster versus cluster 5 was assessed using Fisher's Exact Test. Assistance with bioinformatics analysis was received from Dr. Tim Somervaille.

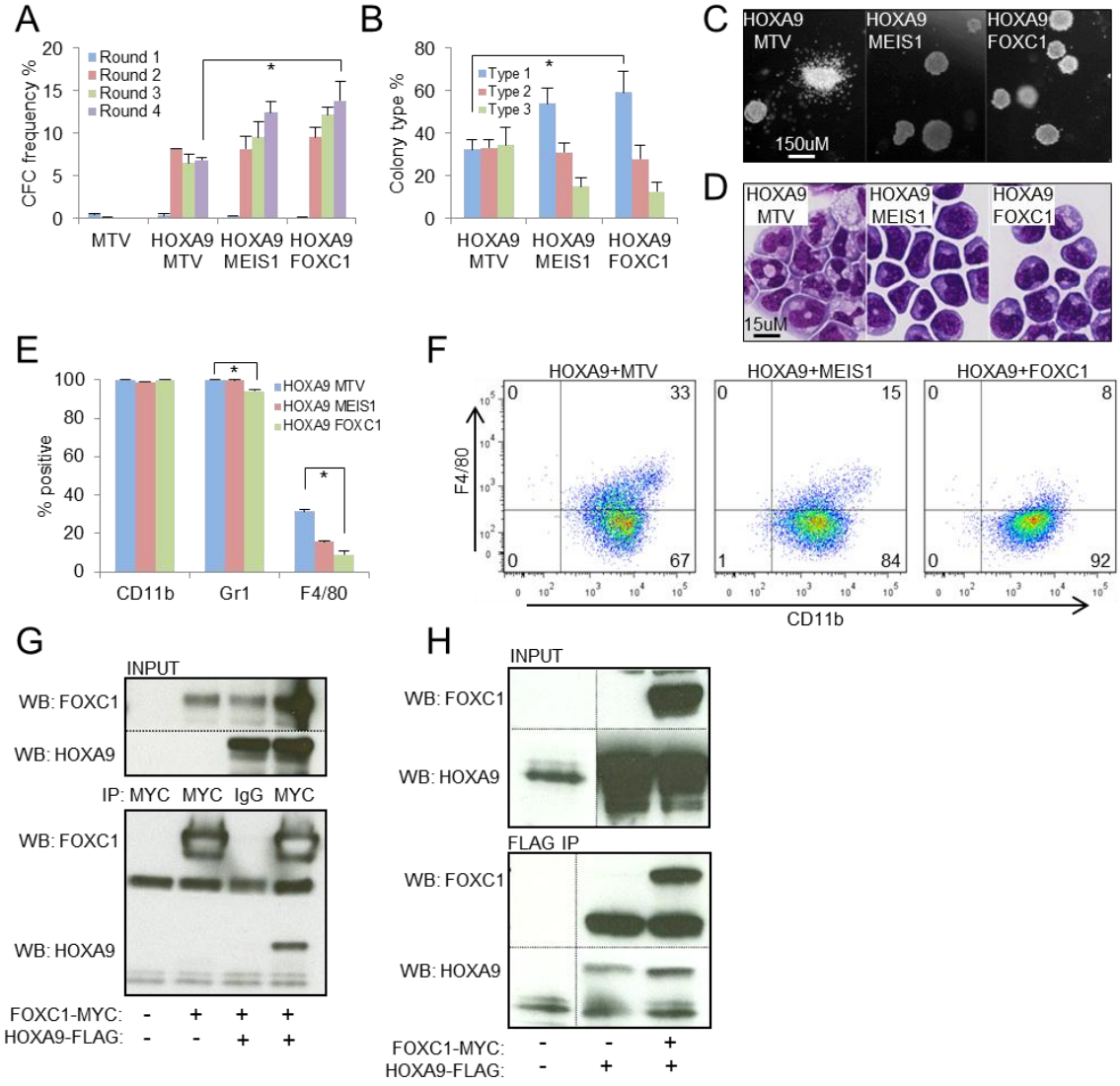
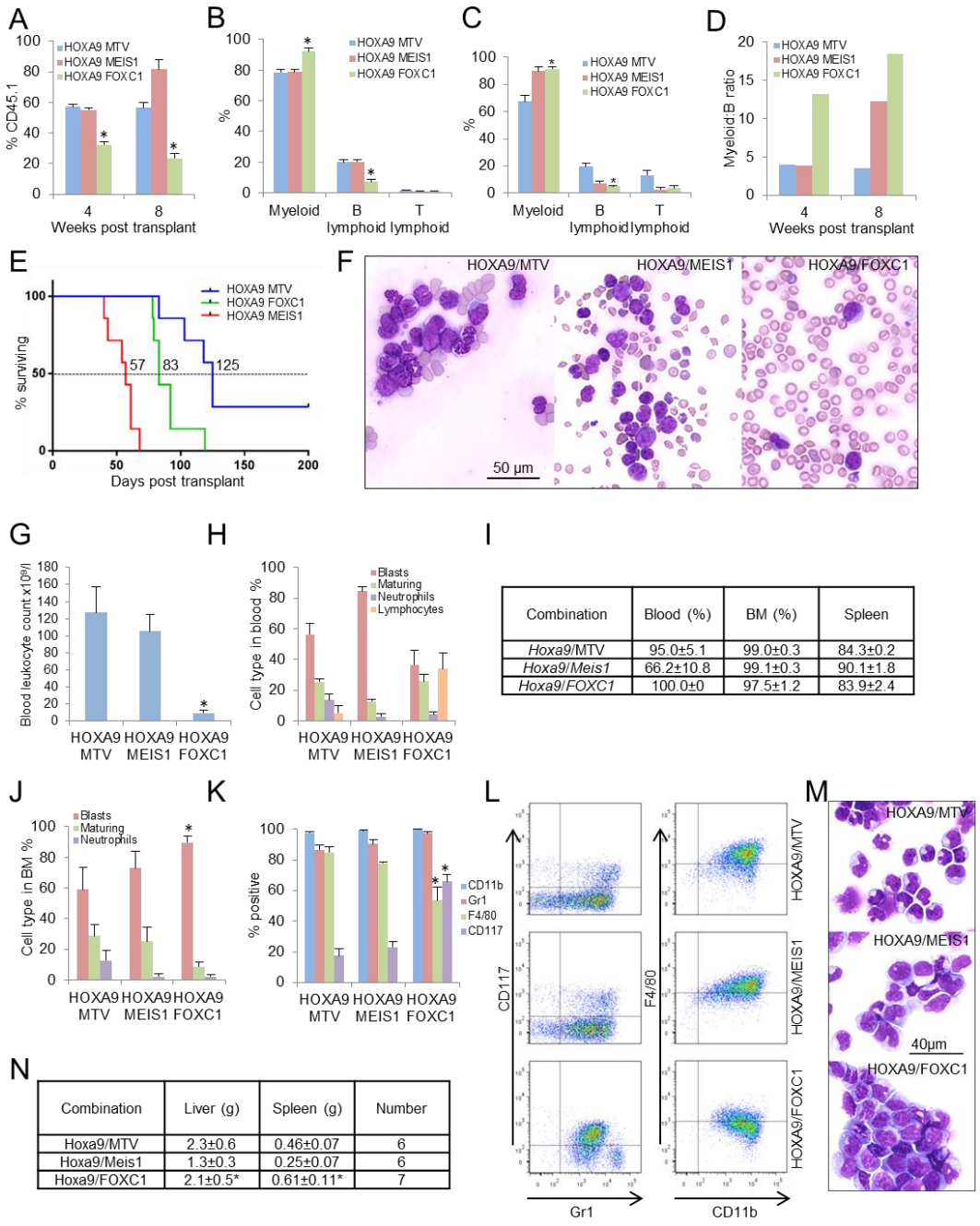


Figure 14. FOXC1 collaborates with HOXA9 to enhance clonogenic potential and differentiation block in BM HSPCs. (A) Bar chart shows mean+SEM colony-forming cell (CFC) frequencies after each round of serial replating of murine CD117^{+} BM cells co-transduced with the indicated retroviral or control expression vectors (n=3). MTV = empty vector. (B) Bar chart shows mean+SEM frequencies of colony types after the fourth round of serial replating (n=3). Type I colonies contain poorly differentiate myeloblasts, Type II colonies contain a mixed population of blasts and differentiating myeloid cells and Type III colonies contain terminally differentiated myeloid cells. For A and B, * indicates p<0.05 by one way ANOVA with Fisher's least significant difference *post hoc* test. Representative images show (C) colonies and (D) cytospin preparations from the end the fourth round of replating. (E) Bar chart shows mean+SEM percentage of cells positive for the indicated cell surface markers (as determined by flow cytometry) following six days in liquid culture (n=3). * indicates p<0.001 by one way ANOVA with Fisher's least significant difference *post hoc* test. (F) Representative flow cytometry plots of cells shown in (E). (G&H) MYC-tagged *FOXC1* and FLAG-tagged *Hoxa9* were overexpressed alone or in combination in HEK 293FT cells and lysate subsequently immunoprecipitated with anti-MYC (G) or anti-FLAG (H) antibody. Subsequent western blots show the interaction between FOXC1 and HOXA9.

To determine whether HOXA9 and FOXC1 collaborate in leukaemogenesis, *Hoxa9/*MTV, *Hoxa9*/*FOXC1* and *Hoxa9*/*Meis1* double transduced HSPCs were transplanted into irradiated congenic recipients. Levels of donor:recipient chimerism in blood at four and eight weeks post-transplant were lower in *Hoxa9*/*FOXC1* recipients than in the other two cohorts (Figure 15A). Donor-derived multilineage engraftment was observed in all three cohorts at both time points (Figures 15B and 15C). However, as with the single transduction transplants (Figures 11B-F), there were significantly higher levels proportionately of myeloid engraftment, and lower levels of B-lymphoid engraftment, among recipients of *Hoxa9*/*FOXC1* cells by comparison with mice receiving either *Hoxa9/*MTV or *Hoxa9*/*Meis1* cells (Figures 15B-D). As expected, recipients of *Hoxa9*/*Meis1* cells developed AML more rapidly than recipients of *Hoxa9/MTV* cells (median 57 days versus 125 days; Figure 15E). Strikingly, despite lower donor:recipient chimerism in the blood in the initial post-transplant period, recipients of *Hoxa9*/*FOXC1* cells also succumbed to AML substantially earlier than mice receiving *Hoxa9/MTV* cells (83 days versus 125 days; Figure 15E). By comparison with *Hoxa9/*MTV or *Hoxa9*/*Meis1* recipients, at the point of death *Hoxa9*/*FOXC1* recipients exhibited significantly lower total circulating leukocyte counts, although 30-40% of these were nevertheless blasts (Figures 15F-H). However, *Hoxa9*/*FOXC1* recipients exhibited extensive tissue infiltration with blast cells, high level donor:recipient chimerism in BM and spleen and significantly larger livers and spleens compared to controls (Figures 15I, 15J and 15N). Thus, a key feature of the *Hoxa9*/*FOXC1* murine AMLs was failure of AML blasts to mobilise substantially to the blood in spite of high level involvement of BM, liver and spleen.

In the BM, *Hoxa9*/*FOXC1* recipients exhibited significantly higher blast percentages, significantly higher CD117 expression and significantly lower F4/80 expression (Figures 15J-M). In each cohort and in each case, autopsy demonstrated splenomegaly and hepatomegaly, with spleen weights being significantly higher in *Hoxa9*/*FOXC1* recipients than both other cohorts, and liver weights being significantly higher in *Hoxa9*/*FOXC1* recipients versus *Hoxa9/Meis1* recipients (Figure 15N). Histological analysis of spleen demonstrated that in *Hoxa9/MTV* and *Hoxa9/Meis1* recipients the splenic architecture was maintained with preservation of the white pulp. The red pulp was however expanded by sheets of blast cells, with scattered normal megakaryocytes and erythroid precursors indicating residual but extramedullary haematopoiesis. By comparison, in *Hoxa9*/*FOXC1* recipients the splenic architecture was completely effaced by sheets of blast cells (Figure 15O). In the livers of *Hoxa9/MTV* and *Hoxa9/Meis1* recipients there was periportal and perivenular accumulation of blast cells, more extensive in the former than the latter, with relative sparing of the intervening parenchyma where only single blasts and small groups of blasts were seen within sinusoids. By contrast, in *Hoxa9*/*FOXC1* recipients there was a diffuse infiltrate of blast cells, forming small clusters which expanded the sinusoids. Periportal and perivenular blasts, while present, were less conspicuous than those seen in mice from the other cohorts (Figures 15P and 15Q). Cell cycle analysis of AML cells from both BM and spleen demonstrated a significantly higher fraction of cycling cells in *Hoxa9*/*FOXC1* recipients by comparison with the other cohorts, consistent with a differentiation block at a proliferative progenitor stage (Figures 15R-T).



Combination	Blood (%)	BM (%)	Spleen
Hoxa9/MTV	95.0±5.1	99.0±0.3	84.3±0.2
Hoxa9/Meis1	66.2±10.8	99.1±0.3	90.1±1.8
Hoxa9/FOXC1	100.0±0	97.5±1.2	83.9±2.4

Combination	Liver (g)	Spleen (g)	Number
Hoxa9/MTV	2.3±0.6	0.46±0.07	6
Hoxa9/Meis1	1.3±0.3	0.25±0.07	6
Hoxa9/FOXC1	2.1±0.5*	0.61±0.11*	7

Figure 15. *FOXC1* collaborates with *HOXA9* to accelerate leukaemogenesis.
Murine CD45.1$^+$ CD117$^+$ BM cells were co-transduced with retroviral vectors. 96 hours later 10^6 drug resistant cells were transplanted into CD45.2$^+$ irradiated congenic recipients. (A) Bar chart shows mean+SEM percentage of donor-derived CD45.1$^+$ cells in blood at the indicated times post-transplant. (B&C) Bar charts show mean+SEM percentage contribution of donor-derived cells to each lineage in blood at four (B) and eight (C) weeks post-transplant. * indicates p<0.001 by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of FOXC1$^+$ samples versus MTV samples. (D) Bar chart shows the mean myeloid:B-lineage ratio of donor-derived cells in blood at the indicated times post-transplant. (E) Survival curves of transplanted mice (n=7 per cohort). Median survivals are shown. (F) Representative blood smears from sick mice with leukaemias initiated by expression of the indicated combination of transcription factors. Bar charts show mean+SEM (G) total blood leukocyte count and (H) percentage leukocyte type in blood at death in the indicated cohorts, as determined by haemocytometer counting and morphologic analysis of blood smears respectively (n=3-5 per cohort). (I) Table shows mean±SEM percentage donor:recipient chimerism at death, as determined by flow cytometry (CD45.1 expression; BM and spleen) or by morphologic analysis of blood smears (i.e. all cells with myeloid leukaemia morphology (blasts, maturing, abnormal neutrophils) were assumed to be of donor origin and part of the AML clone). Figure 15 legend is continued on the next page.

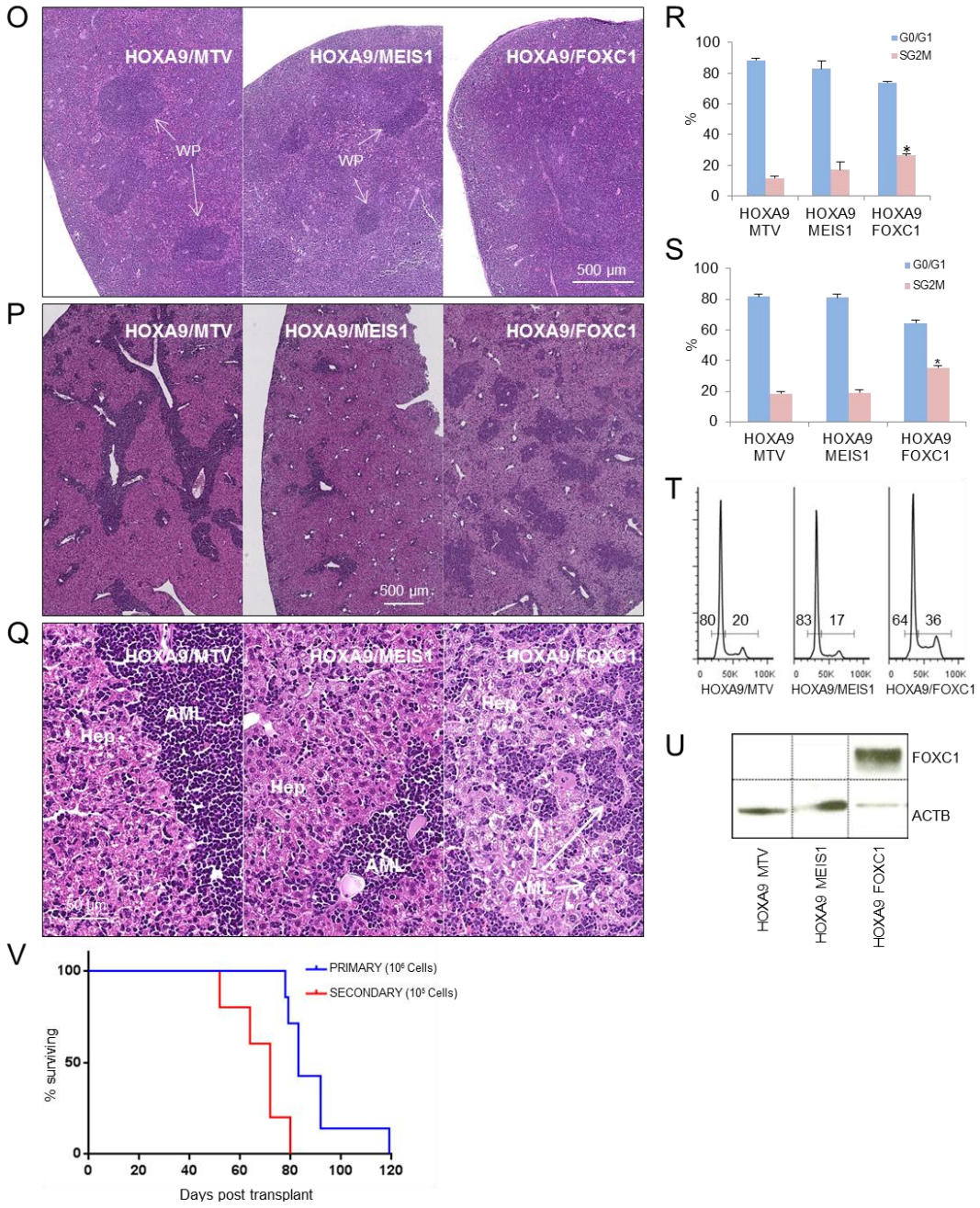


Figure 15 (continued). *FOXC1* collaborates with *HOXA9* to accelerate leukaemogenesis.
(J) Bar chart shows mean+SEM percentage cell type in BM at death (n=3-5 per cohort). (K) Bar chart shows the mean+SEM percentage of donor-derived cells positive for the indicated cell surface markers in BM of leukaemic mice, as determined by flow cytometry. (L) Representative flow cytometry plots from (K). (M) Representative images from (J). (N) Table shows mean±SD organ masses of leukaemic animals at death. * indicates that *Hoxa9/FOXC1* recipient mice organ weights are significantly greater by comparison with those from *Hoxa9/MTV* (spleen only) or *Hoxa9/Meis1* (liver and spleen) recipients ($p<0.05$; unpaired t-test). Images are representative of (O) spleen histology and (P&Q) liver histology from sick mice with leukaemias initiated by expression of the indicated combination of transcription factors. WP = white pulp; AML = leukaemia cells infiltrating liver; Hep = hepatocytes. Bar charts show mean+SEM percentage of BM (R) and spleen (S) cells in the indicated phase of the cell cycle at death. (T) Representative profiles from (R). (U) Western blot shows expression of the indicated proteins in AML splenocytes from representative mice in each cohort. (V) Survival curves of sub-lethally (4.5Gy) irradiated mice secondarily transplanted with 10^5 *Hoxa9/FOXC1* AML cells. Results of primary transplants of 10^6 cells are shown for comparison. For (A), (B), (C), (G), (J), (K), (R) and (S) * indicates $p<0.01$ by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of $FOXC1^+$ samples versus each of the others. Assistance with histopathological analysis was received from Dr. Edmund Cheesman.

As expected, FOXC1 protein was expressed in the *Hoxa9*/*FOXC1* AML cells (Figure 15U), which were also able to initiate leukaemia in secondarily transplanted recipients with shortened latency (Figure 15V).

Together these *in vitro* and *in vivo* data demonstrate that *FOXC1* collaborates with *HOXA9* to increase clonogenic potential and cell cycle progression, enhance a monocyte/macrophage and B-lineage lineage differentiation block and accelerate the onset of symptomatic leukaemia in mice.

3.8. FOXC1 represses a monocyte differentiation programme in leukaemic haematopoiesis

To investigate the consequences of *FOXC1* expression on the transcriptome in murine AML, exon array analysis was performed using flow sorted $CD117^{+}Gr1^{+}$ leukaemia cells recovered from sick mice. Populations with this immunophenotype are enriched for leukaemia-initiating cell activity in *Hoxa9/Meis1* murine leukaemias (Gibbs et al., 2012). In keeping with the observed immunophenotypes of the respective leukaemias (Figures 15K and 15L), analysis of protein coding genes that passed threshold criteria demonstrated that *Hoxa9*/*FOXC1* AMLs clustered separately from *Hoxa9*/*Meis1* and *Hoxa9*/MTV AMLs (which clustered much more closely with one another) (Figure 16A). Genes with the highest mean differential expression (at least two-fold) in *Hoxa9*/*FOXC1* AMLs versus the others formed two groups: Group A genes, which were more highly expressed in the *Hoxa9*/*FOXC1* AMLs versus the others, and Group B genes (the larger set) which were repressed (Figure 16B and Appendix Table 1).

Gene ontology analysis demonstrated significant enrichment within the Group B gene set of biological process terms such as “immune response”, “defence response” and “inflammatory response” (Table 32) indicating a gene set associated with myeloid cells involved in inflammation and immunity. At a similar level of statistical significance there were no enriched terms among the Group A gene set.

To determine whether FOXC1-regulated gene sets in murine AML cells were more highly expressed in monocytes or neutrophils, the relative expression of human homologs of Group A and B genes were evaluated in exon array datasets from primary human cells using gene set enrichment analysis (GSEA) (Appendix Table 2). In keeping with reduced expression of the monocyte/macrophage marker F4/80 in murine *Hoxa9/FOXC1* AML cells (Figures 15K and 15L), this cross-species analysis demonstrated that FOXC1 repressed a monocyte-expressed gene set and promoted expression of a neutrophil-expressed gene set (Figure 16C). Also in keeping with the unexpected ability of the mesenchyme-expressed transcription factor FOXC1 to repress monocyte differentiation, it was observed that phorbol ester treatment of HL60 AML cells (which express FOXC1) led to substantial down regulation of the protein as cells underwent monocytic differentiation, whereas the reverse was the case as cells underwent granulocytic differentiation following all-trans retinoic acid treatment (Figure 16D).

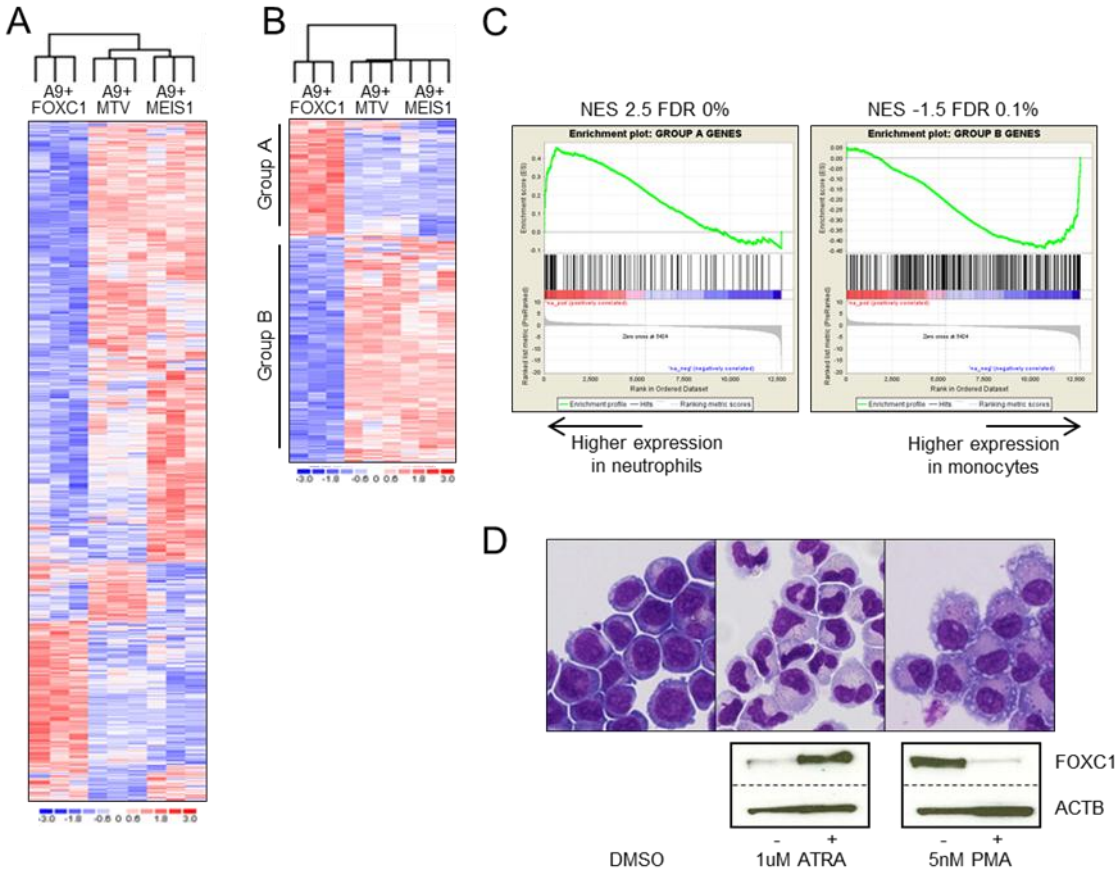


Figure 16. FOXC1 represses a monocyte/macrophage differentiation programme in murine AML. (A) Heat map shows unsupervised hierarchical clustering of 3631 genes passing threshold criteria (expressed (log_2 expression value >4.1), with significantly different expression levels ($p<0.05$, unpaired t-test) in at least one of the three possible pairwise comparisons between AML cohorts). (B) Cluster analysis of 567 protein coding genes that passed threshold criteria (expressed (log_2 expression value >4.1) and with significantly different expression levels ($p<0.05$, unpaired t-test and >2-fold difference) in at least one of the two pairwise comparisons between *Hoxa9/FOXC1* AMLs versus the others). (C) GSEA plots show analyses of enrichment of human homologs of Group A and/or Group B genes in protein coding gene lists ranked using a signal-to-noise metric according to expression in primary human neutrophils versus primary human monocytes. (D) HL60 AML cells were treated for 48 hours with all-trans retinoic acid (ATRA) to promote granulocytic lineage differentiation, phorbol ester (PMA) to promote monocytic differentiation, or DMSO vehicle. Images (upper panels) show cytospin preparations of treated cells and western blots (lower panels) show expression of FOXC1 in the indicated conditions. Assistance with bioinformatics analysis was received from Dr. Tim Somervaille.

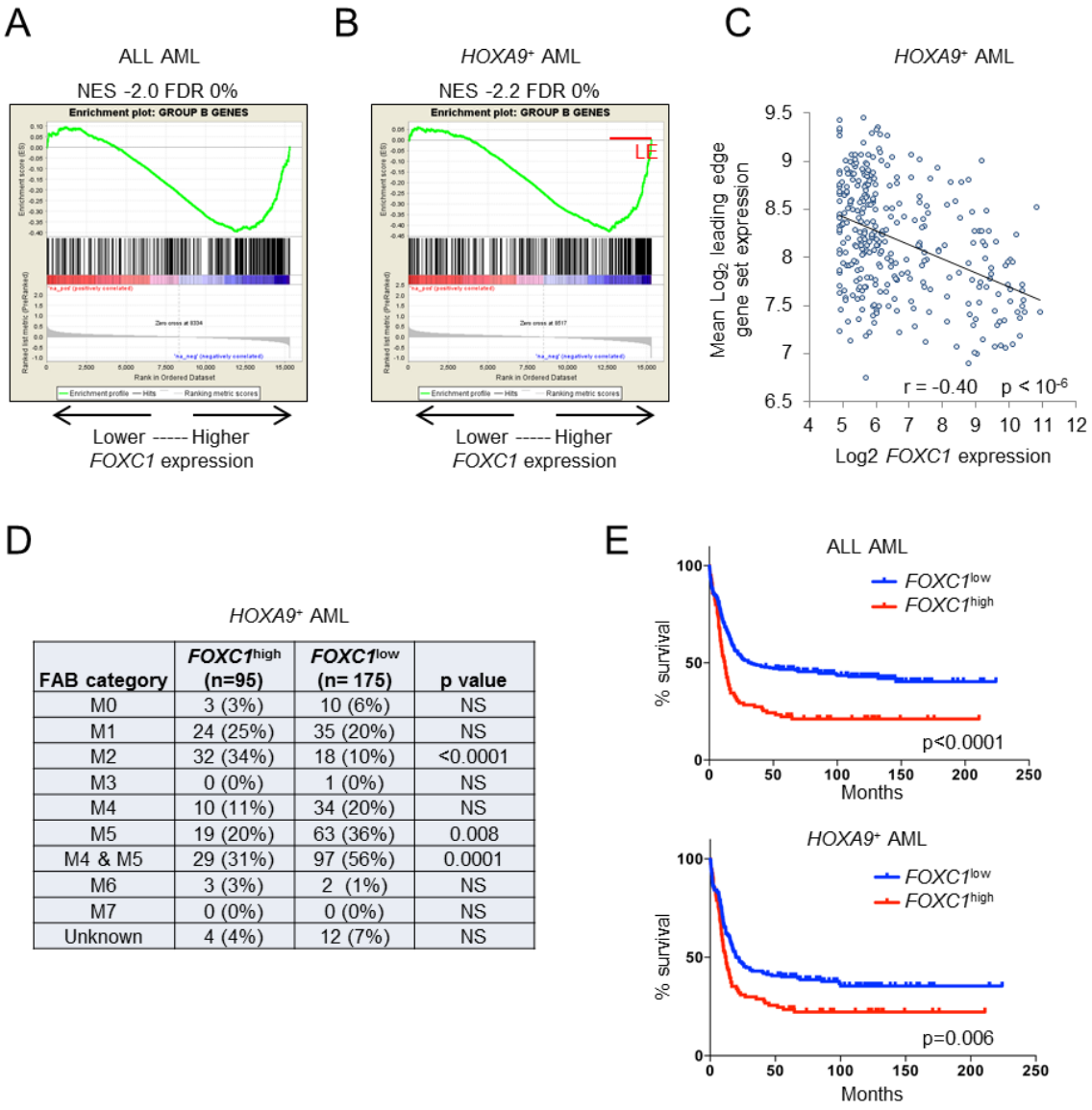
Term	False discovery rate (%)
Immune response	0
Antigen processing and presentation of peptide antigen	0
Defense response	0
Response to wounding	0
Antigen processing and presentation of exogenous peptide antigen	0
Cell activation	0.1
Antigen processing and presentation of exogenous antigen	0.1
Positive regulation of immune system process	0.1
Regulation of cytokine production	0.2
Leukocyte activation	0.3
Inflammatory response	0.5
Antigen processing and presentation	0.5

Table 32. Gene Ontology Biological Process categories enriched in the Group B gene set. Related to Figure 16. Assistance with bioinformatics analysis was received from Dr. Tim Somervaille.

Next, to determine whether a signature of FOXC1 transcriptional activity could be identified in human AML, and to confirm the functional relevance of its derepression in human disease, the expression of human homologs of FOXC1-repressed Group B genes in AML were evaluated using GSEA (Figure 16B and Appendix Table 1). Protein coding genes in *FOXC1*^{high} AMLs versus *FOXC1*^{low} AMLs (Wouters et al., 2009) were ranked using a signal-to-noise ranking metric (Appendix Table 2). GSEA demonstrated highly significant negative enrichment of FOXC1-repressed Group B genes among *FOXC1*^{high} AMLs versus *FOXC1*^{low} AMLs, and this was observed whether all AMLs were considered (Figure 17A; n=461) or just those expressing *HOXA9* (Figure 17B; n=320). In fact, in leading edge analyses (Appendix Table 1), there was a highly significant association of higher *FOXC1* expression with greater repression of Group B genes (Figure 17C). Remarkably, when the morphological classification of *HOXA9*$^{+}$ AMLs was considered, among cases with high *FOXC1* expression there were significantly fewer AMLs of French-American-British (FAB) M4 and M5 subtypes, which exhibit monocytic differentiation, and significantly more of the FAB M2 subtype, which lack it (Figure 17D). Furthermore, patients with *FOXC1*^{high} AML exhibited significantly inferior survival by comparison with *FOXC1*^{low} cases, whether all other AMLs were considered (median 12 versus 32 months) or only those expressing *HOXA9* (median 12 versus 20 months) (Figure 17E). Indeed, in multivariate analyses (Table 33), high *FOXC1* expression was an independent predictor for inferior survival in addition to age, cytogenetic risk score and *NPM1*/*FLT3* mutation status. These data indicate that in human leukaemic haematopoiesis, as in murine leukaemic haematopoiesis, expression of FOXC1 enhances a block to monocyte/macrophage differentiation and leads to inferior survival.

	HR	95% CI	*P*-value
Age	1.016	1.01-1.03	**0.004**
Female (vs male)	1.020	0.80-1.31	0.876
Cytogenetic risk (vs intermediate risk)			**0.005**
Good risk	0.539	0.33-0.87	**0.012**
Poor risk	1.289	0.93-1.79	0.129
Mutation positive (vs wild type):			
FLT3/NPM1:			**0.001**
FLT3^{-}/NPM1^{+}	0.342	0.20-0.60	**<0.001**
FLT3^{+}/NPM1^{-}	1.146	0.81-1.62	0.439
FLT3^{+}/NPM1^{+}	0.780	0.53-1.15	0.207
CEBPA	0.932	0.59-1.47	0.763
IDH1	0.826	0.49-1.38	0.467
IDH2	0.714	0.45-1.15	0.162
KRAS	1.643	0.38-7.13	0.507
NRAS	0.969	0.36-1.48	0.884
FOXC1 high expression (vs low)	1.784	1.29-2.46	**<0.001**

Table 33. Multivariate analysis of overall survival in AML.
Table shows multivariate analysis of overall survival of AML patients from Wouters et al. (2009) (for whom survival data were available (n=458)) using a Cox proportional hazards regression model. Hazard ratios (HR), 95% confidence intervals (CI) and probability values are provided for all variables included in the model. For variables with ≥3 categories the first P-value indicates that variable's overall effects on survival. Although certain variables (e.g. poor risk cytogenetics; FLT3+/NPM1- mutation status) demonstrated significant negative impacts on survival in univariate analyses, and despite hazard ratios indicating effects in the expected directions, when corrected for all covariates their independent prognostic significance was not retained, likely due to the large number of variables included in this model. Age was analysed as a continuous variable, all others as categorical variables. Assistance with statistical analysis was received from Dr. Daniel Wiseman.



