The Role Of The Forkhead Transcription Factor FOXC1 In Acute Myeloid Leukaemia

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Biology, Medicine and Health

2021

Fabrizio Simeoni

Cancer Research UK Manchester Institute
School of Medical Sciences

LIST OF CONTENT

Word count: 39,903

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

CHAPTER 3: RESULTS – GENERATION OF FOXC1 ANTIBODY

CHAPTER 4: RESULTS – THE ROLE OF FOXC1 IN AML

LIST OF TABLES

Large data tables are available on request.

LIST OF APPENDICES

.

LIST OF ABBREVIATIONS

5hmC	5-hydroxymethylcytosine
5mC	5-methylated cytosines
ACTB	Actin beta
ADPN	Adiponutrin gene
AGM	Aorta-gonad-mesonephros
ALK	ALK Receptor Tyrosine Kinase
AMBIC	Ammonium hydrogen carbonate
AML	Acute myeloid leukaemia
APC	Allophycocyanin
APL	Acute promyelocytic leukaemia
ARID3A	AT-Rich Interaction Domain 3A
ASXL1	Additional sex combs like transcriptional regulator 1
ATAC	Assay for Transposase-Accessible Chromatin
ATO	Arsenic trioxide
ATRA	All-trans retinoic acid
AXL	AXL Receptor Tyrosine Kinase
BB	Biobank
BCA	Bicinchoninic acid assay
BCL2	BCL2 Apoptosis Regulator 2
BLBC	Basal-like breast cancer
BM	Bone marrow
CAR	CXCL12-abundant reticular cell
CBFB	Core-binding factor subunit beta
cDNA	Complementary DNA
CEBPA	CCAAT enhancer binding protein alpha
CEBPB	CCAAT enhancer binding protein beta
CEBPG	CCAAT enhancer binding protein gamma
CFU-S	Colony-forming-unit-spleen
ChIP	Chromatin immunoprecipitation
CLP	Common lymphoid progenitor
CML	Chronic myelogenous leukaemia
CMP	Common myeloid progenitor
CSC	Cancer stem cell
CSF1R	Colony stimulating factor 1 receptor
CXCL12	CXC-chemokine receptor 12
CXCR4	CXC-chemokine receptor 4
ddH_20	Double-distilled water
DEK	DEK proto-oncogene

DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DNMT3A	DNA methyltransferase 3 alpha
DNMT3B	DNA methyltransferase 3 beta
dNTP	Deoxyribose nucleoside triphosphate
DSG	Di(N-succinimidyl) glutarate
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EDTA	Ethylenediaminetetraacetic acid
ELF1	E74-Like Factor 1
EGTA	Ethylene glycol-bis(β-aminoethyl ether)- tetraacetic acid
EP300	E1A binding protein P300
ERG	ETS Transcription Factor ERG
ERK1	Extracellular signal regulated kinase 1
ERK2	Extracellular signal regulated kinase 2
ETS	Erythroblast Transformation Specific
ETV6	ETS Variant Transcription Factor 6
EVI1	Ecotropic Viral Integration Site 1
EZH2	Enhancer of zeste 2 polycomb repressive complex 2
FAB	French-American-British
FACS	Fluorescence-Activated Cell Sorting
FBS	Foetal bovine serum
FC	Flow cytometry
FDA	Food and Drug Administration
FDR	False discovery rate
FHD	Forkhead domain
FLT3	Fms related tyrosine kinase 3
FLT3-ITD	FLT3 internal tandem duplications
FOS	c-Fos proto oncogene
FOXC1	Forkhead box C1
FOXM1	Forkhead box M1
FOXO	Forkhead box O
FPKM	Fragments per kilobase of transcript per million mapped reads
G-CSF	Granulocyte colony-stimulating factor
GATA	GATA-binding factor
GFI1	Growth-factor independent 1

GM-CSFG	Granulocyte/macrophage colony-stimulating factor
GMP	Granulocyte-macrophage progenitor
GREAT	GREAT predicts functions of cis-regulatory regions
GSEA	Gene set enrichment analysis
H1	H1 Histone Family
H2A	H2A Histone Family Member
H2B	H2B Histone Family Member
H3	H3 Histone Family
H3K27ac	Histone H3 lysine 27 acetylation
H3K27me3	Histone H3 lysine 27 trimethylation
H3K36me3	Histone H3 lysine 36 trimethylation
H3K4me1	Histone H3 lysine 4 monomethylation
H3K4me3	Histone H3 lysine 4 trimethylation
H3K9me3	Histone H3 lysine 9 trimethylation
H4	H4 Histone Family
HD	Homeodomain
HDAC1	Histone Deacetylase 1
HDAC2	Histone Deacetylase 2
HEK	Human embryonic kidney
HOX	Homeobox
HSC	Haematopoietic stem cell
HSCT	Haematopoietic stem cell transplantation
HSLB	High salt lysis buffer
HSPC	Haematopoietic stem and progenitor cell
ICC	Immunocytochemistry
ID	Inhibitory domain
IDH1	Isocitrate dehydrogenase 1
IDH2	Isocitrate dehydrogenase 2
IF	Immunofluorescence
IHC	Immunohistochemistry
IKZF1	IKAROS Family Zinc Finger 1
IL-3	Interleukin 3
IL-6	Interleukin 6
IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRF8	Interferon-γ-responsive transcription factor 8
IRX3	Iroquois-class homeodomain protein 3
ITD	Internal tandem duplication
JAK	Janus family kinases
JUN	JUN proto-oncogene

JUNB	JUNB proto-oncogene
KD	Knockdown
KIT	proto-oncogene c-KIT
KLF2	Krüppel-like factor 2
KLF4	Krüppel-like factor 4
KLF6	Krüppel-like factor 6
KLS	Kit+Lin-Sca-1+
KO	Knockout
KRAS	KRAS proto-oncogene
LB	Luria Broth
LDS	Lithium dodecyl sulphate
LEPR	Leptin receptor-expressing cell
LMPP	Lymphoid-primed multipotent progenitor
LSC	Leukaemia stem cells
LT	Long term
LTR	Long terminal repeats
MAX	MYC Associated Factor X
MBCF	Molecular Biology Core Facility
MCRC	Manchester Cancer Research Centre
MDS	Myelodysplastic syndrome
MEIS1	Myeloid Ecotropic Viral Integration Site 1 Homolog
MEK1	Mitogen-activated protein kinase/ERK kinase 1
MEK2	Mitogen-activated protein kinase/ERK kinase 2
MEP	Megakaryocyte-Erythroid progenitor
M-CSF	Macrophage colony-stimulating factor
MLL	Mixed lineage leukaemia
MNC	Mononuclear cell
MPN	Myeloproliferative neoplasm
MPP	Multipotent progenitors
MRC	Medical Research Council (UK)
MRD	Minimal residual disease
MSCV	Murine stem cells virus
MYB	MYB proto-oncogene
MYC	MYC proto-oncogene
NEB	New England Biolabs
NGS	Next-generation sequencing
NK	Natural killer
NKX2-3	NK2 homeobox 3
NP40	Nonyl-phenoxypolyethoxylethanol 40
NPM1	Nucleophosmin 1

NRAS	Neuroblastoma RAS viral
NS	Not significant
NTC	Non-targeting control
PAGE	Polyacrylamide gel electrophoresis
PB	Peripheral blood
PBS	Phosphate buffered saline
PE	Phycoerythrin
PEI	Polyethylenimine
PI3K	Phosphatidylinositol 3-kinases
PKC	Protein kinase C
Plat-E	Platinum-E
PML	Promyelocytic leukaemia
PU.1	PU box-binding transcription factor
qPCR	Quantitative polymerase chain reaction
RARA	Retinoic acid receptor alpha
RIME	Rapid immunoprecipitation mass spectrometry of endogenous proteins
RNA	Ribonucleic acid
RPMI	Roswell Park Memorial Institute
RT	Reverse transcription
RT-PCR	Real-time polymerase chain reaction
RUNX1	RUNX family transcription factor 1
RUNX1T1	RUNX partner transcriptional co-repressor 1
SCF	Stem cell factor
SCL	Stem cell leukemia gene
SDS	Sodium dodecyl sulphate
SEM	Standard error of the mean
SLAM	Signaling lymphocyte activation molecule
SLS	Scientific laboratory Supplies
SM	Staining medium
SMAD	SMAD Family Member
SPI1	Spi-1 proto-oncogene
ST	Short term
STAT	Signal transducer and activator of transcription
SV40	Simian vacuolating virus 40
SWI/SNF	SWItch/Sucrose Non-Fermentable
TAD	Transactivation domains
TAE	Tris acetate EDTA
TET2	Tet methyl-cytosine dioxygenase 2
TF	Transcription factor
TLE3	Transducin-Like Enhancer Of Split 3

TNN	Tris HCl NP40
TP53	Tumour protein P53
TPO	Thrombopoietin
TSS	Transcription start site
TST	Thiostrepton
UPL	Universal Probe Library
UV	Ultraviolet
WHO	World Health Organisation
WT1	Wilms tumour protein 1
α-MEM	Minimum essential medium with alpha modification

The University of Manchester
Fabrizio Simeoni
PhD Thesis

The Role Of The Forkhead Transcription Factor FOXC1 In Acute Myeloid Leukaemia

ABSTRACT

A differentiation block is the cardinal pathologic feature of acute myeloid leukemia (AML) but the underlying mechanisms are incompletely understood. Despite absent expression in normal hematopoietic lineages, the Forkhead family transcription factor FOXC1, which is a critical regulator of normal mesenchymal and mesodermal differentiation, is highly expressed in ~20% of cases of AML where it confers a block to monocyte/macrophage lineage differentiation and inferior outcome. Through integrated proteomics and bioinformatics approaches, I have discovered that FOXC1 interacts with RUNX1 through its Forkhead DNA binding domain and that the two factors co-occupy a discrete set of primed and active enhancers distributed close to monocyte/macrophage differentiation genes. FOXC1 stabilises association of RUNX1, HDAC1 and the Groucho family repressor protein TLE3 at these sites to limit enhancer activity: FOXC1 knockdown induced loss of repressor proteins, gain of CEBPA binding, enhancer acetylation and upregulation of nearby genes, including KLF2. Furthermore, it triggered genome-wide redistribution of RUNX1, TLE3 and HDAC1 from enhancers to promoters leading to repression of self-renewal genes including *MYC* and *MYB*. My studies highlight RUNX1 and CEBPA transcription factor swapping at enhancers and promoters as a key feature of leukemia cell differentiation, and reveal that FOXC1 prevents this by stabilising enhancer binding of a RUNX1/HDAC1/TLE3 transcription repressor complex, to oncogenic effect.

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

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, trademarks 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=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses.

ACKNOWLEDGEMENTS

Firstly, I would like to thank Professor Tim Somervaille, my PhD supervisor, for inviting me to join the Leukaemia Biology group at CRUK Manchester Institute. His faith, support and guidance over the last five years have been essential in these early stages of my scientific career and I will always be grateful for that. As I continue my scientific training, I know that Tim will continue to be there to share his experience, knowledge and insight. I would also like to thank Dr Georges Lacaud, my PhD advisor and Dr Esther Baena, my PhD co-supervisor, for helpful discussions relating to my project and career progression.

I feel really lucky to have worked alongside to all the members of the Leukaemia Biology Group, past and present. First, I must thank Dr Gary Spencer for all the things I have learnt from him during the last five years and for putting a lot of faith and trust in me. I would also like to thank Alba Maiques, Emma Williams, Isabel Romero, Mark Williams, John Chadwick, Francesco Camera, Bettina Wingelhofer, Naseer Basma and Bradley Revell, for all the good times spent together in the lab. Special thanks to Fabio Amaral, for all the assistance provided for analysing my data and for the help in preparing the manuscript.

Thanks to Deepti Wilkes for her efforts in the MCRC Biobank and for providing me with samples to work with. Thanks to Jason Caroll and Evangelia Papachristou for the help in processing the RIME samples.
The work within this thesis would not have been possible without all of the core facilities and members of staff at CRUK Manchester institute. In particular, the molecular biology facility has been essential for this work and I thank everyone who made it possible. Thanks to the CRUK Manchester institute and Richard Marais for funding my PhD.

Most of all, I would like to thank my family for their ever-present support and love and my best friends that made this experience in Manchester amazing. I wouldn't be as happy as I am now without you all. THANKS!

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 from Professor Tim Somervaille and Dr. Fabio Amaral.

Chapter 1: Introduction

1.1 Haematopoiesis

Haematopoiesis is the physiological process by which all the blood cellular components are formed. In mammals, blood cells originate from the mesodermal layer of the embryo and their production arises in at least two waves: one within the yolk sac (primitive haematopoiesis) and the second within the embryo proper (definitive haematopoiesis) (Orkin and Zon, 2008). The primary function of primitive haematopoiesis is to support the rapid growth of the embryo by generating erythroid cells to facilitate tissue oxygenation. Definitive haematopoiesis is, on the contrary, located in different sites. It is first observed in the aorta-gonad-mesonephros (AGM) region and subsequently involves the colonisation of distinct anatomical regions, such as 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). In a healthy adult person, approximately 10^{11}-10^{12} new blood cells are produced every day, derived from a rare mostly quiescent cell population called haematopoietic stem cells (HSCs) (Jagannathan-Bogdan and Zon, 2013).

HSCs are clonogenic cells that are capable of self-renewal (the production of additional HSCs) and differentiation (the process by which they give rise to all blood cell lineages). A major obstacle in characterising this population is that the cells are extremely rare, accounting for 0.003% of all the cells in the bone marrow (Morrison and Spradling, 2008). The adult bone marrow is the primary site of haematopoiesis and HSCs depend on perivascular and endosteal microenvironmental niches for the regulation of self-renewal and differentiation. Niche cells produce critical factors such as stem cell factor (SCF), chemokines (CXCL12) and adhesion molecules regulating the fate of the HSCs, in particular, their maintenance and their localisation to the BM (Morrison and Scadden, 2014).

1.1.1 Classical model of haematopoiesis

The classical model of haematopoiesis was first developed through mouse transplantation studies and proposes that blood cells are essentially divided into two branches: myeloid and lymphoid. The myeloid lineage includes fully differentiated, short-lived cell types such as granulocytes, monocytes, erythrocytes and megakaryocytes. The lymphoid branch is formed by T, B and natural killer (NK) cells. Despite this diversity, all blood cells arise from a common stem cell with multipotent capacity. Human HSCs usually divide only once or twice a year, producing highly proliferative progenitor cells that become lineage-restricted and eventually terminally differentiate into mature progeny (Doulatov et al., 2012). A simplified scheme of the classical model is shown in Figure 1.

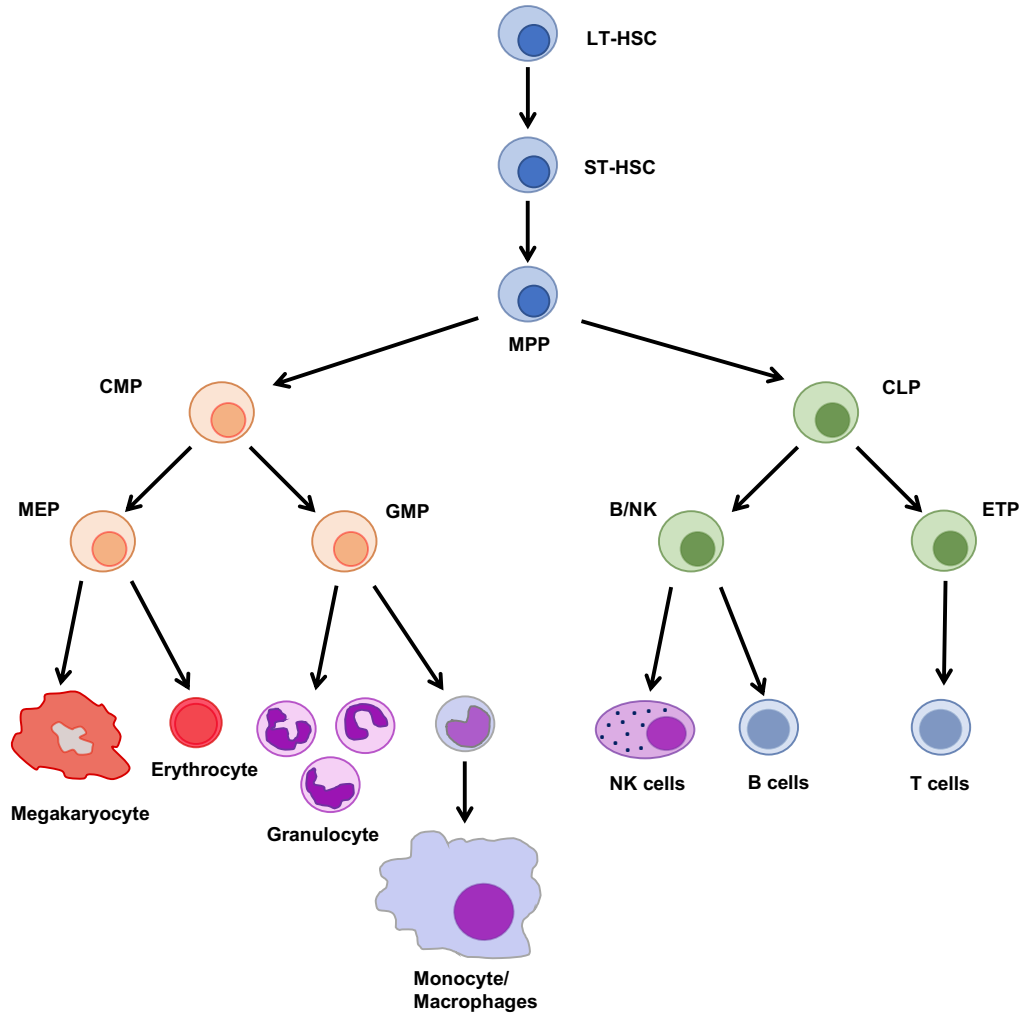


Figure 1. The haematopoiesis hierarchy.

Schematic representation of lineage commitment and differentiation during normal human haematopoiesis. LT-HSC - long-term haematopoietic stem cell; ST-HSC - short-term haematopoietic stem cell; MPP - multipotent progenitor; CMP - common myeloid progenitor; CLP - immature lymphoid progenitor; MEP - megakaryocyte/erythroid progenitor; GMP - granulocyte/monocyte progenitor; B/NK - B/NK cell progenitor; ETP - earliest thymic progenitor; NK - natural killer cells. Image adapted from Doulatov et al. (2012).

In this hierarchical model, HSCs go through successive symmetric and asymmetric divisions, leading to progressive loss of their self-renewal capacity and the emergence of lineage commitment characteristics. The HSC compartment is divided into two subsets according to their capacity for self-renewal. The long-term HSCs (LT-HSCs) are the most immature and quiescent HSCs. Mouse BrdU cell-labelling studies estimated that they divide once every 30-50 days (Kiel et al., 2007), while an even more primitive subpopulation of LT-HSCs is found to divide only five times in a mouse's lifetime (Wilson et al., 2008). LT-HSCs are not easy to study in humans because label-retaining cell studies cannot be performed. Despite these difficulties, human LT-HSCs have been assessed with other techniques such as flow cytometry and the vast majority are in the G_0 phase of the cell cycle (Shepherd et al., 2004). Other studies have estimated a time of division for human LT-HSCs of once every 175-350 days (Catlin et

al., 2011). It is clear then that mouse and human LT-HSCs cannot be the source of daily blood production. Instead, they are thought to serve as a reservoir in case of injury. In fact, when the bone marrow is damaged, they normally resume dividing in order to repopulate the niche and regenerate multipotent progenitors which in turn repopulate mature blood cell lineages. Short-term HSCs (ST-HSCs) by contrast still retain multipotency, but they are actively cycling cells and have a lower self-renewal capacity.

The next step of the haematopoietic hierarchy is the generation of non-self-renewing multipotent progenitors (MPPs) with multilineage differentiation potential. The earliest commitment decision occurs downstream of the MPPs, when the cells start proliferating to meet the enormous daily need of blood cell production and diverge into the common myeloid progenitors (CMP) and the common lymphoid progenitors (CLPs). The former produces the myeloid branch of the blood system, including granulocytes (eosinophils, neutrophils, basophils and mast cells), monocytes/macrophages, erythrocytes and megakaryocytes. The latter instead gives rise to B, T and NK cells, representing the lymphoid part of the system (Doulatov et al., 2012). CMPs retain potency to generate either myeloid or erythroid cells following progression into one of two more specific compartments: the megakaryocyte/erythroid progenitors (MEPs) and the granulocyte/monocyte progenitors (GMPs).

1.1.2 Updated models of haematopoiesis

While the classical view of the haematopoietic hierarchy has been of great value, it is an oversimplification of the real process. The robust development of advanced single-cell technologies and improved cell sorting strategies have started to change our understanding of blood cell production and have uncovered significant diversity of potential within the HSC compartment. Current models suggest that the HSCs are more likely to be a heterogeneous group of cells with varying developmental capacities and different lymphoid/myeloid potentials, rather than a homogeneous group of cells with similar characteristics. For example, a novel subpopulation of progenitors was identified by three different research groups. These cells can generate lymphoid and myeloid cells but not erythrocytes, and for this reason, they were named lymphoid-primed multipotent progenitors (LMPPs) (Adolfsson et al., 2001). LMPPs are characterised by the expression of lymphoid marker genes and FMS-like tyrosine kinase (FLT3) but retain myeloid differentiation potential (Adolfsson et al., 2001; Forsberg et al., 2006). The identification of this population goes against the binary concept of the classical model of haematopoiesis and shows that the true situation is more complex. Therefore, the classical textbook model of a lymphoid–myeloid dichotomy has required revision.

Moreover, most of the knowledge that we have about this complicated system was derived from transplantation assays using irradiated recipient mice. Thus, it may be more representative of blood development under stress-like conditions such as BM

injury rather than normal steady-state haematopoiesis. Recent studies using genetic labelling strategies to mark blood production have yielded some surprising insights. This work has shown that most ST-HSCs and MPPs are already biased towards an early myeloid or lymphoid lineage and these are the main contributors of normal haematopoiesis in steady state (Sun et al., 2014). Meanwhile, LT-HSCs instead have a more critical role in repopulating the bone marrow, particularly following its ablation.

1.1.3 HSC niche

HSCs must be maintained throughout life in order to form progenitor cells and to regenerate all the mature blood cells after stresses such as infection or blood loss. They are mainly present in the BM in adult mammals, although spleen and liver can transiently be haematopoietic sites in response to stress. The bone marrow is a semi-solid tissue located in the central cavity of the bones (Figure 2) and it is formed by a variety of different cell types. The blood cells within the marrow are primarily HSCs and haematopoietic progenitors, whereas the most common non-blood cell types present in the BM are endothelial cells, osteoblasts, adipocytes, nerve fibres, and multiple populations of mesenchymal stromal cells (Sugiyama and Nagasawa, 2012). It is widely accepted that the interactions and crosstalk between HSCs and the niche cells are essential contributors to HSC fate. In particular, niche cells are important for the maintenance of HSCs and their location within the niche itself (Desterke et al., 2015). Here, I will focus in more detail on the role of the bone marrow niche in haematopoiesis.

Two well-defined HSC niche compartments have been described: the 'osteoblastic' niche and the 'vascular' niche. LT-HSCs are located towards the outer edge of the BM and they reside in the endosteal niche. Paracrine signalling from bone-lining cells, including osteoblasts, osteoclasts and osteo-lineage progenitor cells, appears to be essential for maintaining HSCs and keeping them in a quiescent state (Zhang et al., 2003). Moreover, the hypoxic environment of the niche enforces metabolic dormancy and minimises mutations arising from oxidative stress (Testa et al., 2016).

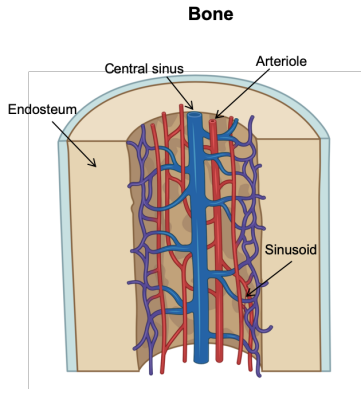


Figure 2. The human bone marrow.

Schematic representation of the bone marrow.

The second significant site of HSC regulation is the vascular niche, which is located around sinusoidal and arteriolar networks and is populated by endothelial cells (Hooper et al., 2009). This niche is highly oxygenated and it is occupied by actively cycling ST-HSCs. Endothelial and mesenchymal cells within the niche are capable of producing multiple HSC-regulatory factors such as stem cell factor (SCF), CXCL12, E-selectin and Notch ligands (Butler et al., 2010). Several studies have pointed to the critical role of BM stromal cells in HSC maintenance. Depletion of most of the cells in the BM, such as osteoblasts (Visnjic et al., 2004), CXCL12-abundant reticular cells (CAR cells) (Omatsu et al., 2010), leptin receptor-expressing cells ($LEPR^+$ cells) (Decker et al., 2017) and many others, lead to HSC activation and subsequently to a strong reduction in their number. CAR cells, for example, are important for the production of two key factors, SCF and CXCL12. CXCL12 promotes HSC maintenance and retention in the BM by activating signalling through CXC-chemokine receptor 4 (CXCR4), which is expressed by HSCs (Zou et al., 1998).

1.1.4 Identification and isolation of HSCs

The first suggestion of the existence of cells with stem cell properties came from Till and McCulloch's experiments in 1961. They discovered that when mouse bone marrow cells were intravenously injected into irradiated recipient mice, a small proportion produced visible colonies in the host spleen. Moreover, the number of colonies was directly correlated to the number of cells injected. The unidentified cells capable of forming these colonies were termed CFU-S (colony-forming-unit-spleen). Additionally, each individual spleen colony was composed of diverse terminally differentiated myeloid cells, and a proportion of cells were capable of forming CFU-S upon transplantation into a secondary recipient (Till and McCulloch, 1961) These findings showed that CFU-S cells displayed multi-lineage differentiation potential and some self-renewal properties. They were initially considered to represent stem cells, although subsequent studies indicated that in fact they were downstream of true LT-HSCs (Becker et al., 2014).

A major thrust of haematopoietic research since the 1960s has been purifying and characterising HSCs, but many difficulties arose due to the similarity of their morphological features with those of white blood cells such as mature lymphocytes. Work started with studies in mice and almost 30 years ago, murine HSC populations of ever increasing purity, were isolated (Spangrude et al., 1988). Their isolation and purification were accomplished due to the development of flow cytometric technology and the increased availability of monoclonal antibodies for specific cell surface markers. Indeed, HSC identification relies on the success of flow cytometry machines to separate cell populations based upon their cell surface marker expression. Thus, for years scientists were looking for unique cell surface molecules expressed by HSCs but not

other bone marrow cells, including closely related hematopoietic progenitor cell counterparts.

Unfortunately, no unique marker that segregates murine HSCs from other cell populations has been found, but a combination of cell surface markers has proved useful for their purification. Almost all HSC isolation methods revolve around cell surface expression of the markers c-Kit and Sca-1; HSCs are negative for markers of mature blood lineages (typically B220, CD4, CD8, Gr-1, Mac-1 and Ter-119). Although this c-Kit+Lin-Sca-1+(KLS) phenotype can enrich for hematopoietic stem cell populations, it still includes progenitor cells in addition to HSCs (Spangrude et al., 1988). Murine HSCs can also be isolated using the SLAM family markers CD150 (also known as Slamf1), CD48 (Slamf2) and CD224 (Slamf4) as $CD150^{+}CD244\text{-}CD48^{-}$ cells. These markers have an important role in confirming the purity of HSCs during sorting experiments. For example, the expression of CD244 is a good marker of MPP and more restricted progenitors and it is not expressed by HSCs, allowing better separation of these two cell types (Kiel et al., 2005).

Similar to murine HSCs, human HSC isolation requires the use of different surface markers and CD34, expressed by only 5% of all blood cells, was the first antigen used (Civin et al., 1984). However, it has been shown that the CD34+ population is heterogeneous. In fact, most CD34+ cells are already lineage-restricted progenitors and thus HSCs are rare (Andrews et al., 1989). *In vivo* experiments using NOD/SCID mice suggested that human HSCs were enriched in a population of CD34+ cells lacking co-expression of CD38 as compared to CD34+ CD38+ cells (Bhatia et al., 1997). In fact, CD38 expression has been used to enrich for ST-HSCs. These cells are able to repopulate the haematopoietic system of NOD/SCID mice in xenoengraftment studies, but only in the short-term (Hogan et al., 2002). CD34+ CD38+ cells are almost all cycling, while CD34+CD38- cells are predominantly quiescent and enriched for LT-HSCs.

In recent years, two labs have proposed other cellular surface makers in order to better isolate human HSCs. The Weissman laboratory demonstrated that CD34+ CD38- CD90+ CD45- Lin- fraction as a highly enriched HSC population, although functional heterogeneity remains (Seita and Weissman, 2010). John Dick's laboratory identified integrin-alpha 6 (also known as CD49f) as a cell surface marker that shows greater enrichment for HSCs compared to CD90 in this population (Notta et al., 2011). Thus, the most recent proposition for a highly enriched human HSC population is the CD34+ CD38- CD49f+ CD45- Lin- fraction.

1.1.5 Molecular basis of myeloid differentiation

The generation of all mature blood cells starting from a single HSC is a highly regulated process and it is associated with two fundamental aspects: reduction of the self-renewal

potential and the acquisition of lineage-specific identity. The molecular basis of blood development and cell differentiation has become a major focus of study for research groups all around the world, particularly with regard to the clinical implications and the possible impact on many aspects of the medical field (transplantations, stem cell therapies and novel druggable targets for many diseases).

Transcription factors (TFs) and epigenetic modifiers are critical regulators of this process, due to their ability to regulate gene expression in a fast, precise and effective manner. Given the focus of my thesis on acute myeloid leukaemia, I will concentrate on the myeloid differentiation branch of haematopoiesis and the essential TFs that drive the highly-connected networks controlling the production of the different mature myeloid cell types (myelopoiesis). Figure 3 shows a schematic representation of myeloid differentiation.