FAB category	*FOXC1*^{high} (n=95)	*FOXC1*^{low} (n= 175)	p value
M0	3 (3%)	10 (6%)	NS
M1	24 (25%)	35 (20%)	NS
M2	32 (34%)	18 (10%)	<0.0001
M3	0 (0%)	1 (0%)	NS
M4	10 (11%)	34 (20%)	NS
M5	19 (20%)	63 (36%)	0.008
M4 & M5	29 (31%)	97 (56%)	0.0001
M6	3 (3%)	2 (1%)	NS
M7	0 (0%)	0 (0%)	NS
Unknown	4 (4%)	12 (7%)	NS

Figure 17. FOXC1 represses a monocyte/macrophage differentiation programme in human AML. (A&B) GSEA plots show analyses of enrichment of human homologs of Group A and/or Group B genes in protein coding gene lists ranked using a signal-to-noise metric according to expression in (A) *FOXC1*^{high} (n=100) versus *FOXC1*^{low} (n=290) primary AML samples (Wouters et al., 2009) or (B) *HOXA9* expressing *FOXC1*^{high} (n=95) versus *FOXC1*^{low} (n=175) primary AML samples. (C) Scatter plot shows expression of *FOXC1* versus mean log_2 expression for the leading edge gene set shown in (B) (see also Appendix Table 1) in *HOXA9* expressing AML (n=320). (D) Analysis of morphological classification of *HOXA9* expressing *FOXC1*^{high} (n=95) versus *FOXC1*^{low} (n=175) primary AML samples. Statistical significance for the indicated comparisons was assessed using Fisher's Exact Test. (E) Survival curves of patients with *FOXC1*^{high} versus *FOXC1*^{low} AML. Assistance with bioinformatics analysis was received from Dr. Tim Somervaille and analysis of survival data from Dr. Daniel Wiseman.

3.9. FOXC1 regulates expression of *KLF4*

The transcription factor KLF4 positively regulates monocyte differentiation (Feinberg et al., 2009) and exhibits anti-proliferative and tumour suppressor activity in B-cell malignancies (Kharas et al., 2007). Consistent with the observation that *FOXC1* expression is inversely associated with monocytic morphologic classification of human AML (Figure 17D; Wouters et al., 2009), *KLF4* expression was significantly lower in *FOXC1*^{high} human AML versus *FOXC1*^{low} AML (Figure 18A). *Klf4* expression was also significantly lower in murine *Hoxa9/FOXC1* AMLs by comparison with *Hoxa9*/MTV and *Hoxa9/Meis1* AMLs (Figure 18B). Functionally, in human THP1 AML cells, *FOXC1* knockdown led to *KLF4* up regulation (Figure 18C) and in murine CD117^{+} HSPCs forced expression of *FOXC1* reduced *Klf4* expression, had no effect on expression of *Hoxa9* and modestly increased *Meis1* expression (Figure 18D). Forced expression of *KLF4* in murine *Hoxa9/FOXC1* AML cells significantly reduced both clonogenic cell frequencies and colony size (Figures 18E-G) through reduction of the proportion of cells in the SG_2M phase of the cell cycle (Figure 18H). These data suggest that direct or indirect repression of *KLF4* by FOXC1 is one significant contributing factor to the phenotypic appearances of murine and human FOXC1^{+} AMLs.

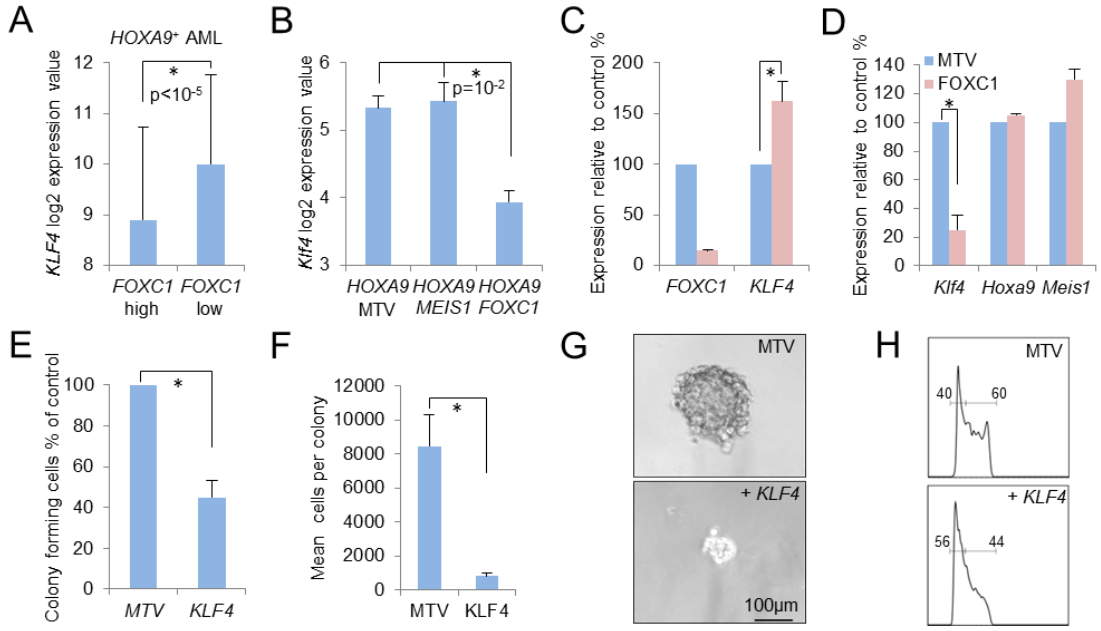


Figure 18. FOXC1 regulates *KLF4*.
Bar charts show (A) mean+SD log_2 array expression value for *KLF4* in *HOXA9*-expressing *FOXC1*^{high} (n=95) and *FOXC1*^{low} (n=175) human AML cases from Wouters et al. (2009) and (B) mean+SEM log_2 array expression values for *Klf4* in the indicated murine leukaemias (n=3 per cohort). Statistical significance was assessed using respectively an unpaired t-test or one way ANOVA with Fisher's least significant difference *post hoc* test. (C) Bar chart shows mean+SEM relative expression of the indicated genes 72 hours after initiation of *FOXC1* KD using construct KD3 or a non-targeting control vector (NTC) (n=3) (see also Figure 5). (D) Bar chart shows mean+SEM relative expression of the indicated genes 72 hours after retroviral infection of CD117^{+} HSPCs with *FOXC1*-expressing or control retroviral vectors, and 48 hours following drug selection (n=3) (see also Figure 10). (E-G) *Hoxa9/FOXC1* AML cells were infected with *KLF4*-expressing or control retroviral vectors and cultured in semi-solid medium. Bar charts show (E) mean+SEM relative colony forming cell frequencies and (F) mean+SEM cells per colony in *KLF4*-expressing cells versus controls (n=3). For (C)-(F) * indicates p<0.05 with an unpaired t-test for the indicated comparisons. (G) Representative image from (E). (H) Representative cell cycle profiles from (E).

3.10. Loss of Polycomb-mediated repression promotes *FOXC1* derepression

In normal human $CD34^+$ HSPCs, the *FOXC1* gene occupies a DNA hypomethylated and histone H3K27 trimethylated region of the genome (Figure 19 and 22A) (Zhou et al., 2011). Interestingly, despite absent or very low level expression, the locus exhibits hypersensitivity to DNase treatment, as well as the presence of histone H3K4 methylation marks and acetylation of H2A, H2B, H3 and H4 residues, suggesting a lack of chromatin compaction (Figure 19). In K562 leukaemia cells, *FOXC*1 is also not expressed and sits in chromatin with similar features to that seen in $CD34^+$ cells. Importantly, there is significant binding of Polycomb Repressive Complex (PRC) 2 components EZH2 and SUZ12 across the locus correlating with the presence of H3K27me3 marks. In addition, there are also co-localised binding peaks of PRC1 components RNF2, CBX2 and CBX8 (Figure 20). In HeLa cells where *FOXC1* is expressed, EZH2 binding to the locus is absent (Figure 21). In order to validate the results of these ChIP-sequencing studies with respect to the *FOXC1* promoter, ChIP for H3K27me3 and H3K4me3 histone modifications at the *FOXC1* promoter region in THP1 AML cells, and as a comparator K562 cells, was performed (Figure 22B). These analyses confirmed that the *FOXC1* promoter region in THP1 AML cells, where the gene is expressed, exhibits greater trimethylation of H3K4 and reduced trimethylation of H3K27 by comparison with the promoter region in K562 cells, where it is not expressed (Figure 22B).

These data raise the possibility that in normal haematopoietic cells transcriptional silence of *FOXC1* is maintained by PRC. To address this normal human $CD34^+$ cells from multiple donors were treated with PRC inhibitors. Cells treated with the EZH2 inhibitor GSK343 (Verma et al., 2012) exhibited a significant increase in expression of *FOXC1*, but not *HOXA9* which lacks significant H3K27 trimethylation in $CD34^+$ cells (Figures 22C). In separate experiments, cells treated with UNC1999, a dual EZH1 and EZH2 inhibitor (Konze et al., 2013), exhibited a more extensive fold-increase in *FOXC1* expression, but again expression of *HOXA9* was unaffected (Figure 22D). PRT4165, a PRC1 E3 ubiquitin ligase inhibitor (Ismail et al., 2013), modestly enhanced *FOXC1* but not *HOXA9* expression (Figure 22D). These data indicate that continued repression of *FOXC1* in the haematopoietic system is mediated by PRC2 and imply that loss of its activity at this genomic locus contributes to its derepression in AML.

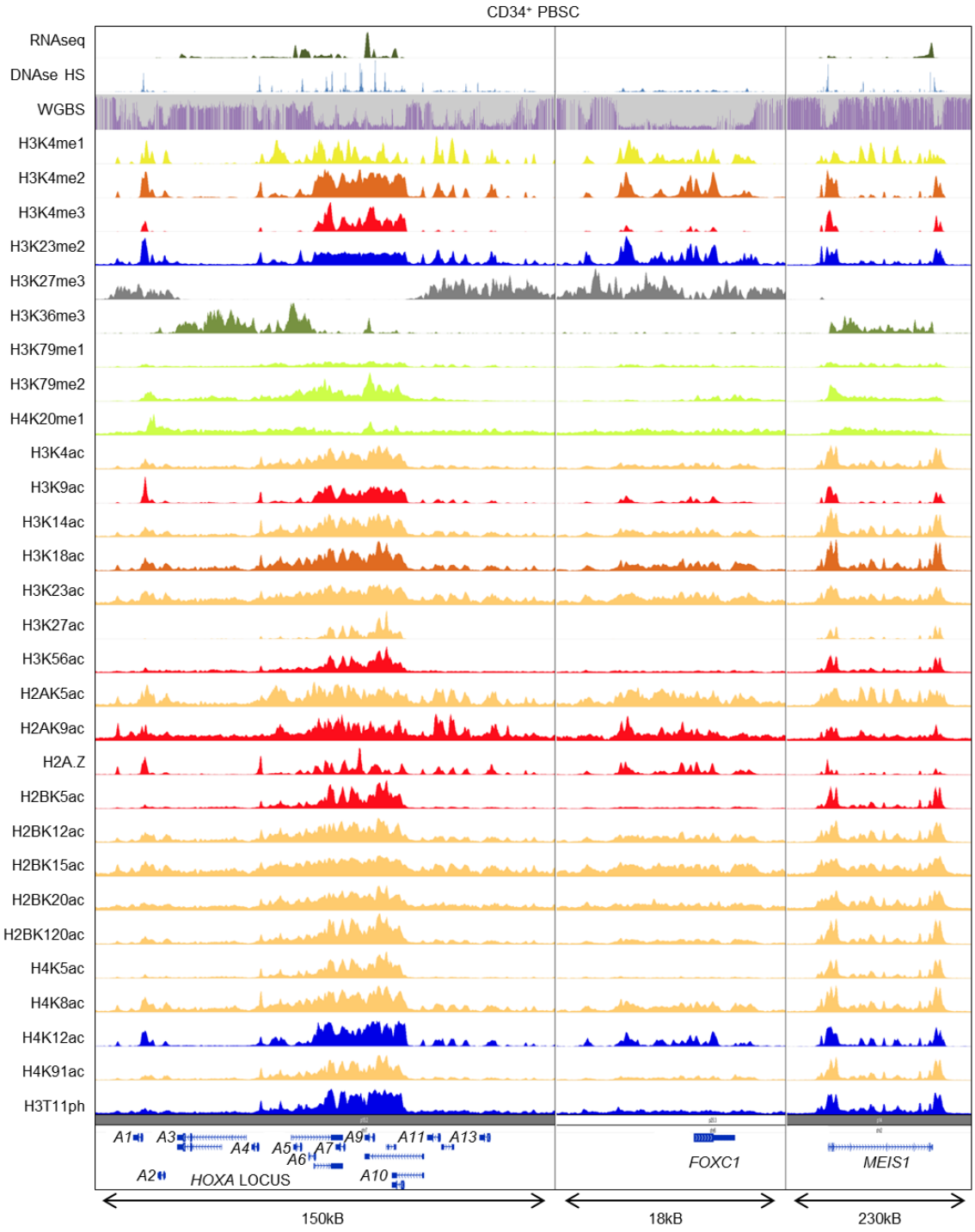


Figure 19. Organisation of chromatin at the *FOXC1* locus in normal CD34$^+$ stem and progenitor cells. High throughput sequencing tracks are shown for the indicated modifications across the *HOXA* locus and surrounding *FOXC1* in normal human CD34$^+$ stem and progenitor cells. The chromatin organisation surrounding the expressed gene *MEIS1* is shown for comparison. All data are from the ENCODE project and available through the Washington University Epigenome Browser (Zhou et al., 2011). Assistance with analysis of ENCODE data was received from Dr. Tim Somervaille.

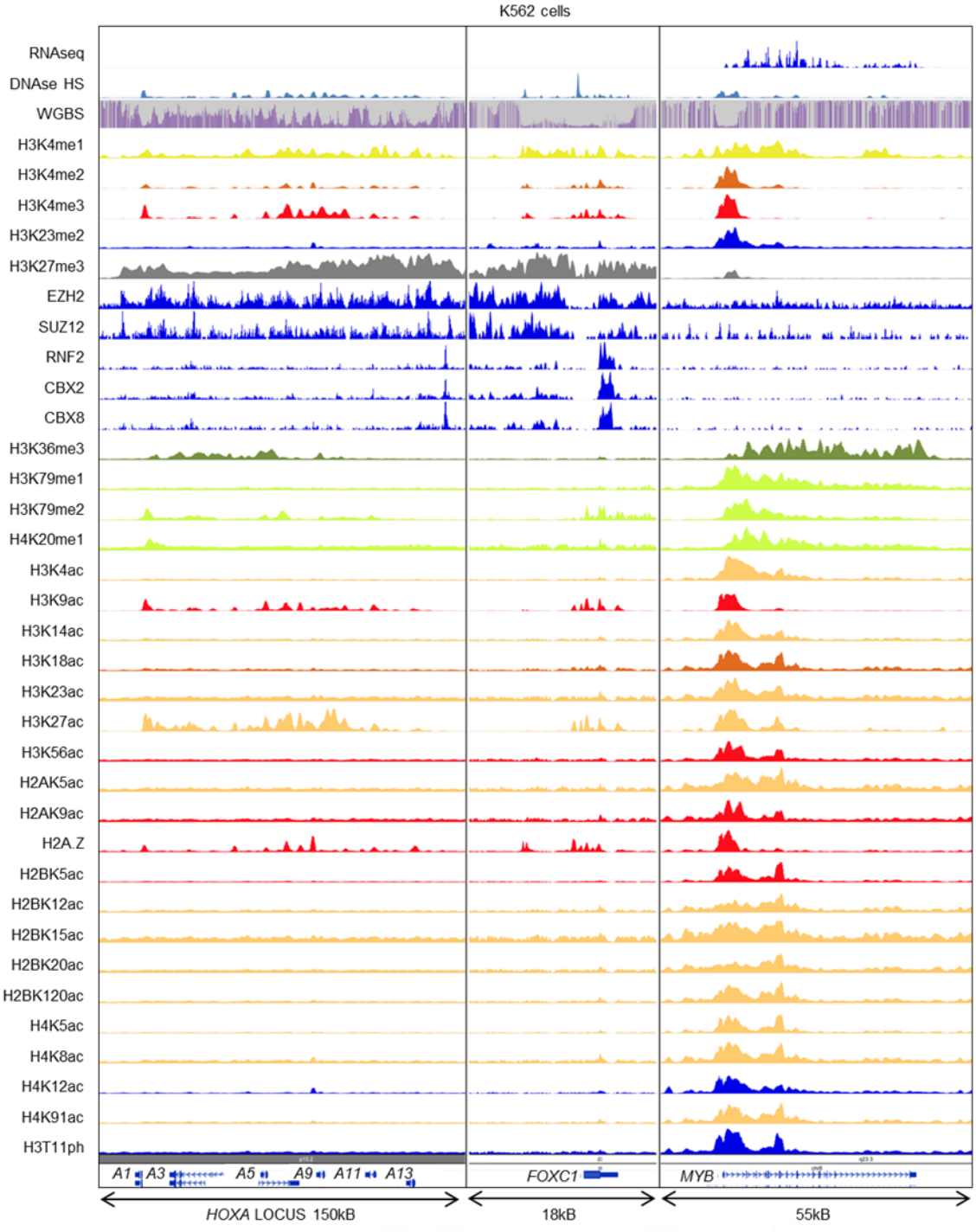


Figure 20. Organisation of chromatin at the *FOXC1* locus in K562 cells.
High throughput sequencing tracks are shown for the indicated modifications across the *HOXA* locus and surrounding *FOXC1* in K562 cells. The chromatin organisation surrounding the expressed gene *MYB* is shown for comparison. All data are from the ENCODE project and available through the Washington University Epigenome Browser (Zhou et al., 2011). Assistance with analysis of ENCODE data was received from Dr. Tim Somervaille.

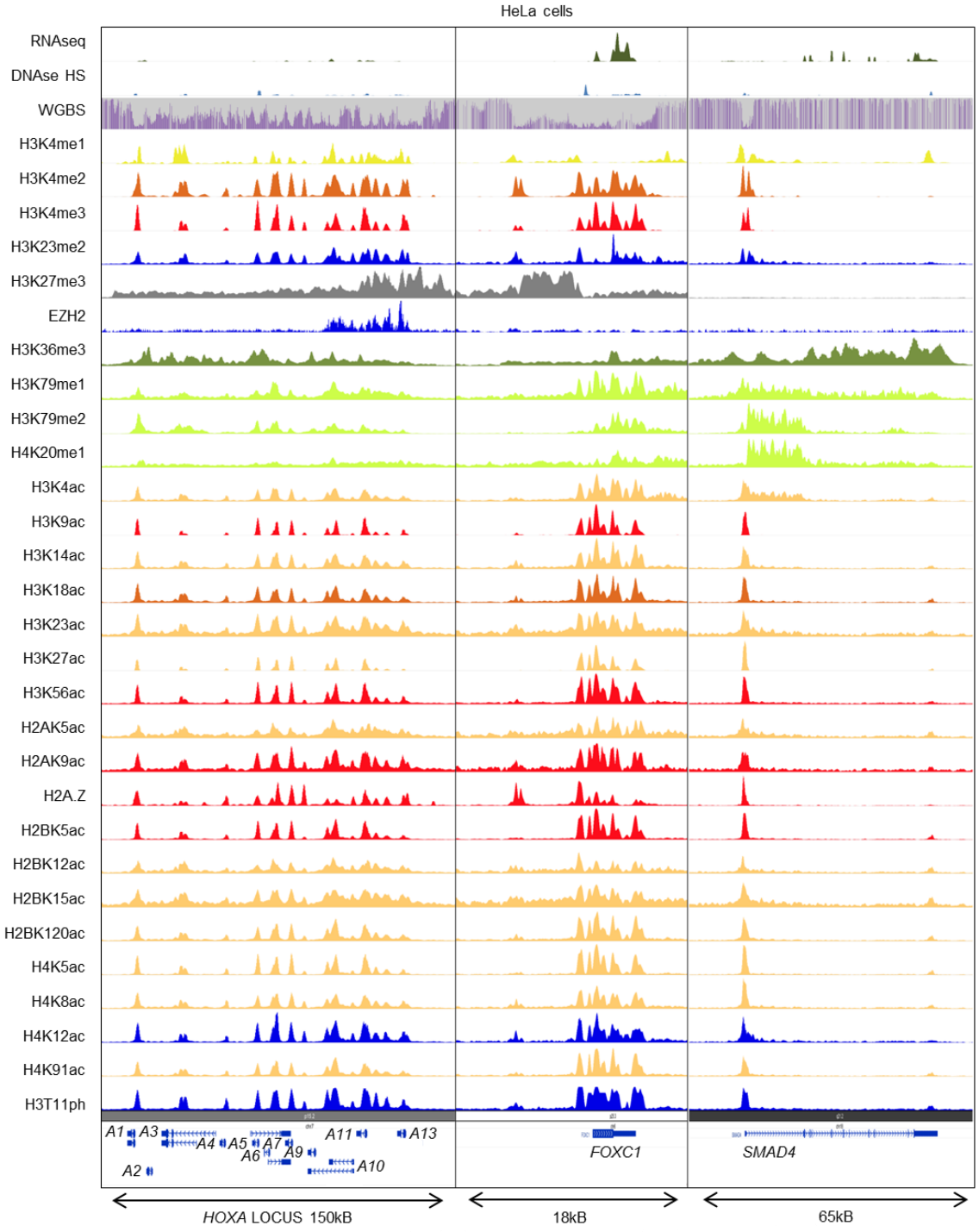
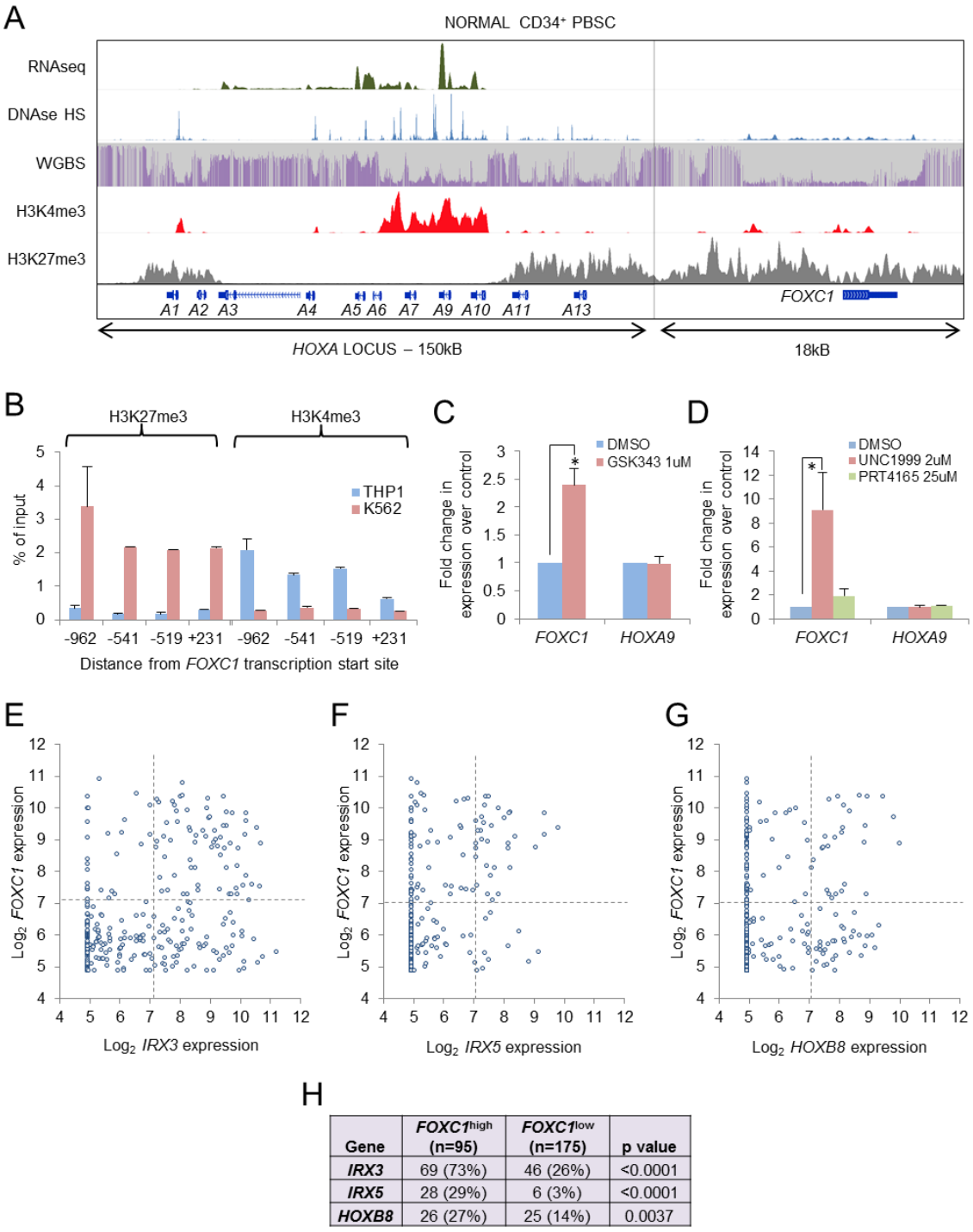


Figure 21. Organisation of chromatin at the *FOXC1* locus in HeLa cells.
High throughput sequencing tracks are shown for the indicated modifications across the *HOXA* locus and surrounding *FOXC1* in HeLa cells. The chromatin organisation surrounding the expressed gene *SMAD4* is shown for comparison. All data are from the ENCODE project and available through the Washington University Epigenome Browser (Zhou et al., 2011). Assistance with analysis of ENCODE data was received from Dr. Tim Somervaille.



H

Gene	*FOXC1*^{high} (n=95)	*FOXC1*^{low} (n=175)	p value
IRX3	69 (73%)	46 (26%)	<0.0001
IRX5	28 (29%)	6 (3%)	<0.0001
HOXB8	26 (27%)	25 (14%)	0.0037

Figure 22. Derepression of *FOXC1* in normal CD34$^+$ cells is induced by PRC inhibition.
(A) Image shows high throughput sequencing tracks from the ENCODE consortium for the *HOXA* locus and *FOXC1*. (B) Bar chart shows ChIP analysis at regions respective to the *FOXC1* transcriptional start site using antibodies against the indicated histone modifications in the indicated cell lines. (C&D) CD34$^+$ cells from separate normal donors (n=3 and 4 respectively) were treated for four or five days respectively with PRC inhibitors in serum-free liquid culture. Graphs show expression of *FOXC1* and *HOXA9* following treatment with the indicated inhibitors. (E-G) Scatter plot shows expression of *FOXC1* versus *IRX3* (E), *IRX5* (F) and *HOXB8* (G) in 320 cases of *HOXA9*-expressing AML. (H) Table shows numbers of *HOXA9*-expressing *FOXC1*^{high} and *FOXC1*^{low} AML cases from Wouters et al. (2009) which also expressed *IRX3*, *IRX5* or *HOXB8* (at log_2 expression value >7.1). Statistical significance for the indicated comparisons was assessed using Fisher's Exact Test. Assistance with analysis of ENCODE data was received from Dr. Tim Somervaille.

To ascertain whether derepression of silenced, Polycomb-marked transcription factor genes is widespread in AML, or a locus-specific phenomenon, the set of transcription factor genes exhibiting a similar epigenetic and transcriptional pattern to *FOXC1* in normal $CD34^+$ cells (i.e. minimally or not expressed and with high H3K27me3) were identified using ENCODE data (n=253 genes; Appendix Table 3) (Zhou et al., 2011). Of the 230 genes represented by probesets on the U133 Plus 2.0 array, only three (*IRX3*, *IRX5* and *HOXB8*) were both expressed significantly in AML (i.e. in the top 20% of probesets in at least 5% of 461 cases) and unexpressed in mature blood cell lineages (as determined by RNA sequencing of peripheral blood mononuclear cells; Zhou et al., 2011). The Iroquois homeobox factors IRX3 and IRX5 have roles in skeletal, cardiac and neural development. While there was a strong, positive association of increased expression of each of these factors with high *FOXC1* expression, and in particular *IRX3*, the correlation was not absolute (Figures 22E-H). Thus, tissue inappropriate expression of *FOXC1* in human AML is a locus-specific phenomenon rather than part of a generalised failure of Polycomb-mediated silencing of repressed genes.

Finally, it was considered whether *FOXC1* derepression might be associated with mutations in Polycomb complex components, or intergenic mutations close to *FOXC1* which introduce an enhancer element, as has recently been reported for *TAL1* in T-acute lymphoblastic leukaemia (Mansour et al., 2014). There was no association of *FOXC1* expression in AML with mutations in PRC1 or PRC2 components (Table 34) and comparison of whole genome sequencing data (Cancer Genome Atlas Research Network, 2013) from six patients with high *FOXC1* expression versus six with absent *FOXC1* expression revealed no genomic loci ±1MB from the transcription start site of *FOXC1* which were consistently altered either by indels or single nucleotide variations in either group (Appendix Table 4). This indicates that derepression of *FOXC1* in AML is not driven by genetic mutation of Polycomb components or *FOXC1* local regulatory regions.

Polycomb complex	Mutated gene	*FOXC1*^{low} (n=101)	*FOXC1*^{high} (n=36)	p value
PRC2	*SUZ12*	2	1	NS
	EZH1	0	0	NS
	EZH2	6	0	NS
	EED	0	2	NS
PRC1	*PCGF1*	0	0	NS
	PCGF2	1	0	NS
	PCGF3	0	0	NS
	BMI1	0	0	NS
	PCGF5	0	0	NS
	PCGF6	0	0	NS
	PHC1	0	0	NS
	PHC2	0	0	NS
	PHC3	0	0	NS
	CBX2	0	0	NS
	CBX4	0	0	NS
	CBX6	0	0	NS
	CBX7	1	0	NS
	CBX8	0	0	NS
	RING1	0	0	NS
	RNF2	1	0	NS
	TOTAL	**11**	**3**	**NS**

Table 34. Absence of association of *FOXC1* expression with Polycomb component mutations, amplifications or deletions in AML. Expression data from 163 AML cases with exome or whole genome sequencing were analysed using cBioPortal (Cerami et al., 2012; Cancer Genome Atlas Research Network, 2013). Cases were divided into *FOXC1*^{low} and *FOXC1*^{high} groups based on $\log_2$ expression values of ≤6.25 or >8 respectively. Statistical significance for the indicated associations was assessed using Fisher's Exact Test.

3.11. Frequent deprepression of the mesenchymal transcription factor gene *FOXC1* in acute myeloid leukaemia: Discussion.

Derepression of *FOXC1* is a frequent phenomenon in human AML, with approximately 20% of cases exhibiting levels of expression among the top 20% of probesets in the largest published AML array series (Wouters et al., 2009). Observed frequencies were higher still in studies focused on the AML stem and progenitor compartment (Saito et al., 2010; Kikushige et al., 2010; Goardon et al., 2011). More specifically, in bulk AML samples, high *FOXC1* expression is seen in ~40% of patients with an *NPM1* mutation and ~50% of those with dual *NPM1* and *FLT3-ITD* mutations (Wouters et al., 2009). The overall frequency of high *FOXC1* expression in human AML exceeds significantly the frequency of mutations in, for example, *IDH1* and *IDH2*, and translocations affecting *MLL*.

Differentiation block is a characteristic feature of AML and the consequence of derepressed *FOXC1* expression is to enhance monocyte/macrophage and B-cell lineage blocks, as evidenced by the murine and human functional studies presented, and bioinformatics analyses. The observation that human AMLs expressing high levels of *FOXC1* are less likely to exhibit morphologic classifications associated with monocyte lineage differentiation (i.e. FAB M4 and M5) and more likely to exhibit morphological classifications associated with granulocyte differentiation (i.e. FAB M2) is particularly significant and strongly suggests that derepressed *FOXC1* influences morphologic differentiation in human AML.

The almost exclusive association of high *FOXC1* expression with high HOXA/B locus expression in human AML suggests that FOXC1 collaborates with HOX to enhance leukaemogenesis. This suggestion is confirmed by murine *in vitro* and *in vivo* analyses, which show that co-expression of *FOXC1* with *Hoxa9* enhances clonogenic potential and proliferation, significantly shortens leukaemic latency and generates leukaemias with pathologic features which are distinct from those seen in *Hoxa9/MTV* and *Hoxa9/Meis1* AMLs. Critically, patients exhibiting high *FOXC1* expression levels also exhibit inferior survival, emphasising the prognostic significance of derepression at this locus. The association of *FOXC1* expression with HOX locus expression implies either that FOXC1 only exerts its phenotypic effects in the presence of HOX, or that HOX transcription factors are required for expression of *FOXC1*, or both. In favour of the former is the observation that forced expression of *FOXC1* in murine $CD117^{+}$ HSPCs fails to up regulate *Hoxa9* expression and induces only a transient myeloid differentiation block with enhanced clonogenic potential. In this setting, presumably as HOX gene expression is down regulated during the course of normal differentiation, FOXC1 loses its collaborating partner(s) and the phenotype is extinguished. By contrast in AML, where HOX gene expression is sustained through multiple mechanisms, FOXC1 exerts its phenotypic effects continuously. Of note, the cyclin-dependent kinase inhibitor gene *CDKN1A* (coding for p21) is one of the Group B genes repressed by FOXC1 in murine leukaemias, suggesting a potential mechanism for the enhanced clonogenic potential of *Hoxa9*/*FOXC1* double transduced cells and accelerated onset of AML. It is interesting that *FOXC1* expression in AML

collaborates with the consequence of a number of distinct genetic mutations (i.e. high level HOX gene expression) rather than any specific mutation subset in exclusivity.

The observations that FOXC1 blocks monocytic lineage (and *in vivo* B-lineage) differentiation, and that it collaborates with HOXA9 to enhance clonogenic potential, emphasise that derepression of this mesenchymal transcription regulator in the haematopoietic system has haematologic rather than mesenchymal consequences. This is despite its lack of expression in normal haematopoiesis. There was, for example, no whole scale up regulation of a mesenchymal gene programme in the *Hoxa9/FOXC1* murine AMLs, although *VCAM1* was one notable exception (Appendix Table 1). Expression of this vascular adhesion molecule may explain the relative failure of murine *Hoxa9/FOXC1* AML cells to mobilise to the blood by comparison with *Hoxa9/MTV* and *Hoxa9/Meis1* AML cells. Indeed some genes such as *FN1* and *VIM* which are induced by ectopic *FOXC1* expression in breast cancer cells (Bloushtain-Qimron et al., 2008) are repressed by FOXC1 in murine leukaemia cells. The findings in the current study differ significantly from observations in solid tumours such as those of liver, pancreas and breast where FOXC1 has been causally implicated in promotion of the epithelial-to-mesenchymal transition characteristic of metastasis (Bloushtain-Qimron et al., 2008; Xia et al., 2013; Yu et al., 2013). In these studies, induced *FOXC1* expression has the expected activity of a mesenchymal regulator: promotion of mesenchymal cellular phenotypes such as enhanced migration, invasion and metastasis. In AML, by contrast, tissue inappropriate expression of *FOXC1* interferes with normal tissue activity by blocking differentiation and enhancing proliferation rather than conferring on it a mesenchymal phenotype.

This raises the question as to how FOXC1 blocks monocytic differentiation in AML. One possibility is that it acts as a dominant negative inhibitor of one or more forkhead transcription factors whose normal function is to promote monocyte differentiation, although none have been well described to date. While diverse forkhead box transcription factors exhibit highly conserved central DNA binding forkhead domains, they have very different flanking and transactivation sequences. ENCODE RNA sequencing data demonstrate that a number of forkhead factors are expressed in normal human $CD14^{+}$ monocytes including *FOXJ3*, *FOXP1*, *FOXN3*, *FOXO1* and *FOXO3*. *FOXP1* has been shown to have an essential role in normal monocyte differentiation and macrophage function *in vivo* (Shi et al., 2008), raising the possibility that FOXC1 may interfere with the physiological function of FOXP1 in monocytes (discussed further in Chapter 5). An alternative possibility is that FOXC1 directly represses through localised binding expression of one or more key transcriptional regulators of monocytic differentiation. For example, expression of *Egr1*, *Klf4* and *Mef2c* is significantly supressed in *Hoxa9*/*FOXC1* AML cells, and all these genes code for critical positive regulators of monocyte differentiation (Laslo et al., 2006; Feinberg et al., 2007; Schuler et al., 2008). Functional analyses demonstrated that FOXC1 represses expression of *KLF4* and show that restoration of *KLF4* expression in *Hoxa9*/*FOXC1* AML cells inhibits proliferation and clonogenic potential. Whether FOXC1-mediated gene repression occurs through binding at multiple enhancer

or promoter sites across the genome or more specifically through interactions with a master regulator such as PU.1 remains unclear.

The basis for the collaboration of FOXC1 with HOXA9 likewise remains unclear. The somewhat similar immunophenotypes and array expression profiles of murine *Hoxa9/MTV* AMLs versus *Hoxa9/Meis1* AMLs are in keeping with the known function of MEIS1 in stabilising the interaction of HOXA9 with DNA (i.e. enhancing its potency) (Shen et al., 1997). The fact that *Hoxa9*/*FOXC1* AMLs appear so different in terms of their immunophenotypes, transcriptomes and histologies implies that FOXC1 may confer on HOXA9 an alternative activity, or perhaps misdirect its binding to distinct enhancer sets. Indeed, FOXC1 and HOXA9 can physically interact, at least when they are exogenously expressed in HEK 293FT cells. The known activity of forkhead proteins as pioneer transcription factors may be relevant in this regard (Lam et al., 2013).

FOXC1 occupies a DNA hypomethylated, DNase hypersensitive genomic locus in $CD34^+$ cells which is marked by both H3K4 and H3K27 trimethylation. This suggests the chromatin surrounding this gene is relatively decompacted, and such epigenetic configurations have previously been termed bivalent. Bivalent genes are transcriptionally repressed by Polycomb and in keeping with this, the expression of *FOXC1* in normal human $CD34^+$ cells could be induced by treatment of cells with inhibitors of PRC. The most effective was the dual EZH1/EZH2 inhibitor UNC1999, suggesting EZH1 may compensate for loss of activity of EZH2 when a pure EZH2 inhibitor is used. Nevertheless, the mechanisms underlying loss of Polycomb repression at a single locus in AML require further investigation. Of over two hundred genes with similar chromatin and transcriptional configurations to *FOXC1* in $CD34^+$ cells only *HOXB8*, *IRX3*, and *IRX5* were similarly derepressed in a significant proportion of AML cases, indicating that while there is locus-specific loss of Polycomb activity in a substantial proportion of AML cases there is no generalised failure of suppression of Polycomb marked genes.

Finally, these findings may have therapeutic consequences. For example, in basal-like breast cancer, high *FOXC1* expression renders cells more susceptible to pharmacological inhibition of NFκB (Wang et al., 2012). Further studies may uncover specific therapeutic targets or approaches in this frequent sub-group of human AML.

Chapter 4. The role of *IRX3* in acute myeloid leukaemia

4.1. Introduction

The results outlined in the previous chapter describe the frequent occurrence and functional relevance of *FOXC1* derepression in AML. Moreover, they highlight that the derepression of silenced, Polycomb-marked transcription factor genes is not widespread in AML, but is rather a locus-specific phenomenon. Of 230 transcription factor genes that were analysed, only three (*HOXB8*, *IRX3* and *IRX5*) were found to exhibit derepression akin to *FOXC1* in AML, i.e. they are expressed significantly in AML but not in normal blood cells. *HOXB8* was in fact the first HOX gene shown to be aberrantly expressed in leukaemia cells (Blatt et al., 1988). *Hoxb8* overexpression induces a myeloid differentiation block that is growth factor-dependent and co-operates with IL3 to rapidly cause overt AML (Blatt et al., 1988; Perkins et al., 1990). In contrast, there are no reports implicating *IRX3* or *IRX5* in normal or leukaemic haematopoiesis.

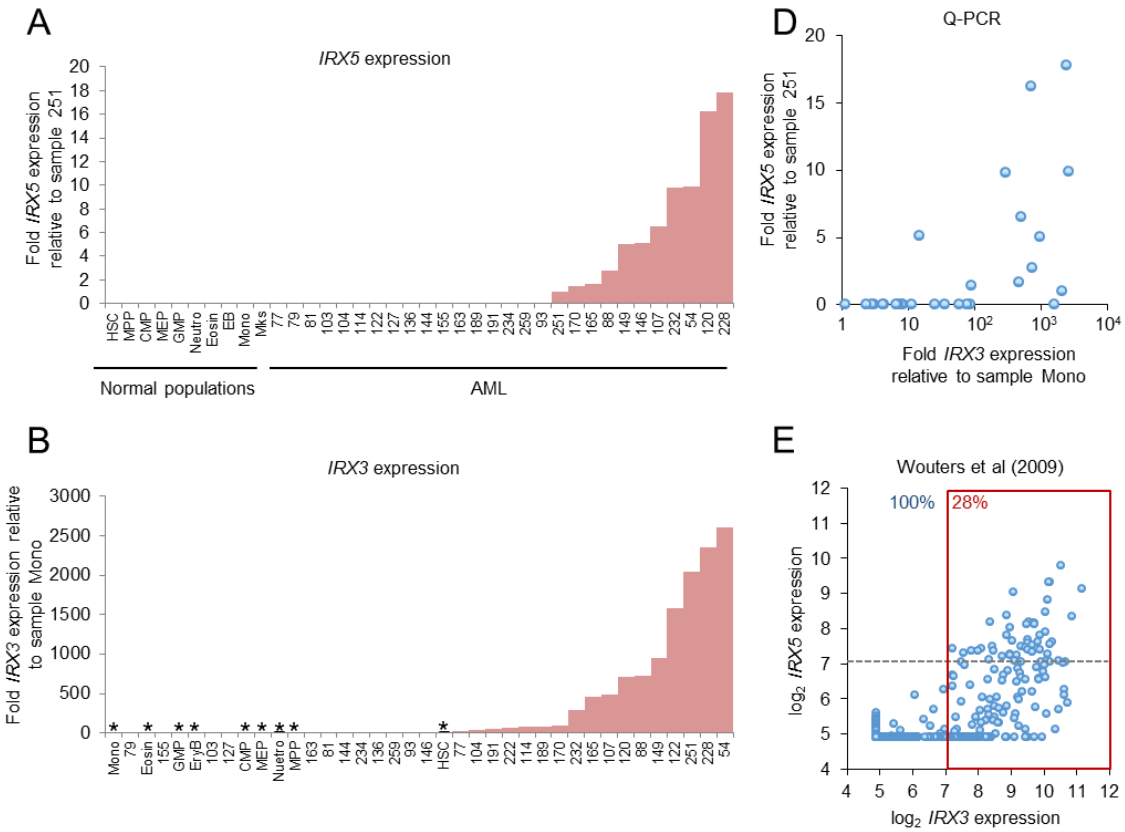
IRX3 and *IRX5* are members of the Iroquois homeobox (*IRX*) gene family of transcription factors and are located in an IRX cluster on human chromosome 16, along with their fellow family member *IRX6*. The IRX genes belong to the three-amino acid loop extension (TALE) superclass of homeobox genes, which also includes the MEIS and PBX gene families (Mukherjee and Bürglin, 2007). IRX proteins are involved in embryonic patterning and specify the identity of a number of body territories during development via the spatial and temporal regulation of their target genes (Cavodeassi et al., 2001). For example, in the chick embryo, sonic hedgehog signalling (Shh) regulates the expression of *Irx3*, as well as a number of other homeodomain transcription factors, to functionally subdivide the neural tube (Briscoe et al., 2000). *Irx3* and *Irx5* exhibit overlapping expression patterns during embryonic development in multiple tissues, including the limb buds and the heart (Cohen et al., 2000; Christoffels et al., 2000; Houweling et al., 2001). Functional redundancy exists between these genes as, although *Irx3* or *Irx5* null mice are viable and fertile, mice lacking both *Irx3* and *Irx5* die in utero due to severe heart defects (Gaborit et al., 2012). A similar phenotype is observed in the Fused toes mouse in which *Irx6* is additionally deleted, and these mice also exhibit severe deformations to the craniofacial and forebrain structure as well as developmental limb abnormalities (Peters et al., 2002). In addition to lethal heart defects, *Irx3*/*Irx5* double knockout mice fail to form a proper hind limb, exhibiting loss of the tibia and the first digit as well as severe femoral hypoplasia (Li et al., 2014). Although mice null for *Irx3* or *Irx5* alone do not display any heart defects in the embryonic stage, heart defects do arise in adulthood (Zhang et al., 2011; Costantini et al., 2005). *Irx3* null mice additionally display a 25-30% reduction in body weight due to loss of fat mass and increased basal metabolic rate with browning of the white adipose tissue (Smemo et al., 2014). These results are consistent with observations in humans linking aberrant *IRX3* expression with increased risk of obesity and type 2 diabetes (Ragvin et al., 2010; Smemo et al., 2014). *IRX3* is expressed in the adult heart, pancreas, lung, spinal cord and the brain, particularly within the hypothalamus and cerebellum (Smemo et al., 2014; Ragvin et al., 2010). Humans with inherited homozygous mutations in *IRX5* exhibit the Hamamy syndrome which is characterised by craniofacial and skeletal abnormalities as well as congenital heart defects (Hamamy et al., 2007; Bonnard et al., 2012).

Studies implicating *IRX3* or *IRX5* in cancer, let alone leukaemia, are limited. *Irx3* has been shown to be hypermethylated and inactivated in a murine model of prostate cancer (Morey et al., 2006). In contrast, similar hypermethylation of *IRX3* was in fact found to correlate with its overexpression in brain tumours (Ordway et al., 2006). Knockdown studies in prostate, breast and colon cancer cell lines suggest that *IRX5* may be required for their proliferation and survival (Myrthue et al., 2008).

Given the absence of information regarding any role for *IRX3* or *IRX5* in normal or leukaemic haematopoiesis, the expression of these transcription factors in AML was initially characterised and subsequently the functional role of *IRX3* in AML was investigated.

4.2. *IRX3* and *IRX5* are expressed in human AML

As discussed, *IRX3* and *IRX5* are significantly expressed in AML and this expression is positively correlated with *FOXC1* expression (Figures 22E, 22F and 22H) (Wouters et al., 2009). Analysis of ENCODE RNA sequencing data from peripheral blood mononuclear cells indicates that *IRX3* and *IRX5* are not expressed in mature blood lineages (Zhou et al., 2011). In order to more thoroughly interrogate the expression of *IRX3* and *IRX5* in normal human haematopoietic cells, qPCR was performed on the flow sorted populations of BM HSPCs and terminally differentiated cells described above (Figures 23A-D) (Huang et al., 2014). *IRX5* expression was absent in all normal BM populations tested (Figure 23A). Similarly, *IRX3* expression was detected at very low levels in most normal BM populations (Figure 23B). Surprisingly, very high level *IRX3* expression was detected in 2/3 megakaryocyte populations tested (Figure 23C). This level of expression detected by qPCR was unexpected as previous data generated through exon array analysis indicated *IRX3* was expressed at low levels in megakaryocytes; it may therefore be possible that these data represent artefact. In the bulk AML blast samples, *IRX3* transcripts were detected at high level (greater than 250-fold increase over expression levels in the lowest expressing sample) in 10/29 (34%) and at intermediate level in 5/29 (17%) samples (50-250-fold increase over expression levels in the lowest expressing sample) (Table 35). *IRX5* transcripts were detected in 11/28 (39%) bulk AML blast samples tested and were undetectable in the remaining 17 samples (Figure 23A). The *IRX5* expression levels that were detected however were comparable to that of the high level *IRX3* expressers, as determined by comparison of their ΔCt values. There was a strong positive correlation between *IRX3* and *IRX5* expression in AML: of the 11 *IRX5* expressing samples, 9 expressed *IRX3* at high levels and 1 expressed *IRX3* at intermediate level (Figure 23D). The strong association of *IRX5* expression with that of *IRX3* in AML was confirmed by the expression dataset from the study by Wouters et al (2009) (Figure 23E). In this study, all of the 44 $IRX5^{high}$ AMLs also co-expressed high levels of *IRX3*, whereas only 44/159 (28%) of the $IRX3^{high}$ AMLs concurrently expressed high levels of *IRX5* (Figure 23E). These data demonstrate that, although both *IRX3* and *IRX5* are highly expressed in AML, *IRX3* is more broadly expressed than *IRX5*, with the latter rarely, if ever, expressed in the absence of *IRX3*. Additionally, with the possible exception of *IRX3* in the megakaryocyte lineage, both *IRX3* and *IRX5* are only minimally expressed during normal haematopoiesis.



C

***IRX3* expression in megakaryocyte populations**

Normal sample	ΔCt value	Fold increase in expression vs. mono
1	4.6	2170
2	16.2	28
3	2.9	11907
Mean	**7.9**	**4701.6**
SEM	**4.2**	**3655.6**

Figure 23. Expression of *IRX3* and *IRX5* in human AML.
(A&B) Bar charts show the relative expression of (A) *IRX5* and (B) *IRX3* in bulk primary human AML samples (n=28 and 29 respectively) and prospectively sorted normal human cell populations (n=3/4 separate individuals per cell type; Huang et al., 2014). See also Table 35. In (B) normal cell populations are marked with an asterisk. HSC, hematopoietic stem cell ($CD34^{+}38^{-}90^{+}45RA^{-}Lin^{-}$); MPP, multipotent progenitor; CMP, common myeloid progenitor; GMP, granulocyte-macrophage progenitor; MEP, megakaryocyte-erythrocyte progenitor; EryB, erythroblast; Eosin, eosinophils; Mono, monocytes; Neut, neutrophils. AML sample numbers refer to the Biobank identifier. (C) Table shows the *IRX3* ΔCt values for three samples of megakaryocyte populations tested, each from separate donors. The fold increase in expression of *IRX3* versus the monocyte population is also displayed. (D&E) Scatter plots show the expression of *IRX5* versus *IRX3* in primary AML patient samples as determined by (D) quantitative PCR (n=28; see also Table 35) or (E) array expression values from the indicated study. For (E) percentages in blue text indicate proportion of $IRX5^{high}$ samples exhibiting high *IRX3* expression. Percentages in red text indicate the proportion of $IRX3^{high}$ samples (in the red box) additionally exhibiting high *IRX5* expression (above the dotted grey line).

ID	Sample	Disease Status	Gender	Age	Cytogenetics
54	Blood	2nd Relapse	Male	75	47,XY,+11[1]/48,sl,+8[7]/49,sdl,+4[2]
77	BM	Diagnosis	Female	67	Normal
79	BM	Diagnosis	Female	65	47,XX,der(3)t(3;9;14;4)(q2?1;q34;q24;q21),der(4)t(3;9;14;4),del(7)(q22q32),add(8)(q24),der(9)del(9)(p2?3)t(3;9;14;4),der(14)t(3;9;14;4),+18,add(19)(q13.3)x2[9]/46,XX[1]
81	Blood	1st Relapse	Male	64	Normal
88	BM	Refractory	Female	64	46,XX,del(12)(p1?2p1?3)[18]/46,XX[2]
93	BM	Diagnosis	Male	59	Normal
103	BM	Diagnosis	Male	51	Normal
104	BM	1st Relapse	Female	31	46,XX,t(6;9;11)(p2?1;p22;q23)[6]/45,idem,der(15)t(15;17)(p11.2;q11.2),-17[4]
107	BM	1st Relapse	Female	16	45,XX,del(7)(q11.2q3?2),t(8;21)(q22;q22),der(12)t(12;18)(p11.2;q11.2),-18[10]/46,XX[1]
114	BM	Diagnosis	Male	56	Normal
120	BM	Diagnosis	Female	68	Normal
122	Blood	Diagnosis	Male	66	Normal
127	BM	Diagnosis	Female	63	46,XX,inv(16)(p13q22)[8]/46,XX[2]
136	BM	Diagnosis	Female	61	Normal
144	Blood	Diagnosis	Male	71	Normal
146	BM	Diagnosis	Male	52	46,Y,?t(X;6)(p22.1;q?25),t(15;17)(q22;q11.2)[9]/46,XY[1]
149	BM	Diagnosis	Female	49	46,XX,t(15;17)(q22;q11.2)[7]/46,sl,-6,add(16)(q12),+mar[3]/46,XX[3]
155	Blood	Diagnosis	Male	63	Normal
163	BM	Diagnosis	Male	20	45,X,-Y,t(8;21)(q22;q22)[8]/46,XY[2]
165	BM	Diagnosis	Female	48	46,XX,t(8;22)(p11;q13),del(9)(q13q32)[10]
170	BM	1st Relapse	Male	50	46,XY,add(6)(p22)[8]/46,XY[2]
189	BM	Diagnosis	Female	76	Normal
191	BM	Diagnosis	Female	38	Normal
222	BM	Diagnosis	Female	80	Normal
228	BM	Diagnosis	Male	54	46,XY,t(1;2)(q2?4;p1?3)[9]/46,XY[1]
232	Blood	Diagnosis	Female	85	Normal
234	Blood	Diagnosis	Male	84	47,XY,+13[3]/93~94,idemx2[cp3]/92~94<4n>,XXYY,+X,der(5)t(1;5)(q21;q31),+13[cp5]/46,XY[4]
251	BM	Diagnosis	Male	16	46,XY,t(6;9)(p22;q34)[9]/46,XY,der(6)t(6;9),der(9)t(6;9)del(9)(q21q34)[2]
259	BM	Diagnosis	Male	75	47,XY,+8[10]

Table 35. Karyotype of 29 Manchester Cancer Research Centre Biobank AML samples analysed for *IRX3* expression. Samples in red exhibit high *IRX3* expression. BM = bone marrow blasts.; sl = stemline; sdl = sideline; der = derivative; t = translocation; del = deletion; add = additional material of unknown origin; idem = denotes the stemline karyotype in a subclone; inv = inversion; mar = marker chromosome; cp = composite karyotype

4.3. *IRX3* expression in human AML is associated with mutations in *NPM1* and internal tandem duplications of *FLT3*

Given that *IRX3* is more significantly expressed in AML compared to *IRX5*, its expression was examined in more detail using the array data from recent studies. Firstly, *IRX3* expression was analysed in AML LSCs versus normal HSPCs using the datasets published by Saito et al. (2010), Kikushige et al. (2010) and Goardon et al, (2011). High level *IRX3* expression (where values for *IRX3* were among the top 25% of protein-coding gene probesets) was observed in 33%, 58% and 19% of samples analysed in each study respectively (Figures 24A-C). The qPCR analysis of bulk AML samples was confirmed using expression data from two recently published studies (Wouters et al., 2009; Cancer Genome Atlas Research Network, 2013). *IRX3* was expressed at high level (i.e. with a probeset expression value among the top 25% of protein coding gene probeset expression values) in 159/461 (34%) of presentation samples from the Wouters dataset (Figure 24D). Similarly 49/163 (30%) of samples from the TCGA dataset also exhibited high expression (Figure 24E) (Cancer Genome Atlas Research Network, 2013).

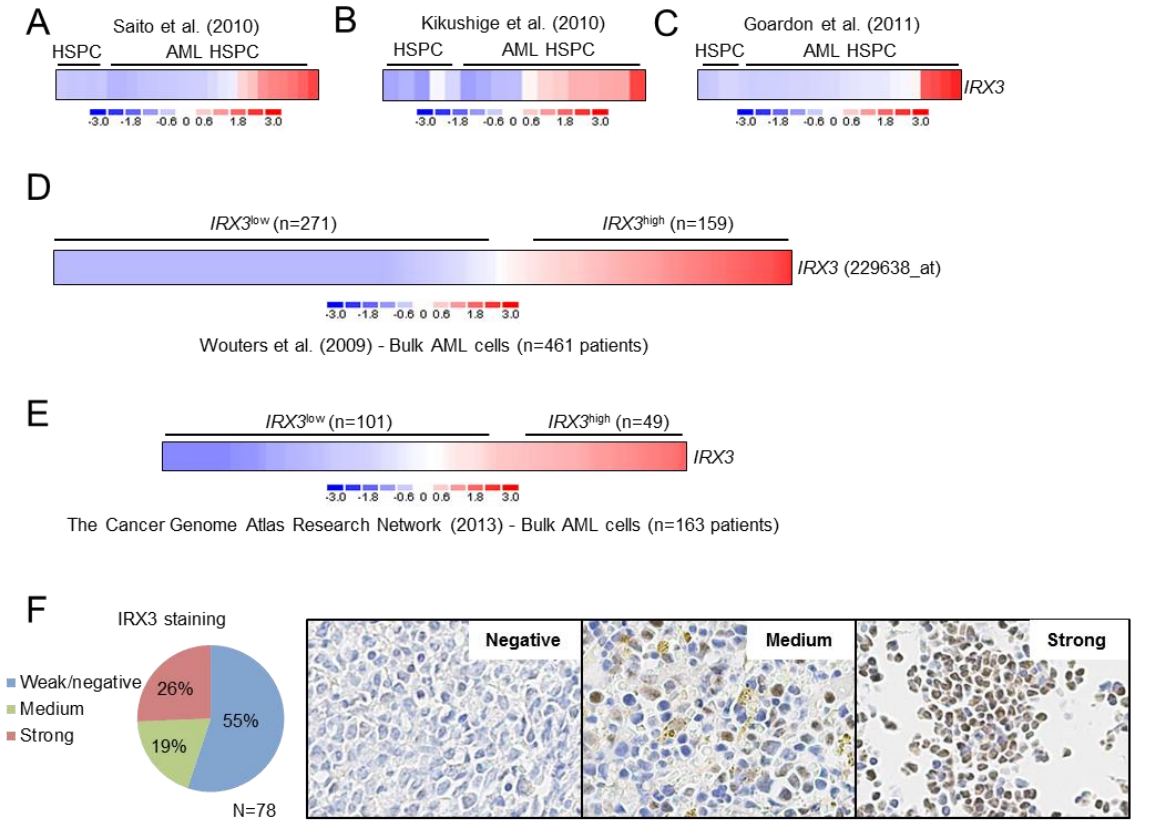


Figure 24. *IRX3* expression in human AML.
(A-C) Heat maps show the expression of *IRX3* in AML HSPCs versus normal HSPCs in the indicated studies. Colour scale indicates the standardised expression level. The definition of HSPCs in each study was (A) $CD34^{+}CD38^{-}$ for both normal (n=5) and AML cells (n=21); (B) $CD34^{+}CD38^{-}Lin^{-}$ for normal cells (n=5) and $CD34^{+}CD38^{-}$ for AML cells (n=12), and (C) $CD34^{+}CD38^{-}CD90^{+}CD45RA\text{-}Lin^{-}$ for normal cells (n=5) and GMP-like (n=21) AML samples. (D-E) Heat maps show (D) expression of *IRX3* in 461 bulk AML samples (data extracted from Wouters et al., 2009). Samples were divided into $IRX3^{low}$ and $IRX3^{high}$ groups based on $\log_2$ expression values of <6.1 or >7.1 respectively (probeset 229638_at) which approximate to expression values at the 40^{th} and 25^{th} percentiles of all protein coding gene probeset expression values; and (E) expression of *IRX3* in 163 bulk AML samples (data extracted from cBioPortal; Cerami et al., 2012; Cancer Genome Atlas Research Network, 2013). Samples were divided into $IRX3^{low}$ and $IRX3^{high}$ groups, also based on $\log_2$ expression values of <6.1 or >7.1 respectively. Colour scale indicates standardised expression level. (F) Immunohistochemical staining of tissue microarrays representative of 78 AML patients. Pie chart (left panel) shows the summary of the intensity of IRX3 staining in each patient sample. Immunoreactivity was semi-quantitatively scored for intensity (either weak, medium or strong) and intensity multiplied by percentage staining to generate an H-score (final range of 0-3 per core). IRX3 positivity was classed as strong, medium or weak/negative based on H-scores of ≥0.5, <0.5 but ≥0.1 and <0.1 respectively. Example images of staining intensity are shown in the right panel. Assistance with bioinformatics analysis was received from Dr. Tim Somervaille. TMAs were made by Eleni Tholouli and Richard Byers. Assistance with TMA analysis was received from Katalin Boros.

In the TCGA study, high *IRX3* expression was strongly and positively associated with the presence of an *NPM1* mutation and a *FLT3* internal tandem duplication, but negatively associated with mutations in *TP53* or *RUNX1* (Table 36). In the Wouters study, high *IRX3* expression was strongly (i.e. p<0.0001) and positively associated with the presence of an *NPM1* mutation, a *FLT3* internal tandem duplication or the acute promyelocytic leukaemia subtype (APL; FAB subtype M3) (Table 37) (Wouters et al., 2009). There were weaker positive associations with intermediate cytogenetic risk, normal karyotype, the presence of an *IDH1* mutation, and the presence of a t(6;9) or 11q23 translocation. High *IRX3* expression was negatively associated with the presence of chromosome 5 or 7(q) loss, the presence of t(8;21) or inv(16), the presence of an *NRAS* mutation or a double *CEBPA* mutation, and the presence of high *MECOM* expression (Table 37) (Wouters et al., 2009). IRX3 expression was also detected by immunohistochemical staining of tissue microarrays (TMAs) representative of 78 AML patients (Figure 24F). Strong IRX3 expression was detected in 20/78 (26%) and medium levels of expression was detected in 15/78 (19%) of AML samples. Of note, no IRX3 staining was detected in megakaryocytes that were clearly present in the BM cores on the TMAs. This suggests that the extraordinarily high levels of IRX3 expression detected by qPRC in 2/3 samples above may be a result of an unknown artefact.

Mutated gene	*IRX3*^{low} (n=101)	*IRX3*^{high} (n=49)	p value
DNMT3A	25	14	NS
TET2	9	6	NS
NPM1	16 (16%)	24 (50%)	<0.0001
IDH1	8	6	NS
IDH2	13	2	NS
RUNX1	15 (15%)	1 (2%)	0.02
TP53	12 (12%)	0 (0%)	0.009
WT1	4	5	NS
PTPN11	2	4	NS
KIT	5	0	NS
EZH2	3	0	NS
HNRNPK	1	0	NS
U2AF1	4	1	NS
SMC1A	5	0	NS
SMC3	5	1	NS
BRINP3	3	2	NS
PHF6	2	3	NS
STAG2	2	3	NS
RAD21	2	3	NS
FLT3 ITD	11 (11%)	17 (34%)	0.0008
FLT3 D835X	6	4	NS
FLT3 ITD and *NPM1* mutation	3 (3%)	12 (24%)	0.0001

Table 36. Association of *IRX3* expression with genetic features in AML.
Expression data from 163 AML cases with exome or whole genome sequencing were analysed using cBioPortal (Cerami et al., 2012; The Cancer Genome Atlas Research Network, 2013). Cases were divided into *IRX3*^{low} and *IRX3*^{high} groups based on $\log_2$ expression values of <6.1 or >7.1 respectively. Statistical significance for the indicated associations was assessed using Fisher's Exact Test.

Characteristic		*IRX3*^{low} (n=271)	*IRX3*^{high} (n=159)	p value
Gender		137M:134F	79M:80F	NS
Age		44 (range 15-61)	43 (range 16-60)	NS
Cytogenetic risk	Good	70	24	NS
	Intermediate	137 (51%)	104 (65%)	0.004
	Poor	59	26	NS
	Unknown	5	5	NS
Karyotype	Trisomy 8	10	8	NS
	11q23	1 (0%)	7 (4%)	0.005
	5 or 7(q) loss	25 (9%)	4 (3%)	0.009
	Complex	12	3	NS
	Normal	98 (36%)	78 (49%)	0.01
	Other	34	16	NS
	t(6;9)	1 (0%)	5 (3%)	0.03
	t(8;21)	33 (12%)	1 (1%)	<0.0001
	inv(16)	31	2 (1%)	<0.0001
	t(15;17)	3 (1%)	17 (11%)	<0.0001
	abn(3q)	2	0	NS
	minus 9q	4	1	NS
	t(9;22)	1	0	NS
	Unknown	16	17	NS
Gene mutations	*CEBPA* double	24 (9%)	0 (0%)	<0.0001
	IDH1	12 (4%)	19 (12%)	0.006
	IDH2	21	15	NS
	*FLT3*ITD	52 (19%)	66 (42%)	<0.0001
	*FLT3*TKD	25	21	NS
	NRAS	34 (13%)	7 (4%)	0.006
	KRAS	2	2	NS
	NPM1	44 (16%)	80 (50%)	<0.0001
	*FLT3*ITD and *NPM1*	21 (8%)	44 (28%)	0.001
High *MECOM*expression		24 (9%)	2 (1%)	0.001
FAB	M0	11	2	NS
	M1	52	41	NS
	M2	63	38	NS
	M3	3 (1%)	20 (13%)	<0.0001
	M4	63 (23%)	15 (9%)	0.0003
	M5	59	33	NS
	M6	2	3	NS
	Unknown	18	7	NS

Table 37. Association of *IRX3* expression with diagnostic and genetic features in AML Array data from 461 AML cases where extended genotyping analysis was available were downloaded and analysed (Wouters et al., 2009). Cases were divided into *IRX3*^{low} and *IRX3*^{high} groups based on $\log_2$ expression values of <6.1 or >7.1 respectively (probeset 229638_at). These values approximate to expression values at the 40th and 25th percentiles of all protein coding gene probeset expression values. Statistical significance for the indicated comparisons was assessed using Fisher's Exact Test, except for age where an unpaired t-test was used. FAB= French-American-British NS = not significant. Assistance with bioinformatics analysis was received from Dr. Tim Somervaille.

4.4. *IRX3* sustains the clonogenic potential and differentiation block in AML cells

To investigate whether *IRX3* expression in AML makes a functional contribution to transformation, knockdown (KD) experiments were performed in human THP1 AML cells, which exhibit the highest levels of *IRX3* expression amongst all cell lines tested (Figure 25A). *IRX3* KD led to loss of clonogenic potential which was due to induction of differentiation and onset of apoptosis (Figures 25B-G). Following initiation of KD there was up regulation of the myeloid differentiation markers CD11b and CD14 (Figures 25E), morphological differentiation (Figure 25F) and apoptosis (Figure 25G), although there was no substantial evidence of G1 arrest (Figure 23H). To confirm that the observed phenotype was an on-target consequence of *IRX3* KD, similar experiments were performed using a KD construct targeting the 3'UTR region of *IRX3* (KD6) in a THP1 cell line constitutively expressing *IRX3* cDNA lacking this regulatory region. IRX3 forced expression and reconstitution following KD was confirmed by qPCR and western blotting (Figures 25I and 25J). Reconstitution of IRX3 in THP1 cells restored clonogenic potential to levels comparable with control cells (Figures 25K), indicating an on-target phenotype. KD of *Irx3* in murine MLL-AF9 AML cells likewise led to loss of clonogenic potential and induction of differentiation (Figure 26A-C). KD of *IRX3* in other human *IRX3* expressing AML cell lines, representative of a variety of molecular subtypes, also led to loss of clonogenic potential in some, but not all cases (Figure 27A). The clonogenic potential of normal human $CD34^+$ cells (which express *IRX3* at very low levels) was largely unaffected following infection with the same *IRX3* KD construct (Figure 27B). However, the number of cells with burst-forming unit erythroid (BFU-E) potential was reduced (Figure 27B), for reasons that are currently unclear. These data demonstrate that *IRX3* expression contributes to the oncogenic potential of human AML cells by maintaining differentiation block and clonogenic potential.

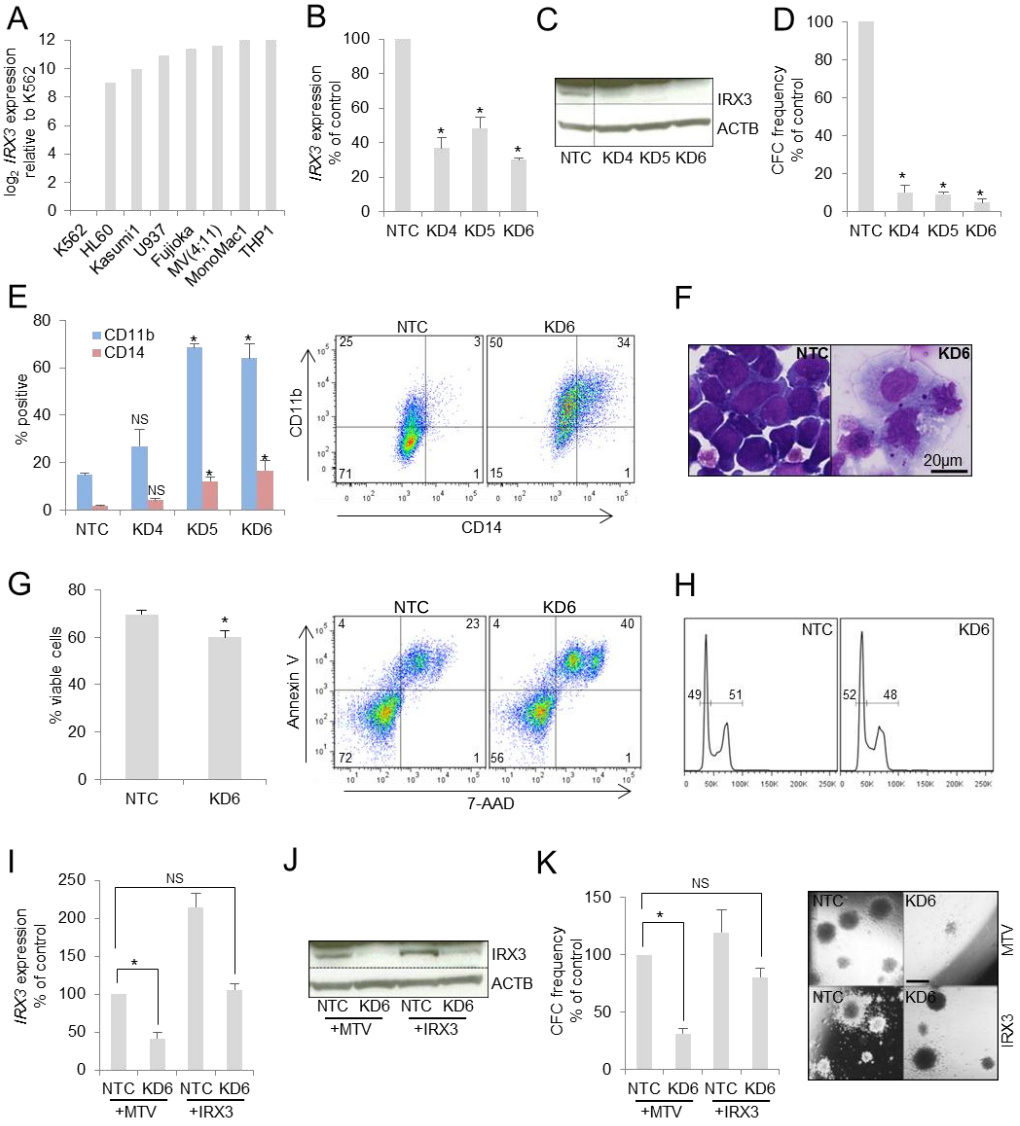


Figure 25. IRX3 sustains the differentiation block and clonogenic potential of human THP1 AML cells. (A) Bar chart shows *IRX3* expression relative to K562 cells across a panel of myeloid leukaemia cell lines as determined by qPCR. (B-K) Human THP1 AML cells were infected with lentiviral vectors targeting *IRX3* for KD (KD4, KD5 and KD6) or a non-targeting control vector (NTC). (B) Bar chart shows mean+SEM relative transcript expression in KD versus control cells (n=3). (C) Western blot shows expression of the indicated proteins in the indicated conditions. (D) Bar chart shows the mean+SEM colony-forming cell (CFC) frequencies of KD cells relative to control cells enumerated after ten days in semi-solid culture (n=4). *indicates $p<0.001$ by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of each KD sample versus NTC. (E) Bar chart (left panel) shows mean+SEM percentage of cells positive for the indicated cell surface markers, as determined by flow cytometry analysis six days following initiation of KD (n=4). Representative flow cytometry plots for KD6 (right panel) are also shown. (F) Representative images of cytospins of cells from (E). (G) Bar chart (left panel) shows mean+SEM proportion of viable cells as determined by Annexin-V/7-AAD analysis seven days following initiation of KD (n=3). *indicates $p<0.05$ by an unpaired t test. Representative flow cytometry plots (right panel) are also shown. (H) Panels show representative cell cycle profiles of viable control and KD cells performed six days following initiation of KD. (I) Bar chart shows mean+SEM relative transcript expression in THP1 AML cells expressing either *IRX3* or a control retroviral vector (MTV) in *IRX3* KD6 cells relative to control cells (n=3). (J) Western blot shows expression of the indicated proteins in the indicated conditions. (K) Bar chart (left panel) shows mean+SEM CFC frequencies of THP1 AML cells expressing either *IRX3* or a control retroviral vector (MTV) in *IRX3* KD6 cells relative to control cells. Colonies were enumerated after ten days in semi-solid culture (n=4). Image (right panel) shows representative colonies. For B, E, I and K *indicates $p<0.05$ by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of each KD sample versus NTC. NS = not significant.