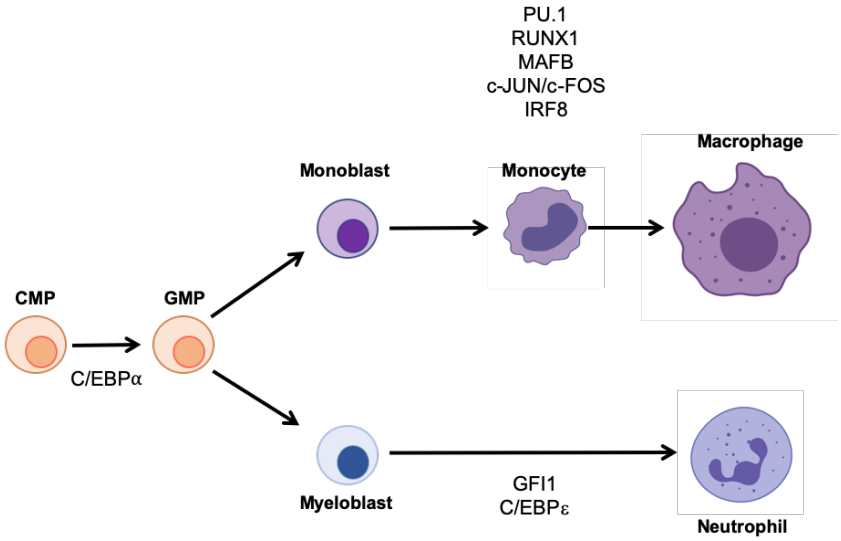


Figure 3. The roles of transcription factors in myelopoiesis.

The most critical regulators of myeloid differentiation are PU.1 and CEBPα, which function in a dose-dependent manner.

The formation of myeloid cells from a single HSC is organised by a relatively small number of TFs: PU.1 (Klemsz et al., 1990), CCAAT/enhancer binding proteins (CEBPα/CEBPβ and CEBPε) (Zhang et al., 1997), growth-factor independent 1 (GFI1) (Hock et al., 2003), interferon-γ-responsive transcription factor 8 (IRF8) (Holtschke et al., 1996) and the runt-related transcription factor 1 (RUNX1) (Okuda et al., 1996). Other factors, such as JUNB (Passegue et al., 2001), FOS (Shafarenko et al., 2004) and MYC (Johansen et al., 1999) are also essential for this process and it is the cooperation of all of these TFs that finely regulates the transition of HSCs into the various myeloid cell types. These transcription factors are able to regulate many myeloid differentiation genes, such as those encoding receptors for macrophage colony-stimulating factor (M-CSF), granulocyte colony-stimulating factor (G-CSF) and the granulocyte/macrophage colony-stimulating factor (GM-CSF) (Tenen et al., 1997).

HSCs are situated at the apex of the haematopoietic hierarchy and their initial formation during embryogenesis is dependent on the expression of the *RUNX1* gene. RUNX1 is a master regulator throughout haematopoiesis and orchestrates (together with SCL) the generation of the HSC pool from earlier stem cells of the mesoderm (Lorsbach et al., 2004). *Runx1* null mice are embryonic lethal and have no detectable haematopoiesis other than primitive erythroid cells, indicating that *Runx1* is essential for the emergence of definitive haematopoiesis (Ichikawa et al., 2004). However, *Runx1* knockout (KO) in adult haematopoietic cells showed that it is partially dispensable at an adult stage. Loss of RUNX1 in adult mice does not substantially alter the frequency of LT-HSCs, but it leads to more lineage-specific defects such as inhibition of CLP production, blocked B and T cell differentiation and reduced platelet formation (Ichikawa et al., 2004). Apart from the stem cell compartment, *RUNX1* is broadly expressed in most adult blood cells and also has a function in myeloid differentiation. In fact, RUNX1 promotes the expression of important myeloid genes including *IL-3*, *CSF1R*, *SPI1* and *CEBPA* (Follows et al., 2003). Notably, RUNX1 is also able to physically interact with different myeloid TFs (PU.1, CEBPα, ERG, GATA1 and GATA2) to finely regulate gene expression (Ran et al., 2013).

Following the haematopoietic hierarchy, the formation of the earliest myeloid transcriptional network relies on the transcription factor PU.1. PU.1 is encoded by the *SPI1* gene and it is a member of the ETS family of TFs and its expression is restricted to blood cells (Klemsz et al., 1990). *Spi1* -/- mice completely lack B cells and mature myeloid cells. Recent studies have shown that PU.1 is required for HSC repopulation and its ablation precludes the formation of CMPs and CLPs (Scott et al., 1994). Therefore, the formation of the first myeloid-bias cell is dependent on this TF, and the level of its expression is another key aspect influencing blood development. Low level of PU.1 is associated with B-cell development, while high levels are necessary for macrophage development (Nutt et al., 2005). On the contrary, T-cell differentiation and erythroid development require downregulation of the *SPI1* gene (Rosenbauer et al., 2006).

The next fate restriction occurs at the passage from CMP to GMP, which is orchestrated by CEBPα. CEBPα is a TF containing a common bZIP DNA-binding domain, shared by all six C/EBP family members (Radomska et al., 1998). *Cebpa* KO mice have a normal number of CMP but completely lack GMP cells and all subsequent granulocytic stages (Zhang et al., 1997). Interestingly, ablation of CEBPα in GMP cells does not affect differentiation into the granulocytic lineage (Zhang et al., 2004). This suggests that CEBPα is crucial for the formation of GMP cells but it is partially dispensable for later stages of granulocytic differentiation.

Following the GMP stage, there is a decision to make: commit to either monocyte/macrophage or granulocyte lineages. This step is again highly regulated by the myeloid transcription factor PU.1 in collaboration with its binding partners. In

particular, PU.1 in collaboration with IRF8 dictates differentiation towards the monocyte/macrophage lineage. IRF8 is only one of the many binding partners of PU.1 that help the latter to enforce the monocyte/macrophage commitment. The others include CEBPα, RUNX1, FOS and JUN, which are all able to orchestrate this process (Tamura et al., 2015).

On the other hand, expression of *CEBPE* and *GFI1* is necessary for granulocytic differentiation. GFI1 is a transcriptional repressor that shares nearly identical zinc-finger domains and a transcriptional repressor domain with its homolog GFI1b. *Gfi1* -/- mice lack neutrophils (Lidonnici et al., 2010). Similar effects were observed in *Cebpe* KO mice with an additional impact on the monocyte lineage (Pundhir et al., 2018).

Taken together, the TFs make up a highly coordinated and ordered gene regulatory network driving the complex process of myelopoiesis. Deregulations and alterations of this network underpin the development of numerous diseases, including blood cancer.

1.2 Acute Myeloid Leukaemia (AML)

Acute myeloid leukaemia is a genetically heterogeneous neoplastic disorder, characterised by a rapid, uncontrolled expansion of immature cells of the myeloid lineage (blasts) in the bone marrow and blood. It is estimated that each year 18,880 new cases are diagnosed in the United States alone and around 10,000 will die of this disease (Zhou and Chng, 2014). AML is the most common acute leukaemia in adults and it primarily occurs in older adults. In fact, the incidence of AML increases with age and the median age at diagnosis is 67 years (Ferrara and Schiffer, 2013). However, paediatric AML also represents 15-20% of cases of all paediatric leukaemia cases (de Rooij et al., 2015),

In AML, the uncontrolled expansion of immature myeloid cells is defined by a block to the normal differentiation program and by an increase in proliferation. The major consequences are disruption of normal haematopoiesis and BM failure as malignant blasts replace healthy haematopoietic tissue. As a result, mature myeloid lineage cells, which include leukocytes, erythrocytes and megakaryocytes cannot be formed and therefore symptoms of AML consist of increased infections, fatigue, and unusual bleeding and bruising caused by neutropenia, anaemia and thrombocytopenia, respectively (Kumar, 2011).

1.2.1 Classification of AML

For any type of cancer, understanding the stage and the characteristics of the disease plays a central role in patient outcome and therapeutic strategy. Therefore, for a neoplastic disorder as heterogeneous as AML, classification into different subgroups is

of huge biological and clinical relevance (Bennett et al., 1976). AML is commonly classified either by the French-American-British (FAB) system or by the World Health Organization (WHO) classification.

The FAB classification uses morphology, cytochemistry and degree of cellular maturation and differentiation to allocate cases into one of eight subtypes, designated M0 to M7. This classification identified one subtype, M3 or acute promyelocytic leukaemia (APL), which was subsequently recognised as a discrete disease entity with a unique therapeutic option and a good prognosis. Table 1 shows the eight subtypes of the FAB classification.

FAB classification		
Subtype	**Description**	**Differentiation**
M0	Acute myeloblastic leukaemia with minimal differentiation	Myeloid
M1	Acute myeloblastic leukaemia without maturation	Granulocytic
M2	Acute myeloblastic leukaemia with maturation	Granulocytic
M3	Acute promyelocytic leukaemia	Granulocytic
M4	Acute myelomonocytic leukaemia	Granulocytic
M5	Acute monocytic leukaemia	Monocytic
M6	Acute erythroid leukaemia	Erythroblastic
M7	Acute megakaryocytic leukaemia	Megakaryocytic

Table 1. FAB classification of acute myeloid leukaemia.
Bennet et al. 1976

Patients with undifferentiated AML (M0) generally have a worse prognosis but this morphological phenotype has lost significance in the era of molecular genetic classification. Subtypes M1 and M2 show predominantly granulocytic differentiation and differ from one another in the level and nature of granulocytic maturation, while M4 shows both a granulocytic and a monocytic phenotype, M5 shows monocytic differentiation, M6 shows erythroblastic differentiation and M7 shows megakaryoblastic differentiation (Bennett et al., 1976).

The FAB classification system is still commonly used to group AML in subtypes and although it provides a consistent morphological and cytochemical framework for

classifying AML, it offers minimal prognostic information and has little to no impact on clinical decisions. Moreover, the rapid emergence of genetic and biological studies, due to the increasing development of new molecular biology techniques, has resulted in novel biological and clinical insights that cannot be entirely appreciated using the FAB criteria alone.

Therefore, in 2002 the World Health Organization developed a new system that includes some of these genetic abnormalities in order to better classify AML. This system is based on morphology, immunophenotype, genetics and clinical features, and classifies AML into seven main groups (Table 2) (Vardiman et al., 2009) (Arber et al., 2016).

The first group, “AML with recurrent genetic abnormalities”, includes AML cases with chromosomal rearrangements or gene mutations that are associated with fairly distinct clinical and morphological features. These events often affect genes that encode for transcription factors or epigenetic modifiers, leading to impaired differentiation of one or more myeloid lineages. “AML with myelodysplasia-related changes” includes AML cases arising from previous myelodysplastic syndrome (MDS) or myelodysplastic/myeloproliferative neoplasm (MDS/MPN), as well as AML cases with cytogenetic abnormalities related to myelodysplasia. “AML not otherwise specified” includes cases that do not fulfil the characteristics of the other subgroups and currently accounts for around 30% of all AML cases. These AML cases mainly rely upon the FAB classification system to clinically define the disease.

WHO classification
Acute myeloid leukaemia 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); *MLLT3-MLL* 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*
Acute myeloid leukaemia with myelodysplasia-related changes
Therapy-related myeloid neoplasms
Acute myeloid leukaemia, not otherwise specified AML with minimal differentiation AML without maturation AML with maturation Acute myelomonocytic leukaemia Acute monoblastic/monocytic leukaemia Acute erythroid leukaemia Acute megakaryoblastic leukaemia Acute basophilic leukaemia Acute panmyelosis with myelofibrosis
Myeloid sarcoma
Myeloid proliferations related to Down syndrome Transient abnormal myelopoiesis Myeloid leukaemia associated with Down syndrome
Blastic plasmacytoid dendritic cell neoplasm

Table 2. The WHO classification.

2016 World Health Organisation (WHO) classification of myeloid neoplasms (adapted from Arber et al.2016).

1.2.2 Treatment

The general therapeutic regimen in patients with AML has not changed substantially over the past 30 years. In particular, almost no improvement has been achieved in the survival rates of older patients, especially the ones aged over 75 (Tallman et al., 2005). AML treatment is typically carried out in two main phases: Induction and Consolidation.

Induction therapy

The first part of treatment is aimed at eradicating leukaemia cells in the bone marrow and blood. The intensity of the treatment depends on different factors such as the patient's age and health. These factors provide insights into whether a patient is eligible

for intensive chemotherapy or whether a less intensive type of treatment is recommended. Induction therapy usually involves the use of cytarabine, a cell cycle-specific agent, in combination with the non-cell-cycle specific anthracycline antibiotic daunorubicin.

A complete response is achieved in 60% to 85% of adults who are 60 years old or younger. In patients who are older than 60 years of age, complete response rates are inferior (40% to 50%) (Burnett et al., 2011). Older patients are more likely to have other cytogenetic abnormalities or co-existing clinical conditions that affect AML treatment. However, no other induction strategy has been shown to be superior. When treatments lead to complete remission, patients undergo a post remission therapy.

Consolidation therapy

This phase aims to lower the chances of leukaemia relapse, by killing any remaining leukaemia cells. Post-remission strategies include conventional chemotherapy and chemo-radiotherapy with allogeneic hematopoietic stem cell transplantation (HSCT) (Ferrara and Schiffer, 2013). Chemotherapy often involves receiving regular injections of medication, such as cytarabine, typically at higher dose. Clinical trial studies have shown that chemotherapy in the consolidation phase of the treatment is beneficial in younger and older adults, although the outcomes for people aged over 60 remain highly unsatisfactory (Tallman et al., 2005). Haematopoietic stem cell transplantation, instead, consists of administration of intensive chemotherapy with or without radiation, and injection of haematopoietic stem cells taken from a matched donor (allogeneic). Not all patients are suitable for this strategy and the decision to undergo transplant is made on a case-by-case basis because treatment-related death rates can be between 10-25% (Tallman et al., 2005). 75% of adults who receive this type of treatment achieve complete remission but most of them will ultimately relapse and die. In fact, in most patients the use of HSCT is not enough to eradicate the minimal residual disease (MRD) which will eventually lead to relapse.

Acute promyelocytic leukaemia

Acute promyelocytic leukaemia is a particular subset of AML and is classified M3 by the FAB system. APL is characterised by an abnormal expansion of immature granulocytes called promyelocytes. This type of blood cancer harbours a chromosomal translocation involving the retinoic acid receptor alpha (*RARA*) gene and one of five different partner genes, by far the most common being the promyelocytic (*PML)* gene on chromosome 15 (Grignani et al., 2000). For these reasons, APL is unique among myeloid leukaemias and is treated in a different way from all other types of AMLs. In fact, due to its sensitivity to all-trans retinoic acid (ATRA), APL represents the most curable AML subtype.

Treatment with ATRA targets the PML-RARα protein, ultimately relieving transcriptional repression and promoting differentiation (Wang and Chen, 2008). Unlike chemotherapies, ATRA does not kill leukaemic cells but induces terminal differentiation, leading to spontaneous apoptosis. Arsenic trioxide (ATO) is another drug proved to be effective for this type of cancer, and it is now routinely used in clinic to treat 20%-30% of patients who manifest disease relapse after initial treatment with ATRA and chemotherapy (Liu and Han, 2003).

Therefore, APL represents the first AML subtype treated with an agent that targets a specific genetic abnormality. The great success of ATRA treatment has created a strong interest in identifying the major genetic abnormalities that lead to other types of AML or more generally leukaemia, in order to evaluate and develop targeted therapies for the different AML subtypes (Dohner, 2007). In 2017-2018 a total of seven new targeted therapies have been FDA-approved in AML: the pan-kinase inhibitor misodtaurin for newly diagnosed *FLT3* mutated AML; the FLT3, AXL, and ALK inhibitor gilteritinib for relapsed or refractory *FLT3*-mutated AML; ivosidenib and enasidenib for patients with relapsed or refractory AML with *IDH1* or *IDH2* mutations respectivetly; the BCL2 inhibitor venetoclax in combination with azacitidine or decitabine or low-dose cytarabine for newly-diagnosed in patients older than 75; the Smoothened receptor inhibitor glasdegib in combination with low-dose cytarabine for newly-diagnosed AML in the elderly; gemtuzumab ozagamicin for newly-diagnosed CD33+ AML alone or in combination with conventional chemotherapy (Wingelhofer and Somervaille, 2019).

1.2.3 Clonal architecture of AML

In the 19th century, it was demonstrated that cancer incorporates a wide range of diseases uniformly characterised by substantial microscopic heterogeneity (Dick, 2009). This discovery underpinned the development of two different models to account for tumour heterogeneity: the stochastic model and the cancer stem cell model. The conventional stochastic model assumes that all cancer cells within a tumour adapt and evolve to produce distinct tumorigenic cells, influenced by intrinsic and extrinsic factors. In contrast, the cancer stem cell model suggests the existence of an infrequent, functionally distinct population of primitive tumour cells that has many properties in common with somatic stem cells. These cells, termed cancer stem cells (CSCs), have the capacity to maintain the bulk tumour population and the ability to generate all cell types within a tumour (Dick, 2009). The first and the best evidence supporting the existence of CSCs came from the study of haematological malignancies.

The discovery that cells isolated from the bone marrow of chronic myelogenous leukaemia (CML) patients harbour the same chromosomal abnormality suggested for the first time that leukaemia may arise from the clonal expansion of a single transformed cell in which a genetic mutation has occurred (Wang and Dick, 2005). Subsequently, the

presence of cells within a tumour with stem cell-like properties was suggested around 20 years ago when scientists demonstrated that only a small subset of leukaemic cells from AML patients was capable of extensive proliferation *in vitro* and of xenoengraftment ability *in vivo*. These cells were termed leukaemic stem cells (LSCs) (Bonnet and Dick, 1997). Later on, studies showed that phenotypically immature CD34+CD38- AML blasts from multiple patients and not the more mature CD34+CD38+ subpopulations, reconstituted leukaemic haematopoiesis in NOD/SCID mice. Moreover, this cell population was also able to initiate the disease in secondary recipients, demonstrating self-renewal capacity. These findings suggested that like the normal haematopoietic system, AML is organized as a hierarchy of different classes of cells, in which AML LSCs sustain the bulk of the clone and LSCs remain infrequent, immature and almost quiescent, similar to normal HSCs (Wiseman et al., 2014).

Contrary to these initial reports, later studies showed a more complex picture of the AML hierarchy. Human LSCs have been reported in several phenotypic compartments, suggesting that they are not limited to the CD34+CD38- fraction. In addition, the development of *in vivo* models, including the generation of more profoundly immunodeficient mice and the use of humanised mice, has resulted in enhanced levels of xenoengraftment with the concomitant presumption of increased LSC frequency.

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 HSCs and very frequent indeed (Somervaille and Cleary, 2006). Similar results were found in a model of human *CEBPA*-mutated AML, where N-terminal CEBPα mutated protein has no effect on HSC expansion but lead to the formation of committed myeloid progenitors with the capacity to generate LSCs (Bereshchenko et al., 2009). These murine models exemplify a principle by which LSCs may exhibit the immunophenotypic properties of a downstream myeloid progenitor rather than those of the conventional normal stem cell compartment and more importantly, they may be actively proliferating cells. Taken together, there are two potentially overlapping models for a LSC to arise from normal cells: 1) normal HSCs gain proliferative potential or 2) more lineage restricted progenitors regain self-renewal capacities.

As discussed above, data support both of the models and therefore, AML is defined as a multistep pathway to malignancy, where mutations can occur in the HSC compartment to generate LSCs, or they can affect a progenitor cell with more lineage restricted characteristics which then undergoes transformation to reinitiate self-renewal capacity.

Pre-leukaemic stem cells

The AML hierarchical structure has been shown to be very heterogeneous and the “cell of origin” of AML is a concept that has been slowly changing over the last few years.

Several studies, in fact, led to the concept of the so called pre-leukaemic stem cell (pre-LSCs). The concept of pre-LSCs acknowledges that the initiating mutations in a normal HSC do not necessarily lead to the formation of a fully-functional LSC but they produce a pre-LSC clone with a proliferative advantage that requires additional mutations for the development of full-fledged disease. For example, Majeti's lab conducted genetic analysis on leukaemia cells and residual normal HSCs isolated from different AML patients and found that residual HSCs are a heterogeneous population, with cells carrying different driver mutations. In fact, several mutations in genes encoding for regulators of DNA methylation, chromatin modification and the cohesin complex (e.g. *IDH1*, *IDH2*, *DNMT3A*) were found in the residual HSCs. These findings indicate that certain mutations may represent the initiating event in the process of leukaemogenesis but are not enough to drive the full progression to leukaemia. Instead, these pre-leukaemic HSCs remain functionally normal until they acquire further mutations such as *FLT3*, *KRAS* or *NRAS* that lead to increased proliferation or a block in differentiation (Majeti, 2012)

Another study supporting the existence of the pre-LSCs comes from AML with t(8:21) rearrangement (Miyamoto et al., 2000). The initiating event and the main driver mutation of this leukaemia is the *RUNX1-RUNX1T1* rearrangement. In remission samples, the fusion *RUNX1-RUNX1T1* transcript was still found in monocytes and B cells. This indicates the presence of a normal pre-leukaemic clone that was not eradicated by the treatment and has full multilineage potential and is able to drive normal haematopoiesis without leading to the initiation of leukaemia.

Together, all these findings show that the AML hierarchy and clonal evolution is a very intriguing process, where the initiating oncogenic event may happen at different stages of haematopoiesis. Figure 4 shows a schematic representation of the AML hierarchy and clonal evolution.

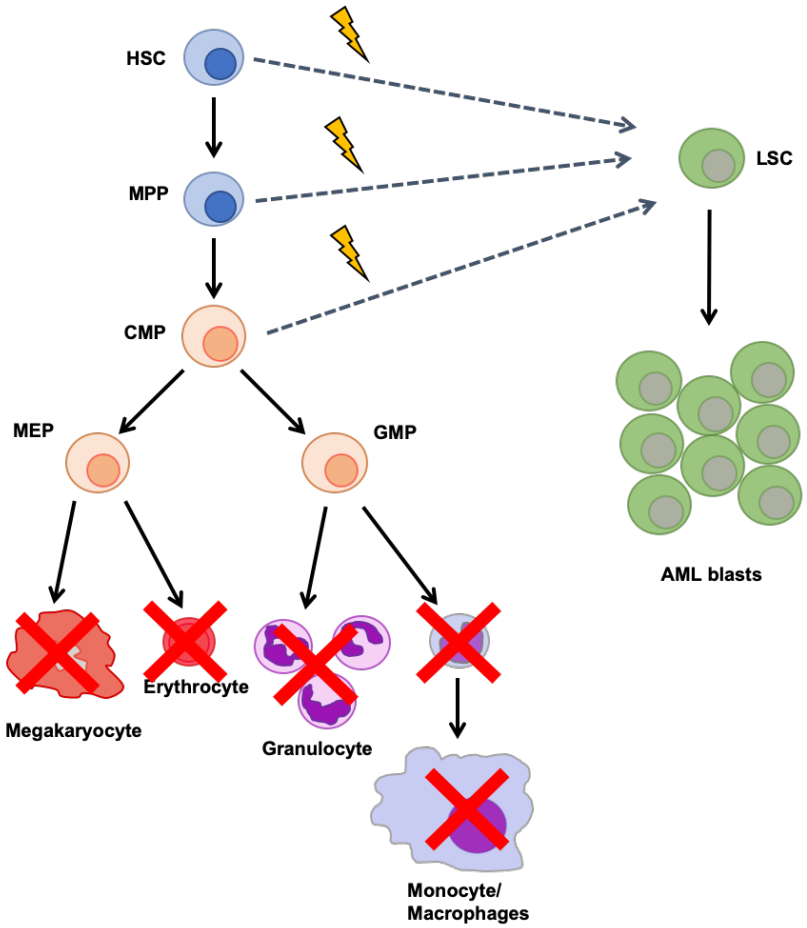


Figure 4. AML hierarchy.

Schematic representation of AML. Normal haematopoietic stem and progenitor cells accumulate genetic mutations and epigenetic changes that 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.

As an overview, initiating mutations may occur in a LT-HSC that over time induces expansion of a mutant pre-leukaemic clone (e.g. *DNMT3A* mutated). The pre-leukaemic clones are usually not malignant and can drive normal haematopoiesis. This phenomenon is called clonal haematopoiesis of indeterminate potential (CHIP), and describes the expansion of a clonal population of HSCs with one or more somatic mutations. The probability of developing haematological malignancies increases when clonal haematopoiesis is present and it is normally linked with ageing (Bowman et al., 2018). These cells, termed pre-leukemic stem cells, give rise to committed progenitors that cycle more frequently and are therefore susceptible to additional genetic or epigenetic events with cooperative functional consequences, leading to the generation of fully transformed LSCs (Weissman, 2005).

On the other hand, certain oncogenic mutations (e.g. *MLL* fusions) are of sufficient potency to confer self-renewal on a downstream myeloid cell; here the

leukaemia initiating cells are the transformed multipotent progenitors and not necessarily the HSCs (Taussig et al., 2008). Despite the different origins of the LSCs, what is functionally important is that these cells have the ability to initiate leukaemia and maintain the tumour burden. This has a profound impact on any clinical applications since LSCs are the major target of therapy.

1.2.4 Molecular Genetics of AML

Leukaemogenesis is a multistep process involving structural and functional changes in multiple genes and ultimately leading to the clonal expansion of immature haematopoietic cells. As already mentioned, transformation can occur in both primitive HSCs and downstream progenitors (Weissman, 2005). For years, an important model of leukaemogenesis was represented by the "two-hit hypothesis", which suggested that AML development requires at least two different types of genetic mutation: one conferring a proliferative advantage (Class 1) and another impairing hematopoietic differentiation (Class 2) (Reilly, 2005). In fact, it is well established that a single mutation alone is not able to cause acute leukaemic transformation. Kelly and Gilliland affirmed that Class 1 mutations alone would only result in myeloproliferative diseases, while Class 2 mutations may lead to myelodysplastic syndromes (Kelly and Gilliland, 2002). It is important to underline that the "two-hit hypothesis" represents just a model created in order to try to summarise a complex system and thus is likely an oversimplification. In fact, other mutations occurring in AML do not belong to either class (such as epigenetic alterations or trisomy 21) but still have a fundamental role in the development of AML (Conway O'Brien et al., 2014).

Present in approximately 50% of AML patients, mutations leading to aberrant activation and proliferation of cellular signalling pathways, referred to as Class 1 mutations, confer a growth advantage on haematopoietic cells (DiNardo and Cortes, 2016) by activating a various number of signal-transduction proteins, including MEK1/2 (mitogen-activated protein kinase/ERK kinase), ERK1/2 (extracellular signal regulated kinase), Janus family kinases (JAKs), signal transducer and activator of transcription (STATs), phosphatidylinositol 3-kinases (PI3K), Akt, protein kinase C (PKC) and SMADs (Miranda and Johnson, 2007). Good examples of Class 1 mutations are *FLT3-ITD*, *NRAS* and *KIT*. On the other hand, Class 2 mutations affect the ability of haematopoietic cells to differentiate and often target transcription factors. The mechanisms by which these mutations act include the inappropriate re-targeting of mutated TFs to oncogenic loci, or disrupting their ability to bind the DNA, which leads to aberrant expression of their targets. *CEBPA*, *RUNX1* mutations and different chromosomal rearrangements fall into the Class 2 category.

As mentioned above, the emergence of new genomic data has unveiled additional classes of mutations that are difficult to group into the previously described

two categories. In particular, the most prominent ones are somatic alterations in genes encoding for epigenetic regulators, which form a separate class, Class 3.

1.2.5 Class 3: epigenetic mutations

The term epigenetics refers to the study of heritable changes in gene expression that occur independently of changes in the primary DNA sequence (Berger et al., 2009). This process is achieved through different chromatin modifications that dictate which genes can be accessed and regulated by the cellular machinery. The basic level of chromatin structure is the "nucleosome", which consists of 147 base pairs of DNA wrapped around an octamer of the four core histones (H2A, H2B, H3 and H4), separated by the linker DNA associated histone H1 (Kornberg, 1974). There are two different types of chromatin regions: heterochromatin, which is a tightly packed form of DNA and primarily contains inactive genes, and euchromatin, which is a lightly packed form of DNA and includes most of the active genes (Bickmore, 2013).

Changes in chromatin structure can be broad, such as the positional sliding of nucleosomes and the replacement of core histones with so-called histone variants, or more localised, occurring on the N-terminal histone tails or directly to the DNA itself. All have important implications for various biological processes including transcription, DNA replication and DNA repair (Gardner et al., 2011). For example, modifications to the histone tail, including acetylation, phosphorylation and ubiquitination, among others, are an important category of epigenetic regulation. The best studied histone tail modifications include lysine acetylation and lysine mono-, di-, or tri-methylation (Wingelhofer and Somervaille, 2019). Histone modifications are strongly associated with transcription activity. For example, histone acetylation at lysine residues is consistently associated with active transcription, whereas histone ubiquitylation is normally associated with repression (Gardner et al., 2011). Conversely, histone methylation is a more complex epigenetic mark and may be associated with activation or repression of transcription. For example, the trimethylation of H3 at lysine 27 (H3K27me3) and 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 (Kouzarides, 2007). Deregulation of the histone code is one of the main features of AML development, associated with aberrant gene expression.

Epigenetic alterations and deregulated expression of chromatin modifiers are frequently observed in cancer and is particularly prominent in haematopoietic malignancies, including AML. These include mutations in tet methyl-cytosine dioxygenase 2 (*TET2*), isocitrate dehydrogenase 1 (*IDH1*), *IDH2*, additional sex combs-like 1 (ASXL1), enhancer of zeste homologous 2 (*EZH2*) and DNA methyl-transferase 3A (*DNMT3A*) (Shih et al., 2012).

DNMT3A is part of a family of three conserved DNA methyltransferases (DNMT1, DNMT3A/3B), that add a methyl group to the fifth carbon of cytosine, which frequently occurs at CpG dinucleotides (Yang et al., 2015). This methylation mark is not easily removed and provides a stable mechanism for silencing gene expression. The role of DNMT3A has been studied in haematopoiesis and *Dnmt3a* -/- mice exhibit LT-HSC expansion (Challen et al., 2012). Moreover, deletion of *Dnmt3a* results in both hypermethylated and hypomethylated regions of DNA, dysregulating expression of crucial genes for self-renewal. Due to its important role in normal blood development, it is not surprising that around 20% of AML patients exhibit *DNMT3A* mutations and they are normally associated with poor prognosis and decreased overall survival (Thol et al., 2011). In AML, somatic alterations occurring in the *DNMT3A* gene involve nonsense or frameshift mutations that lead to a truncated protein or a single missense mutation targeting amino acid R882 ($DNMT3A^{R882}$). The latter has been shown to affect enzymatic activity and affinity for DNA (Bera et al., 2018). *DNMT3A* haploinsufficiency is sufficient to contribute to myeloid transformation and act in a dominant negative fashion to reduce the activity of wild-type DNMT3A. Additionally, it can prompt DNA hypomethylation and increase the binding of activating histone modifiers at enhancer elements. Ultimately, this results in the transcriptional activation of leukaemia-associated genes (e.g. *MEIS1*) resulting in leukaemogenesis (Ferreira et al., 2016).

The *TET2* gene encodes a dioxygenase that oxidises 5-methylated cytosines (5mC) to 5-hydroxymethylcytosine (5hmC), transferring an oxygen atom to the methyl group of 5mC. Once this conversion is made, a series of chemical reactions follow, resulting in DNA demethylation (Rasmussen and Helin, 2016). Loss of function of TET2 has been associated with deregulated gene expression in HSCs, enhancing their self-renewal and promoting aberrant proliferation (Bowman and Levine, 2017). Interestingly, somatic mutations in the *TET2* gene are frequently observed in a wide variety of haematological disease, including both myeloid and lymphoid cancers. Of a particular note, *TET2* is the most frequently mutated gene in MDS, suggesting a critical role in MDS progression (Bowman and Levine, 2017). 7-23% of AML patients present with *TET2* alterations (insertions, deletions, missense, nonsense and frameshift mutations) that result in reduced enzymatic activity. Mutations are most often heterozygous, leading to TET2 haploinsufficiency (Metzeler et al., 2011).

1.2.6 AML with cytogenetic aberrations

Approximately 55% of AML cases present with cytogenetic aberrations, consisting of chromosome alterations such as translocations and inversions. These genetic changes have contributed to AML classification and are important prognostic factors for patient outcome (Dohner, 2007). The majority of abnormal karyotype AML cases are associated with chromosome translocations that lead to gene rearrangement. Table 3 shows the

most common chromosomal aberrations and their corresponding fusion product (Martens and Stunnenberg, 2010). The fusion genes often couple together a transcriptional activator and a repressor complex, leading to the expression of a protein that can still bind DNA through the DNA-binding domain of the transcription activator but has gained repressor function. Therefore, the fusion protein is able to repress genes important for myeloid differentiation, leading to leukaemic transformation. This section briefly covers some of the most common translocations found in AML.

Fusion protein		
Cytogenetics	**Fusion gene**	**Frequency**
t(8;21)	*AML1-ETO*	10%
t(15:17)	*PML-RARA*	10%
inv(16)	*CBFB-MYH11*	5%
der(11q23)	MLL-fusions	4%
t(9:22)	*BCR-ABL1*	2%
t(6;9)	*DEK-CAN*	<1%
t(1;22)	*OTT-MAL*	<1%
t(8;16)	*MOZ-CBP*	<1%
t(7;11)	*NUP98-HOXA9*	<1%
t(12;22)	*MN1-TEL*	<1%
inv(3)	*RPN1-EVI1*	<1%
t(16;21)	*FUS-ERG*	<1%

Table 3. AML-associated oncofusion proteins. Adapted from Martens and Stunnenberg (2010).

Mixed lineage leukaemia (MLL) represents 10% of AML cases harbouring chromosomal abnormalities. In these patients, the MLL gene becomes fused with one of over 50 identified partner genes to create a potent oncogene. Prognosis is often poor for patients harbouring these translocations (Levine, 2013). The MLL DNA-binding domain is able to bring the fusion protein to the DNA while the fusion partner serves as trans-activator, abolishing the normal function of MLL (Slany, 2009). MLL fusions target a specific gene expression program in myeloid leukaemias promoting overexpression of distal *HOXA* genes (*HOXA7-10*) and their cofactor *MEIS1*. Deregulation of the *HOX*

genes and *MEIS*1 ultimately promotes oncogenic properties such as self-renewal and proliferation, spurring leukaemogenesis (Winters and Bernt, 2017).

APL is a biologically and clinically distinct subset of AML associated with chromosomal translocations that result in the fusion of *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. Normally, RARα binds with the retinoid X receptor to form a heterodimer, which acts as a transcriptional repressor recruiting HDAC1. The binding of its ligand, retinoic acid, causes a conformational change that leads to the disruption of the binding to the repressor complex and a switch to a transcription activation activity. Unlike the wild-type variant, PML-RARα is insensitive to physiological concentrations of retinoic acid and therefore maintains interaction with the repressor complex, blocking cell differentiation (Wang and Chen, 2008).

1.2.7 AML with no cytogenetic aberrations

Nearly 45% of AML patients have no cytogenetic alterations. In the last ten years, a great effort has been made to sequence cytogenetically normal AML samples with the aim of finding mutations that explain AML tumour biology and its progression. Many somatic mutations have been identified across several genes, giving us further insight into the genetic complexity of leukaemia, as well as important new prognostic markers. Examples of frequently mutated genes include nucleophosmin 1 (*NPM1*), fms-related tyrosine kinase 3 (*FLT3*), CCAAT/enhancer binding protein alpha (*CEBPA*), myeloid-lymphoid or mixed-lineage leukaemia (*MLL*), neuroblastoma RAS viral, oncogene homolog (*NRAS*), Wilms tumour 1 (*WT1*), and runt-related transcription factor 1 (*RUNX1*) (Mrozek et al., 2007)).

Mutations in the NPM1 gene

55% of AMLs with a normal karyotype carry mutations in *NPM1* that lead to abnormal cytoplasmic localization of the NPM1 protein, altering its normal function. NPM1 is a nucleus-cytoplasm shuttling protein that plays a significant role in many cellular processes such as ribosome biogenesis, centrosome duplication, DNA repair and the regulation of the p53 oncosuppressive pathway (Falini et al., 2007). Mutations in exon 12 of *NPM1* result in the loss of tryptophan residues normally required for NPM1 binding to the nucleoli and in the generation of an additional nuclear export signal motif at the C-terminus (Falini et al., 2007). Approximately 40% of patients with *NPM1* mutations also carry FLT3 internal tandem duplications (*FLT3-ITD*), and several studies have shown that the genotype of mutant *NPM1* without *FLT3-ITD* mutations represents a favourable prognostic marker (Verhaak et al., 2005).

Mutations in the FLT3 gene

The *FLT3* gene encodes a member of the class III receptor tyrosine kinase family and it has been found to be mutated in about 35-45% of normal karyotype AML. Somatic mutations can involve two distinct domains: the juxtamembrane domain and the tyrosine kinase domain. Both genetic changes lead to constitutive activation of the receptor, caused by constitutive receptor autophosphorylation. As a result, FLT3 downstream targets, *STAT5*, *RAS/MAPK* and *PI3K/AKT*, are constantly activated, enhancing cell proliferation (Gilliland and Griffin, 2002). Moreover, the discovery of *FLT3* mutations in AML cases was extremely important from a clinical perspective. In fact, not only are these mutations useful as a prognostic factor, but constitutive autophosphorylation of the FLT3 receptor is an interesting target for molecular therapy. As mentioned in the AML treatment section, midostaurin, a multi-targeted kinase inhibitor, has been FDA approved for newly diagnosed *FLT3* mutated AML. Numerous studies have shown that cytogenetic normal AMLs harbouring *FLT3-ITD* have a significantly inferior outcome compared to patients without *FLT3-ITD* (Falini et al., 2005).

Mutations in key myeloid transcription factors

As discussed in section 1.1.4, myeloid differentiation is a highly regulated process where stage-specific expression and activity of transcription factors play a crucial role in orchestrating the steps that lead to the formation of mature myeloid cells. It is therefore not surprising that disruption of these important TFs results in ectopic activation of proliferation and self-renewal genes and a block in differentiation, leading to AML development. In the next three sections, I will focus on the roles of HOXA9, RUNX1 and CEBPα in leukaemogenesis.

1.2.8 The role of HOXA9 in leukaemogenesis

The clustered homeobox (*HOX*) genes encode transcription factors with key roles in many aspects of development, including controlling the body plan of an embryo along the anterior-posterior axis (Krumlauf, 1994). The *HOX* genes are highly conserved across several species (from Drosophila to Human). The former original *HOX* gene cluster has given rise to 39 genes in mammals through duplications, which are organised into four clusters (A, B, C and D) and located on different chromosomes (Tupler et al., 2001). *HOX* genes encode proteins characterised by a DNA-binding domain (homeodomain, HD) in the C-terminus that contains a conserved tyrosine residue important for the regulation of HOX activity. The ability of HOX proteins to bind to DNA gives them the capacity to regulate gene expression by activating or repressing

DNA transcription. Therefore, the same HOX protein can act as a repressor at one gene and an activator at another (Saleh et al., 2000).

At the N-terminus, HOX proteins include a conserved hexapeptide domain, which is involved in protein-protein interaction. In fact, the presence of this domain gives HOX proteins the capacity to interact with other protein partners, increasing gene regulation ability. The best-known HOX cofactors are transcription factors belonging to PBX and MEIS protein families (Rezsohazy et al., 2015). Some members of the *HOX* gene family have roles in normal haematopoietic development with preferential expression in HSCs and progenitor compartment, while their expression is almost absent in downstream differentiated CD34- cells. In fact, members of the A and B *HOX* gene clusters are most highly expressed in purified subpopulations of human BM cells enriched in stem cells and primitive progenitors, and their expression is absent or reduced in more differentiated cells (Grier et al., 2005). *HOX* genes appear to be essential for maintaining a self-renewing state in the primitive cells of haematopoiesis.

Due to their central role in normal haematopoiesis, deregulation of *HOX* gene function is found in different haematological malignancies and a large proportion of AML subtypes present with this feature. In fact, increased expression of multiple *HOX* genes, as well as the cofactor *MEIS1*, is frequently observed in AML patient samples (Argiropoulos and Humphries, 2007).

HOXA9 is one important member of the HOX transcription factor family that has important roles in segment identity and cell fate during development, but is also essential in adult haematopoiesis (Argiropoulos and Humphries, 2007). *Hoxa9*-deficient mice show severe impairment in the repopulation ability of HSCs, significant deficiencies in myeloid and lymphoid cells and a reduction in the number of circulating CMPs (Lawrence et al., 1997). Conversely, murine *HOXA9* overexpression enhances HSC expansion and myeloid progenitor proliferation. Thus, HOXA9 plays a role in maintaining the self-renewal ability of hematopoietic stem cells and progenitors (Alharbi et al., 2013). Consistent with this, HOXA9 has a key role in the development of AML. Overexpression of *HOXA9* is present in approximately 50% of all AML cases and it is also a predictor of poor prognosis in AML (Lawrence et al., 1999). Many genetic alterations lead to *HOXA9* overexpression suggesting that this gene is part of a common pathway for leukemic transformation. For example, MLL fusion proteins constitutively upregulate *HOXA9*, due to changes in epigenetic markers increasing chromatin accessibility at the *HOXA9* promoter (Ng et al., 2014). Besides MLL fusion leukaemia, upregulation of *HOXA9* is associated with a variety of genetic mutations. For example, cytoplasmic NPM1 is able to upregulate *HOXA9*, *HOXA10* and *MEIS1*, but the precise mechanism is currently unknown (Mullighan et al., 2007). Furthermore, a *HOXA9-NUP98* fusion resulting from chromosomal translocation leads to aberrant gene expression, including *HOX* genes such as *HOXA5*, *HOXA7* and *HOXA9*. Deletions or decreased expression of the EZH2 protein also leads to leukaemia with upregulation of *HOXA9* (Khan et al., 2013).

1.2.9 The role of RUNX1 in leukaemogenesis

RUNX1 is among the most commonly mutated TFs in AML, accounting for 10% of all cases (Gaidzik et al., 2016). This percentage only takes into consideration somatic mutations but as discussed in section 1.2.3, *RUNX1* gene rearrangements are also the most common genetic abnormalities in AML (Speck and Gilliland, 2002). The most common cytogenetic alterations affect *RUNX1* or its essential binding partner *CBFB*, generating a range of chimeric proteins including RUNX1-ETO, RUNX1-EVI1 and CBFβ-SMMHC (Figure 5).

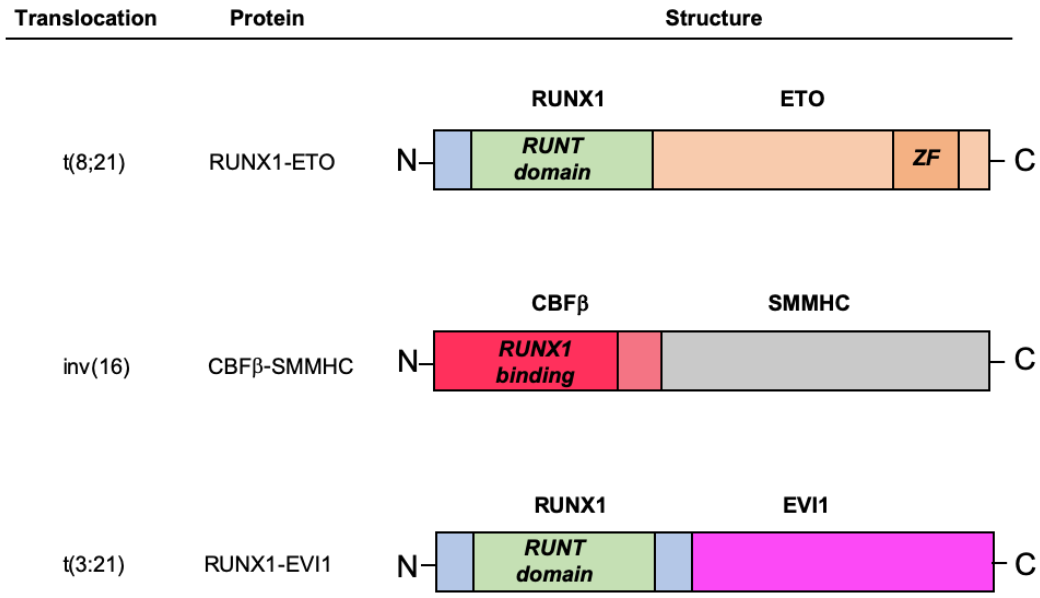


Figure 5. RUNX1 fusion proteins.

RUNX1 and its binding partner CBFβ are frequently rearranged in AML. Figure adapted from Speck and Gilliland (2002).

One of the best-characterised chromosomal translocations in AML is the t(8;21) rearrangement, occurring in approximately 10% of all AMLs. The resulting chimeric protein, RUNX1-ETO, contains the DNA-binding domain of RUNX1 (Runt domain) and almost the entire ETO protein (Miyoshi et al., 1991). The resulting protein lacks the RUNX1 transactivation domain, by which wild-type RUNX1 can recruits co-activators, turning the fusion protein into a strong repressor complex. Chromatin immunoprecipitation (ChIPseq) and immunoprecipitation studies have shown that the NCOR1/HDAC co-repressor complex gets recruited to RUNX1-ETO binding sites near important myeloid genes that are usually activated by wild-type RUNX1, such as *CEBPA* (Pabst et al., 2001; Wang et al., 1998). Moreover, RUNX1-ETO is able to physically interact with several myeloid transcription factors, including wild-type RUNX1, CEBPα

and PU.1, blocking their function and leading to global repression of a myeloid gene expression signature (Vangala et al., 2003). In particular, *CEBPA* has been shown to be an important downstream target of RUNX1-ETO. In fact, *RUNX1-ETO* knockdown (KD) leads to a strong upregulation of *CEBPA* and activation of the myeloid program (Ptasinska et al., 2012). Moreover, numerous studies suggest a strong relationship between the fusion protein and the native RUNX1. Clinical data have shown no inactivating mutation of RUNX1 in t(8;21) AML, suggesting that RUNX1-ETO needs some activity of wild-type RUNX1 in order to function as an oncogene (Ptasinska et al., 2014).

Another chromosomal abnormality that leads to disruption of normal RUNX1 activity is the chromosome 16 inversion, which fuses the RUNX1-binding domain of CBFβ to the coiled-coil domain of SMMHC, generating the *CBFB-MYH11* fusion gene (Castilla et al., 1996). The fusion protein has a dominant negative effect on wild-type CBFβ by disrupting its ability to associate with RUNX1 and stabilise DNA-binding activity (Shigesada et al., 2004).

The third common rearrangement affecting RUNX1 in AML is the t(3;21) translocation, which fuses the DNA binding domain of RUNX1 to the entire EVI1 protein to generate the RUNX1- EVI1 oncoprotein (Nucifora et al., 1994). This chimeric protein also has a dominant negative effect over wild-type RUNX1, by competing for its binding partner CBFβ. RUNX1-EVI1 blocks differentiation by downregulation of myeloid genes including *CTSG*, *CSF1R* and *CEBPA* and by the parallel upregulation of self-renewal genes including *HOXA9* and *MEIS1*.

1.2.10 The role of CEBPα in Leukaemogenesis

CEBPA is a transcription factor with a key role in lineage specification and the differentiation of multipotent myeloid progenitors into neutrophils. The single mRNA transcribed from the *CEBPA* gene is translated into two different proteins, due to the use of an alternative start site usage (Lin et al., 1993). The full-length CEBPα protein (p42) is defined by two transactivation domains (TADs) in the N-terminus, and a DNA binding and dimerization domain at the C-terminus (Lin et al., 1993). The second translation start site leads to the formation of a truncated CEBPα protein (p30), which lacks the first 117 amino acids and therefore does not retain the first TAD. This TAD has been shown to regulate transcriptional activation through interactions with different proteins such as RNA polymerase II preinitiation complex and the TBP/TFIIB complex (Nerlov and Ziff, 1995). The levels of expression of the p42 and p30 isoforms are highly regulated in cells and increased p30 translation is linked with cell proliferation (Calkhoven et al., 2000).

5-10% of AMLs with a normal karyotype are associated with mutations of the *CEBPA* gene (Wouters et al., 2009). *CEBPA* alterations in AML fall into two classes. The first and most common one includes mutations occurring in the 5' end of the gene

and frequently alter the first open reading frame, generating frameshifts that abrogate the full-length p42 isoform but still permit translation of the shorter p30 isoform (Lin et al., 1993). Conversely, the second class comprises mutations that have been identified in the C-terminal part of the protein and occur in the basic region-leucine zipper domain (DNA binding domain/ dimerization domain), leading to a CEPBα protein with decreased DNA-binding capacity or dimerization activity (Gombart et al., 2002). These mutations have been shown to be either mono or biallelic. In general, *CEBPA* mutations predict a more favourable prognosis, at least when both alleles are mutated (Pabst et al., 2009).

The mechanisms through which mutant CEBPα isoforms block myeloid differentiation and induce leukaemia are not completely understood. Mice expressing only p30 as CEBPα protein have GMPs with aberrant self-renewal and develop AML with complete penetrance (Kirstetter et al., 2008) whereas mice expressing CEBPα with C-terminal mutations have only pre-leukaemic features (Porse et al., 2005). The p30 isoform is thought to have a dominant negative function over p42, leading to downregulation of myeloid genes (Schwieger et al., 2004). On the contrary, C-terminal mutations disrupt CEBPα binding to the DNA but the mutated protein is still able to alter the ability of the wild-type protein to activate transcription with cooperation with PU.1 (Kato et al., 2011).

1.3 The Role of Mis-expressed transcription factors in AML

Recent studies in the Somervaille laboratory have led to the discovery of a hitherto unappreciated but very frequent novel pathogenic mechanism in human AML: the tissue-inappropriate derepression of transcription factors (Somerville et al., 2015; Somerville et al., 2018). The authors searched publicly available AML datasets for transcription factor genes highly expressed in LSCs versus normal haematopoietic stem and progenitor cells (HSPCs). They found high expression of different TFs, including *FOXC1*, *IRX3* and *NKX2-3*, in a large number of AML cases. These genes have no known cell-intrinsic role in haematopoiesis and more importantly, they are not normally expressed in any normal blood cell but show inappropriate expression in a significant proportion of AML patients. These findings further led to the discovery that derepression of TFs has a functional role in driving leukaemogenesis and blocking differentiation, the cardinal feature of the disease (Somerville et al., 2015; Somerville et al., 2018). In this thesis, I focus on the role of FOXC1 in AML.

1.3.1 FOXC1

FOXC1 belongs to the Forkhead family of transcription factors, which are characterised by a highly conserved DNA-binding domain of 110 amino acids called the Forkhead domain (FHD) (Jackson et al., 2010). The FHD comprises 4 N-terminal α-helices, 3 β-

strands and 2 C-terminal wings, forming a 'winged helix' structure that has a high affinity for the consensus DNA motif 5'-TGTTTAC-3' (Han et al., 2017). In the past 20 years, more than 100 members of the Forkhead family have been identified and categorised into 19 groups (FOXA to FOXS) on the basis of sequence homology both inside and outside the Forkhead domain (Lam et al., 2013). In mammals, there are 44 *FOX* gene members and the subclasses FOXA, C, O, M and P are the ones that have been well characterised and studied in different biological processes (Figure 6).

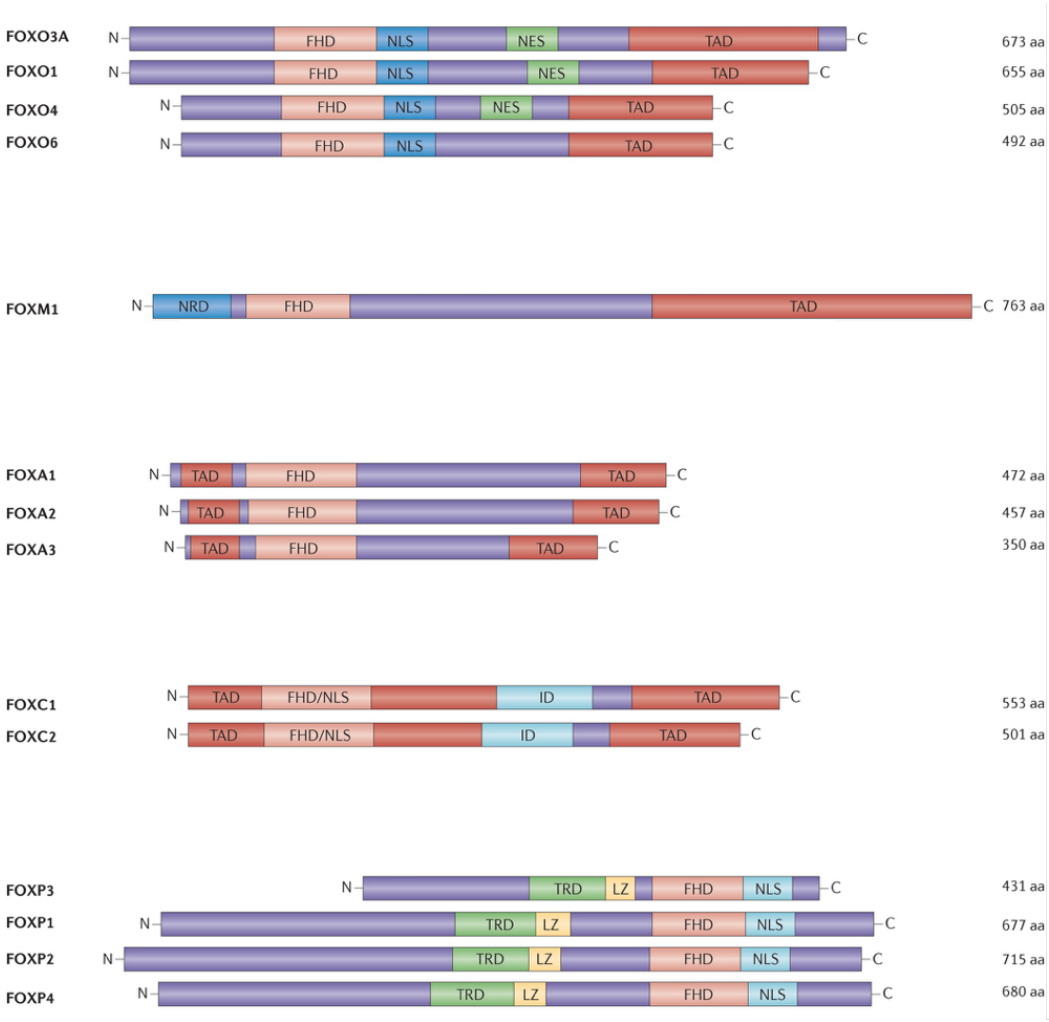


Figure 6. The Forkhead transcription factor family.

As this schematic diagram adapted from Lam et al. (2013) demonstrates, different FOX proteins exhibit considerable diversity in the different domains they contain. However, the presence of the winged-helix forkhead DNA binding domain (FHD) is a unifying feature across the FOX protein family. ID - inhibitory domain; LZ - leucine zipper domain; NES - nuclear export signal; NLS - nuclear localisation signal; TAD - trans-activation domain; TRD - transcriptional repressor domain.

The Forkhead-related factors are fundamental to the regulation of morphogenesis and have several different functions in development such as early establishment of embryonic cell layers, determinations of cell fates in tissues and the control of organogenesis (Kaestner et al., 1993). In the human genome, *FOXC1* is located at a

subtelomeric region at 6p25 and it is normally expressed in paraxial mesoderm, prechondrogenic mesenchyme, neural crest, endothelium and the developing eye and kidney. Despite the high homology and conservation of the Forkhead domain, the remaining part of the FOXC1 protein shares little in common with the other members of the FOX family (Gilding and Somervaille, 2019). Structurally, FOXC1 has two TAD domains that are important for its role as a TF, located at the N-terminus and the C-terminus part of the protein. Moreover, FOXC1 contains an inhibitory domain (ID), which is responsible for protein stability and FOXC1 protein half-life (Berry et al., 2002) (Figure 7).

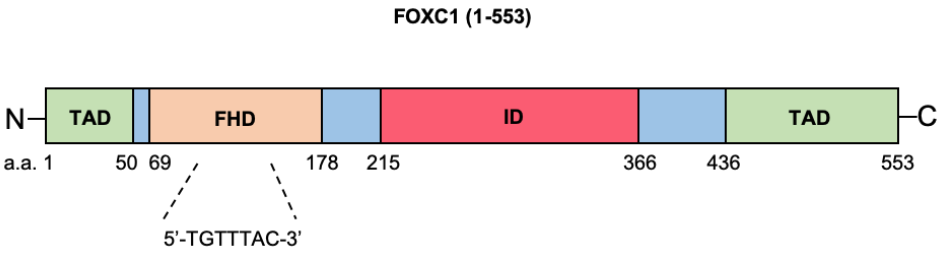


Figure 7. FOXC1 structural domains.

Schematic representation of FOXC1 protein domains. TAD - activating domain; FHD - Forkhead DNA-binding proteins; ID - inhibitory domain. Image adapted from (Gilding and Somervaille, 2019)

Similar to other Forkhead factors, FOXC1 is important throughout development and critically regulates lineage specification and cell fate. For example, *FOXC1* upregulation and expression in the neural crest is required for the correct formation of the neural tube and for promoting an epithelial-mesenchymal transition (Chalamalasetty et al., 2014). As development continues, *FOXC1* expression is essential for the generation of bone and cartilage from the osteogenic and chondrogenic mesenchyme (Sun et al., 2013). In later stages of development, its expression in distinct mesenchymal settings promotes the development of other tissues such as the anterior eye segments, hindbrain, cardiovascular and urinary systems (Kume et al., 2001; Seo et al., 2017).

In accordance with the importance of FOXC1 in mesenchymal differentiation, *Foxc1* null mice die perinatally with bones, heart and kidney abnormalities, hydrocephalus, iris hypoplasia and open eyelids. In humans, inherited haploinsufficiency of FOXC1 due to mutations or deletions causes 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 (Gilding and Somervaille, 2019).

FOXC1 is also an essential determinant of the haematopoietic stem cell niche: its deletion in BM mesenchymal cells leads to marrow hypoplasia. Selective downregulation of *Foxc1* in CAR cells leads to a depletion of the bone marrow niche and a reduction in the number of HSCs (Omatsu et al., 2014). As mentioned above, it is important to note that *FOXC1* is not expressed or required in normal haematopoietic

cells, so these findings shed light on a cell-extrinsic mechanism by which FOXC1 sustains normal blood development.

Overexpression of *FOXC1* has been identified in multiple cancers and can often be linked to poor prognosis. Accumulating evidence demonstrates that FOXC1 plays an essential role in at least 16 different types of cancer, including breast, colorectal, cervical, lung and gastric cancer (Gilding and Somervaille, 2019). FOXC1 function in solid malignancies is quite diverse and it is involved in different aspects of tumour progression such as migration and invasion. High *FOXC1* expression is associated with poor prognosis in a range of solid tumours including hepatocellular carcinoma, pancreatic ductal adenocarcinoma and gastric cancer (Han et al., 2017). Of particular note, a role for FOXC1 has been well described in basal-like breast cancer (BLBC). *FOXC1* KD in various breast cancer cell lines reduces cell proliferation and induces a more differentiated phenotype. Conversely, *FOXC1* overexpression significantly enhanced proliferation of MCF-7 cells, often mediated by upregulation of c-Myc and cyclin D1 (Wang et al., 2017). In addition, *FOXC1* overexpression increases cancer stem cell properties, which are characterised by CD24+CD44+, ALDH+, CD133+ and mammosphere formation (Han et al., 2015).

1.3.2 Forkhead transcription factors in AML

Accumulating evidence suggests that FOX protein deregulation is closely associated with several aspects of AML development, including initiation, progression and drug resistance. For example, the “O” subclass (FOXO1, FOXO3, FOXO4 and FOXO6) is probably the most well-studied of all FOX subfamilies and it has been frequently reported to have a tumour suppressor role in several cancers (Myatt and Lam, 2007). In AML, the consequences of FOXO protein deregulation are quite diverse and the molecular mechanisms underlying their function differ among AML settings. In human t(15:17) AML, FOXO3A has been shown to have a tumour suppressive function. In fact, ATRA treatment reduces phosphorylation of FOXO3A and promotes its translocation into the nucleus, where it upregulates *TRAIL* gene transcription, resulting in cell growth inhibition and induction of apoptosis (Sakoe et al., 2010). The importance of FOXO3A in APL was further confirmed by KD experiments, where *FOXO3A* KD in ATRA-treated NB4 cells inhibits ATRA response and blocks differentiation (Sakoe et al., 2010). FOXO3A also has an oncosuppressor function in *FLT3-ITD* AMLs. The internal tandem duplication of the *FLT3* gene promotes ligand-independent auto-phosphorylation and constitutive tyrosine kinase activity, and therefore, suppression of apoptosis and increased cell division. This signalling pathway activates AKT, which then phosphorylates FOXO3A, relocating it in the cytoplasm. Seedhouse et al. showed that sustained expression of *FLT3-ITD* prevents FOXO3A-mediated apoptosis, and promotes cell survival (Scheijen et al., 2004).