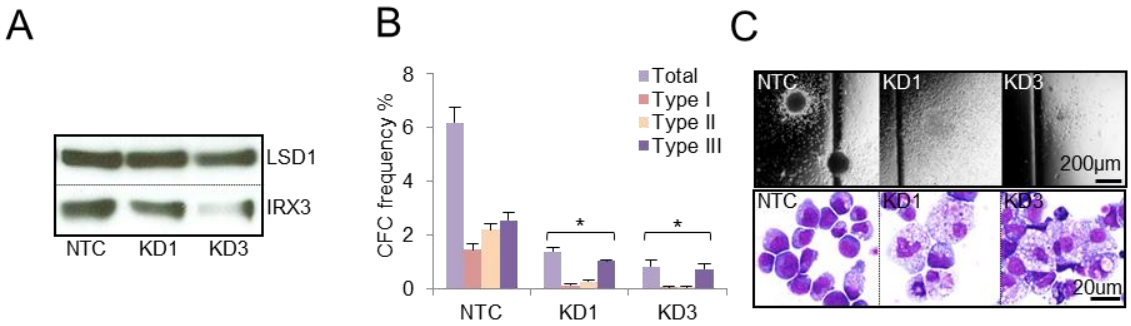


Figure 26. IRX3 sustains the differentiation block and clonogenic potential of murine MLL-AF9 AML cells. Murine MLL-AF9 AML cells were infected with lentiviral vectors targeting *Irx3* for KD (KD1 and KD3) or a non-targeting control vector (NTC). (A) Western blot shows expression of the indicated proteins within the soluble nuclear fraction in the indicated conditions. LSD1 was used as a loading control. (B) Bar chart shows the mean+SEM colony-forming cell (CFC) frequencies enumerated after five to seven days in semi-solid medium (n=3). Type I colonies contain poorly differentiated myeloblasts, Type II colonies contain a mixed population of blasts and differentiating myeloid cells, Type III colonies contain terminally differentiated myeloid cells. (C) Images show representative colonies (top panel) and cytospins (bottom panel) from (B). *indicates $p<0.05$ by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of each KD sample versus NTC.

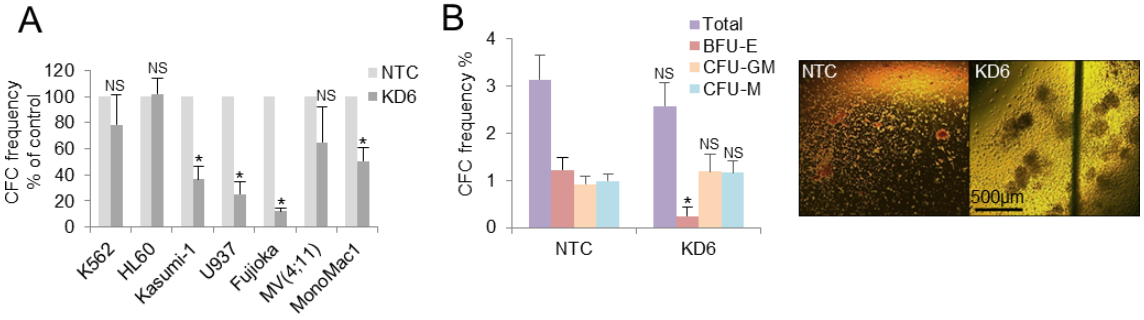


Figure 27. Consequences of *IRX3* KD in human AML cell lines and normal CD34$^+$ cells. Human AML cell lines (A) or normal human CD34$^+$ HSPCs (B) were infected with the *IRX3* KD6 vector or a non-targeting control (NTC). (A) Bar chart shows mean+SEM colony-forming cell (CFC) frequencies of KD cells in the indicated cell lines relative to control cells, enumerated after ten days in semi-solid culture (n=3). *indicates $p<0.05$ using an unpaired t-test. (B) Bar chart (left panel) shows mean+SEM % total and types of CFCs of normal human CD34$^+$ HSPCs (n=3 separate individuals). Colonies were enumerated after 14 days. Image (right panel) shows representative colonies. BFU-E, burst-forming unit erythroid; CFU-GM, colony-forming unit granulocyte/macrophage; CFU-M, colony-forming unit macrophage. * indicates $p<0.05$ by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of KD6 versus NTC . NS = not significant

Analysis of the transcriptional changes prompted by phorbol-12-myristate-13-acetate (PMA)-induced differentiation of THP1 AML cells into macrophages identified IRX3 as one of 78 predicted key transcriptional regulators in this monocytic cell line (Suzuki et al., 2009; Tomaru et al., 2009). Moreover, it was suggested that *IRX3* may be regulated by the transcription factor MYB (Tomaru et al., 2009). With this in mind *MYB* was knocked down in THP1 cells with a number of different constructs with each one reducing *MYB* expression by greater than 50% (Figure 28A). At this same time point significant reductions in *IRX3* transcripts were observed following *MYB* KD for all constructs with the exception of KD3 (Figure 28A). *MYB* KD led to a complete loss of clonogenic potential (Figure 28B). For all but one of the constructs (KD3), evidence of potent differentiation was observed seven days following initiation of *MYB* KD, as demonstrated by FACS analysis for the monocyte/macrophage differentiation marker CD14 (Figure 28C). These data demonstrate that *IRX3* expression is indirectly dependent upon that of MYB.

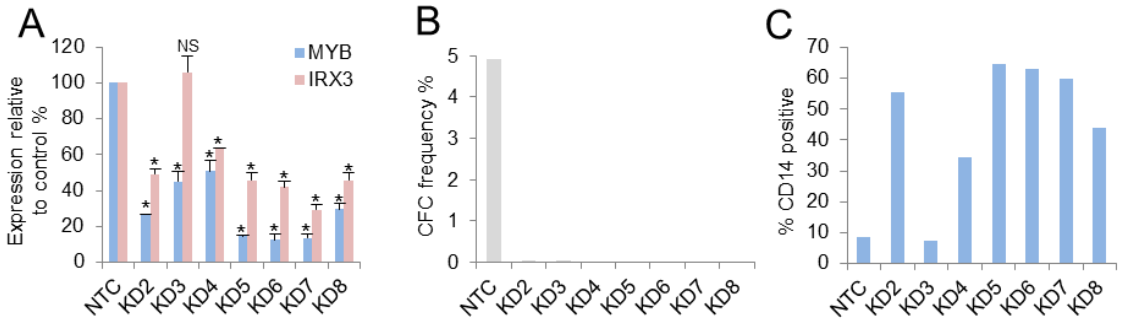


Figure 28. *IRX3* is down regulated upon *MYB* KD-induced differentiation of THP1 AML cells. Human THP1 AML cells were infected with lentiviral vectors targeting *MYB* for KD (KD2-KD8) or a non-targeting control vector (NTC). (A) Bar chart shows mean+SEM relative *MYB* and *IRX3* transcript expression in *MYB* KD versus control cells 72 hours post-infection, 48 hours post-puromycin selection (n=2). (B) Bar chart shows the colony-forming cell (CFC) frequencies enumerated ten days in semi-solid medium (n=1). (C) Bar chart shows percentage of cells positive for CD14 expression, as determined by flow cytometry analysis six days following initiation of KD (n=1). *indicates $p<0.001$ by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of each KD sample versus NTC. NS = not significant.

4.5. *IRX3* enhances the clonogenic and proliferative potential of a myeloid cell with a downstream immunophenotype

To investigate the consequences of *IRX3* forced expression in normal HSPCs, murine $CD117^+$ BM cells were infected with retroviral vectors (Figure 29A). In serial replating experiments using successfully transduced cells, *IRX3*-expressing cells ($IRX3^+$ cells) displayed enhanced colony forming cell frequencies. By comparison with control cells infected with an empty vector (MTV), for which serial replating capacity was extinguished by the third round, $IRX3^+$ cells could be serially replated for at least four rounds (Figure 29B). Comparison of the colony types formed in the third round showed that, whereas MTV cells formed colonies containing only mature cells (Type 3), colonies formed by $IRX3^+$ cells contained both poorly differentiated and mature cells (Figure 29C). Consistent with enhanced colony forming and proliferative potential versus MTV cells, $IRX3^+$ cells grown for seven days in liquid culture contained a higher proportion of cells in the SG_2M phase of

the cell cycle (Figure 29D). Analysis of cytospin preparations of these cells revealed that IRX3$^+$ cells contained a higher proportion of myeloblasts versus mature neutrophils or macrophages (Figure 29E). Despite immature morphology however, immunophenotypic analysis demonstrated that IRX3$^+$ cells displayed higher expression of the mature myeloid differentiation marker F4/80 and lower expression of the progenitor cell marker CD117 (Figures 29F). IRX3$^+$ cells also displayed lower levels of the myeloid marker Gr1 (Figure 29F). In keeping with the elevated F4/80 expression, qPCR analysis of bulk cell populations 72 hours following infection with retroviral vectors demonstrated that IRX3$^+$ cells expressed higher levels of the monocyte/macrophage genes *Klf4* and *MafB* (Figure 29G and 29H). Interestingly, IRX3$^+$ cells also expressed higher levels of the self-renewal-associated transcription factor gene *Hoxa9* (Figure 29I). Up regulation of HOXA9 was also confirmed by western blot (Figure 29J). This reactivation of *Hoxa9* may underlie the enhanced serial replating and proliferative capacity of IRX3$^+$ cells. It should be noted that, as these analyses were performed on bulk populations of cells, it may be that these expression changes are reflective of a heterogeneous population of cells, i.e. the expression of *IRX3* in one cell type may promote differentiation whereas in another (more primitive) cell type it may induce a differentiation block. These data suggest that IRX3 enhances the colony forming potential and proliferative capacity of myeloid cells with a downstream immunophenotype.

To explore this in more detail, murine CD117$^+$ BM cells were again infected with retroviral vectors and grown in selection media for seven days prior to FACS sorting on the basis of their CD11b and F4/80 expression. Three populations were sorted: CD11b$^-$F4/80$^-$ (P1; immature), CD11b^{intermediate}F4/80$^{-/low}$ (P2; maturing), and CD11b^{high}F4/80$^+$ (P3; mature), and their clonogenic potential was assessed by plating in semi-solid media (Figure 30A and 30B). As expected, the clonogenic potential of MTV cells almost entirely resided within the P1 fraction and colonies formed were composed of only mature cells (Figure 30B). In contrast, the clonogenic potential of IRX3$^+$ cells extended to the P2 fraction, and both the P1 and P2 fractions formed more colonies than those of the control (Figure 30B). Moreover, the colonies formed by P1 and P2 IRX3$^+$ cells were mainly composed of poorly differentiated cells (Figure 30B). Consistent with these observations, P1 and P2 IRX3$^+$ cells contained a greater proportion of cells in the SG_2M phase of the cell cycle (Figure 30C) and displayed more immature morphology in cytospin preparations (Figure 30D).

Taken together, these data show that expression of *IRX3* in HSPCs enhances the clonogenic potential and proliferation of both immunophenotypically less mature and mature myeloid cells. They also suggest that *IRX3* expression maintains, or reactivates, the expression of *Hoxa9* which may explain the enhanced serial replating potential of IRX3$^+$ cells.

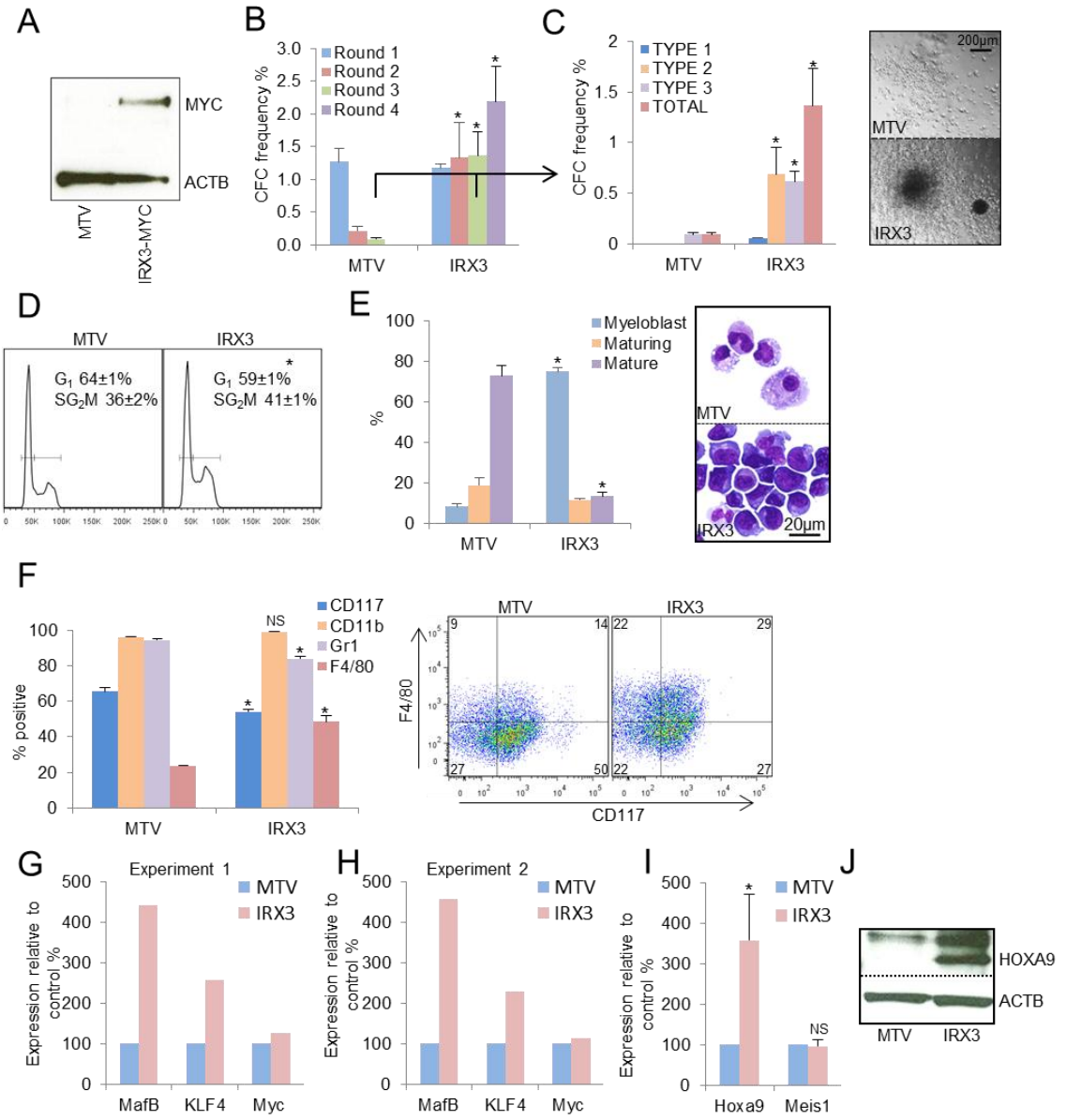


Figure 29. IRX3 enhances the clonogenic and proliferative potential of normal HSPCs.
(A-C) Murine CD117$^+$ BM cells were infected with *IRX3*-expressing or control retroviral vectors and serially replated *in vitro*. (A) Western blot shows MYC-tagged IRX3 expression in CD117$^+$ BM cells 48 hours following drug selection and 72 hours post-spinoculation. (B) Bar chart shows mean+SEM CFC frequencies at the end of each round of serial replating (n=3). (C) Bar chart (left panel) shows mean+SEM types of colonies formed in the second round of culture (n=3). Type 1 colonies contain poorly differentiated cells only; Type 2 colonies contain both poorly differentiated and mature cells; Type 3 colonies contain mature cells only. Image (right panel) shows representative colonies. (D-H) Murine CD117$^+$ BM cells were infected with *IRX3*-expressing or control retroviral vectors and grown in liquid culture. (D) Representative cell cycle profiles of viable control and IRX3$^+$ cells following seven days in liquid culture (n=3). * indicates $p<0.05$ by an unpaired t test. (E) Bar chart (left panel) shows mean+SEM percentage of the indicated cell types in cytospin preparations following seven days in liquid culture (n=3). Representative images (right panel) are shown. (F) Bar chart (left panel) shows mean+SEM percentage of cells positive for the indicated cell surface markers following seven days of liquid culture (n=3). A representative flow cytometry plot (right panel) is shown. (G&H) Bar charts show relative transcript expression for the indicated genes in the indicated conditions 48 hours following drug selection and 72 hours post-spinoculation for two independent experiments. (I) Bar chart shows mean+SEM relative transcript expression for the indicated genes in the indicated conditions 48 hours following drug selection and 72 hours post-spinoculation (n=4). (J) Western blot shows expression of the indicated proteins in the indicated conditions. For B and C * indicates $p<0.05$ and for E, F and G * indicates $p<0.001$ by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of IRX3$^+$ cells versus MTV.

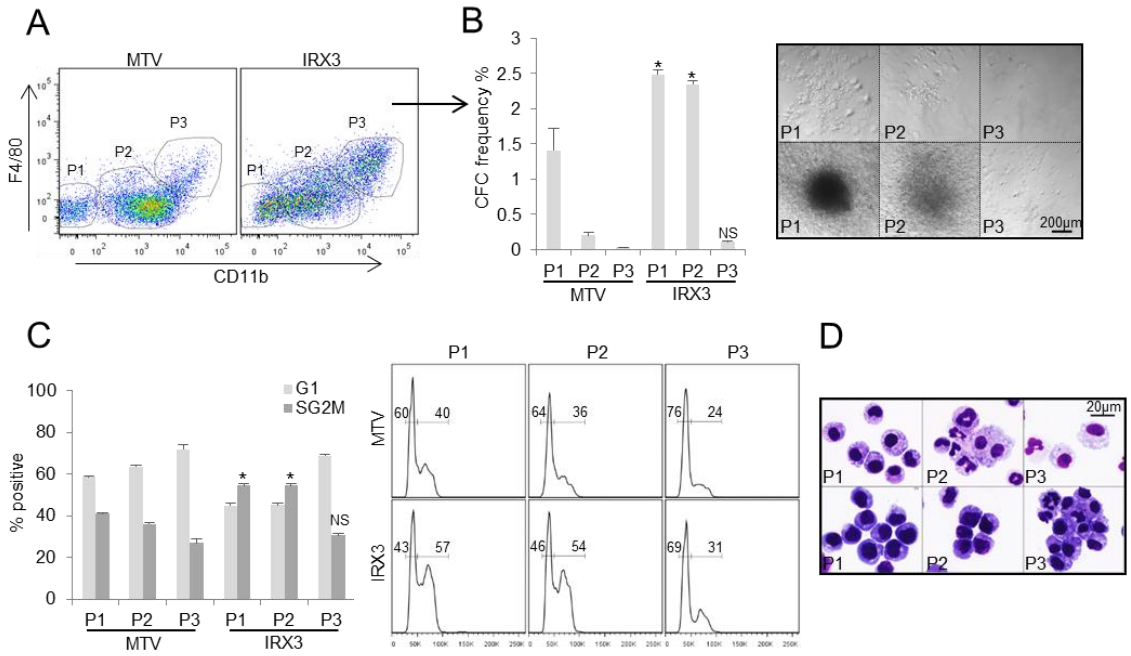


Figure 30. IRX3 enhances the clonogenic and proliferative potential of myeloid cells with a downstream immunophenotype. Murine CD117^{+} BM cells were infected with *IRX3*-expressing or control retroviral vectors and FACS purified into three populations on the basis of their F4/80 and CD11b expression following seven days in liquid culture. (A) Representative flow cytometry plots following showing the gating strategy used. (B) Bar chart (left panel) shows mean+SEM CFC frequencies for each of the three populations per condition enumerated after five days (n=3). Image (right panel) shows representative colonies. (C) Bar chart (left panel) shows mean+SEM percentage of cells in the indicated stage of the cell cycle as determine by propidium iodide staining of FACS purified cells. Representative cell cycle profiles (right panel) are also shown (n=3). (D) Images show representative cytospin images from (C). *indicates p<0.005 by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of IRX3^{+} cells versus MTV for the respective sorted population. NS = not significant.

4.6. IRX3 induces T cell lineage skewing *in vivo*

To investigate the *in vivo* consequences of *IRX3* expression in HSPCs, transplantation experiments were performed as outlined in the previous chapter (see Figure 11). Murine BM HSPCs were infected with retroviral vectors expressing *IRX3*, *Meis1* or *Hoxa9*, or an empty vector (hereafter referred to as IRX3^{+}, MEIS1^{+}, HOXA9^{+} or MTV cells, respectively) and transplanted into irradiated congenic recipients. Assessment of the extent of donor:recipient chimerism in the blood of transplanted mice over 16 weeks demonstrated that at all time points recipients of IRX3^{+} cells exhibited significantly less donor:recipient chimerism in the blood (the BM was not sampled) (Figure 31A). Consideration of the lineage composition of engrafted cells at each time point further revealed IRX3^{+} recipients showed significantly higher levels proportionately of cells with a T cell immunophenotype at the 8 and 12 week time points (Figure 31C, 31D and 31F). This T cell skewing highlighted a discrepancy with the myeloid markers used for initial analyses; the anti-Gr1 antibody used in this experiment, conventionally used to detect Ly-6G positive granulocytes, also reacts with the Ly-6C antigen which is additionally expressed on CD8^{+}CD44^{high} memory type T cells (Matsuzaki et al., 2003). Thus, a significant proportion of IRX3^{+} cells were Gr1^{+}CD11b^{-} (Figures 32A-C). Of note, this immunophenotype was in contrast to the myeloid-skewed HOXA9^{+}

cells (Figure 31F), as well as FOXC1$^+$ cells reported in the previous chapter (see Figure 11F), which were almost entirely double positive for Gr1 and CD11b. Hence, for the analyses displayed in Figure 31B-F, myeloid cells were considered as CD11b$^+$ only. This abnormal Gr1$^+$CD11b$^-$ population was also apparent at 16 weeks (Figure 32C). Thus, expression of *IRX3* in HSPCs reduces donor:recipient chimerism in blood and expands the T cell lineage.

As already shown in Figure 11H, mice transplanted with HOXA9$^+$ cells succumbed to AML with a median latency of 103 days, whereas at the termination of the experiment at 200 days post-transplant none of the mice from the MTV or MEIS1 cohorts had died of a donor-derived haematologic neoplasm or exhibited any features thereof in bone marrow analyses (Figure 32D). In the IRX3 cohort 4/6 mice became sick and died during the experiment. Out of these four mice, one succumbed to a recipient-derived haematological disease and one was found dead before samples could be retrieved for analysis. The remaining two mice succumbed to a T cell leukaemia after 183 days as evidenced by complete infiltration of the bone marrow by donor-derived CD4$^+$CD8$^+$ T cells (Figure 32D and 32E). This is consistent with the abnormal expansion of a T cell clone amongst the IRX3$^+$ recipients. Of note however, flow cytometry analysis of these leukaemic BM T cells showed no positivity for Gr1 expression. This suggests that, if this T cell leukaemia did arise from an abnormal Ly-6C$^+$ clone, the leukaemia cells that ultimately arose lost the expression of this marker. Interestingly, analysis of expression data from a large study investigating the utility of gene expression profiling in the diagnosis of a variety of haematological malignancies (Haferlach et al., 2010) reveals that *IRX3* is significantly expressed in both AML and T cell acute lymphoblastic leukaemia (Figures 32F and 32G). Thus, IRX3 may play a functional role in both of these malignancies. Taken together these data indicate that aberrant *IRX3* expression in HSPCs is sufficient to skew differentiation towards T-cell lineage and initiate T-cell leukaemia in mice.

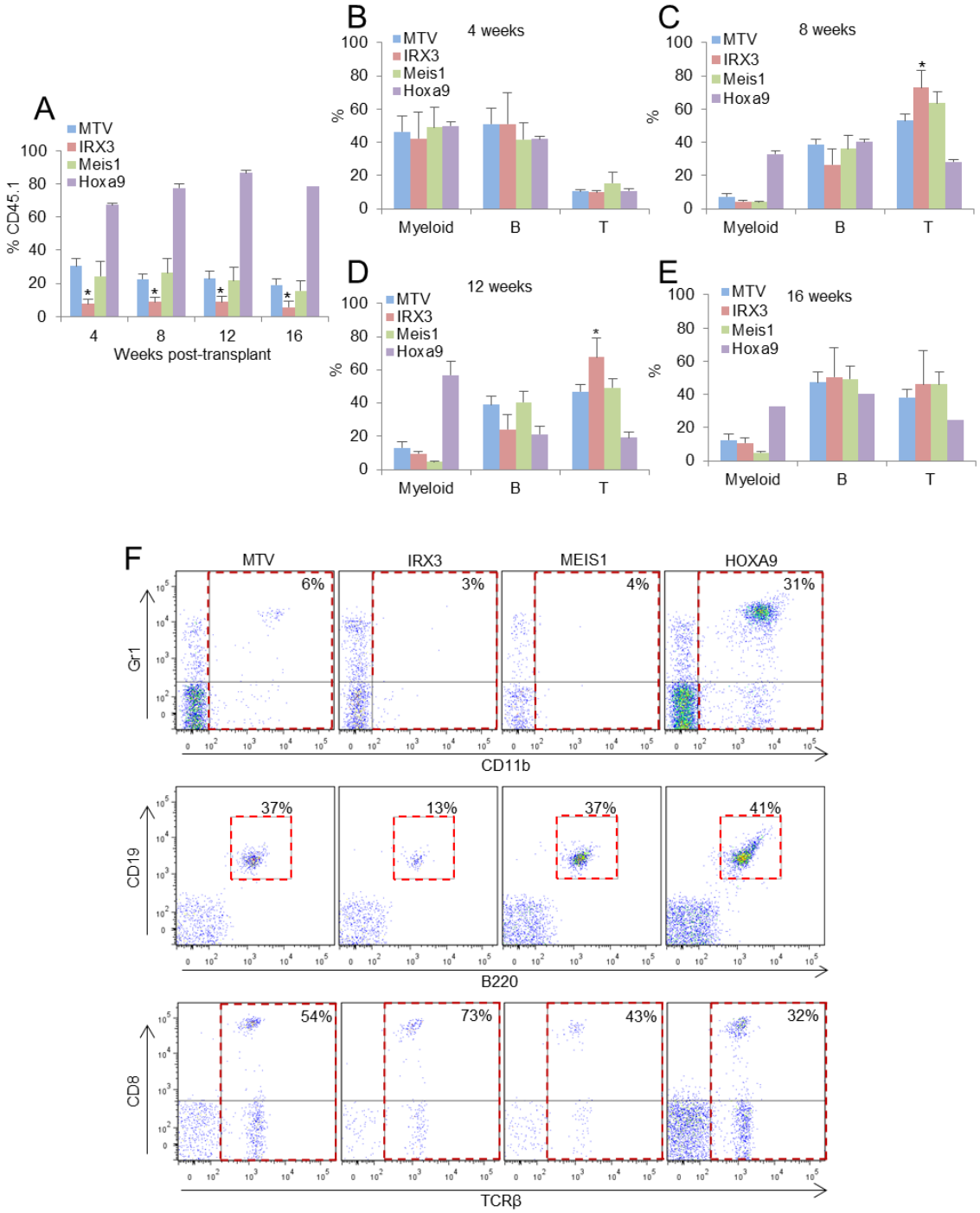


Figure 31. *IRX3* induces T cell lineage skewing *in vivo.*
Murine CD45.1$^+$ CD117$^+$ BM cells were infected with the indicated retroviral vectors; 10^6 drug resistant cells were transplanted into CD45.2$^+$ irradiated congenic recipients 96 hours following spinoculation. (A) Bar chart shows mean+SEM percentage donor-derived CD45.1$^+$ cells in blood at the indicated times post-transplantation. * indicates $p<0.05$ for comparison of IRX3$^+$ recipients versus all others, and at each time point, by one way ANOVA with Fisher's least significant difference *post hoc* test. (B-E) Bar charts show mean+SEM percentage contribution of donor-derived cells to the indicated lineages (myeloid-lineage, CD11b$^+$; B-lineage, CD19$^+$B220$^+$; T-lineage, T-cell receptor β$^+$) at the indicated time points post transplantation. (F) Representative flow cytometry plots indicate the gating strategy for identifying the lineage contributions of donor-derived CD45.1$^+$ cells in blood. Images shown are from eight weeks post-transplantation. For (B-E) *indicates $p<0.05$ for comparison of IRX3$^+$ recipients versus MTV recipients by one way ANOVA with Fisher's least significant difference *post hoc* test.

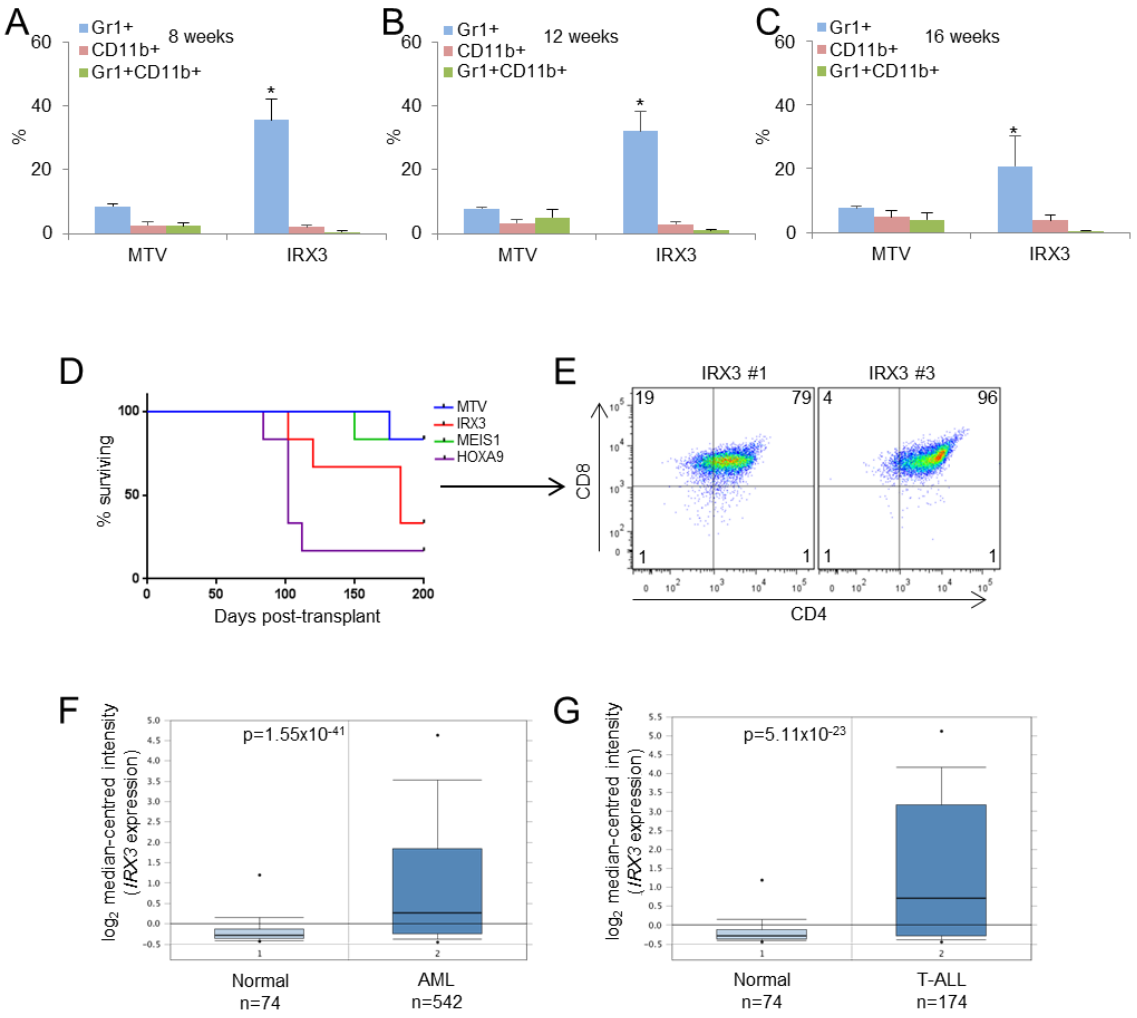


Figure 32. *IRX3* initiates T cell leukaemia in mice.
Murine CD45.1^{+} CD117^{+} BM cells were infected with the indicated retroviral vectors as described in Figure 31. (A-C) Bar charts show mean+SEM percentage contribution of donor-derived cells positive for the indicated cell surface markers at the indicated time points post transplantation. (D) Survival curve of mice transplanted with cells infected with the indicated vectors (n=6-7 mice per cohort). (E) Flow cytometry plots of CD45.1^{+} BM cells at the point of death from the two recipients of IRX3^{+} cells that developed T cell leukaemia. Bar charts show *IRX3* expression in normal (peripheral blood mononuclear cells) versus AML (F) and T cell acute lymphoblastic leukaemia (T-ALL) (G) samples. Bar charts extracted from Oncomine^{TM}. Original data from Haferlach et al. (2010). For (A-C) *indicates $p<0.05$ for comparison of IRX3^{+} recipients versus MTV recipients by one way ANOVA with Fisher's least significant difference *post hoc* test. For (F&G) displayed p values are from an unpaired t test generated by Oncomine^{TM}.

4.7. *IRX3* expression in human AML is associated with high HOX gene expression

Although *IRX3* expression in HSPCs is sufficient to initiate T-cell leukaemia, it is insufficient by itself to induce AML. Given that aberrant *IRX3* expression is observed in AML, it was next investigated whether it might collaborate with other factors to promote myeloid leukaemogenesis. To identify transcription factor genes whose expression is associated with that of *IRX3* in AML, *IRX3*^{high} AMLs were compared with *IRX3*^{low} AMLs using the Wouters et al (2009) data set (Figure 24D, Table 37). Akin to *FOXC1* expression in AML, high level *IRX3* expression is associated with HOX gene expression with *HOXA5* and *HOXA9* found to be the most highly up regulated genes in the *IRX3*^{high} group (Figure 33A). qPCR (Figure 33B) and analysis of expression data (Figure 33C and 33D) confirmed the strong association. By qPCR 9/10 (90%) samples from our cohort with high *IRX3* expression also exhibited high *HOXA9* expression, and 56% of high *HOXA9* expressing AML samples exhibited high *IRX3* expression (Figure 33B). In the published studies, when AML samples from all FAB subtypes were examined, 132/159 (83%) and 35/49 (71%) respectively with high *IRX3* expression also exhibited high *HOXA9* expression (Figures 33C and 33D), and 132/290 (46%) and 35/116 (30%) respectively of high *HOXA9* expressing AML samples exhibited high *IRX3* expression (Figures 33C and 33D). However, high *IRX3* expression is also significantly associated with APL (FAB subtype M3) samples which do not express HOX genes (Table 37). When these samples are excluded from the expression data studies, the association with *HOXA9* is much stronger. In this case, 132/138 (96%) and 35/35 (100%) respectively with high *IRX3* expression also exhibited high *HOXA9* expression (Figures 33E and 33F). The observation that approximately 40% of *HOXA9* expressing AML samples express high levels of *IRX3* provides a strong basis to suggest that, akin to *FOXC1*, *IRX3* collaborates with HOX family transcription factors to enhance leukaemogenesis.

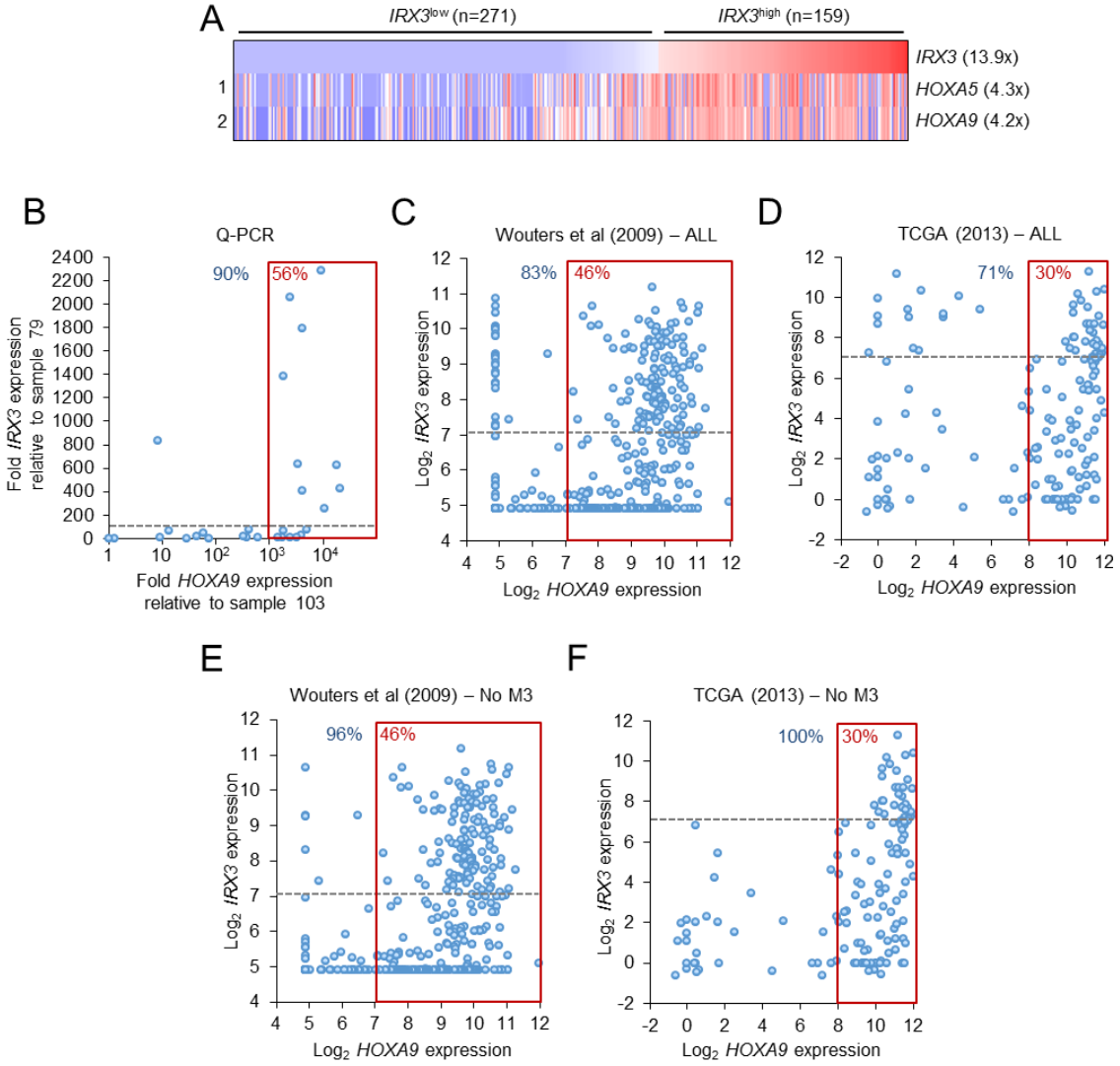


Figure 33. Association of *IRX3* expression with HOX gene expression in human AML.
(A) Heat map shows comparison of *IRX3*^{high} versus *IRX3*^{low} AML cases from Wouters et al. (2009) (see also Figure 24D) indicating the most differentially expressed transcription factor genes, and in brackets their mean fold change. (B-F) Scatter plots show the expression of *IRX3* versus *HOXA9* in primary AML patient samples as determined by (B) quantitative PCR (n=29; see also Table 35) or (C-F) array expression values from the indicated study. Percentages in blue text indicate proportion of *IRX3*^{high} samples exhibiting high *HOXA9* expression. Percentages in red text indicate the proportion of *HOXA9*^{high} samples (in the red box) additionally exhibiting high *IRX3* expression (above the dotted grey line). For (C) and (D) AMLs from all FAB subtypes and shown. For (E) and (F) AML samples categorised as FAB subtype M3 (APL) have been removed.

4.8. *IRX3* antagonises *HOXA9* function in HSPCs *in vitro*

To investigate this question further, murine $CD117^+$ HSPCs were infected in pairwise combinations with retroviral vectors expressing *Hoxa9*, *IRX3* or *Meis1*, or a control vector (to generate *Hoxa9/*MTV, *Hoxa9*/*IRX3* and *Hoxa9*/*Meis1* cells, respectively) and their clonogenic potential was assessed in serial replating assays. As previously described, *Hoxa9* expression induced sustained clonogenic potential of BM HSPCs in serial replating assays, an effect which was enhanced by co-expression of *Meis1* (Figure 34A). Interestingly, the co-expression of *Hoxa9* and *IRX3* also significantly enhanced the clonogenic potential of BM HSPCs versus cells overexpressing *Hoxa9* alone (Figures 34A). Although total colony forming cell frequency was enhanced in *Hoxa9*/*IRX3* cells versus *Hoxa9*/*MTV* cells, the size of the tightly clustered 'Type I' colonies formed was significantly reduced when compared to those formed by *Hoxa9*/MTV and *Hoxa9*/*Meis1* cells (Figure 34B). Consistent with these observations, *Hoxa9*/*IRX3* cells contained fewer cells in the SG_2M phase of the cell cycle by comparison with both *Hoxa9*/MTV and *Hoxa9*/*Meis1* cells (Figure 34C). Furthermore, immunophenotypic analysis demonstrated that *Hoxa9*/*IRX3* cells expressed significantly higher levels of the monocyte/macrophage differentiation marker F4/80 by comparison with *Hoxa9*/MTV and *Hoxa9*/*Meis1* cells (Figure 34D and 34E). The evidence of more immunophenotypic differentiation within the *Hoxa9*/*IRX3* cells is consistent with the observations in the HSPCs expressing *IRX3* alone. However, the observation that *Hoxa9*/*IRX3* cells are less proliferative than the *Hoxa9*/MTV cells is surprising given the enhanced cycling activity of HSPCs expressing *IRX3* alone and the reasons for this currently remain unclear.

4.9. *IRX3* collaborates with *HOXA9* to generate AML with enhanced myeloid differentiation block in mice

To determine the impact of high level *HOXA9* and *IRX3* co-expression on leukaemogenesis *Hoxa9*/MTV, *Hoxa9/IRX3* and *Hoxa9/Meis1* double transduced HSPCs were transplanted into irradiated congenic recipients as previously described. Donor:recipient chimerism analysis at four, eight and 12 weeks post-transplant once more demonstrated an adverse effect of *IRX3* expression on levels of peripheral blood engraftment (Figure 35A). The lineage composition of engrafted cells again demonstrated donor-derived multilineage engraftment in all three cohorts at each time point (with the exception of recipients of *Hoxa9/Meis1* cells who had all succumb to disease prior to the 12 week time point) (Figure 35B-D). As with the single transduction transplants, recipients of *Hoxa9*/MTV cells showed progressive expansion of the myeloid compartment and this effect was further enhanced in the recipients of *Hoxa9/Meis1* cells. In contrast, there were lower levels proportionately of myeloid engraftment, and higher levels of B-lymphoid engraftment, among recipients of *Hoxa9/IRX3* cells by comparison with mice receiving either *Hoxa9*/MTV or *Hoxa9/Meis1* cells at each time point (Figure 35B-D). Thus the co-expression of *IRX3* with *Hoxa9* results in skewing towards the B cell lineage, potentially due to impaired myeloid differentiation.

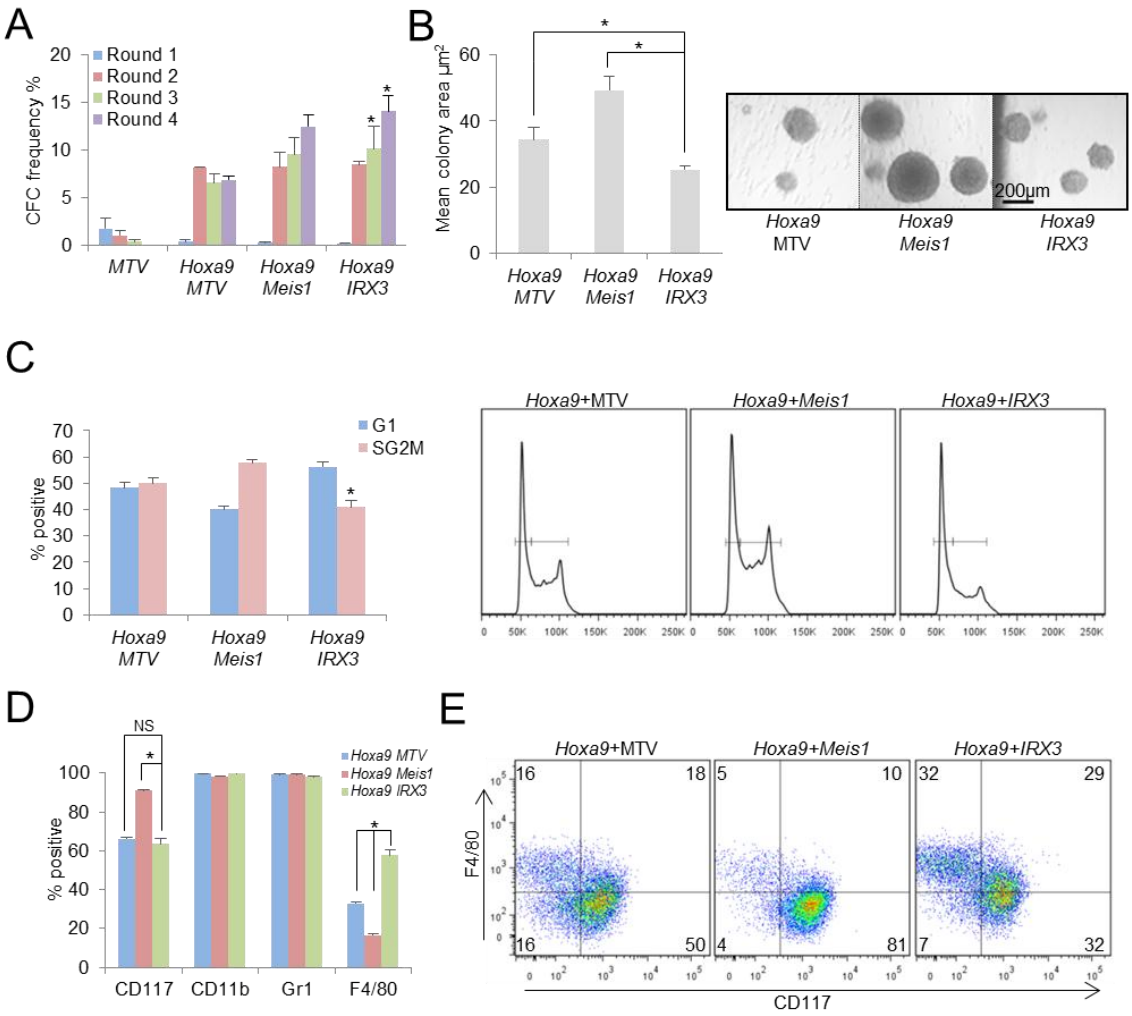
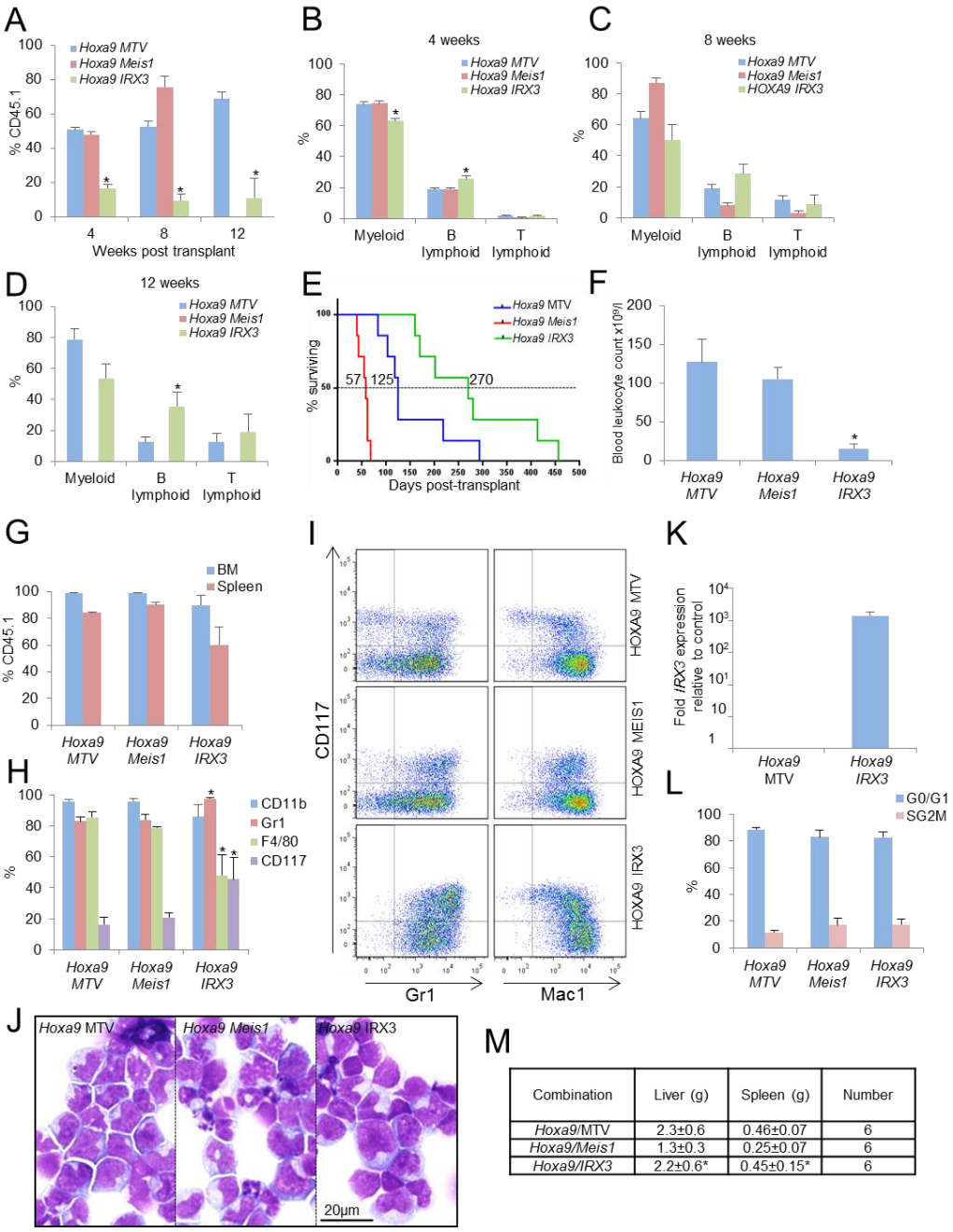


Figure 34. The consequences of *IRX3* and *Hoxa9* co-expression in normal BM HSPCs.
(A) Bar chart shows mean+SEM colony-forming cell (CFC) frequencies after each round of serial replating of murine CD117^{+} BM cells co-transduced with the indicated retroviral or control expression vectors (n=3). MTV = empty vector. *indicates p<0.05 by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of *Hoxa9/IRX3* cells versus *Hoxa9/MTV*. (B) Bar chart (left panel) shows mean+SEM area of Type 1 colonies from (A) and representative images (right panel). *indicates p<0.01 by an unpaired t-test for the indicated comparisons. (C) Bar chart (left panel) shows mean+SEM percentage of cells in the indicated stage of the cell cycle as determine by propidium iodide staining and flow cytometry analysis. Representative cell cycle profiles (right panel) are also shown (n=3). *indicates p<0.05 by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of *Hoxa9/IRX3* cells versus *Hoxa9*/*MTV* and *Hoxa9*/*Meis1*. (D) Bar chart shows mean+SEM percentage of cells positive for the indicated cell surface markers (as determined by flow cytometry) following six days in liquid culture (n=3). (E) Representative flow cytometry plots from (D). *indicates p<0.005 by one way ANOVA with Fisher's least significant difference *post hoc* test for the indicated comparisons. NS = not significant

As shown previously (see Figure 15E), recipients of *Hoxa9/Meis1* cells developed AML more rapidly than recipients of *Hoxa9*/MTV cells (median 57 days versus 125 days). In contrast, the onset of AML was significantly delayed in recipients of *Hoxa9/IRX3* cells compared to mice receiving *Hoxa9/*MTV (median 270 days versus 125 days) (Figure 35E). By comparison with *Hoxa9/*MTV or *Hoxa9*/*Meis1* recipients, at the point of death *Hoxa9*/*IRX3* recipients exhibited significantly lower total circulating leukocyte counts and reduced infiltration of the spleen, although the latter was more variable (Figures 35F and 35G). In the BM however, *Hoxa9*/*IRX3* recipients exhibited extensive donor cell engraftment and evidence of a greater myeloid differentiation block in comparison to recipients of *Hoxa9*/MTV or *Hoxa9/Meis1* cells, as shown by significantly higher CD117 expression and significantly lower F4/80 expression (Figures 35G-J). Importantly, high level *IRX3* expression could still be detected in the *Hoxa9/IRX3* AML BM cells (Figures 35K). Cell cycle analysis of AML cells from the BM demonstrated no difference in the fraction of cycling cells between the three cohorts (Figure 35L). In each cohort and in each case, autopsy demonstrated splenomegaly and hepatomegaly, with spleen and liver weights being comparable between *Hoxa9*/*IRX3* recipients and the *Hoxa9/*MTV recipients (Figure 35M). Histological analysis demonstrated loss of the splenic architecture in all three cohorts, which included loss of the white pulp in *Hoxa9*/*IRX3* recipients (Figure 36A). In the livers of *Hoxa9/IRX3* recipients, akin to *Hoxa9/FOXC1* recipients, there was a diffuse infiltrate of blast cells, forming small clusters which expanded the sinusoids. Periportal and perivenular blasts, while present, were less conspicuous than those seen in mice from the other cohorts (Figure 36B and 36C).

There is a clear disparity between the *in vitro* and *in vivo* data regarding the functional role of *IRX3* in the context of high *Hoxa9* expression. Whereas *in vitro IRX3* appears to antagonise the function of *Hoxa9*, as evidenced by impaired proliferation and increased immunophenotypic differentiation, *in vivo IRX3* actually results in a greater myeloid differentiation block in the presence of *Hoxa9*; this is despite longer disease latency. As the *Hoxa9*/*IRX3* AML cells exhibit sustained high level *IRX3* expression, this suggests that IRX3 is functionally required for the enhanced myeloid differentiation block observed.



Combination	Liver (g)	Spleen (g)	Number
Hoxa9/MTV	2.3±0.6	0.46±0.07	6
Hoxa9/Meis1	1.3±0.3	0.25±0.07	6
Hoxa9/IRX3	2.2±0.6*	0.45±0.15*	6

Figure 35. *IRX3* enhances myeloid differentiation block in collaboration with *HOXA9* in AML.
Murine CD45.1$^+$ CD117$^+$ BM cells were co-transduced with retroviral vectors. 96 hours later 10^6 drug resistant cells were transplanted into CD45$^+$ irradiated congenic recipients. (A) Bar chart shows mean+SEM percentage of donor-derived CD45.1$^+$ cells in blood at the indicated times post-transplant. *p<0.0001 by one way ANOVA with Fisher's least significant difference *post hoc* test (4 and 8 weeks) or by an unpaired t test (12 weeks) for comparison of *Hoxa9/IRX3* cells versus *Hoxa9*/MTV and *Hoxa9/Meis1*. (B-D) Bar charts show mean+SEM percentage contribution of donor-derived cells to each lineage in blood at the indicated time points post-transplant. *p<0.05 by one way ANOVA with Fisher's least significant difference *post hoc* test (4 and 8 weeks) or by an unpaired t test (12 weeks) for comparison of *Hoxa9/IRX3* cells versus *Hoxa9*/MTV. (E) Survival curves of transplanted mice (n=7 per cohort). Median survivals are shown. (F) Bar chart shows mean+SEM total blood leukocyte count at death in the indicated cohorts as determined by haemocytometer counting (n=3-5 per cohort). *p<0.05 by one way ANOVA with Fisher's least significant difference *post hoc* test. (G) Bar chart shows mean+SEM percentage of donor-derived CD45.1$^+$ cells in the BM and spleen at death. (H) Bar chart shows the mean+SEM percentage of donor-derived cells positive for the indicated cell surface markers in BM of leukaemic mice, as determined by flow cytometry. *p<0.05 by one way ANOVA with Fisher's least significant difference *post hoc* test. (I) Representative flow cytometry plots from (H). (J) Images of cytospins from BM cells taken at the point of death. (K) Bar chart shows relative expression of *IRX3* in AML BM cells at the point of death (n=3). (L) Bar chart shows mean+SEM percentage of BM cells in the indicated phase of the cell cycle at death. (M) Table shows mean±SD organ masses of leukaemic animals at death. *indicates that *Hoxa9/IRX3* recipient mice organ weights are significantly greater by comparison with *Hoxa9/Meis1* (liver and spleen) recipients (p<0.05; unpaired t-test).

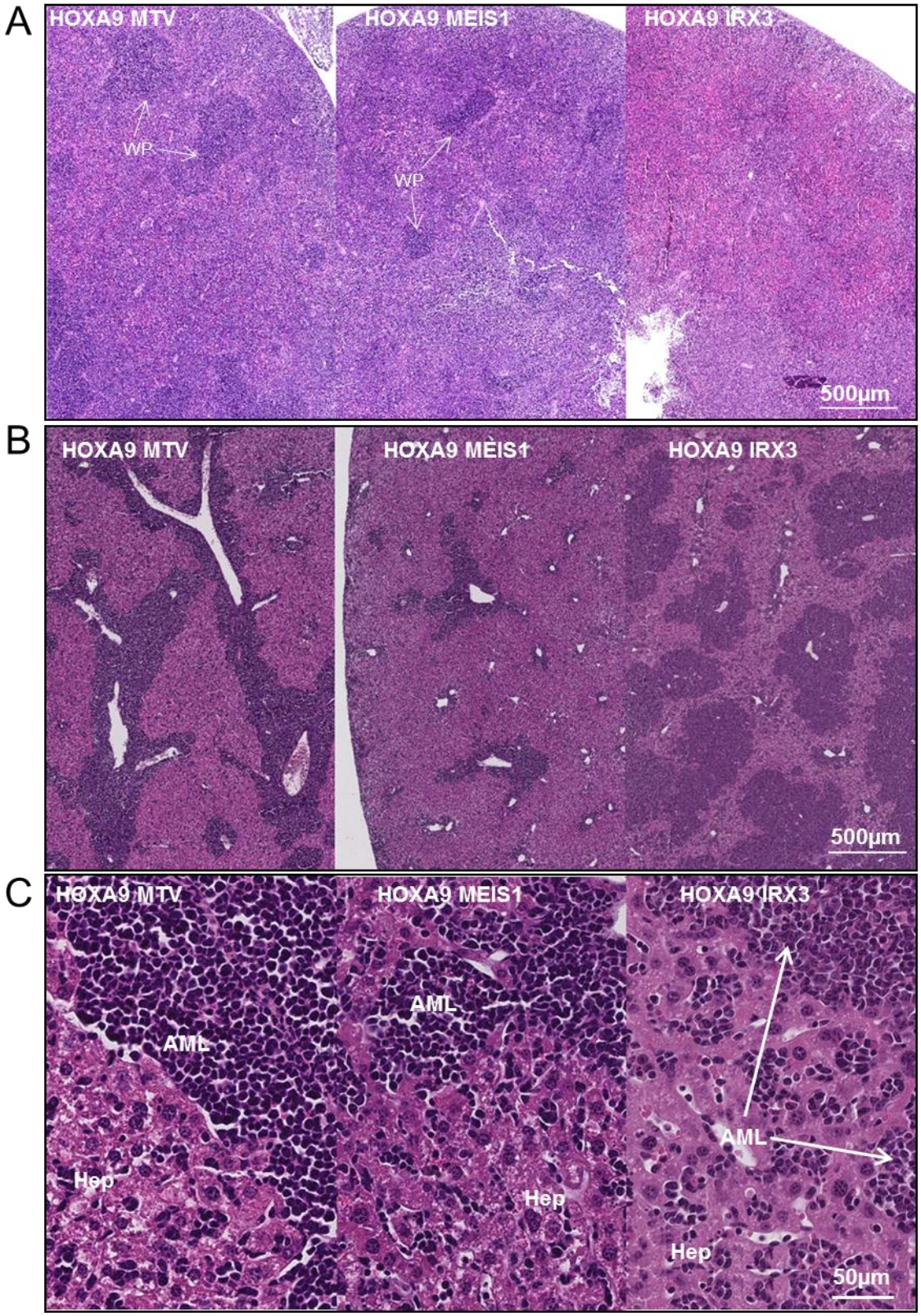


Figure 36. The histopathology of *Hoxa9/IRX3* leukaemias.
Images are representative of (A) spleen histology and (B&C) liver histology from sick mice with leukaemias initiated by expression of the indicated combination of transcription factors. WP = white pulp; AML = leukaemia cells infiltrating liver; Hep = hepatocytes.

4.10. The functional relevance of *IRX3* and *FOXC1* co-expression in myeloid progenitor cells

As already discussed, a strong positive correlation is observed between the expression of *IRX3* and the mesenchymal transcription factor gene *FOXC1* in human AML. To investigate the functional consequences of this tissue inappropriate co-expression, murine $CD117^{+}$ HSPCs were infected in pairwise combinations with retroviral vectors expressing *FOXC1*, *IRX3*, or a control vector (to generate MTV/*FOXC1*, MTV/*IRX3* and *IRX3*/*FOXC1* cells) and their clonogenic potential was assessed in serial replating assays. Consistent with previous experiments, the expression of *FOXC1* alone induced a transient myeloid differentiation block with increased numbers of colonies formed in the second round of the replating assay (Figures 37A and 37B). This enhanced clonogenic potential was extinguished by the third round. In contrast, and as previously shown, the expression of *IRX3* alone enhanced the clonogenic potential of BM HSPCs, enabling their serial replating for at least four rounds (Figures 37A and 37B). Interestingly, the co-expression of *IRX3* with *FOXC1* enhanced the number of second round colonies formed by HSPCs in comparison to those expressing *FOXC1* alone, although this did not reach statistical significance. In comparison to MTV/*FOXC1* cells, there was also evidence of extended replating potential of *IRX3*/*FOXC1* cells up to the fourth round. However, *IRX3*/*FOXC1* cells formed significantly fewer colonies than *MTV*/*IRX3* cells in the third and fourth round (Figure 37A and 37B). Thus, as with cells expressing *FOXC1* alone, it appears that *FOXC1* only transient enhances the clonogenic potential of myeloid cells expressing *IRX3*, after which it actually becomes detrimental. *IRX3/FOXC1* cells expressed significantly higher levels of the monocyte/macrophage differentiation marker F4/80 than MTV/*FOXC1* cells (Figures 37C and 37D). This is consistent with observations made in cells expressing either *IRX3* alone or those concurrently expressing *IRX3* and *HOXA9*. Thus, IRX3 has consistently been shown to induce immunophenotypic differentiation of 'pre-leukaemic' myeloid cells *in vitro*.

Given that patients with *FOXC1*^{high} AML have a significantly poorer outcome than those with *FOXC1*^{low} AML, and that *FOXC1* and *IRX3* are often found co-expressed in AML, the observation that *IRX3* has the capacity to induce immunophenotypic differentiation in certain cell contexts may have important clinical consequences. Indeed, when *FOXC1*^{high} AMLs are divided on the basis of their *IRX3* expression, *IRX3*^{low} patients exhibit considerably inferior survival in comparison to *IRX3*^{high} patients (Figures 37E and 37F). Thus, in the context of high *FOXC1* expression, the expression of *IRX3* appears to exhibit a protective effect in AML.

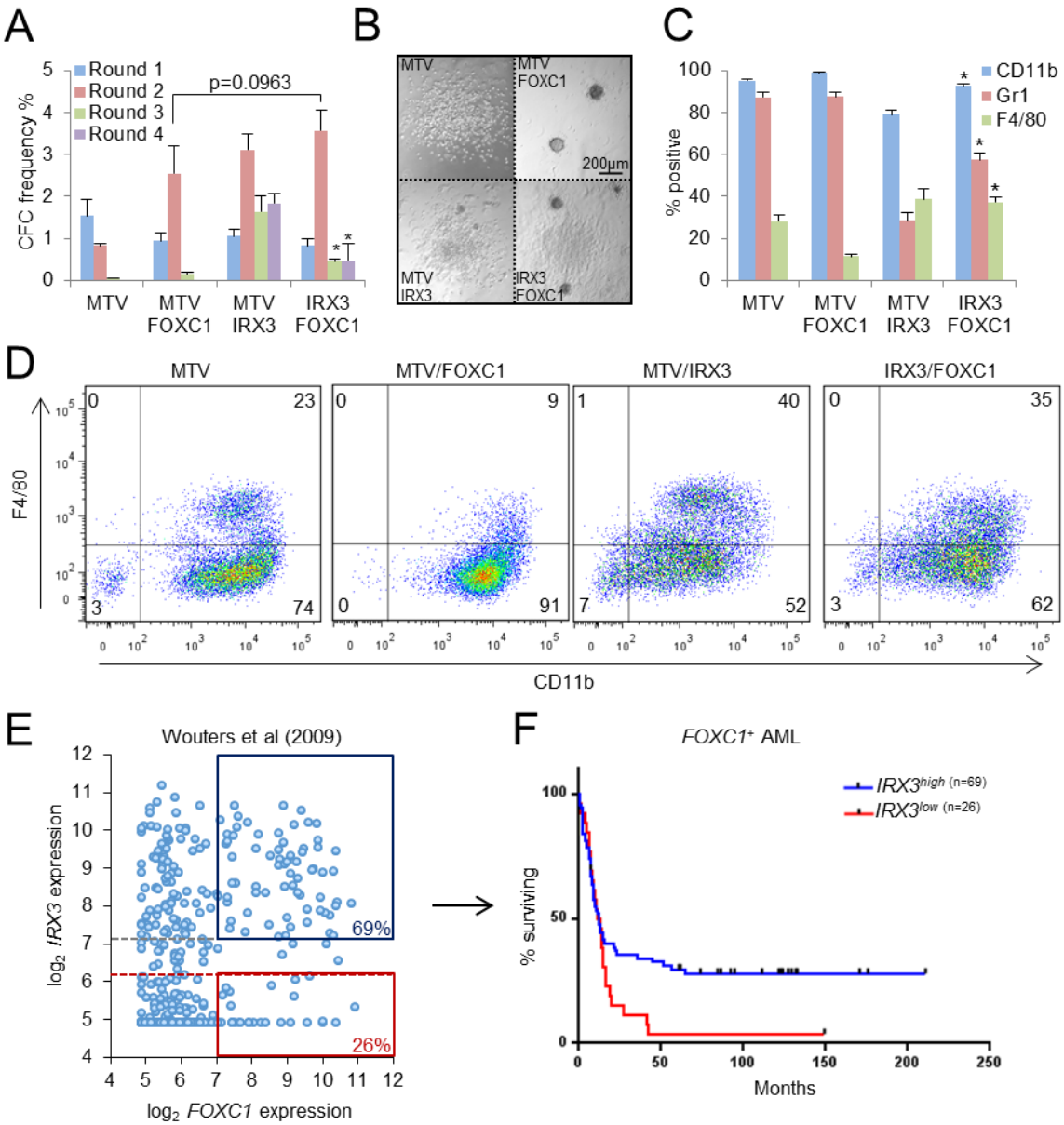


Figure 37. *IRX3* enhances the clonogenic potential of *FOXC1*-expressing HSCPs and the consequences of *IRX3* and *FOXC1* co-expression in AML. (A) Bar chart shows mean+SEM colony-forming cell (CFC) frequencies after each round of serial replating of murine CD117^{+} BM cells co-transduced with the indicated retroviral or control expression vectors (n=3). MTV = empty vector. *indicates p<0.005 by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of *IRX3*/*FOXC1* cells versus MTV/*IRX3* cells. (B) Representative images of colonies from round 2 of the replating assay. (C) Bar chart shows mean+SEM percentage of cells positive for the indicated cell surface markers (as determined by flow cytometry) at the end of round 2 of the replating assay (n=3). *indicates p<0.005 by one way ANOVA with Fisher's least significant difference *post hoc* test for comparison of *IRX3*/*FOXC1* cells versus MTV/*FOXC1* cells. (D) Representative flow cytometry plots from (C). (E) Scatter plot (left panel) shows the expression of *IRX3* versus *FOXC1* in primary AML patient samples as determined array expression values from the indicated study. Percentages in blue text indicate proportion of *IRX3*^{high} samples (above the dotted grey line) exhibiting high *FOXC1* expression (in the blue box). Percentages in red text indicate the proportion of *FOXC1*^{high} samples (in the red box) exhibiting low *IRX3* expression (below the dotted red line). (F) Survival curves of patients with *FOXC1*^{high} AML divided into *IRX3*^{high} and *IRX3*^{low} AML as shown in (E).

4.11. Loss of Polycomb-mediated repression promotes *IRX3* derepression

As already discussed, analysis of ENCODE data revealed that *IRX3* is minimally expressed in normal CD34$^+$ cells and is marked by high levels of H3K27me3 (Zhou et al., 2011). This suggests that, as is the case with *FOXC1*, in normal haematopoietic cells transcriptional silence of *IRX3* is maintained by PRC. To address this normal human CD34$^+$ cells from multiple donors were treated with PRC inhibitors. Whereas cells treated with the EZH2 inhibitor GSK343 (Verma et al., 2012) exhibited a significant increase in expression of *FOXC1*, no increase in *IRX3* or *HOXA9* was observed (Figure 38A). In separate experiments, cells treated with UNC1999, a dual EZH1 and EZH2 inhibitor (Konze et al., 2013), did exhibit significant increase in *IRX3* expression, although this was not as extensive as the increase in *FOXC1* expression observed; again expression of *HOXA9* was unaffected (Figure 38B). PRT4165, a PRC1 E3 ubiquitin ligase inhibitor (Ismail et al., 2013), did not result in any increase in *IRX3* expression (Figure 38B). These data indicate that continued repression of *IRX3* in the haematopoietic system is mediated by PRC2 and imply that loss of the activity of both EZH1 and EZH2 components at this genomic locus is required, and contributes to, its derepression in AML.