In contrast to the tumour suppressor role, recent studies have shed light on a completely novel mechanism of action for FOXO proteins that suggests a possible oncogenic role. In *RUNX1-ETO* leukaemias, increased levels of *FOXO1* are required to sustain the growth of leukaemia cells, enhancing the self-renewal and blocking differentiation. Moreover, genetic inhibition of FOXO1 eradicates preleukaemic and leukaemic HSCs (Lin et al., 2017).

Beyond FOXO subfamily members, FOXM1 is another example of a FOX member that is closely associated with AML progression, maintenance and drug resistance. *FOXM1* is overexpressed in AML cell lines and primary patient samples and it correlates with overall survival (Nakamura et al., 2010). The importance of FOXM1 in AML is underlined by the evidence that *FOXM1* KD results in a block of the cell cycle and proliferation arrest via downregulation of Aurora kinase B, Survivin, Cyclin B1 and an increase in expression of p21 and p27 (Nakamura et al., 2010). Moreover, the addition of the FOXM1 inhibitor thiostrepton (TST) was able to induce apoptosis of MV4–11 and THP1 cells in a dose-dependent manner, suggesting that FOXM1 may be useful as a potential prognostic marker and therapeutic target in AML (Liu et al., 2014). Additionally, FOXM1 also has a role in treatment failure and resistance to chemotherapy. In fact, using transgenic mouse models, Khan et al. showed that *FOXM1* overexpression induces cell survival after chemotherapy, suggesting FOXM1 as a critical mediator in the emergence of resistant leukaemic clones (Khan et al., 2018).

1.3.3 FOXC1 in AML

Aberrant expression of *FOXC1* was recently identified as a feature of AML development. Somerville et al. (2015) searched for transcription regulators expressed in AML LSC but not normal HSPCs and unexpectedly identified the mesenchymal transcription factor gene *FOXC1* as among the most highly up regulated. Through bioinformatics approaches and experiments with primary patient AML samples, they discovered that *FOXC1* is expressed at high level in ~20% of AML cases, almost invariably in association with high *HOXA/B* expression. *FOXC1* derepression in AML was also associated with *FLT3-ITD*, *NPM1* mutations and t(6:9) translocations (Somerville et al., 2015). Knockdown and over expression experiments confirmed that FOXC1 confers a differentiation block and sustains clonogenic activity. In fact, shRNA-mediated depletion of *FOXC1* expression in THP1 cells led to loss of clonogenic potential and induction of morphological and immunophenotypic differentiation (*FOXC1* knockdown was able to downregulate CD117, a stem cell marker, and up regulate CD11b, a myeloid differentiation marker). As expected, *FOXC1* knockdown in normal HSPC had no effect (Somerville et al., 2015). Thus, FOXC1 contributes to the differentiation block in human AML cells.

Further, *in vitro* studies demonstrated that forced *FOXC1* expression in normal murine HSPCs leads to a transient enhancement of clonogenic potential and a myeloid differentiation block. *In vivo*, *FOXC1* overexpression skewed differentiation towards the myeloid lineage and blocked B cell production, although it was insufficient to induce AML on its own (Somerville et al., 2015). Given that, the authors looked for other factors that FOXC1 may collaborate with and demonstrated that *FOXC1* expression strongly associates with high level of *HOXA9*. To investigate whether this association has a role in leukaemogenesis, murine HSPCs were retrovirally infected with either *Hoxa9* alone or in pairwise combinations with *FOXC1*. HSPCs were also infected with retroviral vectors expressing *Hoxa9* and *Meis1* as a control. MEIS1, as already mentioned above, is a known HOXA9 binding partner. As expected, *Hoxa9* overexpression increased the clonogenic potential of BM HSPCs and the co-expression of *Meis1* accentuated this effect.

Compared to Hoxa9 alone, the co-expression of *Hoxa9* and *FOXC1* significantly enhanced the clonogenic potential and myeloid differentiation block of bone marrow HSPCs, as determined by immunophenotyping and colony morphology. Thus, FOXC1 and Hoxa9 collaborate to enhance clonogenic potential and differentiation block in HSPC. *In vivo* studies demonstrated that irradiated congenic mice, transplanted with *Hoxa9*/*FOXC1* transduced HSPCs, developed AML more rapidly and succumbed to AML significantly earlier than mice receiving *Hoxa9* cells. These findings show that FOXC1 collaborates with HOXA9 to accelerate onset of murine AML with an enhanced morphologic, immunophenotypic and functional differentiation block by comparison with leukaemia cells expressing *Hoxa9* alone.

To study the changes in the murine AML transcriptome that resulted from *FOXC1* overexpression, Somerville et al. 2015 performed exon array analysis using flow-sorted CD117+Gr1+ leukaemia cells recovered from sick mice. The *Hoxa9*/*FOXC1* AML transcriptome was found to be significantly different from those of *Hoxa9*/*Meis1* and *Hoxa9*/MTV AMLs ones (which clustered much more closely). Differential gene expression analysis between *Hoxa9*/*FOXC1* expressing leukaemias versus the two control cohorts showed a group of genes, belonging to immune response, defence response, and inflammatory response that were strongly downregulated due to *FOXC1* overexpression. In particular, transcriptomic analysis revealed that FOXC1 was able to repress a monocytic differentiation program, thereby contributing to the generation of AML. Intriguingly, the transcription factor gene *KLF4*, a positive regulator of monocytic differentiation, was found to be repressed upon *FOXC1* expression. Forced expression of *KLF4* was found to be sufficient to rescue the AML phenotype to some degree in *Hoxa9/FOXC1* murine AML cells thus hinting at a phenotypic link between *FOXC1* and *KLF* expression. Critically, this FOXC1-repressed gene set identified in murine leukaemia exhibited quantitative repression in human AML in accordance with FOXC1 expression levels in patient samples, and FOXC1^{high} human AML cases showed

reduced morphological monocytic differentiation and lower survival rates. Therefore, FOXC1 is frequently derepressed with functional effect in human AML.

1.4 Project Aims

AML is a hierarchically-organized, clonal neoplastic disorder sustained by a subpopulation of cells with long-term proliferative potential, often termed LSCs. In recent years, researchers have been working together to understand the genetic, epigenetic, and transcriptional differences between AML cells and their normal cell counterparts. The long-term goal is to develop therapeutic strategies that selectively target leukemia cells. One important method is to prospectively isolate AML cells with immature progenitor immunophenotypes (immunophenotypic LSCs) and compare their transcriptional profiles with normal hematopoietic stem and progenitor cells (HSPC). In view of the important role of transcription factors in many cellular processes such as cell fate regulation, cell maintenance and cell cycle, it was important to compare the expression of these proteins in a normal or malignant context.

The derepression of the mesenchymal transcription factor *FOXC1* in human AML is a new oncogenic mechanism operating in approximately 20% of patients with the disease, and 30% of AML patients who express high level HOX gene expression. However, the mechanism by which FOXC1 confers a differentiation block is unknown. Given the frequency and significance of tissue-inappropriate expression of FOXC1 in AML, the ability to design novel therapies to relieve a FOXC1-imposed differentiation block and promote leukaemia cell differentiation could benefit patient survival.

The goal of the work presented in this thesis is to answer the question, how does misexpression of *FOXC1* in human AML contribute to a monocyte/macrophage differentiation block? Related questions include where does FOXC1 bind to in the leukaemia cell genome, and with what binding partners? Through addressing these and other questions, I hope to understand why derepression of *FOXC1* contributes to inferior survival and outcome in human AML and identify potential therapeutic approaches that may in future be developed for the benefit of patients.

Chapter 2: Materials and Methods

2.1 Cell culture

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

R10

Roswell Park Memorial Institute (RPMI) 1640 medium (Sigma Aldrich) supplemented with 10% foetal bovine serum (FBS), 2mM L-glutamine (Sigma Aldrich) and 10μ/mL penicillin and streptomycin (Sigma Aldrich).

D10

Dulbecco's Modified Eagle Medium (DMEM; Sigma Aldrich) supplemented with 10% foetal bovine serum (FBS), 2mM L-glutamine (Sigma Aldrich) and 10μ/mL penicillin and streptomycin (Sigma Aldrich).

Primary AML culture medium

Minimum Essential Medium with alpha modification (α-MEM) supplemented with 12.5% heat-inactivated FBS, 12.5% heat-inactivated horse serum, 2mM L-glutamine (Sigma Aldrich), 57.2μM β-mercaptoethanol (Sigma Aldrich), 1μM hydrocortisone (Sigma Aldrich) and 20ng/mL IL-3, G-CSF and TPO (PeproTech, Rocky Hill, NJ).

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. Splitting was carried out 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, trypsin was removed by washing cells in culture medium.

HEK 293FT cells

The 293FT cell line is originally derived from human embryonic kidney (HEK) cells and it was optimised for the generation of high titre lentivirus. Constitutive expression of SV40 large T antigen, in a fast-growing clone of these cells (HEK293F) permits episomal

replication of lentiviral plasmids containing the SV40 origin of replication (Life Technologies).

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. The potent EF1α promoter drives high-level expression of these genes, resulting in the production of high titre retrovirus that is capable of infecting murine cells. Plat-E cells were a gift from Toshio Kitamura, University of Tokyo.

2.1.4. Suspension cells

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, confirmed mycoplasma-free and authenticated by short tandem repeat DNA profiling in-house by the Cancer Research UK Manchester Institute Molecular Biology Core Facility.

Fujioka cells

Fujioka cells derived from a child with acute monocytic leukaemia and exhibit a t(10;11) translocation indicative of a *CALM-AF10* fusion, as well as mutations in *NRAS*, *ETV6*, *TP53* and *EZH2* among others (Table 18) (Hirose et al., 1982; Narita et al., 1999). Cells were cultured in R10 medium and were split into fresh medium every 72 hours to maintain cell density below 1 x 10^6 cells/ml.

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, 1-2x10^6 cells were were resuspended in 1mL cold freeze mix (90% FBS, 10% DMSO (Sigma Aldrich)) and transferred to cryovials (Nunc – www.thermofisher.com). Samples were placed in a Mr. Frosty™ freezing container (Nalgene – www.nalgenelabware.com) at -80°C for 24 hours, at which point they were transferred to liquid nitrogen for long-term storage.

2.1.5 Cell thawing

Cell lines

Cryovials were removed from liquid nitrogen and thawed quickly in a 37°C water bath. Cells were then transferred to a suitable centrifuge tube and made up to 10mL with the pre- warmed culture medium and centrifuged for 4 minutes at 290xg. The pellet was resuspended in the appropriate culture medium for plating in a culture flask.

Primary cells

The Manchester Cancer Research Centre (MCRC) Tissue Biobank collects and stores peripheral blood and bone marrow from patients attending The Christie Hospital NHS Foundation Trust. Patients with acute myeloid leukaemia (AML) are routinely approached at presentation and given the opportunity to donate material. Consenting patients typically donate multiple times across the course of their disease providing material from presentation, following therapy and at relapse. Each donating patient is given a unique Biobank number (prefixed with BB) and collections are numbered chronologically with a suffix that indicates the type of sample taken (B for blood, M for bone marrow). For examples, BB475 1M refers to bone marrow taken at the first time point for patient 475. Biobank staff also collect anonymised clinical data which are maintained on a custom designed StarLIMS database (Abbot Informatics, Hollywood, FL). These data were used to identify relevant samples for each experiment.

Peripheral blood and bone marrow samples were collected into lithium heparin and processed by the MCRC Biobank's scientific officer. Mononuclear cells (MNCs) were separated by density gradient centrifugation. Briefly, samples were diluted with phosphate buffered saline (PBS) containing 40u/mL heparin, layered onto an equal volume of Lymphoprep^{TM} (Axis-Shield, Dundee, UK) and centrifuged at 400g for 30 minutes at 20°C with free deceleration. MNCs were aspirated from the Lymphoprep^{TM}/plasma interface, washed with PBS and resuspended in 4.5% human albumin solution with 10% DMSO. Samples were then aliquoted into cryovials, labeled with a unique barcode and cryopreserved.

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. The cells were centrifuged at 200xg for 10 minutes and after discarding the supernatant, the pellet was washed twice in PBS and cells were resuspended in the primary AML medium.

Cell counting and Trypan Blue exclusion

Live cells were enumerated by mixing 10µL of cell suspension with 10µL Trypan Blue

stain (Thermofisher, Waltham, MA) and loading 10µL of the mixture on a Neubauer BS.748 haemocytometer chamber (Hawksley – www.hawksley.co.uk). Viable cells were identified using light microscopy on the basis of their ability to exclude Trypan Blue.

2.1.6 Cell viability assays

$5x10^3$ cells were placed in each well of a flat-bottomed 96-well plate (Corning) with R10 medium. Plates were incubated for 72hr at 37°C. 20µl of 140µg/mL resazurin (Sigma Aldrich) was added to each well. Plates were then incubated for a further four hours and read using a POLARstar Omega plate reader (BMG Labtech, Aylesbury, UK).

2.1.7 Compounds

Compounds were dissolved in dimethyl sulfoxide (DMSO), aliquoted and stored at -20°C. Final DMSO concentrations were <0.5% in all experiments. The compound used in this study is Ro5-3335 (219506; Sigma-Aldrich).

2.2 Ethical approval for the use of human tissue

All primary human AML samples used in this study were obtained from the MCRC Tissue Biobank; their collection was approved by the South Manchester Research Ethics Committee. Sample use was authorised by the MCRC Tissue Biobank's scientific sub-committee, with the informed consent of donors. Where donations were sufficiently plentiful, the Biobank permitted release of a proportion of cells prior to cryopreservation, enabling analysis of fresh leukaemic cells.

2.3 Manufacture of lentiviral and retroviral particles and infection of mammalian cells

2.3.1. 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 is able to condense DNA into positively-charged particles and facilitate its uptake into target cells by endocytosis. The day prior to transfection, 293FT or Plat-E cells, for the manufacture of lentiviruses or retroviruses respectively, were plated in 10cm dishes (BD Biosciences – www.bdbiosciences.com) at a density of $4.5x10^6$ cells per dish in 9mL D10. 24 hours leter, the cells were typically at ~90% confluence. For the transfection, 21µg PEI was firstly diluted in 500µL room temperature, serum-free DMEM medium 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 production of lentiviral particles	
Lentiviral vector	4μg
pCMVΔ8.91 (containing *gag*/*pol* genes)	2μg
pMDG.2 (containing *env* gene)	1μg
Serum-free DMEM	To 500μl

Table 4. 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 production of retroviral particles	
Retroviral vector	7μg
Serum-free DMEM	To 500μl

Table 5. 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.

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

To increase transduction efficiency, Polybrene® (Millipore, Billerica, MA) was added to all viral supernatants to a final concentration of 8μg/mL. Polybrene® aids the infection process by neutralising the negative charge repulsion between viral particles and cell membranes.

2.3.3 Human cell lines

For lentiviral infection of human cell lines $1.5x10^6$ cells were resuspended in 6mL viral supernatant in a 6-well tissue culture plate (Corning) and centrifuged for 30 minutes at 900xg and 37°C (a process called "spinfection"). After centrifugation, cells were incubated at standard conditions overnight. The following morning, cells were centrifuge and resuspended in 10 ml of R10 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 2.14. "FACS analysis and sorting".

2.3.4 Primary human AML cells

For lentiviral infection of primary human AML cells $0.5\text{-}1x10^6$ cells were resuspended in 6mL viral supernatant in a 6-well tissue culture plate (Corning) and centrifuged for 30 minutes at 900xg and 37°C (a process called "spinfection"). After centrifugation, cells were incubated overnight in AML culture medium with cytokines, without stromal support. The following morning, cells were spinoculated a second time as above. 24 hours following spinoculation 3μg/mL puromycin (Sigma-Aldrich) was added for 72 hours to select for successfully transduced cells prior to further manipulation.

2.4 RNA extraction and quantitative PCR (qPCR)

2.4.1 RNA extraction

RNA extraction was performed using the RNeasy® Plus Micro kit (for 5x105 cells or less) or Mini kit (for greater than 5x105 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 5minute 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.

2.4.2 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 reaction	
Components	**Volume**
10x RT buffer	2µl
25x dNTPs (100mM)	0.8µl
10x RT Random Primers	2µl
MultiScribeTM Reverse Transcriptase (50U/µL)	1µl
Nuclease-free H_2O	4.2µl
Total	10µl

Table 6. 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 7. 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.

2.4.3 qPCR assay and data analysis

qPCR reactions were performed in MicroAmp® optical 384-well reaction plates and analysed using a QuantStudio® 5 PCR system (Applied Biosystems). Reactions were performed in triplicate and included primers for *β-Actin* (*ACTB*) as a housekeeping gene. Primers were designed using the Universal Probe Library (UPL) Assay Design Center (Roche, Basel, Switzerland) and purchased from Integrated DNA Technologies (Coralville, IA) or 20x TaqMan® primer/probe assays (Life Technologies) were purchased. Raw fluorescence data was converted to Ct values using the Thermo Fisher Cloud facility (Waltham, MA) and normalised to *ACTB*. The set-up for each qPCR assay was as follows:

qPCR mix for 20x TaqMan® assays	
Components	**Volume(µl)**
2x TaqMan® Fast Universal PCR Mastermix	5
20x TaqMan® assay	0.5
cDNA	25ng
Nuclease-free H_2O	Up to 10µl

Table 8. qPCR mix for 20x TaqMan® assays.

qPCR mix for specific oligonucleotides and UPL probes	
Components	**Volume(µl)**
2x TaqMan® Fast Universal PCR Mastermix	5
Forward primer (10µl)	0.25
Reverse primer (10µl)	0.25
Gene-specific probe (10µM)	0.25
cDNA	25ng
Nuclease-free H_2O	Up to 10µl

Table 9. qPCR mix for specific oligonucleotides and UPL probes.

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

Table 10. Thermal cycling conditions for the qPCR reaction the 7900HT system.

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

TaqMan® primer/probe assays		
Gene	**Species**	**TaqMan Gene Expression Assay**
CEBPA	Human	Assay ID Hs00268872_s1
GFI1	Human	Assay ID Hs00382207_m1
HOXA10	Human	Assay ID Hs00172012_m1
JUN	Human	Assay ID Hs01103582_s1
MYB	Human	Assay ID Hs00920556_m1
MYC	Human	Assay ID hs00153408_m1
RUNX1	Human	Assay ID Hs01021970_m1

Table 11. List of the TaqMan® primer/probe assays used for qPCR.

Listed below 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**	**Prob #**
ACTB	Human	Forward: ATTGGCAATGAGCGGTTC Reverse: GGATGCCACAGGACTCCAT	11
ADPN	Human	Forward: GGGTGACCTCACAGGTGTTC Reverse: ACTGGCATTTGGGACCTG	32
ARID3A	Human	Forward: CCACGGCGACTGGACTTA Reverse: GCTGAACAAGTCATCCAGGAAT	68
CBFB	Human	Forward: ACTGGATGGTATGGGCTGTC Reverse: AAGGCCTGTTGTGCTAATGC	3
CD14	Human	Forward: CAAGTAGATTCTCTGGGATATAAGGAA Reverse: CCTCCTCTGTGAACCCTGAT	69
CD86	Human	Forward: GGAATGCTGCTGTGCTTATG Reverse: GAATGTTACTCAGTCCCATAGTGC	54
CEBPE	Human	Forward: CTCTGCGCGTTCTCAAGG Reverse: GCCGGTACTCAAGGCTATCTT	8
ELF1	Human	Forward: TGTTGTCCAACAGAACGACCT Reverse: GGAAAAATAGCTGGATCACCA	88
ETV6	Human	Forward: CCCTGCGCCACTACTACAA Reverse: TGATTTCATCTGGGGTTTTCA	12
FOS	Human	Forward: AAGGAGAATCCGAAGGGAAA Reverse: GTGTATCAGTCAGCTCCCTCCT	46
FOSB	Human	Forward: AAGAGGTACAGCGGCATCC Reverse: CGTTCCAACAATGGCAAAGT	4
FOXC1	Human	Forward: TGAACGGGAATAGTAGCTGTCA Reverse: GGACGTGCGGTACAGAGAC	11
GATA2	Human	Forward: GATGAGCATCCTGCGAGTG Reverse: CACACAGCACATCCACCCTA	10
IKZF1	Human	Forward: CAATGTGCTCATGGTTCACAA Reverse: GTTGCCCTTCTGGGTGAAT	47
IRX3	Human	Forward: AAAAGTTACTCAAGACAGCTT Reverse: GGATGAGGAGAGAGCCGATA	57
KLF2	Human	Forward: TGGTCTGGTTGCTTGGAACT Reverse: CTGCCCTCCATCAAACTCTC	52
MAX	Human	Forward: CGGTGGGTACAAGATGACG Reverse: CTTGCGGGTGCTTTCTACA	39
SPI1	Human	Forward: CCACTGGAGGTGTCTGACG Reverse: CTGGTACAGGCGGATCTTCT	51
STAT3	Human	Forward: GAGCAGAGATGTGGGAATGG Reverse: CGGTCTCAAAGGTGATCAGG	17
TLE3	Human	Forward: TGGTGAGCTTTGGAGCTGT Reverse: ACATGGAATGAGTACGCTGGT	65

Table 12. **List of the UPL primers and probes used for qPCR.**

ΔCt values relative to ACTB or Actb were assessed using SDS software v2.1 (Applied Biosystems).

2.5 Western blotting

2.5.1 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 4°C for 15 minutes. Samples were then centrifuged at 20,000xg, 4°C for 15 minutes to pellet cell debris and the supernatant was collected. Lysates were stored at -80°C.

2.5.2 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 95°C 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 volts for approximately 1 hour, until the lowest molecular weight marker was close to the bottom of the gel.

2.5.3 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.

2.5.4 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 Na2HPO4, 16.8g KH2PO4, 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 4°C overnight. Primary and secondary antibodies were either diluted in "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 1 hour at room temperature. After 3x 10 minutes 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 (AmershamTM 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:

Antibodies used for western blotting	
Target	**Manufacturer**
ACTB	MAB150; Merck Millipore
KLF2	09-820; Merck Millipore
RUNX1	8529; Cell Signaling Technology
STAT3	124H6; Cell Signaling Technology
Myc tag	2276; Cell Signaling Technology
CEBPA	C15410225; Diagenode
CBFB	ab33516; Abcam
TLE3	ab94972; Abcam
CEBPE	14271-1-AP; Proteintech
FOXC1	in-house genereted
IRX3	in-house genereted
MYB	sc-74512: Santa Cruz

Table 13. List of primary antibodies used for western blotting.

2.6 Subcellular fractionation

Subcellular fractionation was performed using the Subcellular Protein Fractionation Kit for Cultured Cells (ThermoFisher) using manufacturer's instructions.

2.7 Immunoprecipitation followed by western blotting

Immunoprecipitation (IP) experiments were performed using lysate generated from Fujioka AML cells exogenously expressing MYC-tagged-expressing vectors or with an empty vector control. Fujioka AML cells were transfected as described in section 2.3.1. After antibiotic selection, cells were washed twice with ice-cold PBS and lysed in 1mL (for 10 million cells) 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, 4°C for 15 minutes. Samples were then centrifuged at 20,000xg, 4°C 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 5µg of the appropriate antibody (anti-MYC-tag (2276; Cell Signaling), anti-RUNX1 (ab2398; Abcam) or anti-FOXC1) 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 4°C for 1 minute, combined with the prepared lysates and rotated at 40rpm, 4°C 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 70°C 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 2.5.

2.8 Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes

Fujioka cells (1 x 10^8) were grown in RPMI. The media were then removed and replaced with PBS containing 2mM Di(N-succinimidyl) glutarate (DSG) and crosslinked for 30 mins. PBS + GSG were then removed and cells were washed twice in PBS. Cell were then further crosslinked in PBS containing 1% formaldehyde for 10 min. Crosslinking was quenched by adding Glycine to a final concentration of 0.2 M. The cells were washed with ice-cold PBS. The nuclear fraction was extracted by first resuspending the pellet in 10 ml of LB1 buffer (50 mM HEPES-KOH [pH 7.5], 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40 or Igepal CA-630, and 0.25% Triton X-100) for 10 min at 4C. Cells were pelleted, resuspended in 10 ml of LB2 buffer (10 mM Tris-HCL [pH 8.0], 200 mM NaCl, 1 mM EDTA, and 0.5 mM EGTA), and mixed at 4C for 5 min. Cells were pelleted and resuspended in 300 ml of LB3 buffer (10 mM Tris-HCl [pH 8], 100 mM

NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate, and 0.5% N-lauroylsarcosine) and sonicated using a Bioruptor Pico (Diagenode) for 8 cycles, 30 sec ON, 30 sec OFF settings. Triton X-100 was added at a 10% concentration, and lysate was centrifuged for 10 min at 20,000 rcf to purify the debris. The supernatant was then incubated with magnetic beads Dynabeads (Protein G), Invitrogen, Carlsbad, CA) pre-bound with 10 μg antibody, and immunoprecipitation (IP) was conducted overnight at 4C. The beads were washed ten times in 1 ml of RIPA buffer and twice in 100 mM ammonium hydrogen carbonate (AMBIC) solution. For the second AMBIC wash, the beads were transferred to new tubes. RIME samples were prepared and analysed by mass spectrometry as described (Glont et al., 2019; Mohammed et al., 2016; Papachristou et al., 2018). Briefly, the proteins bound to the beads were digested by adding trypsin (Pierce) prepared in 100mM ammonium bicarbonate buffer. Samples were incubated overnight at 37°C followed by a second-step of digestion the next day for four hours. Samples were acidified with the addition of 5% formic acid and purified using C18 columns according to manufacturer's instructions (Harvard Apparatus, Cambridge, UK). After purification, samples were dried with SpeedVac and reconstituted in 15μl of 0.1% formic acid. A volume of 5μl of each sample was injected on the Dionex Ultimate 3000 UHPLC system coupled with the Q-Exactive mass spectrometer. The full MS scan on Q-Exactive was at 70K resolution and the MS2 scans were performed at 35K resolution with collision energy 28% and isolation window 2.0Th. For the HCD data processing, the SequestHT search engine implemented on Proteome Discoverer 1.4 software was used with Precursor Mass Tolerance 20ppm and Fragment Mass Tolerance 0.02Da. Dynamic modifications were oxidation of M (+15.995Da) and deamidation of N/Q (+0.984Da).

2.9 Chromatin immunoprecipitation (ChIP) and next generation sequencing

Fujioka cells were transduced with lentiviral particles for NTC and FOXC1 KD. The day after, the cells were selected by adding puromycin (3μg/mL) in the medium and incubated for 3 days. Following this incubation, viable cells were counted and cross-linked. We crosslinked cells using 1% formaldehyde for H3K4me1, H3K27Ac, H3K4me2, SPI1 and FOXC1. For CEBPA, SMARCC2, EP300, HDAC1, RUNX1 and TLE3, cells were first crosslinked using the ChIP Cross-link Gold (C01019027; Diagenode) for 30 mins in PBS with 1mM MgCl2 and then with 1% formaldehyde for 10 min. The reaction was stopped by incubation for 5 minutes with 0.125M glycine. Cell pellets were washed twice with cold PBS containing protease inhibitors (Complete EDTA- free tablets, Roche, Basel, Switzerland). 100 million cells were used per ChIP, as described in the protocol reported by Lee et al. (2006). Briefly, nuclear lysates were sonicated using a Bioruptor Pico (Diagenode) for 8 cycles, 30 sec ON, 30 sec OFF settings. Immunoprecipiation was performed overnight at 20 rpm and 4°C, with 100ul

magnetic beads (Dynabeads (Protein G), Invitrogen, Carlsbad, CA) per 10µg antibody. ChIP-grade antibodies were used as follows:

ChIP antibodies:	
Antibody	**Company**
anti-FOXC1	in-house generated antibody
anti-monomethyl H3K4	ab8895; Abcam
anti-dimethyl H3K4	ab7766; Abcam
anti-acetyl H3K27	Ab4729; Abcam
anti-HDAC1	ab46985; Abcam
RUNX1	ab23980; Abcam
anti-EP300	ab14984; Abcam
anti-TLE3	ab94972; Abcam
SPI1	2258; Cell Signaling
CEBPA	C15410225; Diagenode
SMARCC2	12760; Cell signaling

Table 14. **Antibodies used for ChIP.**

After washing six times with RIPA buffer (50mM HEPES pH 7.6, 1mM EDTA, 0.7% Na deoxycholate, 1% NP-40, 0.5M LiCl), chromatin IP-bound fractions were extracted at 65°C for 30min with elution buffer (50mM TrisHCl pH8, 10mM EDTA, 1% SDS) vortexing frequently. RNAseA (1mg/ml) and proteinase K (20mg/ml) were used to eliminate any RNA or protein from the samples. Finally DNA was extracted using phenol:chloroform:isoamyl alcohol extraction and precipitated with ethanol (adding 2 volumes of ice-cold 100% ethanol, glycogen (20µg/µl) and 200mM NaCl) for at least 1 hour at -80°C. Pellets were washed with 70% ethanol and eluted in 50ul 10mM TrisHCl pH8.0.

ChIP DNA samples were prepared for sequencing using the Microplex Library Preparation Kit (Diagenode) and 1ng ChIP DNA. Libraries were size selected with AMPure beads (Beckman Coulter) for 200-800 base pair size range and quantified by Q-PCR using a KAPA Library Quantification Kit. ChIPseq data were generated using the NextSeq platform from Illumina with 2x75bp Hi Output. Reads were aligned to the human genome (GRCh38) using BWA-MEM version 0.7.15 (Li and Durbin, 2009). Reads were filtered using Samtools (version 0.1.9) (Li et al., 2009) to keep only reads that mapped to standard chromosomes and Bedtools version 2.25.0 (Quinlan and Hall, 2010) to remove reads mapped to blacklisted regions defined by ENCODE (http://mitra.stanford.edu/kundaje). Peaks were called with Model-based Analysis of ChIP-Seq version 2 (MACS2) using the following parameters -f BAMPE, --keep-dup 5 to keep only paired-end reads with up to 5 duplicates (Zhang et al., 2008). Annotation of peaks was performed with Homer version 4.10 (Heinz et al., 2010).

The sequencing datasets are available at:
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159693.