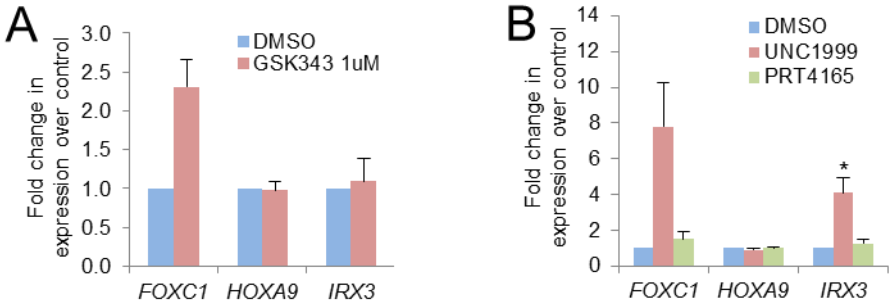


Figure 38. Derepression of *IRX3* in normal CD34$^+$ cells is induced by PRC inhibition.
(A & B) CD34$^+$ cells from separate normal donors (n=3 and 4 respectively) were treated for four or five days respectively with PRC inhibitors in serum-free liquid culture. Graphs show expression of *IRX3*, *FOXC1* and *HOXA9* following treatment with the indicated inhibitors. *indicates $p<0.005$ for comparison of *IRX3* expression in drug treated cells to DMSO controls by one way ANOVA with Fisher's least significant difference *post hoc* test.

4.12. The role of *IRX3* in acute myeloid leukaemia: Discussion

High level expression of *IRX3* is frequently observed in AML, with approximately one in three patients exhibiting significant expression of this transcription factor gene; this is an even more frequent phenomenon than the derepression of *FOXC1* in AML. Akin to *FOXC1*, high level *IRX3* expression is also observed in a significant proportion of AML stem and progenitor cells and is strongly associated with HOX gene expression. It is therefore unsurprising to find that there is also a strong association of high *FOXC1* expression with high *IRX3* expression in AML. This is also reflected in the observations that patients with high levels of *IRX3* are much more likely to exhibit mutations in *NPM1* and internal tandem duplications of *FLT3*, which are associated with high *FOXC1* and HOX gene expression. However, high level *IRX3* expression is also strongly associated with the FAB subtype M3 (acute promyelocytic leukaemia) suggesting that, in contrast to *FOXC1*, *IRX3* may have the potential to display distinct functional consequences in the absence of HOX gene expression.

With respect to normal haematopoiesis, knockdown of *IRX3* in human $CD34^+$ cells led to a significant reduction in the number of erythroid burst-forming units, implicating *IRX3* in erythropoiesis. However, *IRX3* expression was not detected in FACS-purified erythroid precursors. The potential role of *IRX3* in erythropoiesis requires further investigation.

AML is characterised by a myeloid differentiation block. Although retroviral overexpression of *IRX3* in normal HSPCs results in immunophenotypic differentiation, there is a clear functional differentiation block as evidenced by immature morphology and the extended proliferative potential of these cells. Multiple lines of evidence suggest that *IRX3* contributes to a differentiation block in transformed AML cells. This has been shown by knockdown experiments in human and murine leukaemia cell lines and by *in vivo* experiments through the co-expression of *IRX3* and *Hoxa9*. Despite the extended disease latencies of mice transplanted with *Hoxa9*/*IRX3* cells, the leukaemias that arise display a greater degree of myeloid differentiation block compared to those initiated by both Hoxa9/MTV and *Hoxa9*/*Meis1* cells. It is known that *HOXA9* overexpression induces expansion of the HSC compartment with perturbed myeloid differentiation that ultimately progresses to AML (Thorsteinsdottir et al., 2002). In the present study, leukaemias initiated by *Hoxa9*/*IRX3* cells were preceded by skewed differentiation towards the B cell lineage. Lineage skewing may arise when an upstream multi-potent cell exhibits an impaired ability to generate cells committed to a particular lineage. In this case, it may be inferred that, in the presence of *IRX3* expression, the differentiation of *HOXA9*-overexpressing HSCs towards the myeloid lineage is blocked at an early stage, skewing output towards the B cell lineage. This would result in the accumulation of immature myeloid progenitor cells which would be vulnerable to subsequent oncogenic hits that ultimately result in transformation. This may explain the development of AML with features of enhanced myeloid differentiation block that was observed.

The notion that IRX3 collaborates with HOXA9 to enhance myeloid differentiation block in AML is supported by expression data in patient samples, whereby high *IRX3* expression is almost exclusively associated with high HOX gene expression. Although the mechanism by which *IRX3* collaborates with *HOXA9* in AML remains to be discovered, other TALE family members, such as MEIS1 and PBX3, directly bind to and enhance the potency of HOXA9, resulting in enhanced AML pathogenesis (Shen et al., 1997; Thorsteinsdottir et al., 2001; Li et al., 2013). Thus, the collaboration of IRX3 with HOXA9 in AML may be owing to a shared characteristic with other members of this superclass of homeobox proteins.

IRX3 and *FOXC1* show a strong tendency to be co-expressed in human AML. When *FOXC1* and *IRX3* are expressed together in BM HSPCs there is evidence of collaboration with a transient increase in clonogenic potential observed in round 2 of the colony forming assay. However, akin to the expression of *FOXC1* alone, this effect is largely extinguished upon serial replating. A key feature of both *FOXC1*^{high} and *IRX3*^{high} (non-APL) AML cases is an intimate association with high level HOX gene expression. This would suggest that both FOXC1 and IRX3 require sustained high level HOX gene expression to exhibit their phenotypic effects in AML. The expression of *IRX3* alone has been shown to sustain *Hoxa9* expression in BM HSPCs. However, this is still insufficient to significantly enhance the replating potential of HSPCs additionally expressing *FOXC1*. This implies that the levels of *Hoxa9* expression induced by IRX3 are inadequate to sustain the phenotypic effects of high level *FOXC1* expression and/or FOXC1 requires other collaborating partners that are not sustained by *IRX3* expression alone.

An interesting, and somewhat counterintuitive observation regarding the role of *IRX3* in AML is that, whereas *in vivo* it collaborates with *Hoxa9* to induce murine leukaemias displaying a more profound differentiation block, *in vitro* it induces greater immunophenotypic differentiation. One possible explanation for this is that IRX3 may be playing opposing functional roles in blood cells at various stages of development. For example, in immature stem and progenitor cells the expression of *IRX3* may be contributing to an enhanced myeloid differentiation block. In contrast, in maturing cells the expression of *IRX3* may actually promote monocytic differentiation. Thus, the cell or origin may be a critical determinant of the functional output of *IRX3* expression. More experiments are clearly required to add proof to this speculation. However, it is interesting that the expression of *IRX3* in the context of high level *FOXC1* expression appears to have a protective role with respect to overall survival in AML. The potential ability of IRX3 to induce distinct phenotypic effects dependent upon the cellular context may be important in this regard.

The overexpression of *IRX3* in isolation is sufficient to enhance the proliferative potential of normal HSPCs as well induce a functional myeloid differentiation block. These data support the idea that aberrant *IRX3* expression plays a functional role in a myeloid cell context. However, a striking feature of the transplantation experiments reported above is that the expression of *IRX3* is sufficient to induce T cell leukaemia, with 2/6 (33%) mice succumbing to disease with a latency of 183 days. These results, coupled with the observation that *IRX3* is highly expressed in T-ALL as

well as AML (Haferlach et al., 2010), indicate that a functional role for *IRX3* in T-ALL has also hitherto gone unappreciated. Interestingly, *Irx3* has been shown to be regulated by MLL in murine fibroblasts (Schraets et al., 2003), suggesting that deregulation of MLL activity may in turn deregulate *IRX3* expression in these haematological malignancies. Given the frequent high level expression of *IRX3* in AML (and T-ALL), further studies into the biological activity of *IRX3* and the identification of collaborating partners may uncover novel therapeutic targets for the treatment of these haematological diseases.

Chapter 5. Discussion

The results outlined in this thesis report the identification and functional characterisation of two genes encoding transcriptional regulators which are frequently expressed at high levels in AML. The mesenchymal forkhead transcription factor gene *FOXC1* and the Iroquois homeobox gene *IRX3* are expressed in at least 20% and 30% of AML cases respectively. High level expression of both *FOXC1* and *IRX3* is also observed at significant frequencies in studies focusing on the AML LSC compartment, indicating the functional relevance of these transcription factors within the putative critical cellular element of this disease. The expression of *FOXC1* and *IRX3* is most strongly associated with mutations in *NPM1* and internal tandem duplications of *FLT3*, two of the most commonly occurring genetic lesions in AML. Critically, *FOXC1* and *IRX3* are either not expressed or only minimally expressed during normal haematopoiesis, indicating that their derepression in AML occurs in a tissue-inappropriate context. It was therefore crucial to demonstrate functional significance of this aberrant expression in AML. Indeed, both FOXC1 and IRX3 were found to be functionally required to maintain, and contribute to, the myeloid differentiation block that is a hallmark feature of this disease. The functional significance of the derepression of the mesenchymal gene *FOXC1* in haematopoietic tissue is particularly striking, as it was found to influence morphologic differentiation and overall survival of AML patients.

5.1. The mesenchymal transcription factor FOXC1 blocks monocyte differentiation in AML

It is clear that the frequent expression of *FOXC1* observed in AML results in significant functional consequences including enhanced monocyte/macrophage differentiation block and inferior survival. However, the mechanism through which FOXC1 induces this differentiation block remains to be determined. As already discussed, it may interfere with the function of other forkhead transcription factors that may normally regulate monocyte differentiation, for example by dominant negative activity. Out of the forkhead transcription factors that are expressed in normal human $CD14^+$ monocytes, *FOXP1* has been shown to have an essential role in normal monocyte differentiation and macrophage function *in vivo* (Shi et al., 2008). Akin to FOXC1, FOXP1 protein levels are also reduced in HL60 cells following PMA-induced monocyte differentiation, and the induction of monocyte differentiation in response to PMA is attenuated by the overexpression of *FOXP1* (Shi et al., 2004). FOXP1 has been proposed to suppress monocyte differentiation through the binding and repression of the *CSF1R* (c-fms) gene, which encodes M-CSF receptor (Shi et al., 2004). The consensus binding motifs for FOXC1 (GTAAATA) and FOXP1 (GTAAACA) are very similar, suggesting that high level expression of FOXC1 in AML may induce monocyte/macrophage differentiation block through a similar mechanism.

Alternatively, FOXC1 may directly repress expression of *KLF4*, which lies further upstream of c-fms in the regulation of monocytic differentiation (Feinberg et al., 2007). KLF4 promotes monocyte/macrophage differentiation at the expense of granulocytic differentiation, which is in complete contrast to FOXC1. Interestingly, the expression levels of *Spi1* (also known as *Sfpi1*), which encodes the master regulator PU.1 and lies upstream of *Klf4* in the regulation of monocyte differentiation (Feinberg et al., 2007), remain unchanged in murine *Hoxa9/FOXC1* AMLs by

comparison with *Hoxa9*/MTV and *Hoxa9/Meis1* AMLs. The functional experiments described above, coupled with the observations that *KLF4* expression inversely correlates with that of *FOXC1* in human AML, strongly suggest that FOXC1 functions to suppress *KLF4* and block monocyte/macrophage differentiation. Thus, the derepression of the mesenchymal transcription factor gene *FOXC1* in AML may exhibit its striking functional consequences by suppressing monocyte differentiation at multiple levels during myeloid differentiation (Figure 39). Further experiments are critical to understanding the precise mechanisms through which FOXC1 induces differentiation block in AML.

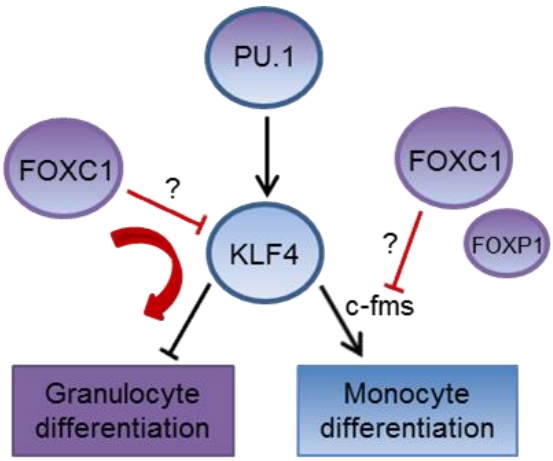


Figure 39. FOXC1 blocks monocyte differentiation in AML.
Schematic diagram represents potential mechanisms through which FOXC1 may block monocyte differentiation and promote granulocytic differentiation in AML.

5.2. FOXC1 and IRX3 collaborate with HOXA9 to enhance differentiation block in AML

The expression of both *FOXC1* and *IRX3* is strongly associated with HOX gene expression in AML, in particular *HOXA9*. *HOXA9* is overexpressed in ~50% of AML cases and is associated with poor prognosis (Golub et al., 1999; Abramovich et al., 2005). The up-regulation of HOX genes is a common pathway in leukaemic transformation, with multiple upstream genetic and epigenetic lesions resulting in the activation of these genes (Collins and Hess, 2015). It is now known that at least 30% of *HOXA9*^{high} AMLs additionally express high levels of *FOXC1* or *IRX3*, with around a quarter of cases expressing both of these genes. Moreover, the expression of either *FOXC1* or *IRX3* (with the exception of APL cases) occurs almost exclusively in the presence of HOX genes. The intimate association of high *FOXC1* and *IRX3* expression with that of *HOXA9* in human AML suggests that they collaborate to enhance leukaemogenesis. Indeed, *Hoxa9*/*FOXC1* or *Hoxa9*/*IRX3* double transduced BM stem and progenitor cells generate murine leukaemias which display an enhanced myeloid differentiation block and, with respect to the *Hoxa9/FOXC1* leukaemias, occur with an accelerated onset. The basis for the collaboration of FOXC1 and IRX3 with HOXA9 is currently unclear. However, it is known that HOX proteins rely upon transcription factors termed cofactors and collaborators to modulate their binding affinity and specificity, enabling their diverse and tissue-specific functions (Mann et al., 2009). HOXA9 itself binds

predominantly at enhancer regions in haematopoietic cells, mediating its transcriptional effects by interacting with “enhanceosomes” composed of HOX cofactors, such as MEIS1, and lineage specific transcription factors, such as CEBPA and RUNX1 (Huang et al., 2012). Thus, it may be that the tissue-inappropriate expression of *FOXC1* and/or *IRX3* enables these transcription factors to act as ‘out of context’ collaborators, conferring upon HOXA9 an alternative activity, or perhaps misdirecting its binding to distinct enhancer regions. In support of this hypothesis are the contrasting transcriptomes of the *Hoxa9*/*FOXC1* leukaemias in comparison to the *Hoxa9*/MTV and *Hoxa9*/*Meis1* leukaemias, suggesting that HOXA9 regulates a distinct set of genes in the presence of FOXC1. Indeed, FOXC1 and HOXA9 are capable of DNA-independent physical interaction when co-expressed in HEK 293FT cells.

It is interesting that *IRX3* is a member of the TALE superclass of homeobox genes, which also encode well established HOX cofactors, such as PBX and MEIS proteins (Ladam et al., 2014; Longobardi et al., 2014). With respect to AML, the role of MEIS1 as a critical cofactor of HOXA9 is well established (Kroon et al., 1998; Thorsteinsdottir et al., 2002). More recently, the TALE protein PBX3 has additionally been identified as a critical cofactor of HOXA9 in AML (Li et al., 2013), a function it may execute through interaction with, and stabilisation of, MEIS1 (Garcia-Cuellar et al., 2015). Such studies have led to the identification of novel therapeutic strategies for the potential targeting of particular subsets of AML. For example, the small molecule inhibitor HXR9 has been shown to target the HOX/PBX interaction and impair the growth of leukaemia cells with elevated levels of *HOXA* and *PBX3* gene expression (Li et al., 2013; Morgan et al., 2007). It is interesting that *MEIS1* and *PBX3*, in contrast to *IRX3*, are both highly expressed in normal HSPCs, perhaps reflecting their roles as HOX cofactors in normal haematopoiesis. Thus, the ability of IRX3 to exacerbate the myeloid differentiation block of HOXA9-induced leukaemias represents the first report of a HOX/TALE collaboration that is unique to malignant haematopoiesis. Whether IRX3 interacts with HOXA9 to act as a *bona fide* HOXA9 cofactor in AML is currently unknown. However, further investigation into the potential interaction between HOXA9 and IRX3, or indeed between HOXA9 and FOXC1, may lead to the discovery novel therapeutic targets in AML.

5.3. Loss of Polycomb-mediated repression occurs at specific loci in AML

The continued repression of both *FOXC1* and *IRX3* expression in normal HSPCs is dependent upon the activity of Polycomb, particularly PRC2, as treatment of normal $CD34^+$ cells with a dual EZH1/2 inhibitor results in their up-regulation. However, it has also been shown that there is no generalised failure of suppression of Polycomb marked genes in AML. On the contrary, loss of Polycomb activity was found to occur significantly only at specific loci in this disease, namely at *FOXC1*, *IRX3*, *IRX5* and *HOXB8*. Furthermore, the derepression of *FOXC1* in AML was found not to be driven by genetic mutation of Polycomb components or local regulatory regions. Although the mechanisms underlying loss of Polycomb repression specifically at these loci in AML require further investigation, they do suggest these genes are somehow targeted for derepression. Interestingly, a recent study identified the up-regulation of *FOXC1* and *IRX3* as part of a common

gene expression signature associated with FLT3-ITD$^+$ AML (Cauchy et al., 2015). This gene expression signature was also associated with a common open chromatin signature mapped by DNase hypersensitivity sites (DHSs). The proximity of these DHSs positively correlated with the expression of genes in this signature and they were enriched for specific transcription factor binding motifs, including those of RUNX1 and CEBPA. Given the known association of HOXA9 with these lineage specific transcription factors in myeloid cells (Huang et al., 2012), this study lends support to the hypothesis that HOXA9 may play a role in the derepression of *FOXC1* and *IRX3* in AML.

Another recent study has also described a mechanism through which *IRX3* and *IRX5* become derepressed in mesenchymal adipocyte precursor cells and functionally contribute to an increased risk of obesity in humans (Claussnitzer et al., 2015). A single nucleotide variant (SNV) associated with increased obesity risk, found within an intron of the FTO gene which lies upstream of *IRX3* and *IRX5* on chromosome 16, was shown to disrupt a highly conserved ARID5B binding motif. Claussnitzer et al (2015) elegantly show that this SNV results in loss of ARID5B-mediated repression of *IRX3* and *IRX5* during early adipocyte differentiation. The ability of somatically-acquired mutations in non-coding regulatory regions to introduce or interfere with enhancer activity and drive expression of distal oncogenes has previously been reported in T-ALL (Mansour et al., 2014). Although such a mechanism was ruled out for the derepression of *FOXC1*, whether such a mechanism results in the derepression of *IRX3* (and *IRX5*) in AML warrants further investigation.

As already alluded to, the strong association of HOX gene expression with that of *FOXC1* and *IRX3* may functionally implicate HOX proteins in the derepression of these genes. Interestingly, the forced expression of *HOXA9*/*HOXA10* in human CD34$^+$ cord blood cells has been shown to up-regulate *IRX3* expression (Ferrell et al., 2005). Moreover, *Irx5* has been shown to be positively regulated and an immediate downstream target of Hoxb4 in early vertebrate development (Theokli et al., 2003), which is a potent enhancer of HSC self-renewal (Antonchuk et al., 2001; Buske et al., 2002). Thus, HOX gene expression may contribute to the continued aberrant expression of *IRX3* and *IRX5* in human AML. However, any potential regulation of *IRX3/5* expression by HOXA/B proteins must be dependent upon the cellular context, as HOXA/B genes are expressed in normal HSPCs whereas the expression of *IRX3/5* is only observed in leukaemic haematopoiesis.

In addition to the striking association of *FOXC1* and *IRX3* expression with that of *HOXA9* in AML, the expression of *FOXC1* and *IRX3* also show a strong positive correlation. Again, this implies a mechanistic link concerning the derepression of these genes and/or functional collaboration between these transcription factors in the expansion of the leukaemic clone. In support of the latter is the observation that the expression of *IRX3* in FOXC1$^+$ BM HSPCs enhances their serial replating capacity. However, this effect was only modest and transient and therefore suggests that additional factors are required for the sustained phenotypic effects of FOXC1 and IRX3 in AML.

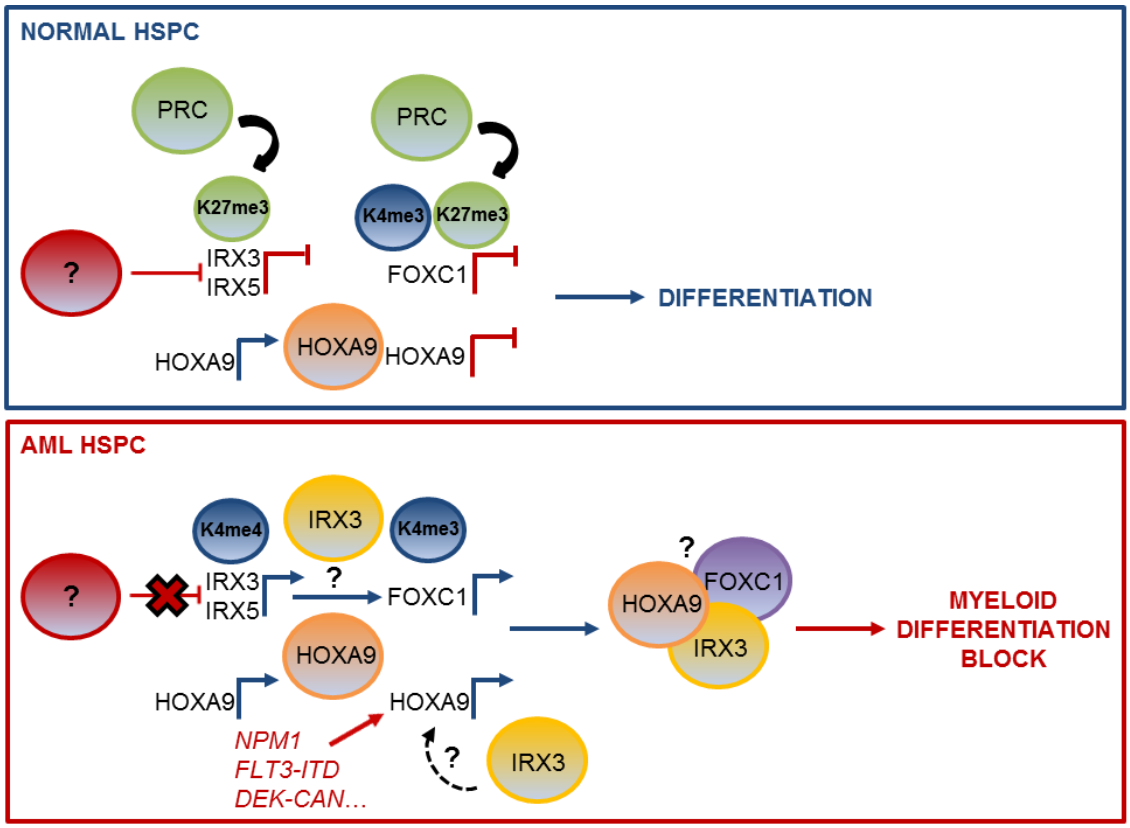


Figure 40. Functional derepression of *IRX3* and *FOXC1* in association with HOXA9 in AML. Schematic diagram represents potential mechanisms of derepression of *IRX3* and *FOXC1* in AML. In normal HSPCs (top panel) *IRX3* and *FOXC1* are bound and their expression repressed by Polycomb repressive complexes (PRC). An unknown repressor (red shaded circle) may contribute to the repression of *IRX3* and *IRX5*. *HOXA9* expression is down-regulated during differentiation. In AML HSPCs (bottom panel) loss of Polycomb-mediated repression is associated with the inappropriate expression of *IRX3* and *FOXC1*. Loss of repressor activity may contribute to the derepression of *IRX3* and *IRX5*. *IRX3* may in turn contribute to derepression of *FOXC1*, possibly through collaboration with HOXA9. High level HOX gene expression is maintained through multiple mechanisms (e.g. mutations in *NPM1* and *FLT3-ITD*) and IRX3 may help sustain *HOXA9* expression. FOXC1 and IRX3 collaborate with HOXA9 to exacerbate the differentiation block in AML. K27me3, H3K27me3; K4Me3, H3K4me3.

Given the intimate expression of these transcription factors with HOX genes in AML, it will be interesting to assess the functional effects of sustained expression of *HOXA9*, *FOXC1* and *IRX3* in BM HSPCs, particularly *in vivo* to ascertain whether this further accelerates and exacerbates the onset of leukaemia in mice. An interesting hypothesis pertaining to the mechanism of derepression of *FOXC1* in AML is that *IRX3* expression may be permissive for that of *FOXC1*. Interestingly, TALE proteins have been shown to recruit histone modifying enzymes to target gene loci during embryogenesis, poising them for activation prior to transcription initiation. This is associated with reduced H3K27me3 and increased H3K27ac at promoter regions and the sequential binding of HOX proteins results in transcriptional activation (Choe et al., 2014). It would be therefore interesting to determine whether IRX3 is capable of inducing such epigenetic changes at *FOXC1* regulatory regions and whether subsequent forced expression of *HOXA9* is sufficient to induce sustained, high level *FOXC1* expression. Figure 40 summarises the potential mechanisms of *FOXC1* and *IRX3* derepression in AML.

5.4. Conclusions

The aims of the studies outlined above were to firstly identify novel transcriptional regulators that are significantly expressed in AML LSCs but not normal HSPCs, hence indicative of a specific functional requirement in leukaemic versus normal haematopoiesis. This analysis initially resulted in the identification of the mesenchymal transcription factor gene *FOXC1*, and subsequently the homeobox gene *IRX3*, as potential novel transcriptional regulators in AML. Further analysis of large, carefully curated expression studies (Wouters et al., 2009; Cancer Genome Atlas Research Network, 2013), as well as our own primary AML patient samples, confirmed high level expression of *FOXC1* and *IRX3* in ~20% and ~30% of AML cases respectively. Secondly, following the identification of candidate genes, a variety of functional studies, including knockdown experiments in cell lines, retroviral overexpression in mouse BM HSPCs and transplantation experiments, highlighted important functional contributions of FOXC1 and IRX3 in AML. Bioinformatics analyses helped to decipher the role of the mesenchymal transcriptional regulator FOXC1 as a critical factor in maintaining monocyte differentiation block in human AML. Moreover, the expression of *FOXC1* and *IRX3* is strongly associated with HOX gene expression in AML and both genes have been shown to collaborate with HOXA9 to enhance myeloid differentiation block in murine leukaemias. Finally, the continued repression of *FOXC1* and *IRX3* in normal HSPCs was shown to be dependent upon Polycomb activity and locus-specific loss of this activity was identified as a mechanism that contributes to the tissue-inappropriate derepression of these transcriptional regulators to functional effect in AML.

List of References

Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., Malinge, S., Yao, J., Kilpivaara, O., Bhat, R.*, et al.* (2009). Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood *114*, 144-147.

Abdel-Wahab, O., Adli, M., LaFave, L. M., Gao, J., Hricik, T., Shih, A. H., Pandey, S., Patel, J. P., Chung, Y. R., Koche, R.*, et al.* (2012). ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. Cancer Cell *22*, 180-193.

Abramovich, C., and Humphries, R. K. (2005). Hox regulation of normal and leukemic hematopoietic stem cells. Curr Opin Hematol *12*, 210-216.

Adli, M., Zhu, J., and Bernstein, B. E. (2010). Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. Nat Methods *7*, 615-618.

Adolfsson, J., Borge, O. J., Bryder, D., Theilgaard-Mönch, K., Astrand-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.*, 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.

Akalin, A., Garrett-Bakelman, F. E., Kormaksson, M., Busuttil, J., Zhang, L., Khrebtukova, I., Milne, T. A., Huang, Y., Biswas, D., Hess, J. L.*, et al.* (2012). Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. PLoS Genet *8*, e1002781.

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.

Anjos-Afonso, F., Currie, E., Palmer, H. G., Foster, K. E., Taussig, D. C., and Bonnet, D. (2013). CD34(-) cells at the apex of the human hematopoietic stem cell hierarchy have distinctive cellular and molecular signatures. Cell Stem Cell *13*, 161-174.

Antonchuk, J., Sauvageau, G., and Humphries, R. K. (2001). HOXB4 overexpression mediates very rapid stem cell regeneration and competitive hematopoietic repopulation. Exp Hematol *29*, 1125-1134.

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

Argiropoulos, B., Yung, E., and Humphries, R. K. (2007). Unraveling the crucial roles of Meis1 in leukemogenesis and normal hematopoiesis. Genes Dev *21*, 2845-2849.

Arthur, D. C., Berger, R., Golomb, H. M., Swansbury, G. J., Reeves, B. R., Alimena, G., Van Den Berghe, H., Bloomfield, C. D., de la Chapelle, A., and Dewald, G. W. (1989). The clinical significance of karyotype in acute myelogenous leukemia. Cancer Genet Cytogenet *40*, 203-216.

Austin, H., Delzell, E., and Cole, P. (1988). Benzene and leukemia. A review of the literature and a risk assessment. Am J Epidemiol *127*, 419-439.

Ayton, P. M., and Cleary, M. L. (2003). Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9. Genes Dev *17*, 2298-2307.

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

Bach, C., Buhl, S., Mueller, D., Garcia-Cuellar, M. P., Maethner, E., and Slany, R. K. (2010). Leukemogenic transformation by HOXA cluster genes. Blood 115, 2910-2918.

Bannister, A. J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res *21*, 381-395.

Barabé, F., Kennedy, J. A., Hope, K. J., and Dick, J. E. (2007). Modeling the initiation and progression of human acute leukemia in mice. Science *316*, 600-604.

Baum, C. M., Weissman, I. L., Tsukamoto, A. S., Buckle, A. M., and Peault, B. (1992). Isolation of a candidate human hematopoietic stem-cell population. Proc Natl Acad Sci U S A *89*, 2804-2808.

Baylin, S. B., and Jones, P. A. (2011). A decade of exploring the cancer epigenome - biological and translational implications. Nat Rev Cancer *11*, 726-734.

Becker, A. J., McCulloch, E. A., and Till, J. E. (1963). Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. Nature *197*, 452-454.

Bedi, A., Zehnbauer, B. A., Collector, M. I., Barber, J. P., Zicha, M. S., Sharkis, S. J., and Jones, R. J. (1993). BCR-ABL gene rearrangement and expression of primitive hematopoietic progenitors in chronic myeloid leukemia. Blood *81*, 2898-2902.

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

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R., and Sultan, C. (1985a). Criteria for the diagnosis of acute leukemia of megakaryocyte lineage (M7). A report of the French-American-British Cooperative Group. Ann Intern Med *103*, 460-462.

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R., and Sultan, C. (1985b). Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French-American-British Cooperative Group. Ann Intern Med *103*, 620-625.

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

Benveniste, P., Frelin, C., Janmohamed, S., Barbara, M., Herrington, R., Hyam, D., and Iscove, N. N. (2010). Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential. Cell Stem Cell *6*, 48-58.

Berenson, R. J., Bensinger, W. I., Hill, R. S., Andrews, R. G., Garcia-Lopez, J., Kalamasz, D. F., Still, B. J., Spitzer, G., Buckner, C. D., and Bernstein, I. D. (1991). Engraftment after infusion of CD34+ marrow cells in patients with breast cancer or neuroblastoma. Blood *77*, 1717-1722.

Berger, S. L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. Genes Dev *23*, 781-783.

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.

Bernt, K. M., Zhu, N., Sinha, A. U., Vempati, S., Faber, J., Krivtsov, A. V., Feng, Z., Punt, N., Daigle, A., Bullinger, L.*, et al.* (2011). MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. Cancer Cell *20*, 66-78.

Bhatia, M., Bonnet, D., Murdoch, B., Gan, O. I., and Dick, J. E. (1998). A newly discovered class of human hematopoietic cells with SCID-repopulating activity. Nat Med *4*, 1038-1045.

Bhatia, M., Wang, J. C., Kapp, U., Bonnet, D., and Dick, J. E. (1997). Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. Proc Natl Acad Sci U S A *94*, 5320-5325.

Bhatia, R., Holtz, M., Niu, N., Gray, R., Snyder, D. S., Sawyers, C. L., Arber, D. A., Slovak, M. L., and Forman, S. J. (2003). Persistence of malignant hematopoietic progenitors in chronic myelogenous leukemia patients in complete cytogenetic remission following imatinib mesylate treatment. Blood *101*, 4701-4707.

Blatt, C., Aberdam, D., Schwartz, R., and Sachs, L. (1988). DNA rearrangement of a homeobox gene in myeloid leukaemic cells. EMBO J *7*, 4283-4290.

Bloushtain-Qimron, N., Yao, J., Snyder, E. L., Shipitsin, M., Campbell, L. L., Mani, S. A., Hu, M., Chen, H., Ustyansky, V., Antosiewicz, J. E., et al. (2008). Cell type-specific DNA methylation patterns in the human breast. Proc Natl Acad Sci U S A *105*, 14076-14081.

Bonnard, C., Strobl, A. C., Shboul, M., Lee, H., Merriman, B., Nelson, S. F., Ababneh, O. H., Uz, E., Guran, T., Kayserili, H., et al. (2012). Mutations in IRX5 impair craniofacial development and germ cell migration via SDF1. Nat Genet *44*, 709-713.

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.

Bosma, G. C., Custer, R. P., and Bosma, M. J. (1983). A severe combined immunodeficiency mutation in the mouse. Nature *301*, 527-530.

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.

Branco, M. R., Ficz, G., and Reik, W. (2012). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat Rev Genet *13*, 7-13.

Briscoe, J., Pierani, A., Jessell, T. M., and Ericson, J. (2000). A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. Cell *101*, 435-445.

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.

Bullinger, L., Ehrich, M., Döhner, K., Schlenk, R. F., Döhner, H., Nelson, M. R., and van den Boom, D. (2010). Quantitative DNA methylation predicts survival in adult acute myeloid leukemia. Blood *115*, 636-642.

Burnett, A. K., Milligan, D., Goldstone, A., Prentice, A., McMullin, M. F., Dennis, M., Sellwood, E., Pallis, M., Russell, N., Hills, R. K.*, et al.* (2009). The impact of dose escalation and resistance modulation in older patients with acute myeloid leukaemia and high risk myelodysplastic syndrome: the results of the LRF AML14 trial. Br J Haematol *145*, 318-332.

Burnett, A., Wetzler, M., and Löwenberg, B. (2011a). Therapeutic advances in acute myeloid leukemia. J Clin Oncol *29*, 487-494.

Burnett, A. K., Hills, R. K., Milligan, D., Kjeldsen, L., Kell, J., Russell, N. H., Yin, J. A., Hunter, A., Goldstone, A. H., and Wheatley, K. (2011b). Identification of patients with acute myeloblastic leukemia who benefit from the addition of gemtuzumab ozogamicin: results of the MRC AML15 trial. J Clin Oncol *29*, 369-377.

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S. M., Reth, M., Höfer, T., and Rodewald, H. R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature *518*, 542-546.

Buske, C., Feuring-Buske, M., Abramovich, C., Spiekermann, K., Eaves, C. J., Coulombel, L., Sauvageau, G., Hogge, D. E., and Humphries, R. K. (2002). Deregulated expression of HOXB4 enhances the primitive growth activity of human hematopoietic cells. Blood *100*, 862-868.

Cancer Genome Atlas Research Network. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 368, 2059-2074.

Cancer Research UK. (2015) Acute myeloid leukaemia (AML) incidence statistics. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-aml/incidence#heading-Zero. [Accessed: 17th June 2015].

Cao, R., and Zhang, Y. (2004a). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol Cell *15*, 57-67.

Cao, R., and Zhang, Y. (2004b). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. Curr Opin Genet Dev *14*, 155-164.

Castaigne, S., Pautas, C., Terré, C., Raffoux, E., Bordessoule, D., Bastie, J. N., Legrand, O., Thomas, X., Turlure, P., Reman, O.*, et al.* (2012). Effect of gemtuzumab ozogamicin on survival of adult patients with de-novo acute myeloid leukaemia (ALFA-0701): a randomised, open-label, phase 3 study. Lancet *379*, 1508-1516.