2.9.1 ChIP sequencing normalization between experiments

To normalize ChIP signal between control and FOXC1 knockdown Fujioka cells, reads surrounding the absolute summit of 160,041 transcription factor binding peaks were counted. The binding peak sets used were (i) all CEBPA peaks in control Fujioka cells (n=36,856), (ii) all RUNX1 peaks in control Fujioka cells (n=34,180), (iii) all SPI1 peaks in control Fujioka cells (n=34,717), (iv) all FOXC1 peaks in control Fujioka cells (n=18,745), (v) all RUNX1 peaks in FOXC1 knockdown Fujioka cells (n=17,589) and (vi) the coordinates of all MYB binding peaks in THP1 AML cells (n=17,954) (Maiques-Diaz et al., 2018). For histone modifications reads were counted ±1,000 base pairs each side of each absolute summit; for all other ChIP sequencing experiments reads were counted ±300 base pairs. For each of the 160,041 value pairs, a fold change in ChIP signal between control and FOXC1 knockdown conditions was calculated. The list of 160,041 value pairs was then ranked from high to low in Excel based on the number of reads in the control condition. The normalized read count surrounding each peak in the FOXC1 knockdown condition was the total knockdown read count multiplied by the mean of the 2499 subsequent fold change values in the rank ordered list as well as the value for that peak. This “2500 value running mean” approach to normalization is superior to normalization using total mapped reads because it accounts for variations in fold change in ChIP signal according to peak strength and also excludes background reads. Relative increases and decreased in ChIPseq signal strength were confirmed using ChIP PCR.

2.9.2 ChIP PCR

For ChIP quantitative PCR, assays were performed in 384-well MicroAmp optical reaction plates using Taqman Fast Universal PCR Mastermix (Life Technologies) and Universal Probe Library System designed primers and probes (Roche). Signal was detected using an ABI PRISM 7900HT Sequence Detection System (Life Technologies). ChIP qPCR antibodies were: anti-RUNX1 (ab23980), anti-SP1 (ab13370) (both from Abcam) and anti-KLF2 (09-820, Merck Millipore). Primers and probes used were:

ChIP qPCR primers		
Group A	**Primer**	**Prob #**
1	F: AGAGTGGCACCAGCCTACAG R: GTCTTGCAAAACCGGAAGC	87
2	F: CCAGTTCAACCACATCCTGA R: AACCAGTATGAAATGGAGCAAAA	72
3	F: GGGGCAGTGTACCTGGAAG R: CCCAGATACCAAGGGGTGA	27
Group C	**Primer**	**Prob #**
1	F: CGCACACACACAGCAAAAG R: AAGGGCTAGAAGTACAGCTGAGAT	34
2	F: CACGAGCTCGATGTGTCG R: TGAAAGGGAAACAGAAAGTCG	30
3	F: CGTCTTAATACAGTGAAAGAATTGAGG R: CGTTCCCGTCTTTCAAACC	77

Table 15. ChIP qPCR primers.

2.10 Assay for Transposase-Accessible Chromatin using sequencing (ATAC seq)

Fujioka cells were transduced with lentiviral particles for NTC and FOXC1 KD. The day after, the cells were selected by adding puromycin (3µg/mL) in the medium and incubated for 3 days. Following this incubation, the Assay for Transposase Accessible Chromatin (ATACseq) protocol (Buenrostro et al., 2013) was performed using 50,000 viable Fujioka cells. Cell pellets were re-suspended in 50µl lysis buffer (10mM Tris-HCL pH7.4, 10mM NaCl, 3mM MgCl2, 0.1% IGEPAL CA-630) and nuclei were pelleted by centrifugation for 10 minutes at 500 x g. Supernatant was discarded and the nuclei were re-suspended in 25µl reaction buffer containing 2µl of Tn5 transposase and 12.5ul TD buffer (Nextera Sample preparation kit from Illumina). The reaction was incubated for 30 minutes at 37°C and 300rpm, and purified using the Qiagen MinElute Kit. Library fragments were amplified using 1x NEBNext High-Fidelity PCR master mix and 1.25µM of custom PCR primers and conditions (Buenrostro et al., 2013). The PCR reaction was monitored to reduce GC and size bias by amplifying the full libraries for five cycles and taking an aliquot to run for 20 cycles using the same PCR cocktail and 0.6x SYBR Green. The remaining 45ul reaction was amplified for additional cycles as determined by qPCR. Libraries were finally purified using the Qiagen MinElute Kit. Libraries were size selected with AMPure beads (Beckman Coulter) for 200-800 base pair size range and quantified by Q- PCR using Kapa Library Quantification Kit (Kapa Biosystems). ATACseq data were generated using the NextSeq platform from Illumina with a 2x75bp High Output.

Sequencing reads were quality checked using FASTQC (Andrews, 2010). Any adapter sequences present in the data were removed using Cutadapt (Martin, 2012). The cleaned and trimmed FASTQ files were mapped to the hg38 reference assembly using BWA (Li and Durbin, 2009) and processed using Samtools (Li et al, 2009). The data were cleaned for duplicates, low mapping quality reads (i.e. MAPQ<30), non-uniquely

mapped reads, not properly paired reads and reads mapped to non- conventional chromosomes and mitochondrial DNA.

2.11 Colony Forming Unit (CFU) Assays

For clonogenic assays of primary human AML cells, cells were cultured in methylcellulose medium (H4320, Stem Cell Technologies) at a starting density of 10^4/mL with 20ng/mL IL-3, G-CSF and TPO (PeproTech, Rocky Hill, NJ). When required, puromycin was added at a concentration of 3µg/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 (2-5x10^3/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 2x103-2x104/mL and the same growth factors and selection antibodies as previous rounds were maintained.

2.12 Cytospin analyses

2-5x10^4 cells were resuspended in 150µL PBS and, through centrifugation at 60xg 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).

2.13 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).

2.14 Flow cytometry analysis

2.14.1 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).

2.14.2 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 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): CD14-FITC, CD11b-PE, CD86-PE-Cy7 and CD33-APC.

2.14.3 Apoptosis

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.

2.14.4 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 -20°C 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 37°C 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 100°C for 15 minutes to eliminate any contaminating DNase prior to sterile filtration.

2.14.5 FOXC1 intracellular staining

FOXC1 intracellular staining was performed using the eBioscience™ Foxp3/ Transcription Factor Staining Buffer Set (Thermo Scientific). Cells were fixed adding the Foxp3 Fixation/Permeabilization working solution and incubated for 60 minutes at 4°C. Following fixation, cells were centrifuged at 400 x g for 5 minutes at room temperature and supernatant was discarded. 200 µL of Permeabilization Buffer were added to each well and samples were centrifuged samples at 400 x g for 5 minutes at room temperature and supernatant was discarded. This step was repeated twice. Cell pellet were then resuspended in 100 µL of Permeabilization Buffer. Without washing, FOXC1 antibody was added and incubated for 60 minutes at room temperature. Cells were washed twice with 200 µL of Permeabilization Buffer and incubated with a PE- anti goat secondary antibody and incubated for 30 minutes. Cells were washed twice with 200 µL of Permeabilization Buffer and resuspended in an appropriate volume of SM Buffer.

2.15 Molecular methods

2.15.1 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	**Volume**
5x GC or HF buffer	10µl
10mM dNTPs	2.5µl
Forward primer (10µM)	2µl
Reverse primer (10µM)	2µl
Template DNA	10ng
DMSO	1.5µl
Phusion® DNA polymerase	0.5µl
Nuclease-free H_2O	Up to 50µl

Table 16. 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			
Steps	**Temperature (°C)**	**Duration (seconds)**	**Number of cycles**
Initial denaturation	98	30	1
Denaturation	98	10	35
Annealing	Mean T_m of primers ±3°C	45	
Extension	72	30/Kb	
Final extension	72	600	1
Hold	4	∞	∞

Table 17. Thermal cycling conditions for PCR with Phusion® DNA polymerase.

PCR products were immediately visualised on a 1% agarose gel as described below in section 2.15.3.

2.15.2 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 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.

2.15.3 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).

2.15.4 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 50°C 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.

2.15.5 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 16°C. As controls, mock ligations containing only insert or only vector were also performed.

2.15.6 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 -80°C 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 2.16.1 and 2.16.2. “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 42°C 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 37°C 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 37°C. 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.

2.16 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.

2.16.1 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 2.17.1. „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 37°C in an SM30 orbital incubator (Edmund Büler GmbH). 1mL of the bacterial cell culture was removed and stored temporarily at 4°C 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 -20°C.

2.16.2 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 37°C 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 37°C 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 -20°C.

2.16.3 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).

2.16.4 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 350ng of template DNA with 2.5µM of the appropriate sequencing primer in a final volume of 20µL within a 0.5mL microcentrifuge tube.

2.16.5 DNA sequencing primers

DNA sequencing primers used in this study are listed below:

PGEMTeasy	T7:TAATACGACTCACTATAGGG
	SP6:TATTTAGGTGACACTATAG
EF1a	5':CCTCAGACAGTGGTTCAAAGT
	3':ATTGAGATGCATGCTTTGCA
PLKO	5':GACTATCATATGCTTACCGT

2.17 Bacterial culture methods

2.17.1 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.

2.17.2 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.

2.17.3 Bacterial growth conditions

Bacteria on LB agar plates were kept at 37°C 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 37°C in an SM30 orbital incubator (Edmund Büler GmbH).

2.18 Lentiviral expression vectors

2.18.1 Vector maps

Vectors maps are shown in Chapter 6.

2.18.2 Lentiviral vectors for gene expression knockdown

Lentiviral vectors for knockdown experiments (non-targeting control pLKO.1 (SHC002), FOXC1 KD3 pLKO.2 (TRCN0000235693) and a lentiviral vector for expression of FOXC1 cDNA resistant to knockdown) were from Somerville at al., 2015. Lentiviral and retroviral supernatants were prepared and leukemic human and infected with viral particles as previously described (Harris et al., 2012).

To generate lentiviral knockdown constructs, pLKO.1 puro was digested with Age1 and EcoR1 and ligated with HPLC purified oligonucleotides previously annealed by incubating at 98°C for 5 minutes and slowly cooling to room temperature. Oligonucleotide sequences were as follows:

ADPN KD1
F: ccggtggcagcttagccggagttaactcgagttaactccggctaagctgccatttttg
R: aattcaaaaatggcagcttagccggagttaactcgagttaactccggctaagctgcca

ADPN KD2
F: ccggatcgaagaccatgaacgtatactcgagtatacgttcatggtcttcgatttttg
R: aattcaaaaaatcgaagaccatgaacgtatactcgagtatacgttcatggtcttcga

ARID3A KD1
F: ccggccctgaaccaagatcactgaactcgagttcagtgatcttggttcagggttttg
R: aattcaaaaaccctgaaccaagatcactgaactcgagttcagtgatcttggttcaggg

ARID3A KD2
F: ccggtggacttacgaggagcagtttctcgagaaactgctcctcgtaagtccatttttg
R: aattcaaaaatggacttacgaggagcagtttctcgagaaactgctcctcgtaagtcca

CBFB KD1
F: ccgggcgagtgtgagattaagtacactcgagtgtacttaatctcacactcgctttttg
R: aattcaaaaagcgagtgtgagattaagtacactcgagtgtacttaatctcacactcgc

CBFB KD2
F: ccggggagaacagcgacaaacacctctcgagaggtgtttgtcgctgttctccttttg
R: aattcaaaaaggagaacagcgacaaacacctctcgagaggtgtttgtcgctgttctcc

CEBPA KD1
F: ccgggctggagctgaccagtgacaactcgagttgtcactggtcagctccagcttttg
R: aattcaaaaagctggagctgaccagtgacaactcgagttgtcactggtcagctccagc

CEBPA KD2
F: ccgggcacgagacgtccatcgacatctcgagatgtcgatggacgtctcgtgcttttg
R: aattcaaaaagcacgagacgtccatcgacatctcgagatgtcgatggacgtctcgtgc

CEBPE KD1
F: ccggcctttgcctaccctccacatactcgagtatgtggagggtaggcaaaggttttg
R: aattcaaaaacctttgcctaccctccacatactcgagtatgtggagggtaggcaaagg

CEBPE KD2
F: ccgggcagtgaacaaagatagccttctcgagaaggctatctttgttcactgcttttg
R: aattcaaaaagcagtgaacaaagatagccttctcgagaaggctatctttgttcactgc

ELF1 KD1
F: ccgggccacttcaaataggaatcaactcgagttgattcctatttgaagtggcttttg
R: aattcaaaaagccacttcaaataggaatcaactcgagttgattcctatttgaagtggc

ELF1 KD1
F: ccgggcactgtaatcacttcagttactcgagtaactgaagtgattacagtgcttttg
R: aattcaaaaagcactgtaatcacttcagttactcgagtaactgaagtgattacagtgc

ETV6 KD1
F: ccggccataagaacagaacaaacatctcgagatgtttgttctgttcttatggttttg
R: aattcaaaaaccataagaacagaacaaacatctcgagatgtttgttctgttcttatgg

ETV6 KD2
F: ccgggcgccactactacaaactaaactcgagtttagtttgtagtagtggcgcttttttg
R: aattcaaaaagcgccactactacaaactaaactcgagtttagtttgtagtagtggcgc

HOXA10 KD1
F: ccggtcgcccatagacctgtggctactcgagtagccacaggtctatgggcgatttttg
R: aattcaaaaatcgcccatagacctgtggctactcgagtagccacaggtctatgggcga

HOXA10 KD2
F: ccggctcacggacagacaagtgaaactcgagtttcacttgtctgtccgtgagttttttg
R: aattcaaaaactcacggacagacaagtgaaactcgagtttcacttgtctgtccgtgag

GATA2 KD1
F: ccgggacgacaaccaccaccttatgctcgagcataaggtggtggttgtcgtcttttttg
R: aattcaaaaagacgacaaccaccaccttatgctcgagcataaggtggtggttgtcgtc

GATA2 KD2
F: ccggccggcacctgttgtgcaaattctcgagaatttgcacaacaggtgccggttttttg
R: aattcaaaaaccggcacctgttgtgcaaattctcgagaatttgcacaacaggtgccgg

IRX3 KD1
F: ccggcaacgtgctctcgtccgtgtactcgagtacacggacgagagcacgttgttttttg
R: aattcaaaaacaacgtgctctcgtccgtgtactcgagtacacggacgagagcacgttg

IRX3 KD2
F: ccgggtttgtttgtccggttgatttctcgagaaatcaaccggacaaacaaacttttttg
R: aattcaaaaagtttgtttgtccggttgatttctcgagaaatcaaccggacaaacaaac

IKZF1 KD1
F: ccggccgttggtaaacctcacaaatctcgagatttgtgaggtttaccaacggttttttg
R: aattcaaaaaccgttggtaaacctcacaaatctcgagatttgtgaggtttaccaacgg

IKZF1 KD2
F: ccgggccgaagctataaacagcgaactcgagttcgctgtttatagcttcggcttttttg
R: aattcaaaaagccgaagctataaacagcgaactcgagttcgctgtttatagcttcggc

MAX KD1
F: ccggacacacaccagcaagatattgctcgagcaatatcttgctggtgtgtgttttttg
R: aattcaaaaaacacacaccagcaagatattgctcgagcaatatcttgctggtgtgtg

MAX KD2
F: ccggccacagaatatatccagtatactcgagtatactggatatattctgtggttttttg
R: aattcaaaaaccacagaatatatccagtatactcgagtatactggatatattctgtgg

MYB KD1
F: ccgggcatcagaagatgaagacaatctcgagattgtcttcatcttctgatgcttttttg
R: aattcaaaaagcatcagaagatgaagacaatctcgagattgtcttcatcttctgatgc

MYB KD2
F: ccggaacagaatggaacagatgacctcgaggtcatctgttccattctgtttttttg
R: aattcaaaaagcatcagaagatgaagacaatctcgagattgtcttcatcttctgatgc

RUNX1 KD1
F: ccggcctcgaagacatcggcagaaactcgagtttctgccgatgtcttcgaggttttttg
R: aattcaaaaacctcgaagacatcggcagaaactcgagtttctgccgatgtcttcgagg

RUNX1 KD2
F: ccgggaaccactccactgcctttaactcgagttaaaggcagtggagtggttcttttttg
R: aattcaaaaagaaccactccactgcctttaactcgagttaaaggcagtggagtggttc

SPI1 KD1
F: ccggcggatctataccaacgccaaactcgagtttggcgttggtatagatccgttttttg
R: aattcaaaaacggatctataccaacgccaaactcgagtttggcgttggtatagatccg

SPI1 KD2
F: ccggcaagaagaagatccgcctgtactcgagtacaggcggatcttcttcttgttttttg
R: aattcaaaaacaagaagaagatccgcctgtactcgagtacaggcggatcttcttcttg

STAT3 KD1
F: ccgggcaaagaatcacatgccacttctcgagaagtggcatgtgattctttgctttttg
R: aattcaaaaagcaaagaatcacatgccacttctcgagaagtggcatgtgattctttgc

STAT3 KD2
F: ccgggcacaatctacgaagaatcaactcgagttgattcttcgtagattgtgctttttg
R: aattcaaaaagcacaatctacgaagaatcaactcgagttgattcttcgtagattgtgc

TLE3 KD
F: ccggcctatggcttgaacattgaaactcgagtttcaatgttcaagccataggttttttg
R: aattcaaaaacctatggcttgaacattgaaactcgagtttcaatgttcaagccatagg

2.18.3 Lentiviral vectors for overexpression

To generate doxycycline-inducible Fujioka cells, a lentiviral plasmid expressing the rtTA protein under the control of the EF1α promoter was generated by cloning the rtTA-IRES-neomycin expression cassette from pRetroX-Tet-On Advanced (Clontech, Mountain View, CA) into the BamH1/Sal1 sites of pLentiGS (Huang et al., 2014). Fujioka cells constitutively expressing rtTA protein were generated by infection with pLentiGS EF1α-rtTA-IRES-neo followed by neomycin selection (500μg/ml) for two weeks.

To generate the tetracycline inducible FOXC1 lentiviral expression construct, human FOXC1 cDNA was PCR amplified from pcDNA3.1-FOXC1 (a gift from Jane Sowden) using oligonucleotides which introduced coding sequences for a C-terminal GSG linker and Myc tag:

F:cacGAATTCACCATGCAGGCGC
R:cacACTAGTtcacagatcctcttctgagatgagtttttgttcacccgaaccCACAGATCCTCTTCTGA

The product was subcloned into pGEM-T. Sequence was verified and cDNA was excised and ligated into the EcoR1 and Spe1 sites of pLentiGS-minCMV-TET-blasticidin (Huang et al., 2014).

To generate FOXC1 mutants FOXC1 Δ(1-50) and FOXC1 Δ(436-553), FOXC1 cDNA was amplified from pLentiGS-minCMV-TET-blasticidin vector-FOXC1-Myc using the following oligonucleotides:

FOXC1 Δ(1-50)

F:cacgaattcaccatgGCGCACGCCGAGCAGTAC
R: cacACTAGTtcacagatcctcttctgagatgagtttttgttcacccgaaccCACAGATCCTCTTCTGA

FOXC1 Δ(436-553)

F:cacGAATTCACCATGCAGGCGC
R:cacACTAGTtcacagatcctcttctgagatgagtttttgttcacccgaaccCGGAGGCAGAGAGTAGTCGG

The products were then digested using EcoR1 and Spe1 and ligated into the pLentiGS-minCMV-TET- blasticidin vector.

To generate FOXC1 mutants FOXC1 Δ(69-178), FOXC1 Δ(215-366), FOXC1 G165R and FOXC1 F112S, site direct mutagenesis was performed using overlap extension PCR followed by Dpn1 digestion with pLentiGS-minCMV-TET-blasticidin vector-FOXC1-Myc as a template. The following primers (5'-3') were used:

FOXC1 Δ(69-178)

F: cgggccctacacggcggtgaaggaca R: tgtccttcaccgccgtgtagggcccg

FOXC1 Δ(215-366)

F: caacgcgcccggtcagacctccagcg R: cgctggaggtctgaccgggcgcgttg

FOXC1 G165R

F: aacatgttcgagaaccgcagcttcctgcggcgg R: ccgccgcaggaagctgcggttctcgaacatgtt

FOXC1 F112S

F: ccagttcatcatggaccgctccccttctaccggg

R: cccggtagaaggggagcggtccatgatgaactgg

To generate a doxycycline-inducible FOXC1-DBD RUNX1c lentiviral fusion construct, the DNA binding domain (DBD) of FOXC1 was PCR amplified using the following oligonucleotide primers and full length FOXC1 as template.

F:atagaattcgccaccatggaacaaaaactcatctcagaagaggatctgccgcagccgcagcccaag
R: atagctagcacccgaaccgtccttcttcttgaagcgccgccg

The amplicon was subcloned into pGEM-T, sequence verified, excised and cloned into EcoR1/Nhe1 sites of the doxycycline-inducible lentiviral vector pLentiGS-minCMV-TET-blasticidin (Huang et al., 2014). Full length RUNX1c was then PCR amplified using the following oligonucleotides subcloned into pGEM-T, sequence verified, excised and cloned into Nhe1/Cla1 sites of FOXC1 DBD pLentiGS- minCMV-TET-blasticidin:

F: atagctagcatggcttcagacagcatatttgagtc R: ataatcgattcagtagggcctccacacgg

To generate tetracycline inducible KLF2 lentiviral expression construct, human KLF2 cDNA was PCR amplified from Fujioka cells cDNA using oligonucleotides:

F: cacgaattcaccatggcgctgagtgaacccatc R: cactctagatcacatgtgccgtttcatgtgcagc

The product was subcloned into pGEM-T. Sequence was verified and cDNA was excised and ligated into the EcoRI and Xba1 sites of pLentiGS-minCMV-TET-blasticidin vector. rtTA Fujioka cells were then infected with doxycycline-inducible vectors and selected for 10 days in blasticidin (6μg/ml). Cells were maintained in RPMI containing 10% tetracycline-free FBS (Lonza, Basel, Switzerland) in the presence of 250μg/ml neomycin and 6μg/ml blasticidin. Protein expression was induced using 1μg/ml doxycycline (Clontech).

To generate a lentiviral expression for KLF6, human KLF6 cDNA was PCR amplified from Fujioka cells using oligonucleotides which introduced coding sequences for a C-terminal GSG linker and Myc tag:

F:cacgaattcgccaccatggacgtgctccccatgtgc

R:cactctagactacagatcctcttctgagatgagtttttgttcacccgaaccgaggtgcctcttcatgtgcagggcc

The product was sub-cloned into pGEM-T. Sequence was verified and cDNA was excised and ligated into the EcoRI and XbaI sites of pLentiGS-minCMV-TET-puromcyin vector. rtTA Fujioka cells were then infected with doxycycline-inducible vectors and selected for 10 days in puromycin (3μg/ml). Cells were maintained in RPMI containing 10% tetracycline-free FBS (Lonza, Basel, Switzerland) in the presence of 250μg/ml neomycin and 3μg/ml puromycin. Protein expression was induced using 1μg/ml doxycycline (Clontech).

2.19 Next generation sequencing (NGS)

2.19.1 Targeted DNA sequencing with TruSight® Myeloid Panel

Genomic DNA from flow sorted leukaemic blasts was subjected to targeted NGS of 54 genes recurrently mutated in myeloid neoplasms. Libraries were prepared using a

TruSight® Myeloid Sequencing Panel (Illumina, San Diego, CA) and samples sequenced with a NextSeq 500 System (Illumina). Library preparation was performed by Dr Dan Wiseman of the University of Manchester and sequencing performed by staff of the CRUK-MI MBCF.

2.19.2 RNA sequencing

Total RNA was extracted from cells using QIAshredder spin columns and an RNeasy Plus Micro Kit (Qiagen, Manchester, UK). RNA quality was checked using the Agilent Bioanalyzer. Indexed total RNA libraries were prepared with an input of 500ng of total RNA and 10 cycles of amplification using the TruSeq Stranded Total RNA LT Sample Preparation Kit – Set A (with Ribozero Gold) (Illumina, San Diego, CA). Library quality was checked using the Agilent Bioanalyzer. Libraries were quantified by qPCR using the KAPA Library Quantification Kit for Illumina (Kapa Biosystems, Woburn, MA). 1.8 pM pooled libraries were loaded onto the NextSeq 500 and 2x75bp sequencing was carried out using a NextSeq 500/550 High Output v2 kit (Illumina). Reads were aligned to the human genome (GRCh38 and gene annotated with its corresponding GTF files (GENCODE GRCh38) using STAR version 2.4.2a with the settings --outFilterMultimapNmax 20 ,--outFilterType BySJout, --alignSJoverhangMin 8, --quantMode GeneCounts (Dobin et al., 2013). DESeq2 was used to perform differential gene expression analysis and calculate FPKM (fragments per kilobase of transcript per million mapped reads) values for each gene, counting only reads that mapped to exonic regions (Love et al., 2014)

2.20 Antibody production

2.20.1 Generation of pet28a-FOXC1 expression vector

To generate a bacteria expression vector for FOXC1, FOXC1 cDNA (NM_001453) was excised from pcDNA3.1-FOXC1 (a gift from Jane Sowden) using EcoR1 and Xho1 restriction sites and sub-cloned into the EcoR1/Xho1 sites of pET28 expression vector. Successful subcloning and confirmation of the FOXC1 coding sequence was determined by DNA Sanger sequencing.

To generate FOXC1 mutant FOXC1 Δ(69-178), site direct mutagenesis was performed using overlap extension PCR followed by Dpn1 digestion with pET28-FOXC1 as a template. The vector was then transformed into Rosetta BL21 (DE3) Competent Cells (Novagen®). For transformation, 100 μl of competent cells were thawed on ice and 5 μl of vector was added. Cells were kept on ice for 30 minutes and subsequently heat-shocked for 45 seconds at 42°C to allow DNA uptake. After three minutes incubation on ice, 900 μl of LB medium were added to the tube and bacteria were incubated at 37°C for 1 hour in an SM30 orbital incubator (Edmund Büler GmbH - www.edmund-

buehler.de). Then, cells were then spread on LB agar plates containing kanamycin and chloramphenicol antibiotics and incubated overnight at 37°C.

2.20.2 Fusion protein induction

Sterile tips were used to pick cells from a single colony and transfer them into antibiotic containing LB medium and incubated overnight at 37°C. Cells were then diluted to an OD (600 nm) of 0.05 and grown until the OD reached 0.4. The following step was to supplement cells media with 1 mM of IPTG (Isopropyl β-D-1-thiogalactopyranoside) and cells were incubated at 37°C. At 1, 2, 3 and 4 hours after induction, 1 ml of each culture was harvested, resuspended in 20 μl of lysis buffer (Buffer A) (50 mM Tris pH 8.0, 100 mM NaCl, 15 mM MgSO4, 0.2 mM phenylmethane sulfonyl fluoride proteinase inhibitor) and frozen at -20°C overnight. The following day, samples were defrosted at 37°C for 30 minutes. Subsequently, each tube was supplemented with 4 μg of lysozyme and incubated again at 37°C for 30 minutes. Following this second incubation, samples were supplemented with 10 μg/ml DNAse I and left at room temperature for 30 minutes. Next, 1 μl of 10 % sodium deoxycholate and 1% triton were added to each tube and samples were incubated on ice for 15 min. Finally, tubes were centrifuged at 10000 rpm for 15 min and 20 μl of both soluble fraction (supernatant) and inclusion bodies (pellet) were run on acrylamide gel and stained with blue Coomassie. This process allows the determination of the best condition to purify FOXC1 protein and inclusion bodies and 3 hours of induction resulted in the best protein expression.

2.20.3 FOXC1 protein purification

Protein purification by gel electrophoresis (SDS-PAGE) has been used in various applications such as antigen preparation for antibody generation. The system is based on a denaturant electrophoresis which uses a discontinuous polyacrylamide gel and sodium dodecyl sulfate (SDS) to denature the proteins. The SDS confers to all the proteins the same charge-to-mass ratios allowing the separation on the basis of their mass. One liter of E. coli cells was grown until OD reached O.4 and induced with 1 mM IPTG as previously described. Cells were harvested and resuspended in 20 ml of buffer A and frozen at -20°C overnight. Samples were treated as describe in section a 2.20.2 and the final pellet was resuspended in 1.5 ml PBS supplemented 10x NuPAGE® sample reducing agent and 4x NuPAGE® lithium dodecyl sulphate (LDS) sample loading buffer (both from Life Technologies). Samples were then incubated at 95°C for 10 minutes in order to ensure complete unfolding of the protein secondary structure, and then loaded into the acrylamide gel. The gel was constituted by two different components: the stacking gel (4% acrylamide) and the running gel (10% acrylamide), and the tank was filled with electrode buffer (25 mM Tris-Hcl, 192mM Glycin, 0.1%

SDS). Gels were electrophoresed at 80 volts overnight, allowing good protein separation.
Following SDS-page, the proteins were detected through blue Coomassie staining. The gel was disassembled and soaked in about 30 ml of blue Coomassie (GekCode Blue Safe Protein) for around 30 minutes, or until the proteins were clearly visible. The gel was distained in water.

2.20.4 Protein electroelution

Using a sterile scalpel, the protein band corresponding to FOXC1 was excised. The excised band was then placed into a dialysis membrane (Dialysis tubing cellulose membrane - Sigma) and elute overnight at 100 mA. The electroelution was performed into an electrophoresis tank filled with electroelution buffer (81mM Na_2HPO_4 ,770mM NaH_2PO_4, 1mM SDS).
The purified protein solution was then transferred into a new dialysis membrane overnight in order to eliminate the SDS in the solution. In particular, the dialysis buffer used was 81mM Na_2HPO_4 ,770mM NaH_2PO_4.

2.20.5 BCA quantification

Following dialysis, the protein concentration was calculated using the BCA assay (bicinchoninic acid). The procedure was performed using the commercial kit PierceTM BCA Protein Assay Kit (Thermofisher scientific), following manufacturer's instructions. In particular, the optical density was read at 595 nm and the standard curve was designed using BSA.

2.20.6 Animal immunisation

The animal immunisation in goats was executed by the biotechnology supplier "Eurogentec" using the purified recombinant FOXC1 protein. Specifically, 2 μg of purified protein was sent with a concentration of 200μg per ml. The company performed a "Speedy28-Day program of immunisation" (https://secure.eurogentec.com/speedy.html), consisting of a step of immunisation and the collection of the serum after 28 days. The serum was then used to affinity purified anti-FOXC1 antibodies.

2.20.7 Antibody purification

The "AminoLink Plus Immobilization Kit" (Thermo Scientifc) was used for antibody purification according to the manufacturer's instructions. At first, 2 mg of the purified FOXC1 antigen was coupled to the AminoLink column, and subsequently the coupled column was used for the affinity purification of FOXC1 antibody from 5ml of serum. After

elution, the purified antibody was dialysed in PBS and eventually supplemented with an equal volume of glycerol 100% and 0.05% of Sodium azide and stored at -80°C.

2.21 Statistics

For flow cytometry, quantitative PCR and ChIP-PCR, statistical significance was determined using the unpaired, two-tailed Student's t-test when comparing two experimental groups. All tests were performed in Excel (Microsoft) or Prism 8 (GraphPad). P values of <0.05 were considered statistically significant. Survival curves were generated using a Cox proportional hazards model using Prism 8 (Graphpad). The statistical methods used to analyze next generation sequencing data are detailed in the relevant sections of the methods and results chapters. Analysis of next generation sequencing data was performed with the assistance of Professor Tim Somervaille and Dr Fabio Amaral.