Castilla, L. H., Perrat, P., Martinez, N. J., Landrette, S. F., Keys, R., Oikemus, S., Flanegan, J., Heilman, S., Garrett, L., Dutra, A.*, et al.* (2004). Identification of genes that synergize with Cbfb-MYH11 in the pathogenesis of acute myeloid leukemia. Proc Natl Acad Sci U S A *101*, 4924-4929.

Catlin, S. N., Busque, L., Gale, R. E., Guttorp, P., and Abkowitz, J. L. (2011). The replication rate of human hematopoietic stem cells in vivo. Blood *117*, 4460-4466.

Cauchy, P., James, S. R., Zacarias-Cabeza, J., Ptasinska, A., Imperato, M. R., Assi, S. A., Piper, J., Canestraro, M., Hoogenkamp, M., Raghavan, M., et al. (2015). Chronic FLT3-ITD Signaling in Acute Myeloid Leukemia Is Connected to a Specific Chromatin Signature. Cell Rep *12*, 821-836.

Cavodeassi, F., Modolell, J., and Gomez-Skarmeta, J. L. (2001). The Iroquois family of genes: from body building to neural patterning. Development *128*, 2847-2855.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet *10*, 295-304.

Cedar, H., and Bergman, Y. (2011). Epigenetics of haematopoietic cell development. Nat Rev Immunol *11*, 478-488.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2, 401-404.

Challen, G. A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J. S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y.*, et al.* (2012). Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet *44*, 23-31.

Choe, S. K., Ladam, F., and Sagerstrom, C. G. (2014). TALE factors poise promoters for activation by Hox proteins. Dev Cell *28*, 203-211.

Chou, W. C., Chen, C. Y., Hou, H. A., Lin, L. I., Tang, J. L., Yao, M., Tsay, W., Ko, B. S., Wu, S. J., Huang, S. Y.*, et al.* (2009). Acute myeloid leukemia bearing t(7;11)(p15;p15) is a distinct cytogenetic entity with poor outcome and a distinct mutation profile: comparative analysis of 493 adult patients. Leukemia *23*, 1303-1310.

Chowdhury, M., Mihara, K., Yasunaga, S., Ohtaki, M., Takihara, Y., and Kimura, A. (2007). Expression of Polycomb-group (PcG) protein BMI-1 predicts prognosis in patients with acute myeloid leukemia. Leukemia *21*, 1116-1122.

Christoffels, V. M., Keijser, A. G., Houweling, A. C., Clout, D. E., and Moorman, A. F. (2000). Patterning the embryonic heart: identification of five mouse Iroquois homeobox genes in the developing heart. Dev Biol *224*, 263-274.

Chu, S., McDonald, T., Lin, A., Chakraborty, S., Huang, Q., Snyder, D. S., and Bhatia, R. (2011). Persistence of leukemia stem cells in chronic myelogenous leukemia patients in prolonged remission with imatinib treatment. Blood *118*, 5565-5572.

Civin, C. I., Strauss, L. C., Brovall, C., Fackler, M. J., Schwartz, J. F., and Shaper, J. H. (1984). Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. J Immunol *133*, 157-165.

Claussnitzer, M., Dankel, S. N., Kim, K. H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I. S., Beaudry, J. L., Puviindran, V., et al. (2015). FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N Engl J Med.

Cohen, D. R., Cheng, C. W., Cheng, S. H., and Hui, C. C. (2000). Expression of two novel mouse Iroquois homeobox genes during neurogenesis. Mech Dev *91*, 317-321.

Collins, C. T., and Hess, J. L. (2015). Role of HOXA9 in leukemia: dysregulation, cofactors and essential targets. Oncogene. 1-9.

Coombs, C. C., Tavakkoli, M., and Tallman, M. S. (2015). Acute promyelocytic leukemia: where did we start, where are we now, and the future. Blood Cancer J *5*, e304.

Copley, M. R., Beer, P. A., and Eaves, C. J. (2012). Hematopoietic stem cell heterogeneity takes center stage. Cell Stem Cell *10*, 690-697.

Corces-Zimmerman, M. R., Hong, W. J., Weissman, I. L., Medeiros, B. C., and Majeti, R. (2014). Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. Proc Natl Acad Sci U S A *111*, 2548-2553.

Cornelissen, J. J., Gratwohl, A., Schlenk, R. F., Sierra, J., Bornhäuser, M., Juliusson, G., Råcil, Z., Rowe, J. M., Russell, N., Mohty, M.*, et al.* (2012). The European LeukemiaNet AML Working Party consensus statement on allogeneic HSCT for patients with AML in remission: an integrated-risk adapted approach. Nat Rev Clin Oncol *9*, 579-590.

Cortes, J., O'Brien, S., and Kantarjian, H. (2004). Discontinuation of imatinib therapy after achieving a molecular response. Blood *104*, 2204-2205.

Costantini, D. L., Arruda, E. P., Agarwal, P., Kim, K. H., Zhu, Y., Zhu, W., Lebel, M., Cheng, C. W., Park, C. Y., Pierce, S. A., et al. (2005). The homeodomain transcription factor Irx5 establishes the mouse cardiac ventricular repolarization gradient. Cell *123*, 347-358.

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.

Crowther, D., Bateman, C. J., Vartan, C. P., Whitehouse, J. M., Malpas, J. S., Fairley, G. H., and Scott, R. B. (1970). Combination chemotherapy using L-asparaginase, daunorubicin, and cytosine arabinoside in adults with acute myelogenous leukaemia. Br Med J *4*, 513-517.

Cui, K., Zang, C., Roh, T. Y., Schones, D. E., Childs, R. W., Peng, W., and Zhao, K. (2009). Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell *4*, 80-93.

Dahl, R., Iyer, S. R., Owens, K. S., Cuylear, D. D., and Simon, M. C. (2007). The transcriptional repressor GFI-1 antagonizes PU.1 activity through protein-protein interaction. J Biol Chem *282*, 6473-6483.

Daigle, S. R., Olhava, E. J., Therkelsen, C. A., Majer, C. R., Sneeringer, C. J., Song, J., Johnston, L. D., Scott, M. P., Smith, J. J., Xiao, Y.*, et al.* (2011). Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. Cancer Cell *20*, 53-65.

Danet, G. H., Luongo, J. L., Butler, G., Lu, M. M., Tenner, A. J., Simon, M. C., and Bonnet, D. A. (2002). C1qRp defines a new human stem cell population with hematopoietic and hepatic potential. Proc Natl Acad Sci U S A *99*, 10441-10445.

Dawson, M. A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. Cell *150*, 12-27.

de Thé, H., and Chen, Z. (2010). Acute promyelocytic leukaemia: novel insights into the mechanisms of cure. Nat Rev Cancer *10*, 775-783.

Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Massé, A., Kosmider, O., Le Couedic, J. P., Robert, F., Alberdi, A.*, et al.* (2009). Mutation in TET2 in myeloid cancers. N Engl J Med *360*, 2289-2301.

Derolf, A. R., Kristinsson, S. Y., Andersson, T. M., Landgren, O., Dickman, P. W., and Björkholm, M. (2009). Improved patient survival for acute myeloid leukemia: a population-based study of 9729 patients diagnosed in Sweden between 1973 and 2005. Blood *113*, 3666-3672.

Deschler, B., and Lübbert, M. (2006). Acute myeloid leukemia: epidemiology and etiology. Cancer *107*, 2099-2107.

Di Croce, L., Raker, V. A., Corsaro, M., Fazi, F., Fanelli, M., Faretta, M., Fuks, F., Lo Coco, F., Kouzarides, T., Nervi, C.*, et al.* (2002). Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. Science *295*, 1079-1082.

Dick, J. E. (2008). Stem cell concepts renew cancer research. Blood *112*, 4793-4807.

Dick, J. E., and Lapidot, T. (2005). Biology of normal and acute myeloid leukemia stem cells. Int J Hematol *82*, 389-396.

Dicke, K. A., Spitzer, G., and Ahearn, M. J. (1976). Colony formation in vitro by leukaemic cells in acute myelogenous leukaemia with phytohaemagglutinin as stimulating factor. Nature *259*, 129-130.

Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., Ritchey, J. K., Young, M. A., Lamprecht, T., McLellan, M. D.*, et al.* (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature *481*, 506-510.

Döhner, H., Estey, E. H., Amadori, S., Appelbaum, F. R., Büchner, T., Burnett, A. K., Dombret, H., Fenaux, P., Grimwade, D., Larson, R. A.*, et al.* (2010). Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood *115*, 453-474.

Dou, Y., Milne, T. A., Tackett, A. J., Smith, E. R., Fukuda, A., Wysocka, J., Allis, C. D., Chait, B. T., Hess, J. L., and Roeder, R. G. (2005). Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. Cell *121*, 873-885.

Douer, D., and Tallman, M. S. (2005). Arsenic trioxide: new clinical experience with an old medication in hematologic malignancies. J Clin Oncol *23*, 2396-2410.

Doulatov, S., Notta, F., Laurenti, E., and Dick, J. E. (2012). Hematopoiesis: a human perspective. Cell Stem Cell *10*, 120-136.

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

Dunbar, C. E., Cottler-Fox, M., O'Shaughnessy, J. A., Doren, S., Carter, C., Berenson, R., Brown, S., Moen, R. C., Greenblatt, J., and Stewart, F. M. (1995). Retrovirally marked CD34-enriched peripheral blood and bone marrow cells contribute to long-term engraftment after autologous transplantation. Blood *85*, 3048-3057.

Eaves, C. J. (2015). Hematopoietic stem cells: concepts, definitions, and the new reality. Blood *125*, 2605-2613.

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.

Ernst, T., Chase, A. J., Score, J., Hidalgo-Curtis, C. E., Bryant, C., Jones, A. V., Waghorn, K., Zoi, K., Ross, F. M., Reiter, A.*, et al.* (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet *42*, 722-726.

Estey, E. H. (2014). Acute myeloid leukemia: 2014 update on risk-stratification and management. Am J Hematol *89*, 1063-1081.

Faber, J., Krivtsov, A. V., Stubbs, M. C., Wright, R., Davis, T. N., van den Heuvel-Eibrink, M., Zwaan, C. M., Kung, A. L., and Armstrong, S. A. (2009). HOXA9 is required for survival in human MLL-rearranged acute leukemias. Blood *113*, 2375-2385.

Fathi, A. T., and Abdel-Wahab, O. (2012). Mutations in epigenetic modifiers in myeloid malignancies and the prospect of novel epigenetic-targeted therapy. Adv Hematol *2012*, 469592.

Feinberg, A. P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature *301*, 89-92.

Feinberg, M. W., Wara, A. K., Cao, Z., Lebedeva, M. A., Rosenbauer, F., Iwasaki, H., Hirai, H., Katz, J. P., Haspel, R. L., Gray, S., et al. (2007). The Kruppel-like factor KLF4 is a critical regulator of monocyte differentiation. EMBO J *26*, 4138-4148.

Fenaux, P., Preudhomme, C., Laï, J. L., Morel, P., Beuscart, R., and Bauters, F. (1989). Cytogenetics and their prognostic value in de novo acute myeloid leukaemia: a report on 283 cases. Br J Haematol *73*, 61-67.

Fenaux, P., Mufti, G. J., Hellström-Lindberg, E., Santini, V., Gattermann, N., Germing, U., Sanz, G., List, A. F., Gore, S., Seymour, J. F.*, et al.* (2010). Azacitidine prolongs overall survival compared with conventional care regimens in elderly patients with low bone marrow blast count acute myeloid leukemia. J Clin Oncol *28*, 562-569.

Ferrell, C. M., Dorsam, S. T., Ohta, H., Humphries, R. K., Derynck, M. K., Haqq, C., Largman, C., and Lawrence, H. J. (2005). Activation of stem-cell specific genes by HOXA9 and HOXA10 homeodomain proteins in CD34+ human cord blood cells. Stem Cells *23*, 644-655.

Fialkow, P. J., Gartler, S. M., and Yoshida, A. (1967). Clonal origin of chronic myelocytic leukemia in man. Proc Natl Acad Sci U S A *58*, 1468-1471.

Fialkow, P. J., Jacobson, R. J., and Papayannopoulou, T. (1977). Chronic myelocytic leukemia: clonal origin in a stem cell common to the granulocyte, erythrocyte, platelet and monocyte/macrophage. Am J Med *63*, 125-130.

Figueroa, M. E., Lugthart, S., Li, Y., Erpelinck-Verschueren, C., Deng, X., Christos, P. J., Schifano, E., Booth, J., van Putten, W., Skrabanek, L.*, et al.* (2010a). DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer Cell *17*, 13-27.

Figueroa, M. E., Abdel-Wahab, O., Lu, C., Ward, P. S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H. F.*, et al.* (2010b). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell *18*, 553-567.

Gaborit, N., Sakuma, R., Wylie, J. N., Kim, K. H., Zhang, S. S., Hui, C. C., and Bruneau, B. G. (2012). Cooperative and antagonistic roles for Irx3 and Irx5 in cardiac morphogenesis and postnatal physiology. Development *139*, 4007-4019.

Garcia-Cuellar, M. P., Steger, J., Fuller, E., Hetzner, K., and Slany, R. K. (2015). Pbx3 and Meis1 cooperate through multiple mechanisms to support Hox-induced murine leukemia. Haematologica *100*, 905-913.

Gardner, K. E., Allis, C. D., and Strahl, B. D. (2011). Operating on chromatin, a colorful language where context matters. J Mol Biol *409*, 36-46.

Gentles, A. J., Plevritis, S. K., Majeti, R., and Alizadeh, A. A. (2010). Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. JAMA *304*, 2706-2715.

Gerber, J. M., Smith, B. D., Ngwang, B., Zhang, H., Vala, M. S., Morsberger, L., Galkin, S., Collector, M. I., Perkins, B., Levis, M. J.*, et al.* (2012). A clinically relevant population of leukemic CD34(+)CD38(-) cells in acute myeloid leukemia. Blood *119*, 3571-3577.

Gibbs, K. D., Jr., Jager, A., Crespo, O., Goltsev, Y., Trejo, A., Richard, C. E., and Nolan, G. P. (2012). Decoupling of tumor-initiating activity from stable immunophenotype in HoxA9-Meis1-driven AML. Cell Stem Cell *10*, 210-217.

Gilliland, D. G., and Griffin, J. D. (2002). The roles of FLT3 in hematopoiesis and leukemia. Blood *100*, 1532-1542.

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.

Godley, L. A., and Larson, R. A. (2008). Therapy-related myeloid leukemia. Semin Oncol *35*, 418-429.

Goff, D. J., Court Recart, A., Sadarangani, A., Chun, H. J., Barrett, C. L., Krajewska, M., Leu, H., Low-Marchelli, J., Ma, W., Shih, A. Y.*, et al.* (2013). A Pan-BCL2 inhibitor renders bone-marrow-resident human leukemia stem cells sensitive to tyrosine kinase inhibition. Cell Stem Cell *12*, 316-328.

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

Goyal, S., and Zandstra, P. W. (2015). Stem cells: Chasing blood. Nature *518*, 488-490.

Graham, S. M., Jørgensen, H. G., Allan, E., Pearson, C., Alcorn, M. J., Richmond, L., and Holyoake, T. L. (2002). Primitive, quiescent, Philadelphia-positive stem cells from patients with chronic myeloid leukemia are insensitive to STI571 in vitro. Blood *99*, 319-325.

Griffin, J. D., and Löwenberg, B. (1986). Clonogenic cells in acute myeloblastic leukemia. Blood *68*, 1185-1195.

Grignani, F., Valtieri, M., Gabbianelli, M., Gelmetti, V., Botta, R., Luchetti, L., Masella, B., Morsilli, O., Pelosi, E., Samoggia, P.*, et al.* (2000). PML/RAR alpha fusion protein expression in normal human hematopoietic progenitors dictates myeloid commitment and the promyelocytic phenotype. Blood *96*, 1531-1537.

Grimwade, D., Walker, H., Oliver, F., Wheatley, K., Harrison, C., Harrison, G., Rees, J., Hann, I., Stevens, R., Burnett, A., and Goldstone, A. (1998). The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. Blood *92*, 2322-2333.

Grimwade, D., Hills, R. K., Moorman, A. V., Walker, H., Chatters, S., Goldstone, A. H., Wheatley, K., Harrison, C. J., Burnett, A. K., and Group, N. C. R. I. A. L. W. (2010). Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood *116*, 354-365.

Guan, Y., Gerhard, B., and Hogge, D. E. (2003). Detection, isolation, and stimulation of quiescent primitive leukemic progenitor cells from patients with acute myeloid leukemia (AML). Blood *101*, 3142-3149.

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.

Guzman, M. L., Neering, S. J., Upchurch, D., Grimes, B., Howard, D. S., Rizzieri, D. A., Luger, S. M., and Jordan, C. T. (2001). Nuclear factor-kappaB is constitutively activated in primitive human acute myelogenous leukemia cells. Blood *98*, 2301-2307.

Guzman, M. L., Swiderski, C. F., Howard, D. S., Grimes, B. A., Rossi, R. M., Szilvassy, S. J., and Jordan, C. T. (2002). Preferential induction of apoptosis for primary human leukemic stem cells. Proc Natl Acad Sci U S A *99*, 16220-16225.

Guzman, M. L., Rossi, R. M., Karnischky, L., Li, X., Peterson, D. R., Howard, D. S., and Jordan, C. T. (2005). The sesquiterpene lactone parthenolide induces apoptosis of human acute myelogenous leukemia stem and progenitor cells. Blood *105*, 4163-4169.

Haferlach, T., Kohlmann, A., Wieczorek, L., Basso, G., Kronnie, G. T., Bene, M. C., De Vos, J., Hernandez, J. M., Hofmann, W. K., Mills, K. I., et al. (2010). Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group. J Clin Oncol *28*, 2529-2537.

Hamamy, H. A., Teebi, A. S., Oudjhane, K., Shegem, N. N., and Ajlouni, K. M. (2007). Severe hypertelorism, midface prominence, prominent/simple ears, severe myopia, borderline intelligence, and bone fragility in two brothers: new syndrome? Am J Med Genet A *143A*, 229-234.

Harris, W. J., Huang, X., Lynch, J. T., Spencer, G. J., Hitchin, J. R., Li, Y., Ciceri, F., Blaser, J. G., Greystoke, B. F., Jordan, A. M., et al. (2012). The histone demethylase KDM1A sustains the oncogenic potential of MLL-AF9 leukemia stem cells. Cancer Cell 21, 473-487.

Henikoff, S., and Shilatifard, A. (2011). Histone modification: cause or cog? Trends Genet *27*, 389-396.

Hock, H., Hamblen, M. J., Rooke, H. M., Traver, D., Bronson, R. T., Cameron, S., and Orkin, S. H. (2003). Intrinsic requirement for zinc finger transcription factor Gfi-1 in neutrophil differentiation. Immunity *18*, 109-120.

Hogan, C. J., Shpall, E. J., and Keller, G. (2002). Differential long-term and multilineage engraftment potential from subfractions of human CD34+ cord blood cells transplanted into NOD/SCID mice. Proc Natl Acad Sci U S A *99*, 413-418.

Holtz, M. S., Forman, S. J., and Bhatia, R. (2005). Nonproliferating CML CD34+ progenitors are resistant to apoptosis induced by a wide range of proapoptotic stimuli. Leukemia *19*, 1034-1041.

Holyoake, T., Jiang, X., Eaves, C., and Eaves, A. (1999). Isolation of a highly quiescent subpopulation of primitive leukemic cells in chronic myeloid leukemia. Blood *94*, 2056-2064.

Hosen, N., Park, C. Y., Tatsumi, N., Oji, Y., Sugiyama, H., Gramatzki, M., Krensky, A. M., and Weissman, I. L. (2007). CD96 is a leukemic stem cell-specific marker in human acute myeloid leukemia. Proc Natl Acad Sci U S A *104*, 11008-11013.

Houweling, A. C., Dildrop, R., Peters, T., Mummenhoff, J., Moorman, A. F., Ruther, U., and Christoffels, V. M. (2001). Gene and cluster-specific expression of the Iroquois family members during mouse development. Mech Dev 1*07*, 169-174.

Huang, Y., Sitwala, K., Bronstein, J., Sanders, D., Dandekar, M., Collins, C., Robertson, G., MacDonald, J., Cezard, T., Bilenky, M., et al. (2012). Identification and characterization of Hoxa9 binding sites in hematopoietic cells. Blood *119*, 388-398.

Huang, X., Spencer, G. J., Lynch, J. T., Ciceri, F., Somerville, T. D., and Somervaille, T. C. (2014). Enhancers of Polycomb EPC1 and EPC2 sustain the oncogenic potential of MLL leukemia stem cells. Leukemia 28, 1081-1091.

Hughes, T. P., Kaeda, J., Branford, S., Rudzki, Z., Hochhaus, A., Hensley, M. L., Gathmann, I., Bolton, A. E., van Hoomissen, I. C., Goldman, J. M*., et al.* (2003). Frequency of major molecular responses to imatinib or interferon alfa plus cytarabine in newly diagnosed chronic myeloid leukemia. N Engl J Med *349*, 1423-1432.

Huntly, B. J., and Gilliland, D. G. (2005). Leukaemia stem cells and the evolution of cancer-stem-cell research. Nat Rev Cancer *5*, 311-321.

Ikuta, K., and Weissman, I. L. (1992). Evidence that hematopoietic stem cells express mouse c-kit but do not depend on steel factor for their generation. Proc Natl Acad Sci U S A *89*, 1502-1506.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4, 249-264.

Ishikawa, F., Yoshida, S., Saito, Y., Hijikata, A., Kitamura, H., Tanaka, S., Nakamura, R., Tanaka, T., Tomiyama, H., Saito, N.*, et al.* (2007). Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. Nat Biotechnol *25*, 1315-1321.

Ismail, I. H., McDonald, D., Strickfaden, H., Xu, Z., and Hendzel, M. J. (2013). A small molecule inhibitor of polycomb repressive complex 1 inhibits ubiquitin signaling at DNA double-strand breaks. J Biol Chem *288*, 26944-26954.

Ito, M., Hiramatsu, H., Kobayashi, K., Suzue, K., Kawahata, M., Hioki, K., Ueyama, Y., Koyanagi, Y., Sugamura, K., Tsuji, K.*, et al.* (2002). NOD/SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells. Blood *100*, 3175-3182.

Jacobson, R. H., Ladurner, A. G., King, D. S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422-1425.

Ji, H., Ehrlich, L. I., 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.

Jin, L., Hope, K. J., Zhai, Q., Smadja-Joffe, F., and Dick, J. E. (2006). Targeting of CD44 eradicates human acute myeloid leukemic stem cells. Nat Med *12*, 1167-1174.

Jin, L., Lee, E. M., Ramshaw, H. S., Busfield, S. J., Peoppl, A. G., Wilkinson, L., Guthridge, M. A., Thomas, D., Barry, E. F., Boyd, A.*, et al.* (2009). Monoclonal antibody-mediated targeting of CD123, IL-3 receptor alpha chain, eliminates human acute myeloid leukemic stem cells. Cell Stem Cell *5*, 31-42.

Jones, P. A., and Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. Nat Rev Genet *10*, 805-811.

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet *13*, 484-492.

Jurcic,J.G. (2008). Treating Elderly Patients with Acute Myeloid Leukemia: Implications of Prognostic Modeling for the Development of New Therapies. Clinical Leukemia 2, 154-155.

Kamel-Reid, S., and Dick, J. E. (1988). Engraftment of immune-deficient mice with human hematopoietic stem cells. Science *242*, 1706-1709.

Kamminga, L. M., Bystrykh, L. V., de Boer, A., Houwer, S., Douma, J., Weersing, E., Dontje, B., and de Haan, G. (2006). The Polycomb group gene Ezh2 prevents hematopoietic stem cell exhaustion. Blood *107*, 2170-2179.

Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A.*, et al.* (2013). Mutational landscape and significance across 12 major cancer types. Nature *502*, 333-339.

Kantarjian, H. M., Thomas, X. G., Dmoszynska, A., Wierzbowska, A., Mazur, G., Mayer, J., Gau, J. P., Chou, W. C., Buckstein, R., Cermak, J.*, et al.* (2012). Multicenter, randomized, open-label, phase III trial of decitabine versus patient choice, with physician advice, of either supportive care or low-dose cytarabine for the treatment of older patients with newly diagnosed acute myeloid leukemia. J Clin Oncol *30*, 2670-2677.

Karasuyama, H., and Melchers, F. (1988). Establishment of mouse cell lines which constitutively secrete large quantities of interleukin 2, 3, 4 or 5, using modified cDNA expression vectors. Eur J Immunol *18*, 97-104.

Kasper, L. H., Brindle, P. K., Schnabel, C. A., Pritchard, C. E., Cleary, M. L., and van Deursen, J. M. (1999). CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98-HOXA9 oncogenicity. Mol Cell Biol *19*, 764-776.

Kelly, L. M., Liu, Q., Kutok, J. L., Williams, I. R., Boulton, C. L., and Gilliland, D. G. (2002a). 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.

Kelly, L. M., Kutok, J. L., Williams, I. R., Boulton, C. L., Amaral, S. M., Curley, D. P., Ley, T. J., and Gilliland, D. G. (2002b). PML/RARalpha and FLT3-ITD induce an APL-like disease in a mouse model. Proc Natl Acad Sci U S A *99*, 8283-8288.

Kennison, J. A. (1995). The Polycomb and trithorax group proteins of Drosophila: trans-regulators of homeotic gene function. Annu Rev Genet *29*, 289-303.

Kharas, M. G., Yusuf, I., Scarfone, V. M., Yang, V. W., Segre, J. A., Huettner, C. S., and Fruman, D. A. (2007). KLF4 suppresses transformation of pre-B cells by ABL oncogenes. Blood 109, 747-755.

Kiel, M. J., Yilmaz, O. H., Iwashita, T., Terhorst, C., and Morrison, S. J. (2005). SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. Cell *121*, 1109-1121.

Kikushige, Y., Shima, T., Takayanagi, S., Urata, S., Miyamoto, T., Iwasaki, H., Takenaka, K., Teshima, T., Tanaka, T., Inagaki, Y., and Akashi, K. (2010). TIM-3 is a promising target to selectively kill acute myeloid leukemia stem cells. Cell Stem Cell *7*, 708-717.

Kim, G. D., Ni, J., Kelesoglu, N., Roberts, R. J., and Pradhan, S. (2002). Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. EMBO J *21*, 4183-4195.

Klein, C. (2011). Genetic defects in severe congenital neutropenia: emerging insights into life and death of human neutrophil granulocytes. Annu Rev Immunol *29*, 399-413.

Ko, M., Bandukwala, H. S., An, J., Lamperti, E. D., Thompson, E. C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S. B., and Rao, A. (2011). Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proc Natl Acad Sci U S A *108*, 14566-14571.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L., and Wilson, R. K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22, 568-576.

Kon, A., Shih, L. Y., Minamino, M., Sanada, M., Shiraishi, Y., Nagata, Y., Yoshida, K., Okuno, Y., Bando, M., Nakato, R.*, et al.* (2013). Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. Nat Genet *45*, 1232-1237.

Kondo, M., Weissman, I. L., and Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell *91*, 661-672.

Konze, K. D., Ma, A., Li, F., Barsyte-Lovejoy, D., Parton, T., Macnevin, C. J., Liu, F., Gao, C., Huang, X. P., Kuznetsova, E., et al. (2013). An orally bioavailable chemical probe of the Lysine Methyltransferases EZH2 and EZH1. ACS Chem Biol *8*, 1324-1334.

Kosmider, O., Gelsi-Boyer, V., Slama, L., Dreyfus, F., Beyne-Rauzy, O., Quesnel, B., Hunault-Berger, M., Slama, B., Vey, N., Lacombe, C.*, et al.* (2010). Mutations of IDH1 and IDH2 genes in early and accelerated phases of myelodysplastic syndromes and MDS/myeloproliferative neoplasms. Leukemia *24*, 1094-1096.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.
Kreso, A., and Dick, J. E. (2014). Evolution of the cancer stem cell model. Cell Stem Cell *14*, 275-291.

Krivtsov, A. V., and Armstrong, S. A. (2007). MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer *7*, 823-833.

Krivtsov, A. V., Feng, Z., Lemieux, M. E., Faber, J., Vempati, S., Sinha, A. U., Xia, X., Jesneck, J., Bracken, A. P., Silverman, L. B.*, et al.* (2008). H3K79 methylation profiles define murine and human MLL-AF4 leukemias. Cancer Cell *14*, 355-368.

Krivtsov, A. V., Twomey, D., Feng, Z., Stubbs, M. C., Wang, Y., Faber, J., Levine, J. E., Wang, J., Hahn, W. C., Gilliland, D. G.*, et al.* (2006). Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. Nature *442*, 818-822.

Kroon, E., Krosl, J., Thorsteinsdottir, U., Baban, S., Buchberg, A. M., and Sauvageau, G. (1998). Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. EMBO J *17*, 3714-3725.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S.*, et al.* (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet *4*, e1000242.

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

Kume, T., Deng, K. Y., Winfrey, V., Gould, D. B., Walter, M. A., and Hogan, B. L. (1998). The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus. Cell *93*, 985-996.

Ladam, F., and Sagerstrom, C. G. (2014). Hox regulation of transcription: more complex(es). Dev Dyn *243*, 4-15.

Lam, E. W., Brosens, J. J., Gomes, A. R., and Koo, C. Y. (2013). Forkhead box proteins: tuning forks for transcriptional harmony. Nat Rev Cancer *13*, 482-495.

Lansdorp, P. M., Sutherland, H. J., and Eaves, C. J. (1990). Selective expression of CD45 isoforms on functional subpopulations of CD34+ hemopoietic cells from human bone marrow. J Exp Med *172*, 363-366.

Lapidot, T., Pflumio, F., Doedens, M., Murdoch, B., Williams, D. E., and Dick, J. E. (1992). Cytokine stimulation of multilineage hematopoiesis from immature human cells engrafted in SCID mice. Science *255*, 1137-1141.

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.

Larochelle, A., Savona, M., Wiggins, M., Anderson, S., Ichwan, B., Keyvanfar, K., Morrison, S. J., and Dunbar, C. E. (2011). Human and rhesus macaque hematopoietic stem cells cannot be purified based only on SLAM family markers. Blood *117*, 1550-1554.

Laslo, P., Spooner, C. J., Warmflash, A., Lancki, D. W., Lee, H. J., Sciammas, R., Gantner, B. N., Dinner, A. R., and Singh, H. (2006). Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. Cell *126*, 755-766. 18

Lee, J. L., Kim, S., Kim, S. W., Kim, E. K., Kim, S. B., Kang, Y. K., Lee, J., Kim, M. W., Park, C. J., Chi, H. S., et al. (2005). ESHAP plus G-CSF as an effective peripheral blood progenitor cell mobilization regimen in pretreated non-Hodgkin's lymphoma: comparison with high-dose cyclophosphamide plus G-CSF. Bone Marrow Transplant *35*, 449-454.

Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K.*, et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Lee, J. S., Smith, E., and Shilatifard, A. (2010). The language of histone crosstalk. Cell *142*, 682-685.

Lessard, J., and Sauvageau, G. (2003). Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. Nature *423*, 255-260.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565-570.

Ley, T. J., Ding, L., Walter, M. J., McLellan, M. D., Lamprecht, T., Larson, D. E., Kandoth, C., Payton, J. E., Baty, J., Welch, J.*, et al.* (2010). DNMT3A mutations in acute myeloid leukemia. N Engl J Med *363*, 2424-2433.

Ley, T. J., Mardis, E. R., Ding, L., Fulton, B., McLellan, M. D., Chen, K., Dooling, D., Dunford-Shore, B. H., McGrath, S., Hickenbotham, M.*, et al.* (2008). DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature *456*, 66-72.

Li, C., and Wong, W. H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A 98, 31-36.

Li, Li., Piloto, O., Nguyen, H. B., Greenberg, K., Takamiya, K., Racke, F., Huso, D. and Small, D. (2008). Knock-in of an internal tandem duplication mutation into murine FLT3 confers myeloproliferative disease in a mouse model. Blood *111*, 3849-3858

Li, Z., Cai, X., Cai, C. L., Wang, J., Zhang, W., Petersen, B. E., Yang, F. C., and Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood *118*, 4509-4518.

Li, L., Wang, L., Wang, Z., Ho, Y., McDonald, T., Holyoake, T. L., Chen, W., and Bhatia, R. (2012). Activation of p53 by SIRT1 inhibition enhances elimination of CML leukemia stem cells in combination with imatinib. Cancer Cell *21*, 266-281.

Li, Z., Zhang, Z., Li, Y., Arnovitz, S., Chen, P., Huang, H., Jiang, X., Hong, G. M., Kunjamma, R. B., Ren, H., et al. (2013). PBX3 is an important cofactor of HOXA9 in leukemogenesis. Blood *121*, 1422-1431.

Li, D., Sakuma, R., Vakili, N. A., Mo, R., Puviindran, V., Deimling, S., Zhang, X., Hopyan, S., and Hui, C. C. (2014). Formation of proximal and anterior limb skeleton requires early function of Irx3 and Irx5 and is negatively regulated by Shh signaling. Dev Cell *29*, 233-240.

Link, H., Arseniev, L., Bähre, O., Kadar, J. G., Diedrich, H., and Poliwoda, H. (1996). Transplantation of allogeneic CD34+ blood cells. Blood *87*, 4903-4909.

Lo-Coco, F., Avvisati, G., Vignetti, M., Thiede, C., Orlando, S. M., Iacobelli, S., Ferrara, F., Fazi, P., Cicconi, L., Di Bona, E.*, et al.* (2013). Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. N Engl J Med *369*, 111-121.

Longobardi, E., Penkov, D., Mateos, D., De Florian, G., Torres, M., and Blasi, F. (2014). Biochemistry of the tale transcription factors PREP, MEIS, and PBX in vertebrates. Dev Dyn *243*, 59-75.

Löwenberg, B., Ossenkoppele, G. J., van Putten, W., Schouten, H. C., Graux, C., Ferrant, A., Sonneveld, P., Maertens, J., Jongen-Lavrencic, M., von Lilienfeld-Toal, M.*, et al.* (2009). High-dose daunorubicin in older patients with acute myeloid leukemia. N Engl J Med *361*, 1235-1248.

Lund, K., Adams, P. D., and Copland, M. (2014). EZH2 in normal and malignant hematopoiesis. Leukemia *28*, 44-49.

Magee, J. A., Piskounova, E., and Morrison, S. J. (2012). Cancer stem cells: impact, heterogeneity, and uncertainty. Cancer Cell *21*, 283-296.

Majeti, R., Chao, M. P., Alizadeh, A. A., Pang, W. W., Jaiswal, S., Gibbs, K. D., van Rooijen, N., and Weissman, I. L. (2009). CD47 is an adverse prognostic factor and therapeutic antibody target on human acute myeloid leukemia stem cells. Cell *138*, 286-299.
Mandelli, F., Vignetti, M., Suciu, S., Stasi, R., Petti, M. C., Meloni, G., Muus, P., Marmont, F., Marie, J. P., Labar, B.*, et al.* (2009). Daunorubicin versus mitoxantrone versus idarubicin as induction and consolidation chemotherapy for adults with acute myeloid leukemia: the EORTC and GIMEMA Groups Study AML-10. J Clin Oncol *27*, 5397-5403.

Mann, R. S., Lelli, K. M., and Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. Curr Top Dev Biol *88*, 63-101.

Mansour, M. R., Abraham, B. J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A. D., Etchin, J., Lawton, L., Sallan, S. E., Silverman, L. B., et al. (2014). Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science *346*, 1373-1377.

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.

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.

Mardis, E. R., Ding, L., Dooling, D. J., Larson, D. E., McLellan, M. D., Chen, K., Koboldt, D. C., Fulton, R. S., Delehaunty, K. D., McGrath, S. D.*, et al.* (2009). Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med *361*, 1058-1066.