All the spreadsheets for every single experiment are available on request.

Chapter 3: Results – Generation of FOXC1 antibody

3.1 Introduction

The primary role of a transcription factor is to regulate gene expression and therefore dictate cellular identity. Gene expression networks drive the majority of biological processes, allowing cells to explicate their activities. This coordinated network is regulated by a complex transcription machinery formed by interactions between different proteins (e.g. TFs, RNA polymerases, etc..) and sequence-specific DNA elements, called TF-binding sites. Therefore, in order to study the role of a particular TF in a specific biological context, the first and most fundamental analysis is achieved by the identification of the preferential binding sites of a TF to specific DNA sequences and the isolation of the TF-protein interaction network.

Due to their high specificity, antibodies are used throughout biology for the characterisation of protein function and are useful tools to investigate TF biology. A major limitation to investigate the role of FOXC1 in AML was the lack of a reliable, quality commercial antibody for repeat analyses. I, therefore, decided to initiate this project with the production of an "in-house generated antibody".

Antibody production involves different steps. Initially, an antigen is generated and injected into laboratory animals so as to evoke high expression levels of antigen-specific antibodies in the serum. After several repeated immunisations, the blood is collected and antibodies are purified. Polyclonal antibodies are recovered directly from serum (Figure 8A).

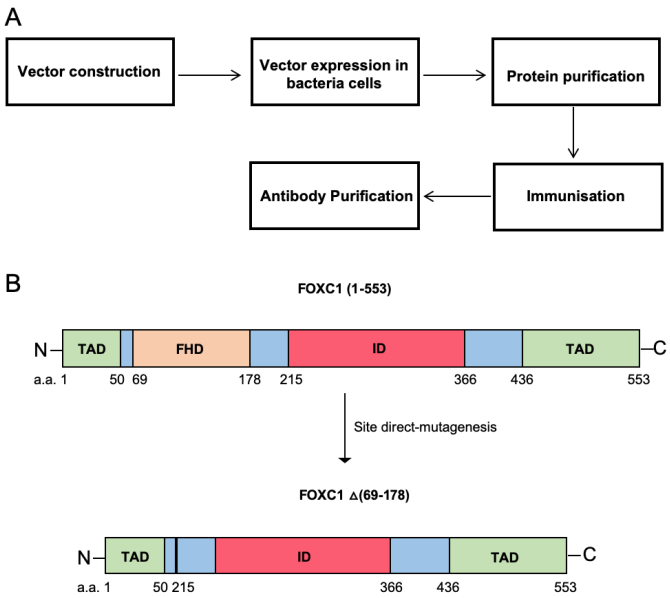


Figure 8. Schematic example of experiment outline.

(A) Experimental outline. (B) Site direct mutagenesis to create FOXC1 Δ(69-178).

3.2 Antibody generation

3.2.1 Expression of recombinant FOXC1 in *E. coli* cells and purification

A polyclonal antibody targeting FOXC1 was generated by animal immunisation using purified recombinant FOXC1 protein as an antigen. Choosing the best possible antigen is one of the most critical steps in any custom antibody project. Ordinarily, researchers either use protein antigens or peptide antigens for this reason. Protein antigens are commonly the first and most highly recommended methodology with regards to excellent antigen production for antibody-generation purposes. The greatest advantage with utilising the full-length protein is that antibodies will be produced against the native conformation of the target and they will bind to various conformational epitopes inside the target protein. As a result, there is a very high probability that antibody selectivity and specificity are enhanced and the generated antibodies will bind the native protein in the target assay. The drawback of this methodology is that since antibodies are being produced against various epitopes, there is a higher probability that antibodies against a portion of these epitopes could recognize comparable protein structures present in different molecules with a certain degree of homology with the target protein.
FOXC1 shows little conservation among members of the FOX transcription factor family apart from the Forkhead domain (FHD) (Jackson et al., 2010). Therefore, to generate a specific FOXC1 antibody, the FHD region was removed from *FOXC1* cDNA by site-directed mutagenesis. The resultant cDNA sequence was then sub-cloned into the pet28A expression vector (Figure 8B). The cloning region of the pet28A plasmid is under the control of a promoter recognised by the RNA polymerase of bacteriophage T7. Next, the pet28A-FOXC1 Δ(69-178) was transformed into BL21(DE3) competent cells. BL21(DE3) is an E. coli B strain and it is deficient for two important proteins: the lon protease and the outer membrane protease OmpT. The lack of these two key proteases reduces degradation of heterologous proteins expressed in the cells. Moreover, the strain is engineered to contain the T7 RNA polymerase gene under the control of an inducible promoter. Therefore, an IPTG-inducible bacteria expression system of exogenous recombinant FOXC1 was generated.

This system guarantees a high degree of accumulation of the product in bacteria cells, a key feature for antigen purification. Commonly, recombinant proteins accumulate intracellularly in insoluble aggregates rather than diffuse in the soluble fraction of the cytoplasm. This aggregation of recombinant proteins overexpressed in bacterial cells are called inclusion bodies and are defined by dimers and multimers of the overexpressed antigen. The presence of the recombinant antigen in the inclusion bodies is advantageous for the following step of antibody purification. In fact, inclusion bodies often contain almost entirely the overexpressed protein. Moreover, the protein is shielded from proteolytic degradation and the amount of proteins contained in these

regions can be 50% or more of the total cellular protein (Garcia-Fruitos, 2010). In order to define whether FOXC1 expression was highly present in the inclusion bodies or in the soluble fraction (cytoplasm), a pilot experiment using 1 mM IPTG was set up. The experiment was performed at three different incubation points (2, 3 and 6 hours).

Total bacteria lysate was extracted and separated using SDS-PAGE along with a non-induced control. In addition to the soluble protein fraction, inclusion bodies proteins were extracted and analysed with the same method. The pilot experiment showed that FOXC1 is strongly expressed in the inclusion bodies. The expression remains quite stable over time and there was no major difference in protein expression between 2, 3 and 6 hours of incubation (Figure 9A and 9B).

Based on these results, I decided to induce the expression of FOXC1 using 1mM IPTG with an incubation of 3 hours at 37°C, and to purify the protein from the inclusion bodies. Following expression of FOXC1 protein using the conditions described above, I purified the recombinant protein by SDS-PAGE. Following electrophoresis and the Coomassie blue staining of proteins in the gel, I excised the band corresponding to the expected molecular weight for recombinant FOXC1 and extracted the protein from the polyacrylamide gel by electroelution. Eventually, I performed dialysis to promote refolding of the protein to its native conformation. In order to evaluate the purity of the extracted protein, I separated an aliquot of the extract by SDS-PAGE and proceeded with Coomassie blue staining (Figure 9C). As illustrated in figure 9C, I was able to obtain a strong band corresponding to FOXC1 protein and no visible protein impurities. Overall, this analysis suggested a good purity level for FOXC1 protein. Eventually, I measured the concentration of the purified protein using the BCA (bicinchoninic acid) assay, obtaining a value of 636 μg/ml (Figure 9D).

Following antigen purification, I sent the purified antigen to the Eurogentec (www.eurogentec.com) company that proceeded with animal immunisation and serum collection.

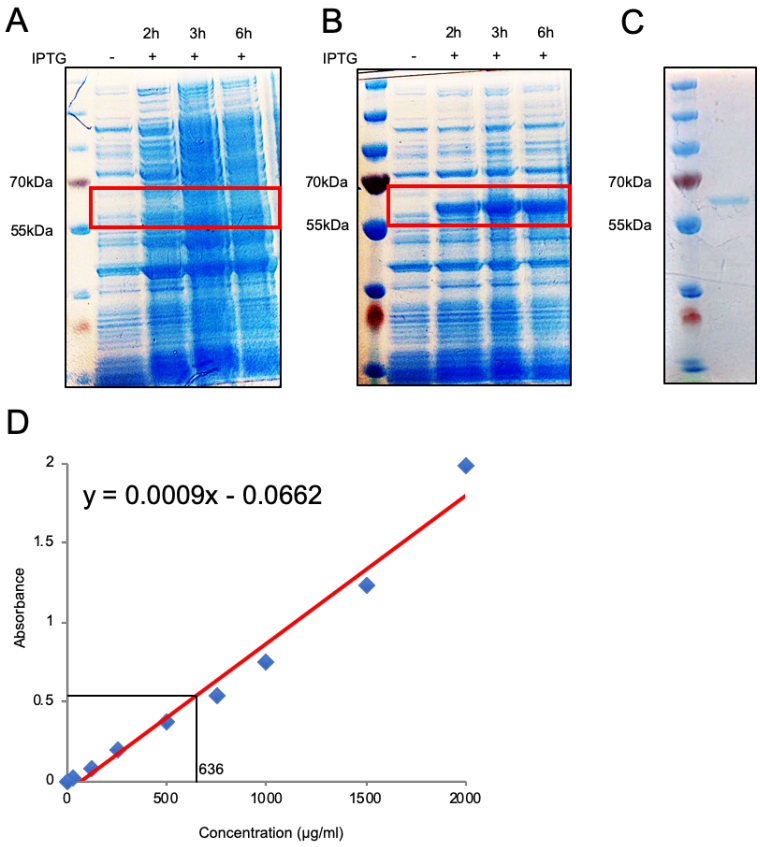


Figure 9. FOXC1 protein purification.

(A) SDS-PAGE of total cell lysate (A) and inclusion bodies (B) from BL21 (DE3) cells expressing FOXC1 Δ(69-178). For both images incubation times are indicated. (C) SDS-PAGE of protein extracted from the polyacrylamide gel after the electroelution. (D) BCA results show protein concentration after purification.

3.2.2 Antibody purification

Once the serum is collected, there are different methods to purify the polyclonal antibodies generated by animal immunisation. These methods range from very crude to highly specific:

Crude – protein precipitation of a subset of total serum molecules includes immunoglobulins;
General – purification of a certain class of antibody, without regard to antigen specificity;
Specific – affinity purification of only those antibodies in the serum that recognise a particular antigen.

Specific affinity purification was chosen and recombinant FOXC1 Δ(69-178) was used as antigen molecule. Firstly, I coupled FOXC1 antigen to “AminoLink resin” (Thermo Scientific) which interacts with the protein of interest by covalent linkage between the

aldehyde groups of the resin and the primary amines of the protein. Subsequently, I used the column to purify the specific antibody for FOXC1 from the serum using a gravity-flow method.

3.3 Evaluation of Antibody function and specificity

The quality of research antibodies has been an issue for decades in the scientific community and once a new antibody is generated, there is a need to demonstrate that it is specific, selective, and reproducible in the context for which it is used.

Therefore, to ensure accurate and consistent results, antibodies need to be validated using different methods, such as including western blot (WB), immunohistochemistry (IHC), immunocytochemistry (ICC), immunofluorescence (IF), ELISA, immunoprecipitation (IP), chromatin immunoprecipitation (ChIP), peptide array, and protein array, among others.

Thus, I next assessed the quality of the generated antibody using four different methodologies.

Western blot validation

To investigate the specificity of the in-house generated FOXC1 antibody, knockdown (KD) experiments were firstly performed in human Fujioka AML cells which exhibit high *FOXC1* expression using a shRNA construct delivered through a lentivirus transduction system. Western blot showed the reduction of FOXC1 expression in cell transduced with the knockdown vector, suggesting that the antibody is indeed able to detect FOXC1 in a western blot assay in a specific manner (Figure 10A).

Immunoprecipitation validation

Next, I investigate whether the in-house generated FOXC1 antibody was able to recognise the native conformation of FOXC1 protein. This second approach was particularly important because the capacity to bind to the native structure of the molecule is an essential step for several methods (ChIPseq, RIME, etc) commonly used to study transcription factor biology.
The immunoprecipitation assay was performed once again in Fujioka cells. Specifically, FOXC1 antibody was immobilised to sepharose beads and this support was used to immunoprecipitate FOXC1 from the total and nuclear protein lysate. Subsequently, I assessed the efficacy and specificity of the immunoprecipitation through a western blot on the immunoprecipitate. As a negative control, I used sepharose beads coupled with

IgG. As illustrated in Figure 10B, the FOXC1 antibody was able to immunoprecipitate FOXC1 both from the total or nuclear lysate, suggesting its utility in assays which rely on an initial immunoprecipitation step.

Immunofluorescence validation

In order to study the intracellular localisation of proteins, cells are usually labeled with specific fluorescent antibodies after cell fixation and permeabilisation. Permeability must destroy the cell membrane sufficiently to allow the antibody to pass through, while retaining the same membrane structure and protein composition.

Fujioka AML cells were firstly infected with *FOXC1* KD lentiviral vector and an NTC (non-targeting control) vector and cultured for three days. Then, followed by fixation and permeabilization with FoxP3 Fix/Perm Buffer, cells were stained with FOXC1 "in-house" generated antibody. A secondary anti-goat-PE antibody was used. Figure 10C shows reduction of FOXC1 mean fluorescence intensity upon its depletion.

RIME validation

To provide further confirmation of the specificity of the antibody, I performed Rapid Immunoprecipitation Mass spectrometry of Endogenous protein (RIME), in Fujioka AML cells. This technique is a rapid and sensitive method for low abundance endogenous proteins that allows the identification of transient interactors. I used $100x10^6$ cells and once the desired number of cells was achieved, these were double cross linked and then lysed to extract the chromatin and associated bound proteins. This fraction was sonicated and then incubated with 10μg of FOXC1 antibody (or IgG control). Following this, 12 robust washing steps were performed in order to reduce noise. The samples were then submitted for Mass spectrometry as described in methods and with the assistance of Jason Carroll's laboratory at the CRUK Cambridge Institute.
FOXC1 was one of the most significantly enriched proteins identified with 11 unique peptides and it is the only FOX family protein identified in the mass spectrometry data (Table 19). This confirmed the specificity of the polyclonal antibody generated and the value of the protocol used. The coverage of the total sequence was 35.44% (Figure 10D), perfectly in line with other successful RIME experiments published by other laboratories (Glont et al., 2019).

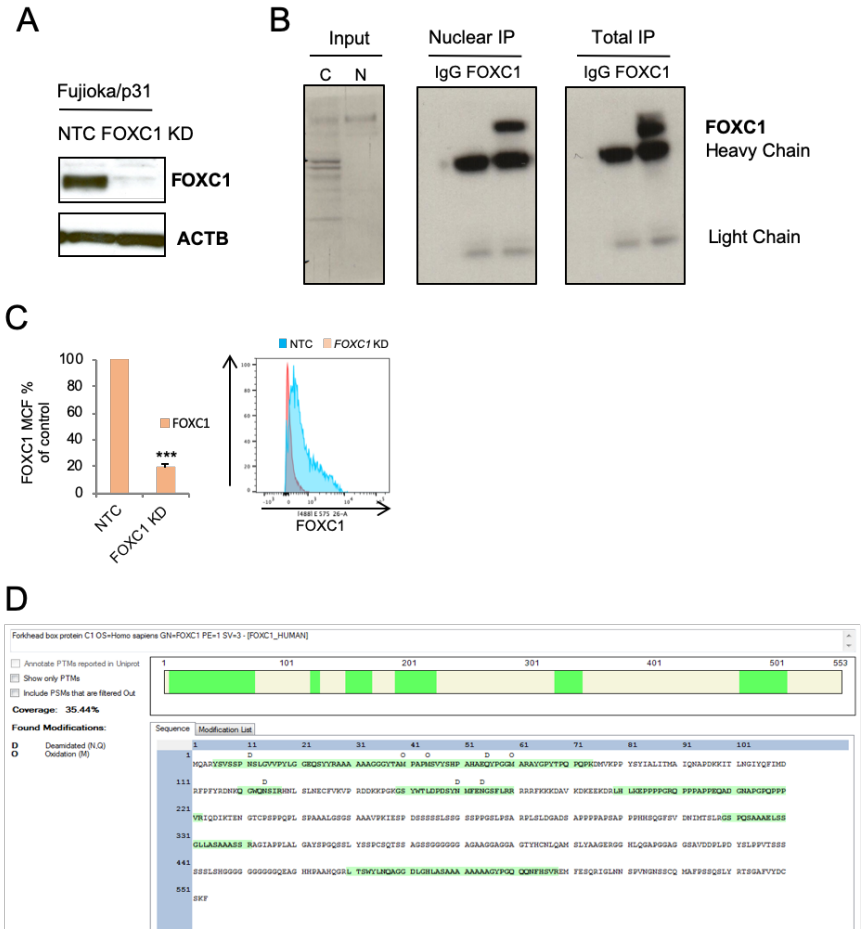


Figure 10. FOXC1 antibody validation.

(A) Human Fujioka AML cells were infected with a lentivirus targeting *FOXC1* for KD or a non-targeting control (NTC). Western blot shows *FOXC1* KD in Fujioka AML cells 72hrs after initiation of KD. (B) Anti-FOXC1 immunoprecipitation in Fujioka AML cells (representative of n=3 experiments). IP, immunoprecipitation; C, cytoplasm; N, nucleus. Western blots show expression of the indicated proteins in the indicated conditions in co-immunoprecipitation experiments in Fujioka AML cells (representative of n=3 experiments). (C) Human Fujioka AML cells were infected with a lentivirus targeting *FOXC1* for KD or a non-targeting control (NTC). Bar chart (left panel) shows mean+SEM FOXC1 mean cell fluorescence (MCF) as determined by flow cytometry in the indicated conditions on Day 5 (n=3). Representative flow cytometry plots (right panel) are also shown. (D) FOXC1 protein coverage in RIME experiments.

This antibody is a rare resource: researchers in related fields have tried and failed to ChIP this protein due to the poor quality of available reagents

Chapter 4: Results – THE ROLE OF FOXC1 IN AML

4.1 Introduction

Acute myeloid leukaemia (AML) is a genetically heterogeneous neoplastic disorder, characterised by an abnormal clonal proliferation of immature cells of myeloid lineage in the bone marrow and blood. Consequently, the cardinal pathologic feature of AML is a block to cell differentiation, which results in a failure of normal haematopoiesis (Wiseman et al., 2014).

AML is hierarchical organised and it is sustained at its apex by a sub-population of cells, called leukaemia stem cells (LSCs), with the ability to self-renew and the capacity to initiate, sustain or regenerate the disease (Wang and Dick, 2005). LSCs are the critical cellular component of AML as they are not only responsible for the maintenance of the disease but also for its restoration following treatment failure, which ultimately leads to disease relapse and death of the patient (Hanekamp et al., 2017). Thus, the elimination of the LSC compartment is essential, and perhaps sufficient, to cure AML.

In the last decade, the majority of AML studies focused on identifying new treatments with the final goal of removing LSCs in AML patients without affecting normal haematopoiesis (Ishikawa et al., 2007). Hence, understanding the differences between normal and leukemic stem cells remains a fundamental focus of leukaemia research. Using comparative transcriptional profiling of prospectively sorted human populations, studies have begun to define the genes and pathways that differentially regulate AML LSCs by comparison with normal haematopoietic stem cells (Goardon et al., 2011; Saito et al., 2010).

In particular, Somerville et al., recently reported that the Forkhead box transcription factor FOXC1 is mis-expressed in approximately 20% of patients with the disease, but it is neither required nor expressed in normal hematopoietic cells (Somerville et al., 2015). This means that *FOXC1* expression in AML is due to a failure of the physiological molecular mechanisms that keep this gene silenced in normal haematopoiesis. Moreover, they discovered that its mis-expression in acute leukaemia has functional consequences, contributing to a block in monocyte/macrophage differentiation and enhancing clonogenic potential and prognostic significance. In fact, younger adult AML patients with *FOXC1* high expression exhibited significantly inferior survival in comparison with *FOXC1* low cases.

Given the importance of FOXC1 as a key transcription factor in AML and the absence of studies documenting the mechanism of action of FOXC1's function, I set out to functionally define the role of this transcription factor. In particular, I set out to identify the molecular mechanisms by which FOXC1 is able to cause a monocyte/macrophage differentiation block, as the core oncogenic feature of acute myeloid leukaemia.

4.2 *FOXC1* expression in AML cell lines and patient samples

To investigate the oncogenic role of FOXC1 in AML, *FOXC1* expression levels were first determined in a panel of AML cell lines and primary AML samples by Q-PCR (Figures 11A and 11B). Of the cell lines tested, the highest *FOXC1* transcript levels were observed in Fujioka cells. These are derived from a child with acute monocytic leukaemia and exhibit a t(10;11) translocation indicative of a *CALM-AF10* fusion, as well as mutations in *NRAS*, *ETV6*, *TP53* and *EZH2* among others (Table 18) (Hirose et al., 1982; Narita et al., 1999). Of the primary samples tested, *FOXC1* transcripts were detected at high level (greater than 500-fold increase over expression levels in the lowest expressing AML sample) in 7/27 (25%) bulk AML blast samples tested (Figure 11A-B; Table 18). Due to the high cell number required for the experiments performed in this thesis, Fujioka cells and BB475 sample were chosen as cellular models to study *FOXC1*^{high} leukaemias.

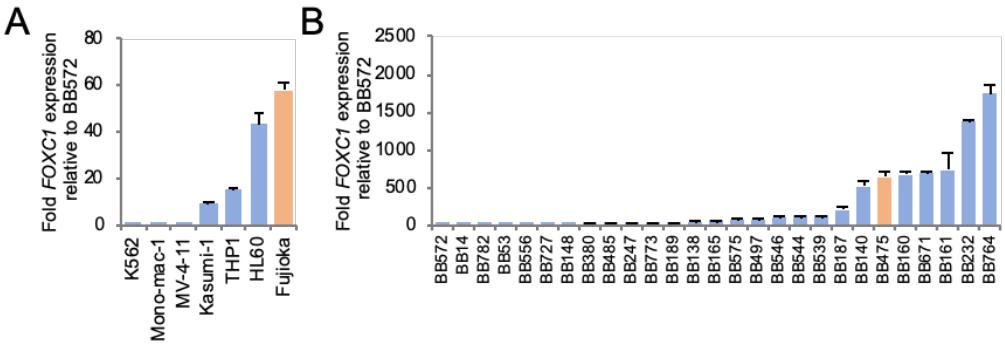


Figure 11. *FOXC1* expression in AML cell lines and patient samples.

(A) Bar chart shows the relative expression levels of *FOXC1* in the indicated cell lines as determined by quantitative PCR. (B) Bar chart shows relative expression of *FOXC1* in bulk primary human AML samples (n=27).

Cell line name or biobank number	BM or PB	Karyotype	Identified mutations by
Fujioka AML cells		50,X,?add(Y)(q11.2),del(2q)(q35q37),del(4)(q2?8),7,der(7)t(3;7)(q12;q32),inv(9)(p13q12)?c,t(10;11)(p12;q21),+13,+13,+del(13)(q12q14),add(18)(q?21),?+add(18)(q?23),+19	EZH2 p.A736fs, TP53 R196X, ETV6 R399C, NRAS G12C, TP53 Y236C
BB14	PB	46,XY [20]	SFSR2 P95R, ASXL1 G646Wfs*, FLT3 D835E, BCORL1 S953X, CEBPA P189del
BB53	PB	46,XY [20]	FLT3-ITD, BPA P112Sfs, K313dup
BB138	BM	46,XX [20]	NPM1 L287fs, NRAS G12A/G12S/G13D, RAD21 W18X, NOTCH1 V2229
BB140	PB	46,XY [20]	FLT3-ITD, NPM1 L287fs
BB148	PB	47,XY,+11[1]/48,sl,+8[7]/49,sdl,+4[2]	KRAS G12V G13D
BB160	PB	46,XY [20]	IDH2 R140Q, BCOR L884P
BB161	PB	46,XY,t(6;11)(q27;q23)[10]/48,idem,+der(6)t(6;11),+21[4]	SRSF2 P95H, ASXL1 T655Pfs*63, IDH2 R140Q, NMP1 W288Cfs*12, GATA2 G200Vfs*18, CEBPA P14L
BB165	BM	46,XX,t(8;22)(p11;q13),del(9)(q13q32)[10]	TET2 G1754R, DNTM3A R55H
BB171	PB	46,XX [20]	IDH1 R132H EZH2 P432LFS*31
BB187	BM	47,XY,+8[5]/46,XY[5]	SRSF2 P95R, IDH1 R132C, DNMT3A R882H, RUNX1 Y414Ffs*187, PHF6 A288_I290del, BCOR P910L
BB189	PM	46,XX [20]	DNMT3A R882L, NPM1 L287fs, FLT-ITD
BB232	PB	46,XX [20]	ASXL1 T655Pfs*63 (44bp ins), IDH1 R132C, DNMT3A R882H, WT1 R302Lfs*3/S313Lfs*70, NOTCH1 A1778V, CELSR2 T1454M, CSMD3 D2372E
BB247	PB	46,XY [20]	DNMT3A S349X, FLT3 D835V, NMP1 W288Cfs*12, GATA2 T354delinsTQ, CDKN2A I27L
BB380	PB	46,XX [20]	TET2 Q913Ffs*11, SH2B3 G451S
BB475	PB	46,XX [20]	IDH2 R140Q, DNMT3A R882H, FLT3-ITD, NPM1 W288C fs*12, RAD21 A544V
BB485	PB	46,XY [20]	IDH1 R132H, DNMT3A S714C, FLT3 D839G, NMP1 W288Cfs*12, KRAS Q61L, PTPN11 D61H
BB497	PM	Failed	DNMT3A R882H, FLT3-ITD, NPM1 W288Cfs*12, WT1 T382Ifs*9
BB539	BM	46,XY [20]	IDH1 R132H, NMP1 W290Efs*10, PTPN11 W290Efs*10
BB544	PM	46,XX [20]	FLT3 D835Y, NMP1 W288Cfs*12
BB546	PB	46,XY [20]	SF3B1 K666N, FLT3-ITD, WT1 V368fs, STAG2 K692R
BB556	PB	46,XY [20]	DNMT3A R882C, NMP1 L287fs, NRAS G13D, STAG2 T626fs
BB572	PB	46,XY [20]	TET2 T1091fs / Q1274E, ASLX1 G642fs, RUNX1 R201P / Y287fs / S318fs , KRAS A59E, BCOR E1185fs, ZRSR2 E79fs
BB575	PM	46,XX [20]	SRSF2 M89V, DNMT3A R882H, FLT3-ITD, NPM1 L287fs, SMC3 L242P
BB671	PB	46,XY [20]	DNMT3A R882H, FLT3-ITD, NPM1 W288Cfs*12
BB727	PB	Inv(16)	FLT3-ITD, FLT3 D835Y
BB764	BM	Failed	DNMT3A F755S, FLT3-ITD, FLT3 D835Y, NPM1 W288Cfs*12, PTPN11 E76K
BB773	PB	46,XY [20]	DNMT3A R882H, FLT3-ITD, NPM1 W288Cfs*12,
BB782	PB	46,XX,t(9;11)(p21.3;q23)[10]	No detected mutation

Table 18. Karyotype of 28 Manchester Cancer Research Centre Biobank AML samples analysed for *FOXC1* expression.

BM = bone marrow blast; PB = peripheral blood blasts. Assistance with sample curation and karyotype was received from Dr. Daniel Wiseman.

4.3 FOXC1 sustains clonogenic potential and differentiation block in AML cells

To confirm that FOXC1 contributed to the differentiation block exhibited by Fujioka cells, knockdown (KD) experiments were firstly performed in human Fujioka cells. Mean transcript levels of *FOXC1*, as determined by q-PCR, were reduced to 15% of control upon shRNA KD (Figure 12A). Western blotting analysis confirm KD efficiency (Figure 12B). Following the initiation of knockdown there was an increase of the myeloid differentiation marker CD86 (Figure 12C) and loss of clonogenic potential (Figure 12E). Moreover, *FOXC1* KD led to morphological differentiation (Figure 12D).

Following the initiation of knockdown there was a reduction in cell expansion and the cell counts at day 4 and 6 were reduced suggesting that these cells rely upon FOXC1 in order to maintain their proliferative capacity (Figure 12F). In keeping with reduced cell expansion, cell cycle analysis at day nine following knockdown reveals a reduced proportion of cells in the SG2M phase of the cell cycle (Figure 12G), as well as an increase in apoptosis (Figure 12H). Taken together, these findings provide strong evidence that *FOXC1* expression is functionally important to maintaining the hallmark properties of AML Fujioka cells, proliferation and differentiation block.

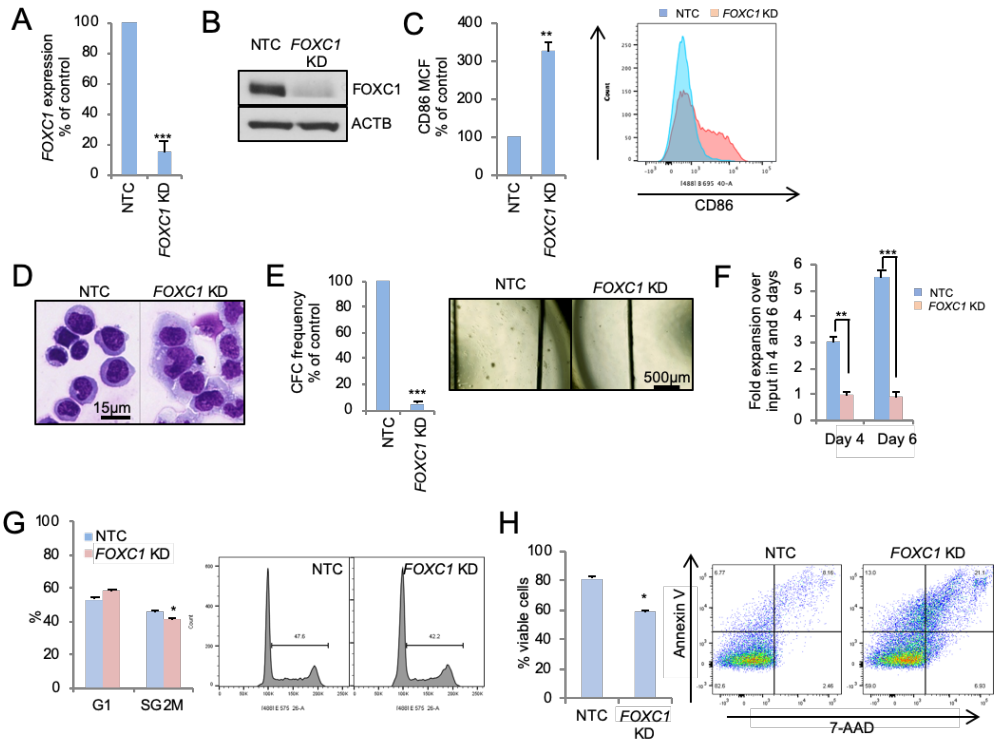


Figure 12. FOXC1 sustains the differentiation block and clonogenic potential of human AML cells.

Human Fujioka AML cells were infected with a lentivirus targeting *FOXC1* for KD or a non-targeting control vector (NTC). (A) Bar chart shows mean+SEM relative expression of *FOXC1* in KD versus control cells (n=3) 72hrs after initiation of KD. (B) Western blot shows *FOXC1* KD in Fujioka AML cells 72hrs after initiation of KD. (C) Bar chart (left panel) shows mean+SEM CD86 mean cell fluorescence (MCF) as determined by flow cytometry in the indicated conditions on Day 5 (n=3). Representative flow cytometry plots (right panel) are also shown. (D) Representative images of cytospins of cells from Day 7. (E) Bar chart (left panel) shows mean+SEM colony-forming cell (CFC) frequencies in the indicated conditions relative to control cells enumerated after twelve days in semi-solid culture (n=3). Image (right panel) shows representative images of colonies. (F) Bar chart shows mean+SEM fold change in cell number at day 4 and 6 in liquid culture (n=3). (G) Bar chart (left panel) shows mean+SEM percentage of cells in G1 or SG2M six days following initiation of KD. Right panels: representative cell cycle profiles. (H) Bar chart (left panel) shows mean+SEM percentage of viable cells as determined by Annexin-V/7-AAD analysis seven days following initiation of KD (n=3). Right panels: representative flow cytometry plots. * indicates $p<0.05$, ** indicates $p<0.01$ and *** indicates $p<0.0001$ for the indicated comparisons by an unpaired t test.

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 (FOXC1 SDM3) engineered by site-directed mutagenesis to generate KD-resistant transcripts. FOXC1 forced expression and resistance to KD was confirmed by western blot (Figure 13A). Expression of KD-resistant FOXC1 in *FOXC1* KD cells completely averted induction of differentiation (Figure 13B) and prevented loss of clonogenic potential, indicating an on-target phenotype (Figure 13C).

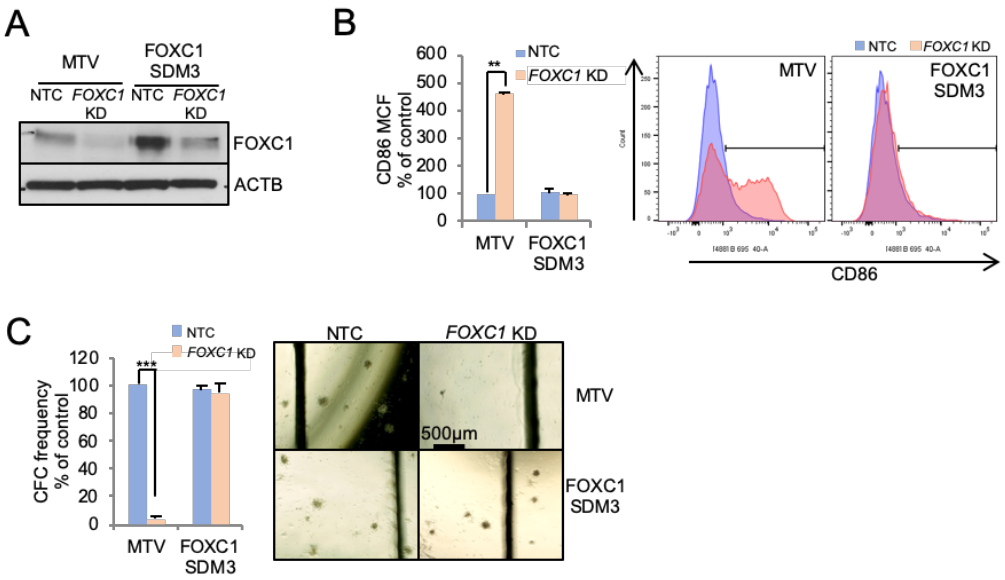


Figure 13. Rescue of clonogenic potential with FOXC1 SDM3.

Human Fujioka AML cells stably expressing *FOXC1* SDM3 or a control retroviral vector (MTV) were infected with the same lentiviral *FOXC1* KD vector as shown in Figure 13 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 percentage of cells positive for CD86 surface marker as determined by flow cytometry analysis in the indicated conditions (n=3). Representative flow cytometry plots (right panel) are also shown. (C) Bar chart (left panel) shows mean+SEM CFC frequencies of Fujioka AML cells expressing either FOXC1 SDM3 or MTV in *FOXC1* KD cells relative to control cells. Colonies were enumerated after fourteen days in semi-solid culture (n=3). Image (right panel) shows representative colonies. * indicates $p<0.05$, ** indicates $p<0.01$ and *** indicates $p<0.0001$ for the indicated comparisons by an unpaired t test.

I then performed similar experiments in *FOXC1*^{high} primary human AML cells from a patient with normal karyotype AML with mutations in *NPM1*, *FLT3*, *DNMT3A* and *IDH2* (BB475; Table 18), with similar results. Following initiation of KD, Q-PCR and western blotting analysis confirm KD efficiency (Figure 14A-B). I observed differentiation, as evidenced by morphology (Figure 14C), increased expression of the monocyte/macrophage lineage differentiation markers CD86, CD11b and CD14 (Figure 14D), reduced clonogenic activity (Figure 15A), a reduced proportion of cells in the SG2M phase of the cell cycle with a G1 arrest (Figure 15B), as well as an increase in apoptosis (Figure 15C).

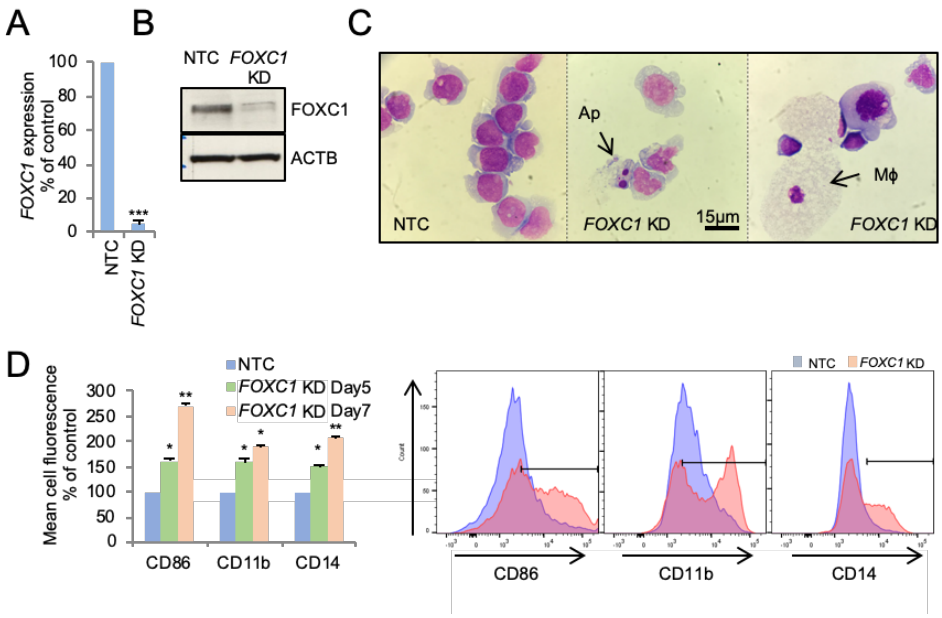


Figure 14. FOXC1 sustains the differentiation block of primary AML cells.

Primary patient AML cells (BB475) were infected with a lentivirus targeting *FOXC1* for KD or a NTC with puromycin drug resistance as selectable marker (n=2). (A) Bar chart shows mean+SEM relative transcript expression in KD versus control cells (n=3). (B) Western blot shows *FOXC1* KD in BB475 AML cells 72hrs following initiation of KD. (C) Representative images of cytospins of cells on Day 7. Ap, apoptotic cell; Mφ, macrophage. (D) Bar chart (left panel) shows mean+SEM percentage of cells positive for CD86, C11b and C14 surface markers as determined by flow cytometry analysis in the indicated conditions (n=3). Representative flow cytometry plots (right panel) at Day 7 are also shown. * indicates $p<0.05$, ** indicates $p<0.01$ and *** indicates $p<0.0001$ for the indicated comparisons by an unpaired t test.

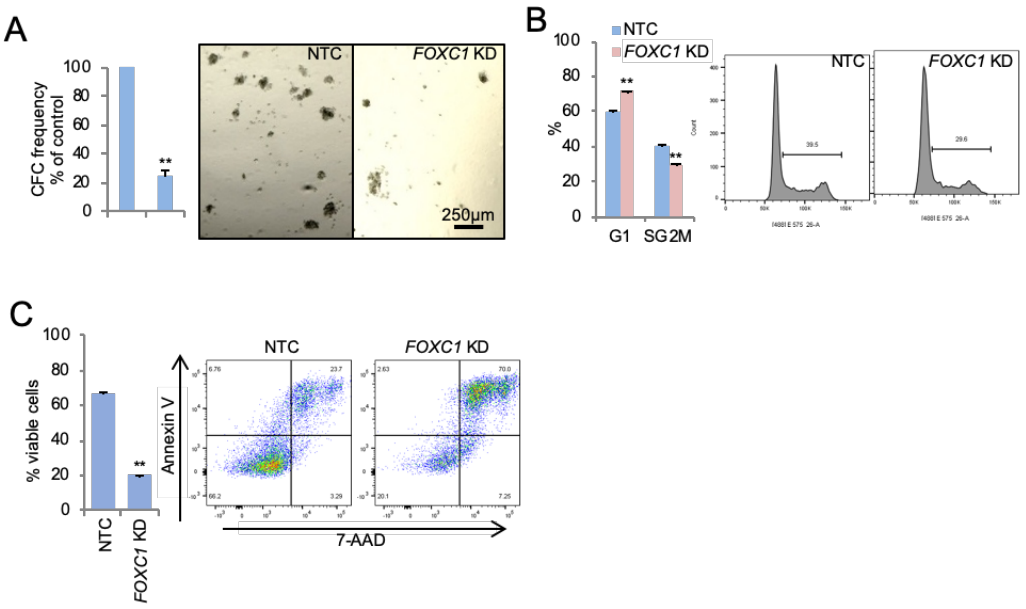


Figure 15. *FOXC1* KD in primary AML cells induces loss of clonogenic potential and apoptosis.

Primary patient AML cells (BB475) were infected with a lentivirus targeting *FOXC1* for KD or a NTC with puromycin drug resistance as selectable marker (n=2). (A) Bar chart shows the mean+SEM colony-forming cell (CFC) frequencies of KD cells relative to control cells enumerated after fourteen days in semi-solid culture (n=3). (B) 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. (C) 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). 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. * indicates $p<0.05$, ** indicates $p<0.01$ and *** indicates $p<0.0001$ for the indicated comparisons by an unpaired t test.

Thus, in support of our prior conclusions and those of others (Assi et al., 2019; Somerville et al., 2015), mis-expressed *FOXC1* confers a differentiation block in human AML cells.

4.4 Identification of chromatin bound FOXC1-interacting proteins

To identify in an unbiased manner FOXC1 interacting proteins with potential functional roles, I performed Rapid Immunoprecipitation Mass spectrometry of Endogenous protein (RIME) (Mohammed et al., 2016). As discussed in Chapter 3, I generated a polyclonal antibody to a version of human FOXC1 engineered to lack the Forkhead domain shared by other Forkhead family transcription factors. This technique is a rapid and sensitive method for low abundance endogenous proteins that allows the identification of transient interactors. I used $100x10^6$ cells and I performed three separate analyses, two in Fujioka cells and a third in primary AML blast cells (BB475), which expressed high levels of *FOXC1* (Figure 11A-B). Cells were freshly thawed and expanded. Once the desired number of cells was achieved, these were double cross linked and then lysed to extract

the chromatin and associated bound proteins. This type of double cross-linking is very effective when you are using ChIP to observe the binding pattern of transcription factors bound directly to DNA or even those found in DNA binding complexes not bound directly to DNA. This fraction was sonicated and then incubated with 10ug of FOXC1 antibody (or IgG control). Following this, 12 robust washing steps were performed in order to reduce non-specific interactors and enhance stringency. The samples were then submitted for mass spectrometry as described in methods and with the assistance of Jason Carroll's laboratory at the CRUK Cambridge institute.

Upon removal of all of the interactions found in the negative controls (IgG), I identified 131 proteins present in all three experiments. I deemed these high confidence FOXC1 interacting proteins (Figure 16A; Table 19). FOXC1 was the only Forkhead family member identified. This confirmed the specificity of the polyclonal antibody generated, the value of the protocol used, and is the only antibody generated which is suitable and effective for immunoprecipitation.

As expected, there was strong enrichment for proteins with Gene Ontology biological process annotations such as "nucleoplasm" ($P=10^{-46}$), "nucleus" ($P=10^{-28}$) and "nuclear chromatin" ($P=10^{-12}$) (Figure 16B), highlighting the capacity of the RIME technology to identify nuclear interactors of a target protein. Moreover, Gene Ontology analysis revealed also a strong enrichment for annotations such as "poly(A) RNA proteins", "protein binding", "RNA binding", "nucleotide binding" and "chromatin binding" (Figure 16C).

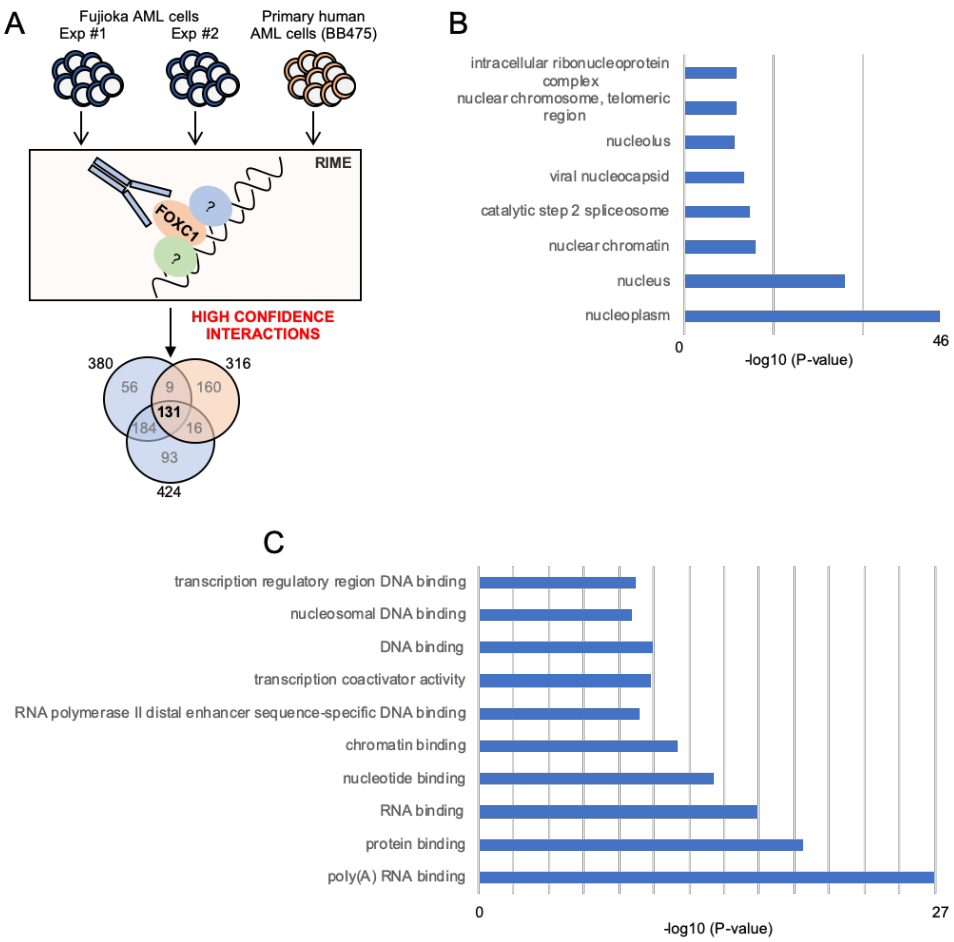


Figure 16. Identification of chromatin bound FOXC1-interacting proteins

(A) Experimental outline. (B-C) Gene ontology analysis using DAVID.

List of high confidence FOXC1 interacting proteins				
FOXC1	EWSR1	LUC7L2	RNF20	TARDBP
A2M	FBL	MATR3	RNF40	TBL1XR1
ADNP	FKBP15	MAX	RPA1	TKT
AHCY	FUBP1	MCM2	RTCB	TLE3
AIP	GATAD2B	MCM3	RUNX1	TMPO
ANP32E	GBE1	MCM5	SEC16A	TOP2B
ANXA6	GSN	MCM6	SF3A3	TPM3
APEX1	GSTP1	MCM7	SFPQ	TPR
ARF3	GTF2I	MSH6	SMARCA2	TUBB4B
ARID1A	H2AFY	MTA2	SMARCA4	USP7
ARID3A	HDGF	NCOR1	SMARCA5	WDR1
BUB3	HMG20A	NELFE	SMARCC2	
C4A	HMGB1	NONO	SMARCD2	
CALM2	HMGB2	NUDT21	SMARCE1	
CAPG	HNRNPA0	NUMA1	SMC1A	
CASP1	HNRNPA1	PEBP1	SMC3	
CBFB	HNRNPH3	PGD	SMCHD1	
CEBPA	HNRNPL	PRKCD	SNRNP200	
CEBPE	HNRNPM	PRKDC	SNRNP40	
CHD4	HNRNPR	PRMT1	SNRNP70	
CTBP1	HNRNPUL1	PRPF19	SNX3	
DDX17	HNRNPUL2	PRPF8	SPI1	
DDX23	HOXA10	PSPC1	SRSF1	
DEK	IKZF1	RAB7A	SRSF7	
DHX15	ILF3	RALY	SSB	
DHX9	IRF2BP2	RBBP7	SSRP1	
ELAVL1	KDM1A	RBM39	STAT3	
ELF1	KHSRP	RBMX	SUB1	
EP300	LIG3	RCC2	SUPT16H	
ETV6	LMNB1	RECQL	TALDO1	

Table 19. List of high confident interactors present in all three RIME experiments.
FOXC1 highlighted in yellow.

I focused my initial interest on the 12 sequence specific transcription factors identified, because transcription factors are critical regulators of differentiation and cell fate (Figure 17A). These transcription factors (ADPN, ARID3A, CEBPB, CEBPA, CEBPE, ELF1, ETV6, HOXA10, IKF1, MAX, RUNX1, SPI1 and STAT3) are potentially key FOXC1 interactors and their established roles in haematopoiesis and AML make them potential targets for further exploration.

RUNX1 has an established importance in AML. It is commonly part of the *RUNX1-ETO* and *RUNX1-EVI1* fusions, however, *RUNX1* mutations also occur in AML (Speck and Gilliland, 2002). STAT3 is also an important TF in AML. *STAT3* siRNA reduces

proliferation of AML cells (Hossain et al., 2014). *CEBPA* is mutated in 10–15% of patients with AML with intermediate risk cytogenetics (Tenen, 2001). The *SPI1* gene is rarely mutated in human AML, but downregulation and/or alterations of its function have been described in different subtypes of AML, in particular in those carrying the fusion oncogenes *RUNX1-ETO* and *PML-RARA* (Sive et al., 2016).

To determine which of these might be functionally linked to the differentiation block conferred by FOXC1, I performed knockdown of each gene in Fujioka cells; I included CBFB, which we also identified as a FOXC1- interacting protein, in view of its coding for the obligate heterodimeric binding partner of RUNX1 (Hart and Foroni, 2002). Knockdown of *RUNX1*, *CBFB*, *CEBPA*, *STAT3* and *CEBPE* using two separate shRNA hairpins for each gene induced upregulation of CD86, which I used as a surrogate marker for upregulation of a differentiation program (Figure 17B-C). Knockdown efficiency was confirmed by q-PCR (Figure 18).

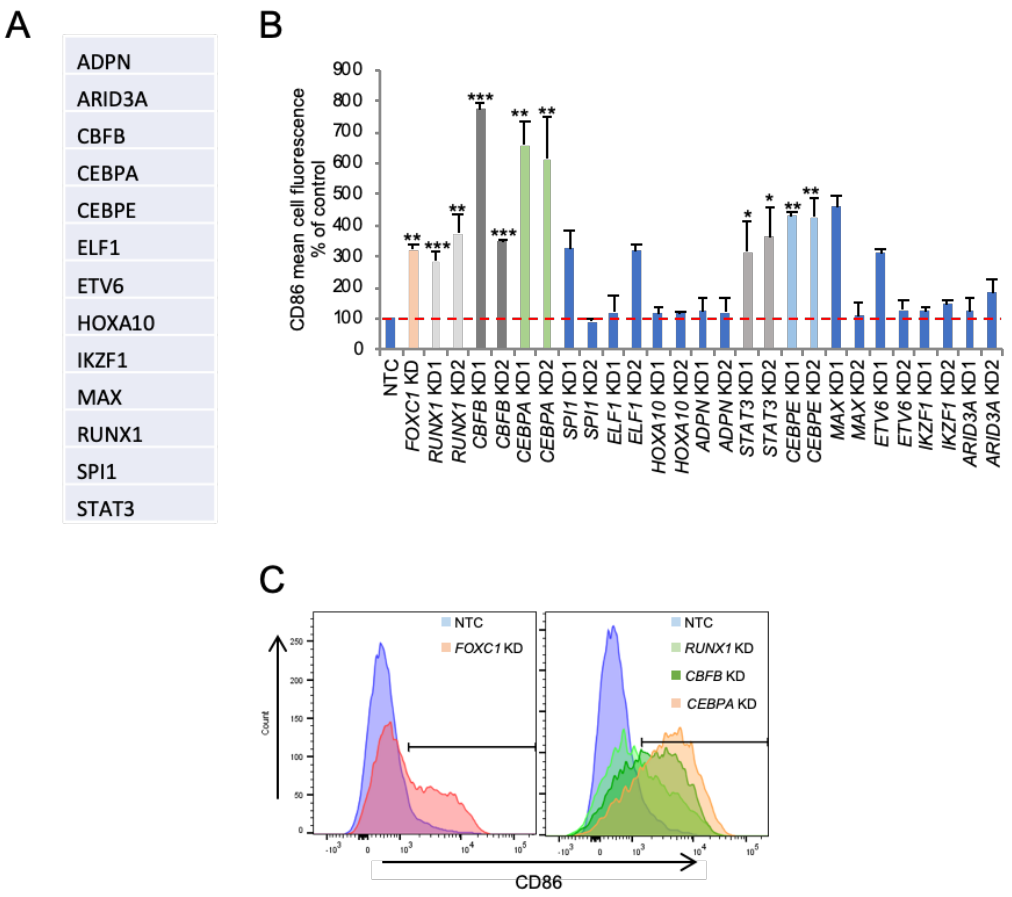


Figure 17. Validation of chromatin bound FOXC1-interacting transcription factors.

Human Fujioka AML cells were infected with lentiviruses targeting the indicated genes for KD, or a non-targeting control (NTC). (A) Table shows FOXC1 interacting transcription factors. (B) Bar chart shows mean+SEM CD86 mean cell fluorescence as determined by flow cytometry analysis on Day 5 (n=3). (C) Representative flow cytometry plots of (B). * indicates $p<0.05$, ** indicates $p<0.01$ and *** indicates $p<0.0001$ for the indicated comparisons by an unpaired t test.

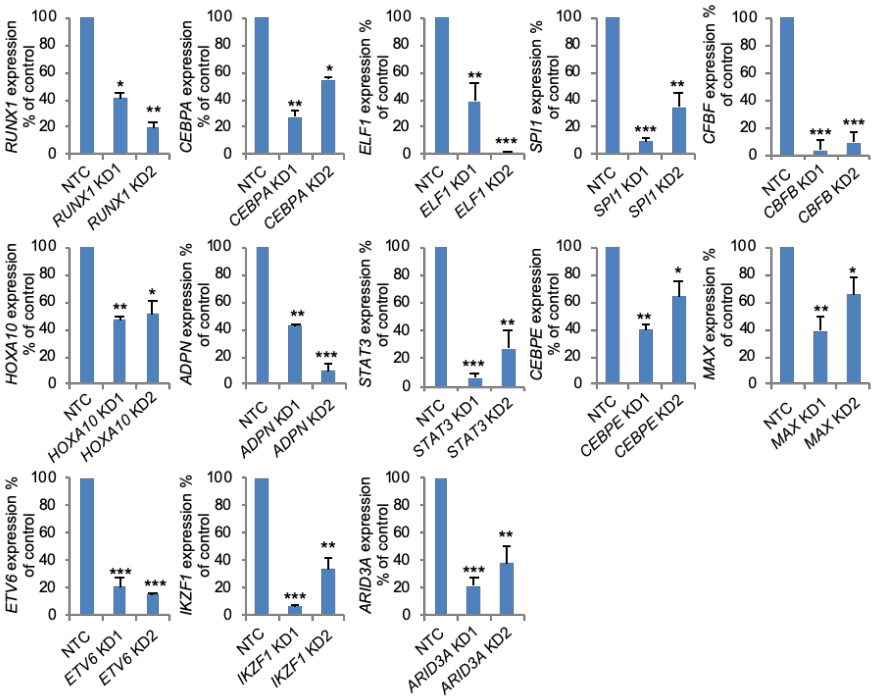


Figure 18. Knockdown of transcription factor genes in Fujioka AML cells.

Graphs show mean+SEM (n=3) relative expression following KD of the indicated genes in human Fujioka AML cells with the indicated lentiviral KD constructs relative to a non-targeting control (NTC). * indicates p<0.05, ** indicates p<0.01 and *** indicates p<0.0001 for the indicated comparisons by an unpaired t test.

Transcription factor interactions with FOXC1 identified by RIME may include those mediated by direct protein:protein interaction as well as those mediated by a short intervening sequence of DNA (Figure 17A). To eliminate the latter, I performed confirmatory co-immunoprecipitation experiments in the presence of benzonase endonuclease to remove DNA and RNA, and noted that only RUNX1, CBFB and CEBPA were pulled down by FOXC1 immunoprecipitation (Figure 19).

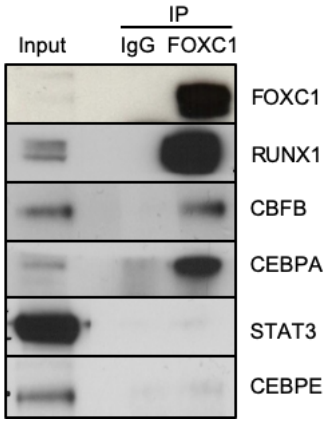


Figure 19. IP- Western blotting confirms FOXC1-associated transcription factors.

Anti-FOXC1 immunoprecipitation in Fujioka AML cells (representative of n=3 experiments). IP, immunoprecipitation.

Furthermore, to map the interaction domain of FOXC1 with RUNX1 and CEBPA, I expressed recombinant FOXC1 proteins in Fujioka AML cells and performed pull-down experiments using FOXC1-Myc full length or deletion proteins (Figure 20A). FOXC1 protein is composed of the classical Forkhead DNA binding domain (FHD), which is responsible for DNA binding, two transactivation domains (TAD), which modulate transcriptional activity and one regulatory domain, which is important for FOXC1 protein stability through post-translation modifications (Gilding and Somervaille, 2019). Co-immunoprecipitation assays using anti-Myc tag antibody releveled that RUNX1 and CEBPA bind to the DNA binding domain of FOXC1. In fact, Δ69-178 deletion completely abolishes the interaction (Figure 20B). It is not surprising that FOXC1 uses its DNA binding domain not only for DNA binding but also to interact with other proteins. In fact, previous work from different laboratories showed that mutations in few residues (e.g. F112 and G165) in the DNA binding domain do not affect FOXC1 DNA-binding ability. Plus, molecular modelling of the FOXC1 FHD predicts that the side chains of these residues point away from the DNA and face opposite to the DNA-binding interface (Huang et al., 2008; Murphy et al., 2004), suggesting the possibility that these residues are not involved in DNA binding but they could be important for protein-protein interactions (Figure 20C).

I then generated two recombinant FOXC1 proteins, harboring some mutations present in Axenfeld-Rieger (AR) syndrome (F112S and G165R) and analysed FOXC1 interaction with RUNX1 and CEBPA through co-IPs. Both mutations strongly reduced FOXC1/RUNX1/CEBPA interaction and in particular FOXC1 G165R completely prevents it (Figure 20D).

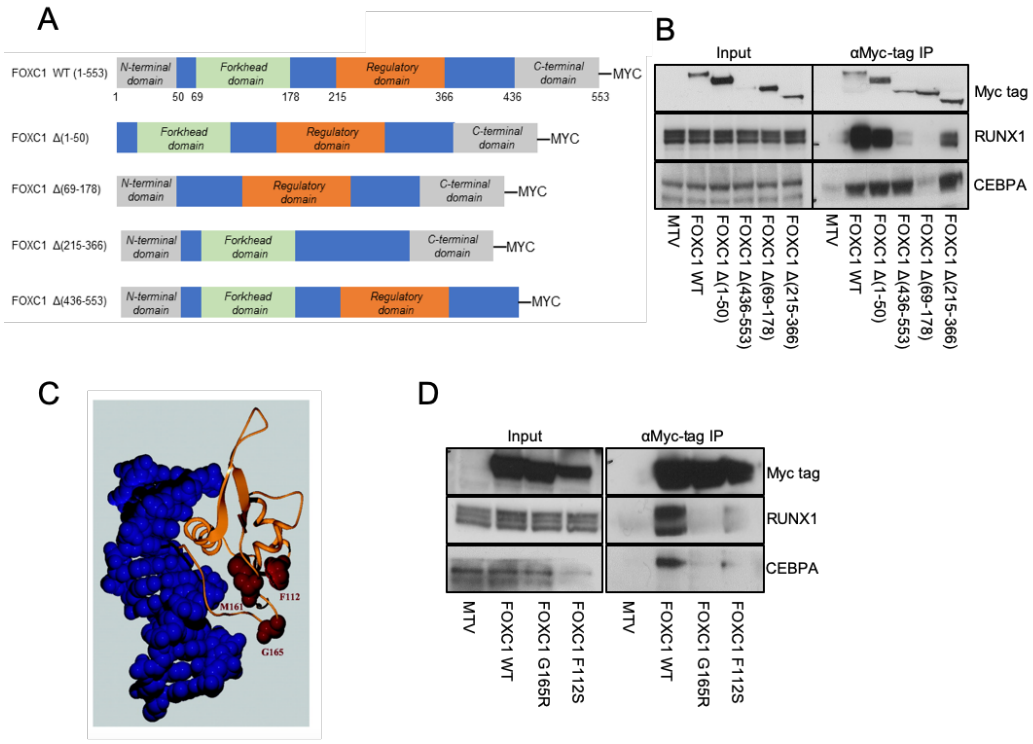


Figure 20. FOXC1 interacts with RUNX1 and CEBPA through its Forkhead domain.