Matsuzaki, J., Tsuji, T., Chamoto, K., Takeshima, T., Sendo, F., and Nishimura, T. (2003). Successful elimination of memory-type CD8+ T cell subsets by the administration of anti-Gr-1 monoclonal antibody in vivo. Cell Immunol *224*, 98-105.

Mayer, R. J., Davis, R. B., Schiffer, C. A., Berg, D. T., Powell, B. L., Schulman, P., Omura, G. A., Moore, J. O., McIntyre, O. R., and Frei, E. (1994). Intensive postremission chemotherapy in adults with acute myeloid leukemia. Cancer and Leukemia Group B. N Engl J Med *331*, 896-903.

Mazzarella, L., Riva, L., Luzi, L., Ronchini, C., and Pelicci, P. G. (2014). The genomic and epigenomic landscapes of AML. Semin Hematol *51*, 259-272.

Mills, A. A. (2010). Throwing the cancer switch: reciprocal roles of polycomb and trithorax proteins. Nat Rev Cancer *10*, 669-682.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M. E., Vasanthakumar, A., Patel, J., Zhao, X.*, et al.* (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell *20*, 11-24.

Moreno-Miralles, I., Pan, L., Keates-Baleeiro, J., Durst-Goodwin, K., Yang, C., Kim, H. G., Thompson, M. A., Klug, C. A., Cleveland, J. L., and Hiebert, S. W. (2005). The inv(16) cooperates with ARF haploinsufficiency to induce acute myeloid leukemia. J Biol Chem *280*, 40097-40103.

Morey, S. R., Smiraglia, D. J., James, S. R., Yu, J., Moser, M. T., Foster, B. A., and Karpf, A. R. (2006). DNA methylation pathway alterations in an autochthonous murine model of prostate cancer. Cancer Res *66*, 11659-11667.

Morgan, R., Pirard, P. M., Shears, L., Sohal, J., Pettengell, R., and Pandha, H. S. (2007). Antagonism of HOX/PBX dimer formation blocks the in vivo proliferation of melanoma. Cancer Res *67*, 5806-5813.

Morita, S., Kojima, T., and Kitamura, T. (2000). Plat-E: an efficient and stable system for transient packaging of retroviruses. Gene Ther *7*, 1063-1066.

Morrison, S. J., and Scadden, D. T. (2014). The bone marrow niche for haematopoietic stem cells. Nature *505*, 327-334.

Morrison, S. J., Wandycz, A. M., Hemmati, H. D., Wright, D. E., and Weissman, I. L. (1997). Identification of a lineage of multipotent hematopoietic progenitors. Development *124*, 1929-1939.

Mrózek, K., Heinonen, K., and Bloomfield, C. D. (2001). Clinical importance of cytogenetics in acute myeloid leukaemia. Best Pract Res Clin Haematol *14*, 19-47.

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.

Mukherjee, K., and Bürglin, T. R. (2007). Comprehensive analysis of animal TALE homeobox genes: new conserved motifs and cases of accelerated evolution. J Mol Evol 65, 137-153.
myeloid leukemia. N Engl J Med *368*, 2059-2074.

Myrthue, A., Rademacher, B. L., Pittsenbarger, J., Kutyba-Brooks, B., Gantner, M., Qian, D. Z., and Beer, T. M. (2008). The iroquois homeobox gene 5 is regulated by 1,25-dihydroxyvitamin D3 in human prostate cancer and regulates apoptosis and the cell cycle in LNCaP prostate cancer cells. Clin Cancer Res *14*, 3562-3570.

Nakamura, T., Largaespada, D. A., Lee, M. P., Johnson, L. A., Ohyashiki, K., Toyama, K., Chen, S. J., Willman, C. L., Chen, I. M., Feinberg, A. P.*, et al.* (1996). Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. Nat Genet *12*, 154-158.

Nakanishi, M., Tanaka, K., Shintani, T., Takahashi, T., and Kamada, N. (1999). Chromosomal instability in acute myelocytic leukemia and myelodysplastic syndrome patients among atomic bomb survivors. J Radiat Res *40*, 159-167.

National Cancer Institute. (2015) SEER Stat facts sheet: Acute myeloid leukaemia (AML). Available from: http://seer.cancer.gov/statfacts/html/amyl.html. [Accessed 17th June 2015].

Nekrasov, M., Wild, B., and Müller, J. (2005). Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. EMBO Rep *6*, 348-353.

Nishimura, D. Y., Swiderski, R. E., Alward, W. L., Searby, C. C., Patil, S. R., Bennet, S. R., Kanis, A. B., Gastier, J. M., Stone, E. M., and Sheffield, V. C. (1998). The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25. Nat Genet *19*, 140-147.

Notta, F., Doulatov, S., Laurenti, E., Poeppl, A., Jurisica, I., and Dick, J. E. (2011). Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. Science *333*, 218-221.

Nowell, P. C. (1976). The clonal evolution of tumor cell populations. Science *194*, 23-28.

O'Brien, S. G., Guilhot, F., Larson, R. A., Gathmann, I., Baccarani, M., Cervantes, F., Cornelissen, J. J., Fischer, T., Hochhaus, A., Hughes, T.*, et al.* (2003). Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. N Engl J Med *348*, 994-1004.

Ogawa, M., Porter, P. N., and Nakahata, T. (1983). Renewal and commitment to differentiation of hemopoietic stem cells (an interpretive review). Blood *61*, 823-829.

Oguro, H., Yuan, J., Ichikawa, H., Ikawa, T., Yamazaki, S., Kawamoto, H., Nakauchi, H., and Iwama, A. (2010). Poised lineage specification in multipotential hematopoietic stem and progenitor cells by the polycomb protein Bmi1. Cell Stem Cell *6*, 279-286.

Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V. M., Su, L., Xu, G., and Zhang, Y. (2005). hDOT1L links histone methylation to leukemogenesis. Cell *121*, 167-178.

Okoniewski, M. J., and Miller, C. J. (2008). Comprehensive analysis of affymetrix exon arrays using BioConductor. PLoS Comput Biol *4*, e6.

Omatsu, Y., Seike, M., Sugiyama, T., Kume, T., and Nagasawa, T. (2014). Foxc1 is a critical regulator of haematopoietic stem/progenitor cell niche formation. Nature *508*, 536-540.

Ooi, S. K., and Bestor, T. H. (2008). The colorful history of active DNA demethylation. Cell *133*, 1145-1148.

Ordway, J. M., Bedell, J. A., Citek, R. W., Nunberg, A., Garrido, A., Kendall, R., Stevens, J. R., Cao, D., Doerge, R. W., Korshunova, Y., et al. (2006). Comprehensive DNA methylation profiling in a human cancer genome identifies novel epigenetic targets. Carcinogenesis *27*, 2409-2423.

Orkin, S. H. (2000). Diversification of haematopoietic stem cells to specific lineages. Nat Rev Genet *1*, 57-64.

Orkin, S. H., and Zon, L. I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. Cell *132*, 631-644.

Osawa, M., Hanada, K., Hamada, H., and Nakauchi, H. (1996). Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. Science *273*, 242-245.

Othus, M., Kantarjian, H., Petersdorf, S., Ravandi, F., Godwin, J., Cortes, J., Pierce, S., Erba, H., Faderl, S., Appelbaum, F. R., and Estey, E. (2014). Declining rates of treatment-related mortality in patients with newly diagnosed AML given 'intense' induction regimens: a report from SWOG and MD Anderson. Leukemia *28*, 289-292.

Owen, C., Barnett, M., and Fitzgibbon, J. (2008). Familial myelodysplasia and acute myeloid leukaemia--a review. Br J Haematol *140*, 123-132.

Park, I. K., Qian, D., Kiel, M., Becker, M. W., Pihalja, M., Weissman, I. L., Morrison, S. J., and Clarke, M. F. (2003). Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. Nature *423*, 302-305.

Paschka, P., Schlenk, R. F., Gaidzik, V. I., Habdank, M., Krönke, J., Bullinger, L., Späth, D., Kayser, S., Zucknick, M., Götze, K.*, et al.* (2010). IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplication. J Clin Oncol *28*, 3636-3643.

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.

Pattabiraman, D. R., and Weinberg, R. A. (2014). Tackling the cancer stem cells - what challenges do they pose? Nat Rev Drug Discov *13*, 497-512.

Pearce, D. J., Taussig, D., Zibara, K., Smith, L. L., Ridler, C. M., Preudhomme, C., Young, B. D., Rohatiner, A. Z., Lister, T. A., and Bonnet, D. (2006). AML engraftment in the NOD/SCID assay reflects the outcome of AML: implications for our understanding of the heterogeneity of AML. Blood *107*, 1166-1173.

Perkins, A., Kongsuwan, K., Visvader, J., Adams, J. M., and Cory, S. (1990). Homeobox gene expression plus autocrine growth factor production elicits myeloid leukemia. Proc Natl Acad Sci U S A *87*, 8398-8402.

Perry, A. M., and Attar, E. C. (2014). New insights in AML biology from genomic analysis. Semin Hematol *51*, 282-297.

Peters, T., Ausmeier, K., Dildrop, R., and Ruther, U. (2002). The mouse Fused toes (Ft) mutation is the result of a 1.6-Mb deletion including the entire Iroquois B gene cluster. Mamm Genome *13*, 186-188.

Pollyea, D. A., Gutman, J. A., Gore, L., Smith, C. A., and Jordan, C. T. (2014). Targeting acute myeloid leukemia stem cells: a review and principles for the development of clinical trials. Haematologica *99*, 1277-1284.

Radulović, V., de Haan, G., and Klauke, K. (2013). Polycomb-group proteins in hematopoietic stem cell regulation and hematopoietic neoplasms. Leukemia *27*, 523-533.

Ragvin, A., Moro, E., Fredman, D., Navratilova, P., Drivenes, O., Engstrom, P. G., Alonso, M. E., de la Calle Mustienes, E., Gomez Skarmeta, J. L., Tavares, M. J., et al. (2010). Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to HHEX, SOX4, and IRX3. Proc Natl Acad Sci U S A *107*, 775-780.

Ray, P. S., Wang, J., Qu, Y., Sim, M. S., Shamonki, J., Bagaria, S. P., Ye, X., Liu, B., Elashoff, D., Hoon, D. S., et al. (2010). FOXC1 is a potential prognostic biomarker with functional significance in basal-like breast cancer. Cancer Res *70*, 3870-3876.

Reya, T., Morrison, S. J., Clarke, M. F., and Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. Nature *414*, 105-111.

Riva, L., Luzi, L., and Pelicci, P. G. (2012). Genomics of acute myeloid leukemia: the next generation. Front Oncol *2*, 40.

Rose, N. R., McDonough, M. A., King, O. N., Kawamura, A., and Schofield, C. J. (2011). Inhibition of 2-oxoglutarate dependent oxygenases. Chem Soc Rev *40*, 4364-4397.

Rosenbauer, F., and Tenen, D. G. (2007). Transcription factors in myeloid development: balancing differentiation with transformation. Nat Rev Immunol *7*, 105-117.

Rousselot, P., Huguet, F., Rea, D., Legros, L., Cayuela, J. M., Maarek, O., Blanchet, O., Marit, G., Gluckman, E., Reiffers, J.*, et al.* (2007). Imatinib mesylate discontinuation in patients with chronic myelogenous leukemia in complete molecular remission for more than 2 years. Blood *109*, 58-60.

Sabbath, K. D., Ball, E. D., Larcom, P., Davis, R. B., and Griffin, J. D. (1985). Heterogeneity of clonogenic cells in acute myeloblastic leukemia. J Clin Invest *75*, 746-753.

Saito, Y., Kitamura, H., Hijikata, A., Tomizawa-Murasawa, M., Tanaka, S., Takagi, S., Uchida, N., Suzuki, N., Sone, A., Najima, Y.*, et al.* (2010). Identification of therapeutic targets for quiescent, chemotherapy-resistant human leukemia stem cells. Sci Transl Med *2*, 17ra19.

Saito, Y., Yuki, H., Kuratani, M., Hashizume, Y., Takagi, S., Honma, T., Tanaka, A., Shirouzu, M., Mikuni, J., Handa, N.*, et al.* (2013). A pyrrolo-pyrimidine derivative targets human primary AML stem cells in vivo. Sci Transl Med *5*, 181ra152.

Sanchez, P. V., Perry, R. L., Sarry, J. E., Perl, A. E., Murphy, K., Swider, C. R., Bagg, A., Choi, J. K., Biegel, J. A., Danet-Desnoyers, G., and Carroll, M. (2009). A robust xenotransplantation model for acute myeloid leukemia. Leukemia *23*, 2109-2117.

Sanz, M. A., Grimwade, D., Tallman, M. S., Lowenberg, B., Fenaux, P., Estey, E. H., Naoe, T., Lengfelder, E., Büchner, T., Döhner, H.*, et al.* (2009). Management of acute promyelocytic leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. Blood *113*, 1875-1891.

Sarry, J. E., Murphy, K., Perry, R., Sanchez, P. V., Secreto, A., Keefer, C., Swider, C. R., Strzelecki, A. C., Cavelier, C., Récher, C.*, et al.* (2011). Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice. J Clin Invest *121*, 384-395.

Sauvageau, M., and Sauvageau, G. (2010). Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. Cell Stem Cell *7*, 299-313.

Sawyers, C. L. (1999). Chronic myeloid leukemia. N Engl J Med *340*, 1330-1340.

Schnabel, C. A., Jacobs, Y., and Cleary, M. L. (2000). HoxA9-mediated immortalization of myeloid progenitors requires functional interactions with TALE cofactors Pbx and Meis. Oncogene *19*, 608-616.

Schiffer, C. A., Lee, E. J., Tomiyasu, T., Wiernik, P. H., and Testa, J. R. (1989). Prognostic impact of cytogenetic abnormalities in patients with de novo acute nonlymphocytic leukemia. Blood *73*, 263-270.

Schraets, D., Lehmann, T., Dingermann, T., and Marschalek, R. (2003). MLL-mediated transcriptional gene regulation investigated by gene expression profiling. Oncogene *22*, 3655-3668.

Schuettengruber, B., Martinez, A. M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. Nat Rev Mol Cell Biol *12*, 799-814.

Schuler, A., Schwieger, M., Engelmann, A., Weber, K., Horn, S., Muller, U., Arnold, M. A., Olson, E. N., and Stocking, C. (2008). The MADS transcription factor Mef2c is a pivotal modulator of myeloid cell fate. Blood *111*, 4532-4541.

Schwartz, Y. B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet *8*, 9-22.

Schwartz, Y. B., Kahn, T. G., Stenberg, P., Ohno, K., Bourgon, R., and Pirrotta, V. (2010). Alternative epigenetic chromatin states of polycomb target genes. PLoS Genet *6*, e1000805.

Seita, J., and Weissman, I. L. (2010). Hematopoietic stem cell: self-renewal versus differentiation. Wiley Interdiscip Rev Syst Biol Med *2*, 640-653.

Shackleton, M., Quintana, E., Fearon, E. R., and Morrison, S. J. (2009). Heterogeneity in cancer: cancer stem cells versus clonal evolution. Cell *138*, 822-829.

Sharma, S., Kelly, T. K., and Jones, P. A. (2010). Epigenetics in cancer. Carcinogenesis *31*, 27-36.

Shen, H., and Laird, P. W. (2013). Interplay between the cancer genome and epigenome. Cell *153*, 38-55.

Shen, W. F., Montgomery, J. C., Rozenfeld, S., Moskow, J. J., Lawrence, H. J., Buchberg, A. M., and Largman, C. (1997). AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. Mol Cell Biol *17*, 6448-6458.

Shepherd, B. E., Guttorp, P., Lansdorp, P. M., and Abkowitz, J. L. (2004). Estimating human hematopoietic stem cell kinetics using granulocyte telomere lengths. Exp Hematol *32*, 1040-1050.

Shi, C., Zhang, X., Chen, Z., Sulaiman, K., Feinberg, M. W., Ballantyne, C. M., Jain, M. K., and Simon, D. I. (2004). Integrin engagement regulates monocyte differentiation through the forkhead transcription factor Foxp1. J Clin Invest *114*, 408-418.

Shi, C., Sakuma, M., Mooroka, T., Liscoe, A., Gao, H., Croce, K. J., Sharma, A., Kaplan, D., Greaves, D. R., Wang, Y., and Simon, D. I. (2008). Down-regulation of the forkhead transcription factor Foxp1 is required for monocyte differentiation and macrophage function. Blood *112*, 4699-4711.

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., Lyons, B. L., Burzenski, L. M., Gott, B., Chen, X., Chaleff, S., Kotb, M., Gillies, S. D., King, M., Mangada, J*., et al.* (2005). Human lymphoid and myeloid cell development in NOD/LtSz-

scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. J Immunol *174*, 6477-6489.

Shultz, L. D., Schweitzer, P. A., Christianson, S. W., Gott, B., Schweitzer, I. B., Tennent, B., McKenna, S., Mobraaten, L., Rajan, T. V., and Greiner, D. L. (1995). Multiple defects in innate and adaptive immunologic function in NOD/LtSz-scid mice. J Immunol *154*, 180-191.
Siegel, R., Naishadham, D., and Jemal, A. (2013). Cancer statistics, 2013. CA Cancer J Clin *63*, 11-30.

Siminovitch, L., McCulloch, E. A., and Till, J. E. (1963). The distribution of colony-forming cells among spleen colonies. J Cell Physiol *62*, 327-336.

Simon, J. (1995). Locking in stable states of gene expression: transcriptional control during Drosophila development. Curr Opin Cell Biol *7*, 376-385.

Simon, J. A., and Kingston, R. E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol *10*, 697-708.

Simon, J. A., and Kingston, R. E. (2013). Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. Mol Cell *49*, 808-824.

Sitnicka, E., Buza-Vidas, N., Larsson, S., Nygren, J. M., Liuba, K., and Jacobsen, S. E. (2003). Human CD34+ hematopoietic stem cells capable of multilineage engrafting NOD/SCID mice express flt3: distinct flt3 and c-kit expression and response patterns on mouse and candidate human hematopoietic stem cells. Blood *102*, 881-886.

Sizemore, S. T., and Keri, R. A. (2012). The forkhead box transcription factor FOXC1 promotes breast cancer invasion by inducing matrix metalloprotease 7 (MMP7) expression. J Biol Chem *287*, 24631-24640.

Slovak, M. L., Kopecky, K. J., Cassileth, P. A., Harrington, D. H., Theil, K. S., Mohamed, A., Paietta, E., Willman, C. L., Head, D. R., Rowe, J. M.*, et al.* (2000). Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study. Blood *96*, 4075-4083.

Smemo, S., Tena, J. J., Kim, K. H., Gamazon, E. R., Sakabe, N. J., Gomez-Marin, C., Aneas, I., Credidio, F. L., Sobreira, D. R., Wasserman, N. F., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature *507*, 371-375.

Somervaille, T. C., and Cleary, M. L. (2006). Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. Cancer Cell *10*, 257-268.

Somervaille, T. C., Matheny, C. J., Spencer, G. J., Iwasaki, M., Rinn, J. L., Witten, D. M., Chang, H. Y., Shurtleff, S. A., Downing, J. R., and Cleary, M. L. (2009). Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. Cell Stem Cell *4*, 129-140.

Spangrude, G. J., Heimfeld, S., and Weissman, I. L. (1988). Purification and characterization of mouse hematopoietic stem cells. Science *241*, 58-62.

Speck, N. A., and Gilliland, D. G. (2002). Core-binding factors in haematopoiesis and leukaemia. Nat Rev Cancer *2*, 502-513.

Spivakov, M., and Fisher, A. G. (2007). Epigenetic signatures of stem-cell identity. Nat Rev Genet *8*, 263-271.

Stein. E., Tallman, M., Pollyea, D.A., Flinn, I.W., Fathi, A.T., Stone, R.M., Levine, R.L., Agresta, S., Schenkein, D., Yang, H. *et al* (2014) Clinical safety and activity in a phase I trial of AG-221, a first in class, potent inhibitor of the IDH2-mutant protein, in patients with IDH2 mutant positive advanced hematologic malignancies. In Proceedings of the 105th Annual Meeting of the American Association for Cancer Research, 2014 Apr 5–9; San Diego, CA. Philadelphia (PA): AACR; 2014. Abstract nr CT103.

Stock, J. K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A. G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat Cell Biol *9*, 1428-1435.

Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

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., and Mesirov, J. P. (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, J., Ramos, A., Chapman, B., Johnnidis, J. B., Le, L., Ho, Y. J., Klein, A., Hofmann, O., and Camargo, F. D. (2014). Clonal dynamics of native haematopoiesis. Nature *514*, 322-327.

Surface, L. E., Thornton, S. R., and Boyer, L. A. (2010). Polycomb group proteins set the stage for early lineage commitment. Cell Stem Cell *7*, 288-298.

Suzuki, M. M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet *9*, 465-476.

Suzuki, H., Forrest, A. R., van Nimwegen, E., Daub, C. O., Balwierz, P. J., Irvine, K. M., Lassmann, T., Ravasi, T., Hasegawa, Y., de Hoon, M. J., et al. (2009). The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. Nat Genet *41*, 553-562.

Takenaka, K., Prasolava, T. K., Wang, J. C., Mortin-Toth, S. M., Khalouei, S., Gan, O. I., Dick, J. E., and Danska, J. S. (2007). Polymorphism in Sirpa modulates engraftment of human hematopoietic stem cells. Nat Immunol *8*, 1313-1323.

Tallman, M. S., Andersen, J. W., Schiffer, C. A., Appelbaum, F. R., Feusner, J. H., Woods, W. G., Ogden, A., Weinstein, H., Shepherd, L., Willman, C.*, et al.* (2002). All-trans retinoic acid in acute promyelocytic leukemia: long-term outcome and prognostic factor analysis from the North American Intergroup protocol. Blood *100*, 4298-4302.

Tallman, M. S., Gilliland, D. G., and Rowe, J. M. (2005). Drug therapy for acute myeloid leukemia. Blood *106*, 1154-1163.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N.*, et al.* (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell *146*, 1016-1028.

Taussig, D. C., Pearce, D. J., Simpson, C., Rohatiner, A. Z., Lister, T. A., Kelly, G., Luongo, J. L., Danet-Desnoyers, G. A., and Bonnet, D. (2005). Hematopoietic stem cells express multiple myeloid markers: implications for the origin and targeted therapy of acute myeloid leukemia. Blood *106*, 4086-4092.

Taussig, D. C., Miraki-Moud, F., Anjos-Afonso, F., Pearce, D. J., Allen, K., Ridler, C., Lillington, D., Oakervee, H., Cavenagh, J., Agrawal, S. G.*, et al.* (2008). Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells. Blood *112*, 568-575.

Tenen, D. G. (2003). Disruption of differentiation in human cancer: AML shows the way. Nat Rev Cancer *3*, 89-101.

Theokli, C., Morsi El-Kadi, A. S., and Morgan, R. (2003). TALE class homeodomain gene Irx5 is an immediate downstream target for Hoxb4 transcriptional regulation. Dev Dyn *227*, 48-55.

Thiel, A. T., Blessington, P., Zou, T., Feather, D., Wu, X., Yan, J., Zhang, H., Liu, Z., Ernst, P., Koretzky, G. A., and Hua, X. (2010). MLL-AF9-induced leukemogenesis requires coexpression of the wild-type Mll allele. Cancer Cell *17*, 148-159.

Thol, F., Bollin, R., Gehlhaar, M., Walter, C., Dugas, M., Suchanek, K. J., Kirchner, A., Huang, L., Chaturvedi, A., Wichmann, M.*, et al.* (2014). Mutations in the cohesin complex in acute myeloid

leukemia: clinical and prognostic implications. Blood *123*, 914-920.

Thorsteinsdottir, U., Kroon, E., Jerome, L., Blasi, F., and Sauvageau, G. (2001). Defining roles for HOX and MEIS1 genes in induction of acute myeloid leukemia. Mol Cell Biol *21*, 224-234.

Thorsteinsdottir, U., Mamo, A., Kroon, E., Jerome, L., Bijl, J., Lawrence, H. J., Humphries, K., and Sauvageau, G. (2002). Overexpression of the myeloid leukemia-associated Hoxa9 gene in bone marrow cells induces stem cell expansion. Blood *99*, 121-129.

Till, J. E., and McCulloch, E. A. (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiat Res *14*, 213-222.

Till, J. E., McCulloch, E. A., and Siminovitch, L. (1964). A stochastic model of stem cell proliferation, based on the growth of spleen-colony forming cells. Proc Natl Acad Sci U S A *51*, 29-36.

Tkachuk, D. C., Kohler, S., and Cleary, M. L. (1992). Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias. Cell *71*, 691-700.

Tomaru, Y., Simon, C., Forrest, A. R., Miura, H., Kubosaki, A., Hayashizaki, Y., and Suzuki, M. (2009). Regulatory interdependence of myeloid transcription factors revealed by Matrix RNAi analysis. Genome Biol *10*, R121.

Trowbridge, J. J., Snow, J. W., Kim, J., and Orkin, S. H. (2009). DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell Stem Cell *5*, 442-449.

Tumer, Z., and Bach-Holm, D. (2009). Axenfeld-Rieger syndrome and spectrum of PITX2 and FOXC1 mutations. Eur J Hum Genet 17, 1527-1539.

Uribesalgo, I., and Di Croce, L. (2011). Dynamics of epigenetic modifications in leukemia. Brief Funct Genomics *10*, 18-29.

Vakoc, C. R., Mandat, S. A., Olenchock, B. A., and Blobel, G. A. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol Cell *19*, 381-391.

Valk, P. J., Verhaak, R. G., Beijen, M. A., Erpelinck, C. A., Barjesteh van Waalwijk van Doorn-Khosrovani, S., Boer, J. M., Beverloo, H. B., Moorhouse, M. J., van der Spek, P. J., Lowenberg, B., and Delwel, R. (2004). Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 350, 1617-1628.

van Rhenen, A., Feller, N., Kelder, A., Westra, A. H., Rombouts, E., Zweegman, S., van der Pol, M. A., Waisfisz, Q., Ossenkoppele, G. J., and Schuurhuis, G. J. (2005). High stem cell frequency in acute myeloid leukemia at diagnosis predicts high minimal residual disease and poor survival. Clin Cancer Res *11*, 6520-6527.

van Rhenen, A., van Dongen, G. A., Kelder, A., Rombouts, E. J., Feller, N., Moshaver, B., Stigter-van Walsum, M., Zweegman, S., Ossenkoppele, G. J., and Jan Schuurhuis, G. (2007). The novel AML stem cell associated antigen CLL-1 aids in discrimination between normal and leukemic stem cells. Blood *110*, 2659-2666.

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., and Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. Nat Rev Genet *10*, 252-263.

Vardiman, J. W., Harris, N. L., and Brunning, R. D. (2002). The World Health Organization (WHO) classification of the myeloid neoplasms. Blood *100*, 2292-2302.

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., and Bloomfield, C. D. (2009). The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood *114*, 937-951.

Verma, S. K., Tian, X., LaFrance, L. V., Duquenne, C., Suarez, D. P., Newlander, K. A., Romeril, S. P., Burgess, J. L., Grant, S. W., Brackley, J. A., et al. (2012). Identification of Potent, Selective, Cell-Active Inhibitors of the Histone Lysine Methyltransferase EZH2. ACS Med Chem Lett *3*, 1091-1096.

Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W., van Schaik, F. M., Varier, R. A., Baltissen, M. P., Stunnenberg, H. G., Mann, M., and Timmers, H. T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58-69.

Vidal, S. J., Rodriguez-Bravo, V., Galsky, M., Cordon-Cardo, C., and Domingo-Domenech, J. (2014). Targeting cancer stem cells to suppress acquired chemotherapy resistance. Oncogene *33*, 4451-4463.

Villa, R., Pasini, D., Gutierrez, A., Morey, L., Occhionorelli, M., Viré, E., Nomdedeu, J. F., Jenuwein, T., Pelicci, P. G., Minucci, S*., et al.* (2007). Role of the polycomb repressive complex 2 in acute promyelocytic leukemia. Cancer Cell *11*, 513-525.

Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J. M*., et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871-874.

Walter, M. J., Shen, D., Ding, L., Shao, J., Koboldt, D. C., Chen, K., Larson, D. E., McLellan, M. D., Dooling, D., Abbott, R*., et al.* (2012). Clonal architecture of secondary acute myeloid leukemia. N Engl J Med *366*, 1090-1098.

Wang, J. C., Doedens, M., and Dick, J. E. (1997). Primitive human hematopoietic cells are enriched in cord blood compared with adult bone marrow or mobilized peripheral blood as measured by the quantitative in vivo SCID-repopulating cell assay. Blood *89*, 3919-3924.

Wang, J. C., Lapidot, T., Cashman, J. D., Doedens, M., Addy, L., Sutherland, D. R., Nayar, R., Laraya, P., Minden, M., Keating, A*., et al.* (1998). High level engraftment of NOD/SCID mice by primitive normal and leukemic hematopoietic cells from patients with chronic myeloid leukemia in chronic phase. Blood *91*, 2406-2414.

Wang, Y., and Leung, F. C. (2004). An evaluation of new criteria for CpG islands in the human genome as gene markers. Bioinformatics *20*, 1170-1177.

Wang, G. G., Cai, L., Pasillas, M. P., and Kamps, M. P. (2007). NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. Nat Cell Biol *9*, 804-812.

Wang, Y., Krivtsov, A. V., Sinha, A. U., North, T. E., Goessling, W., Feng, Z., Zon, L. I., and Armstrong, S. A. (2010a). The Wnt/beta-catenin pathway is required for the development of leukemia stem cells in AML. Science *327*, 1650-1653.

Wang, Z., Iwasaki, M., Ficara, F., Lin, C., Matheny, C., Wong, S. H, Smith, K. S., and Cleary M. L. (2010b). GSK-3 promotes conditional association of CREB and its coactivators with MEIS1 to facilitate HOXmediated transcription and oncogenesis. Cancer Cell *17*, 597-608.

Wang, J., Ray, P. S., Sim, M. S., Zhou, X. Z., Lu, K. P., Lee, A. V., Lin, X., Bagaria, S. P., Giuliano, A. E., and Cui, X. (2012). FOXC1 regulates the functions of human basal-like breast cancer cells by activating NF-kappaB signaling. Oncogene *31*, 4798-4802.

Ward, P. S., Patel, J., Wise, D. R., Abdel-Wahab, O., Bennett, B. D., Coller, H. A., Cross, J. R., Fantin, V. R., Hedvat, C. V., Perl, A. E*., et al.* (2010). The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. Cancer Cell *17*, 225-234.

Weishaupt, H., Sigvardsson, M., and Attema, J. L. (2010). Epigenetic chromatin states uniquely define the developmental plasticity of murine hematopoietic stem cells. Blood *115*, 247-256.

Weissman, I. L. (2000). Translating stem and progenitor cell biology to the clinic: barriers and opportunities. Science *287*, 1442-1446.

Welch, J. S., Ley, T. J., Link, D. C., Miller, C. A., Larson, D. E., Koboldt, D. C., Wartman, L. D., Lamprecht, T. L., Liu, F., Xia, J.*, et al.* (2012). The origin and evolution of mutations in acute myeloid leukemia. Cell *150*, 264-278.

Wheatley, K., Brookes, C. L., Howman, A. J., Goldstone, A. H., Milligan, D. W., Prentice, A. G., Moorman, A. V., and Burnett, A. K. (2009). Prognostic factor analysis of the survival of elderly patients with AML in the MRC AML11 and LRF AML14 trials. Br J Haematol *145*, 598-605.

Wiseman, D. H., Greystoke, B. F., and Somervaille, T. C. (2014). The variety of leukemic stem cells in myeloid malignancy. Oncogene *33*, 3091-3098.

Wong, O. (1995). Risk of acute myeloid leukaemia and multiple myeloma in workers exposed to benzene. Occup Environ Med *52*, 380-384.

Wouters, B. J., Lowenberg, B., Erpelinck-Verschueren, C. A., van Putten, W. L., Valk, P. J., and Delwel, R. (2009). Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. Blood *113*, 3088-3091.

Wu, A. M., Till, J. E., Siminovitch, L., and McCulloch, E. A. (1968). Cytological evidence for a relationship between normal hemotopoietic colony-forming cells and cells of the lymphoid system. J Exp Med *127*, 455-464.

Wu, M., Wang, P. F., Lee, J. S., Martin-Brown, S., Florens, L., Washburn, M., and Shilatifard, A. (2008). Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Mol Cell Biol *28*, 7337-7344.

Wu, H., and Zhang, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell *156*, 45-68.

Wunderlich, M., Chou, F. S., Link, K. A., Mizukawa, B., Perry, R. L., Carroll, M., and Mulloy, J. C. (2010). AML xenograft efficiency is significantly improved in NOD/SCID-IL2RG mice constitutively expressing human SCF, GM-CSF and IL-3. Leukemia *24*, 1785-1788.

Xia, L., Huang, W., Tian, D., Zhu, H., Qi, X., Chen, Z., Zhang, Y., Hu, H., Fan, D., Nie, Y., and Wu, K. (2013). Overexpression of forkhead box C1 promotes tumor metastasis and indicates poor prognosis in hepatocellular carcinoma. Hepatology *57*, 610-624.

Xie, H., Xu, J., Hsu, J. H., Nguyen, M., Fujiwara, Y., Peng, C., and Orkin, S. H. (2014). Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. Cell Stem Cell *14*, 68-80.

Xu, Y., Shao, Q. S., Yao, H. B., Jin, Y., Ma, Y. Y., and Jia, L. H. (2014). Overexpression of FOXC1 correlates with poor prognosis in gastric cancer patients. Histopathology 64, 963-970.

Yang, Y., Wang, W., Cleaves, R., Zahurak, M., Cheng, L., Civin, C. I., and Friedman, A. D. (2002). Acceleration of G(1) cooperates with core binding factor beta-smooth muscle myosin heavy chain to induce acute leukemia in mice. Cancer Res *62*, 2232-2235.

Yokoyama, A., and Cleary, M. L. (2008). Menin critically links MLL proteins with LEDGF on cancer-associated target genes. Cancer Cell *14*, 36-46.

You, J. S., and Jones, P. A. (2012). Cancer genetics and epigenetics: two sides of the same coin? Cancer Cell *22*, 9-20.

Yu, M., Bardia, A., Wittner, B. S., Stott, S. L., Smas, M. E., Ting, D. T., Isakoff, S. J., Ciciliano, J. C., Wells, M. N., Shah, A. M., et al. (2013). Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science *339*, 580-584.

Zhang, B., Strauss, A. C., Chu, S., Li, M., Ho, Y., Shiang, K. D., Snyder, D. S., Huettner, C. S., Shultz, L., Holyoake, T., and Bhatia, R. (2010). Effective targeting of quiescent chronic myelogenous leukemia stem cells by histone deacetylase inhibitors in combination with imatinib mesylate. Cancer Cell *17*, 427-442.

Zhang, S. S., Kim, K. H., Rosen, A., Smyth, J. W., Sakuma, R., Delgado-Olguin, P., Davis, M., Chi, N. C., Puviindran, V., Gaborit, N., et al. (2011). Iroquois homeobox gene 3 establishes fast conduction in the cardiac His-Purkinje network. Proc Natl Acad Sci U S A *108*, 13576-13581.

Zhou, X., Maricque, B., Xie, M., Li, D., Sundaram, V., Martin, E. A., Koebbe, B. C., Nielsen, C., Hirst, M., Farnham, P., et al. (2011). The Human Epigenome Browser at Washington University. Nat Methods 8, 989-990.

Zuber, J., Radtke, I., Pardee, T. S., Zhao, Z., Rappaport, A. R., Luo, W., McCurrach, M. E., Yang, M. M., Dolan, M. E., Kogan, S. C.*, et al.* (2009). Mouse models of human AML accurately predict chemotherapy response. Genes Dev *23*, 877-889.