Fujioka AML cells were infected with lentiviruses expressing coding sequences for full length or domain mutant versions of FOXC1. (A) FOXC1 and domain mutants used. (B and D) Western blots shows expression of the indicated proteins in the indicated conditions in co-immunoprecipitation experiments (representative of n=3 experiments). IP, immunoprecipitation. (C) Molecular modelling of the FOXC1 Forkhead domain from Huang.; et al 2008.

These data demonstrate that FOXC1 interacts with RUNX1 and CEBPA through residues in its Forkhead domain and raise a question as to whether the functional effects of FOXC1 misexpression in AML are mediated through its interaction with one or both of these proteins.

4.5 Genome-wide binding profiles of FOXC1, RUNX1, CEBPA and SPI1

To identify FOXC1 binding sites genome wide and to determine their proximity to RUNX1 and CEBPA binding sites, I performed ChIP sequencing for FOXC1, RUNX1 and CEBPA in Fujioka AML cells. In view of its critical role in myeloid development (Iwasaki et al., 2005), I also performed ChIP sequencing for SPI1 (also known as PU.1). ChIPed DNA was processed for sequencing by NextSeq 500 sequencing system (Illumina), which generated a data set of 50-70 million reads for each protein. The reads were mapped by Dr. Fabio Amaral onto the human reference genome hg38 and a browser link was created for viewing in the UCSC Genome Browser Database.

In Fujioka cells, after excluding blacklisted genomic regions prone to artefact and making use of stringent threshold criteria (called peaks had pileup value ≥50 and fold enrichment over input ≥5), MACS2 (Zhang et al., 2008) identified 18,745 FOXC1 peaks, 34,180 RUNX1 peaks, 36,856 CEBPA peaks and 34,717 SPI1 peaks. MEME-ChIP

(Machanick and Bailey, 2011) confirmed that genomic sequences at the center of transcription factor binding peaks were strongly enriched for the appropriate consensus binding motif (Figure 21A). In all cases the great majority of peaks were distributed over intronic and intergenic regions versus promoter regions (Figures 21B-C), consistent with putative roles at enhancers.

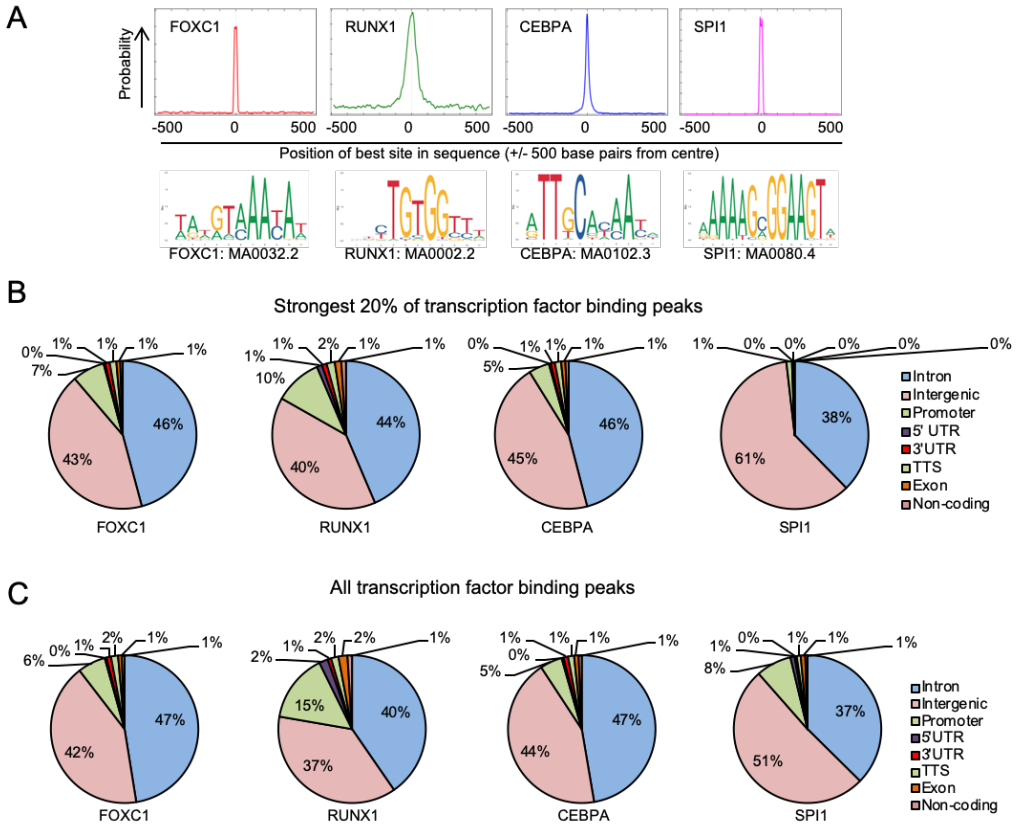


Figure 21. FOXC1 binding peaks are predominantly distributed in intronic and intergenic regions.

(A) MEME-ChIP motif enrichment plots. (B) Pie charts show genome annotations for the strongest 20% of transcription factor binding peaks. (C) Pie charts show genome annotations for all transcription factor binding peaks.

I next performed ChIP sequencing for H3K27Ac and H3K4Me1 in Fujioka cells and categorized the chromatin surrounding each transcription factor binding peak as Active-A ($H3K27Ac^{high}$, $H3K4Me1^{high}$), Active-B ($H3K27Ac^{high}$, $H3K4Me1^{low}$), Primed ($H3K27Ac^{low}$, $H3K4Me1^{high}$) or Silent ($H3K27Ac^{low}$, $H3K4Me1^{low}$) (Figures 22A). H3K27ac is an epigenetic modification of Histone H3 and marks active enhancers and promoters. In contrast, H3K4me1 generally marks enhancers only (Bannister and Kouzarides, 2011).

Considering the strongest 20% of binding peaks by pileup value for each transcription factor, I found that 29% and 41% of CEBPA or RUNX1 peaks respectively were bound

at sites of active chromatin (i.e. Active-A or Active B) but, consistent with its role as a pioneer factor, only 2% of SPI1 peaks. The reverse pattern was observed for silent chromatin with 98%, 35% and 24% of SPI1, CEBPA and RUNX1 peaks respectively bound in these regions. Consistent with pioneer activity, as for FOXA transcription factors, and a dual role in regulating the function of primed and active enhancers, the chromatin distribution of the strongest 20% of FOXC1 binding sites showed an intermediate distribution: 59% were bound to silent chromatin and 20% to active chromatin (Figure 22B). A similar pattern was observed when all transcription factor binding peaks for the four transcription factors were considered (Figure 22C). Many Active- B sites were located at gene promoters (Figure 23A-B), in contrast to the other classes of binding site.

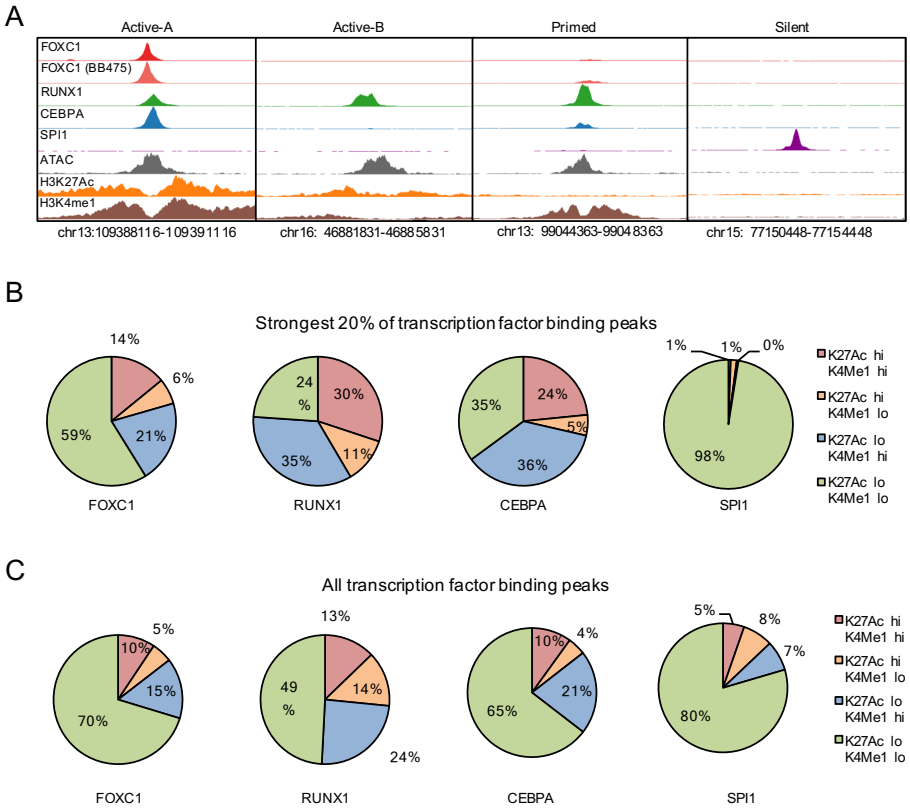


Figure 22. Chromatin distribution of FOXC1, RUNX1, CEBPA and SPI1 binding sites.
(A) Exemplar ChIP-seq tracks. (B) Pie charts show genome annotations for the strongest 20% of transcription factor binding peaks. (C) Pie charts show genome annotations for all transcription factor binding peaks.

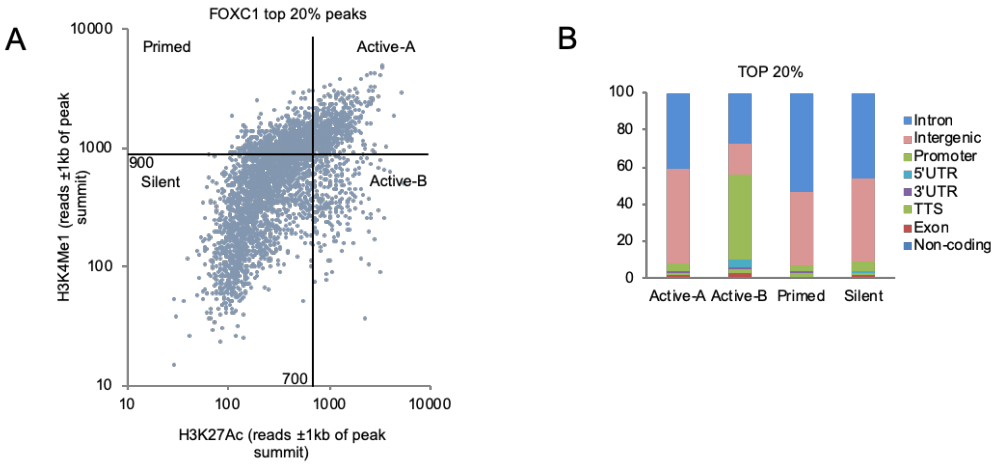


Figure 23. Chromatin distribution of FOXC1 binding sites.

(A) Dot plot shows H3K27Ac versus H3K4Me1 reads ±1kB from the absolute summit of each of the strongest 20% of FOXC1 peaks (n=3,773). This facilitated the annotation of transcription factor binding peaks according to their surrounding chromatin into four categories. (B) Genome region annotations for the strongest 20% of FOXC1 peaks (n=3,773) according to chromatin category shown in (A).

The differences in the strength and distribution of ChIP signal for H3K27Ac and K3K4Me1 surrounding the binding peaks of the four transcription factors are further demonstrated in the line and violin plots shown in Figure 24A. I also performed ATAC-sequencing in Fujioka cells and observed consistent findings: the strongest RUNX1 and CEBPA peaks bound more accessible chromatin whereas the opposite was the case for SPI1 (Figures 24B-C); FOXC1 exhibited an intermediate pattern of association.

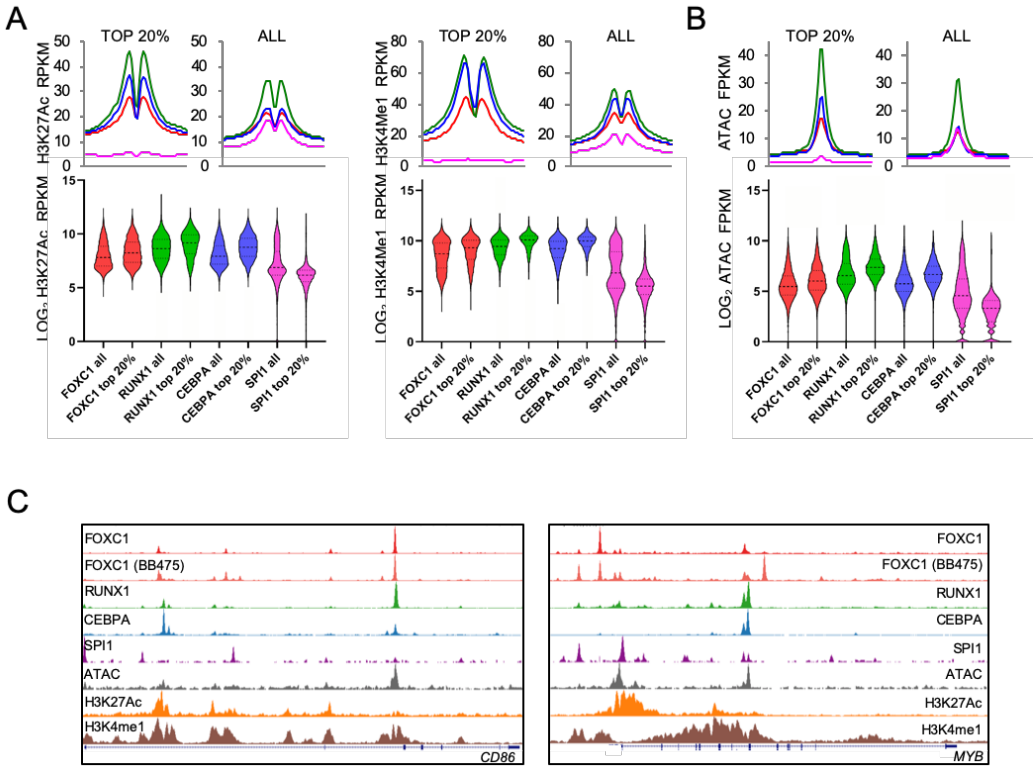


Figure 24. H3K27AC and ATAC signal at FOXC1, RUNX1, CEBPA and SPI1 binding sites.

(A) Graphs (upper panels) show mean ChIP signal for H3K27Ac (left) or H3K4Me1 (right) ±1kB surrounding the indicated sets of transcription factor binding peaks. Violin plots show distribution, median (thick dotted line) and interquartile range (light dotted lines) for ChIP signal. (B) As for (A) but for ATAC-seq signal. FPKM, fragments per kilobase per million mapped reads. (C) Exemplar ChIP-seq tracks.

To confirm a similar distribution of FOXC1 binding sites in Fujioka cells by comparison with primary patient blast cells, I performed FOXC1 ChIP sequencing in a normal karyotype AML sample (BB475). In keeping with the higher expression levels of *FOXC1* in primary versus Fujioka AML cells (Figures 11A-B), MACS2 (Zhang et al., 2008) identified 39,941 FOXC1 peaks (called peaks had pileup value ≥50 and fold enrichment over input ≥5). A representative example of ChIP-seq peaks is shown in Figure 25A. This revealed a well-defined narrow peak profile as expected for a transcription factor.

There was a substantial overlap of FOXC1 binding peaks in the two cell populations with, for example, 85.6% of the strongest 20% of FOXC1 peaks in Fujioka cells being represented in the BB475 primary sample (Figures 25B-C). There was also a strong positive correlation of FOXC1 peak strength in the two samples (Figures 25C-E). As a test of the binding specificity and in order to determine the preferred sequences bound by FOXC1, sequences of these FOXC1 binding regions were screened for overrepresented DNA motifs using MEME-ChIP. Forkhead transcription factors generally bind to DNA through sequences related to the RYAAAYA motif (where R= purine and Y = pyrimidine) (Chen et al., 2013b). The 5'-GTAAACA-3' sequence was found to be the highest-scoring sequence in FOXC1 peaks with p value $=10^{-8400}$ and 10^{-5825} in Fujioka and in BB475 cells, respectively. Notably, the binding sequence was centered on peak summits (+/-10 bps) as expected (Figure 25F). Altogether, these data identified for the first time the binding pattern of FOXC1 in AML, proved that the binding pattern was largely similar in patient AML samples and AML cell lines and revealed a predominant intergenic and intronic binding, suggesting a role in regulating gene transcription through enhancers.

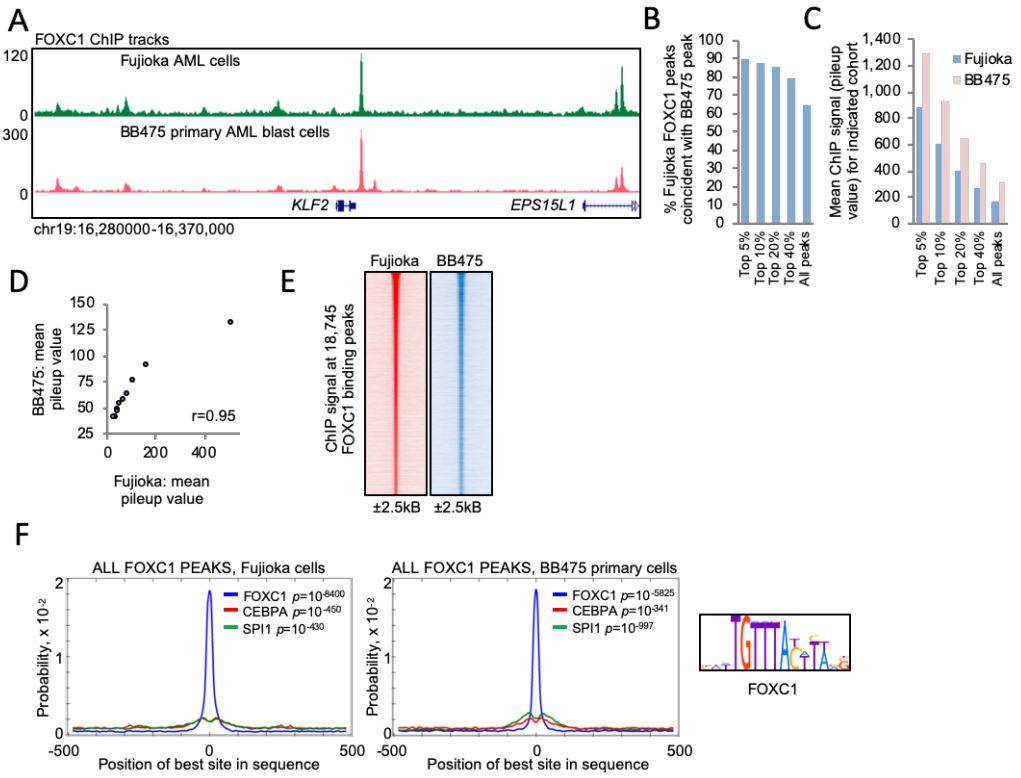


Figure 25. FOXC1 distribution in primary BB475 AML cells.

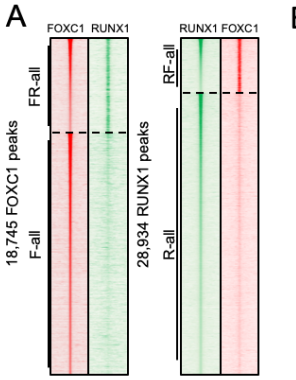
(A) Exemplar ChIPseq tracks. (B) Bar chart shows percentage of FOXC1 binding peaks in Fujioka AML cells in the indicated peak cohorts which are coincident with a FOXC1 binding peak in BB475 primary AML cells. (C) Bar chart shows mean FOXC1 ChIP signal in the indicated peak cohorts in Fujioka cells and at the same genomic locations in BB475 primary AML cells. (D) Scatter plot shows significant correlation between pileup values for specific peaks between two samples. (E) Heatmaps show FOXC1 ChIP signal at FOXC1 binding sites in Fujioka and at the same genomic locations in BB475 primary AML cells, ranked by peak strength. (F) MEME-ChIP motif enrichment plots.

Thus, FOXC1 exhibits a mixed pattern of binding to silent, primed and active chromatin predominantly at intergenic and intronic locations, with largely overlapping binding sites in primary and Fujioka AML cells.

4.6 Close physical interaction of FOXC1 with RUNX1 on chromatin

Our RIME and IP data (Table 19, Figure 19) suggested a strong physical interaction of FOXC1 with RUNX1. I therefore addressed how and where these two factors co-localised with each other on chromatin. Considering first all FOXC1 and RUNX1 binding sites in Fujioka AML cells, we found 5,246 genomic locations where the absolute summit of a FOXC1 peak was 200 base pairs or closer to the absolute summit of a RUNX1 peak (i.e. 28.0% of FOXC1 peaks and 15.3% of RUNX1 peaks) (Figures 26A-B). Considering the strongest 20% of FOXC1 and RUNX1 peaks, we identified 621 sites where a strong FOXC1 peak (pileup value ≥150) was co-located with a strong RUNX1 peak (pileup value ≥200) (termed “FR-20” sites) (Figure 26C). The genome-wide coincidence of FOXC1 and CEBPA peaks was lower (Figure 26B; 26.5% of FOXC1 peaks and 13.3%

of CEBPA peaks coincided; 554 sites of coincident strong FOXC1 and CEBPA binding). There was virtually no genome-wide coincident strong FOXC1 and SPI1 binding (Figure 26B; 8 sites genome-wide).



B

	Peak group	n		Peak group	n		Peak group	n
FOXC1 & RUNX1	FR-20	621	FOXC1 & CEBPA	FC-20	554	FOXC1 & SPI1	FS-20	8
	F-20	3152		F-20	3219		F-20	3765
	R-20	6695		C-20	6809		S-20	6931
	FR-all	5246		FC-all	4900		FS-all	1627
	F-all	13499		F-all	13845		F-all	16848
	R-all	28934		C-all	32056		S-all	33090
	TOTAL F PEAKS	18475		TOTAL F PEAKS	18475		TOTAL F PEAKS	18475
	TOTAL R PEAKS	34180		TOTAL C PEAKS	36956		TOTAL S PEAKS	34717

Pileup values for top 20% peaks: FOXC1 - 150; RUNX1 - 200; CEBPA - 220; SPI1 - 1000

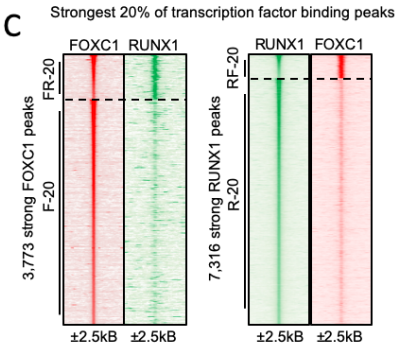


Figure 26. FOXC1 colocalization with RUNX1.

(A) Heatmaps show ChIP signal for RUNX1 at all FOXC1 binding sites (left panel) and FOXC1 at all RUNX1 binding sites (right panel). (B) Table shows the number of binding peaks in each of the indicated categories. (C) Heatmaps show ChIP signal for RUNX1 at strong FOXC1 binding sites (left panel) and FOXC1 at strong RUNX1 binding sites (right panel).

On the assumption that stronger peaks by pileup value were more likely to be functionally relevant (Maiques-Diaz et al., 2018), I focused my attention on an evaluation of the consequences of *FOXC1* knockdown at sites of strong dual FOXC1 and RUNX1 binding. Given the predominant distribution of FOXC1 and RUNX1 peaks at putative enhancers, I excluded sites located at promoter and 5'UTR sequences from the analysis in the first instance. There were 581 such genomic locations (which we term "FR-20 enhancer" sites). As comparators we evaluated strong FOXC1 binding sites without a nearby strong RUNX1 peak ("F-20 enhancer" sites, n=2,911) and vice versa for RUNX1 ("R-20 enhancer" sites, n=5,885). Most F-20 enhancer sites were at regions of silent chromatin whereas the great majority of R-20 and FR-20 enhancer sites were at regions of primed or active chromatin (Figure 27A). A similar pattern was observed when all peaks were considered (Figure 27B).

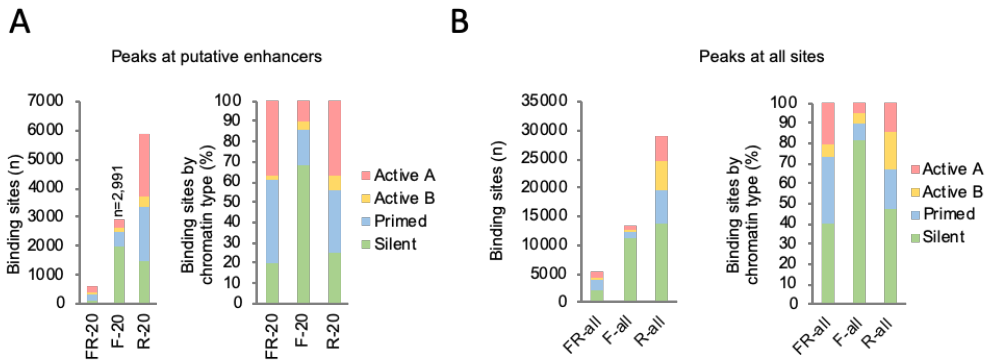


Figure 27. FOXC1 colocalization with RUNX1 at enhancers.

(A) Bar charts show chromatin categories for the indicated classes of strong FOXC1 and RUNX1 binding peaks in Fujioka cells by number (left panel) and proportion (right panel). (B) Bar charts show chromatin categories for the indicated classes of FOXC1 and RUNX1 binding peaks in Fujioka cells by number (upper panel) and proportion (lower panel).

Importantly, consistent with the physical interaction between FOXC1 and RUNX1 stabilising their interaction with chromatin, I noted that at FR-20 enhancer sites there was significantly greater FOXC1 ChIP signal by comparison with F-20 sites (mean±SEM 1672±97 versus 1097±19 reads/600 base pairs; t-test $p=10^{-9}$; Figure 28A). Likewise, at RF-20 enhancer sites there was significantly greater RUNX1 ChIP signal by comparison with R-20 sites (mean±SEM 2076±71 versus 1675±22 reads/600 base pairs; t-test $p=10^{-7}$; Figure 28B). Note that FR-20 and RF-20 refers to the same set of 581 genomic locations where the summits of a strong FOXC1 and a strong RUNX1 binding peak occur within 200 base pairs of each other. However, for FR-20 sites the ChIP signal shown is that surrounding the absolute summit of the FOXC1 binding peak whereas for RF-20 sites it is that surrounding the absolute summit of the RUNX1 peak.

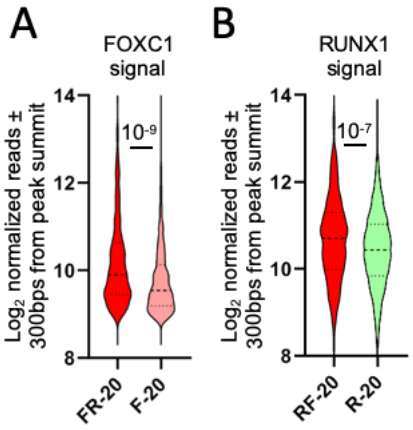


Figure 28. FOXC1 and RUNX1 stabilise each other at enhancers.

(A-B) Violin plots show distribution, median (thick dotted line) and interquartile range (light dotted lines) for ChIP signal for the indicated proteins at sites with strong FOXC1 and RUNX1 binding (FR-20, FOXC1 centered; RF-20 RUNX1 centered), FOXC1 binding (F-20) or RUNX1 binding (R-20) in Fujioka AML cells.

Once more, consistent with the physical interaction between FOXC1 and RUNX1 stabilising their interaction with chromatin, following *FOXC1* knockdown (Figures 12A-B) there was a relative loss of RUNX1 ChIP signal from RF-20 enhancer sites but not R-20 sites (RF-20: mean±SEM 2026±68 (NTC) versus 1640±73 (*FOXC1* KD) reads/600 base pairs; t-test $p=10^{-4}$. R-20: NTC versus KD comparison, p=not significant; Figure 29A). There was also a relative loss of RUNX1 signal from F-20 enhancer sites (F-20: mean±SEM 349±8 (NTC) versus 227±21 (*FOXC1* KD) reads/600 base pairs; t-test $p=10^{-7}$; Figure 29A), further supporting the concept that FOXC1 stabilises RUNX1 binding to chromatin, even where baseline RUNX1 ChIP signal is lower. By contrast there was an increase in relative CEBPA ChIP signal at FR-20 enhancer sites following *FOXC1* KD (mean±SEM 1248±72 versus 1622±170 reads/600 base pairs; t-test p=0.03), but no change at F-20 or R20 enhancer sites (Figure 29B).

To provide additional context to the analysis we performed ChIP sequencing for histone acetyltransferase EP300 and SWI/SNF chromatin remodelling complex protein SMARCC2, both of which we identified as FOXC1 interacting proteins (Table 19), histone deacetylase HDAC1 which is known to be recruited to chromatin by RUNX1, and H3K4Me2. While HDAC2 was identified as a FOXC1 interacting protein in Fujioka AML cells, HDAC1 was identified in BB475 primary AML cells (data not shown). My prior analyses have shown HDAC1 and HDAC2 exhibit overlapping genome-wide binding sites genome wide (data not shown). There was no change in EP300, H3K4Me2 or SMARCC2 ChIP signal at FR-20, F-20 or R-20 enhancer sites following FOXC1 knockdown; no change in H3K27Ac ChIP signal at FR-20 and F-20 sites; and no change in ATACseq signal at F-20 sites (Figures 29C-H). There was a modest relative increase in H3K27Ac ChIP signal at R-20 sites (mean±SEM 648±8 versus 704±9 reads/600 base pairs; t-test $p=10^{-8}$; Figure 29C) as well as a modest relative decrease in ATACseq signal (mean±SEM 155±2 versus 150±2 reads/600 base pairs; t-test p=0.03; Figure 29E). There was also a modest relative decrease in ATACseq signal at FR-20 sites (mean±SEM 170±6 versus 150±5 reads/600 base pairs; t-test p=0.02; Figure 29E).

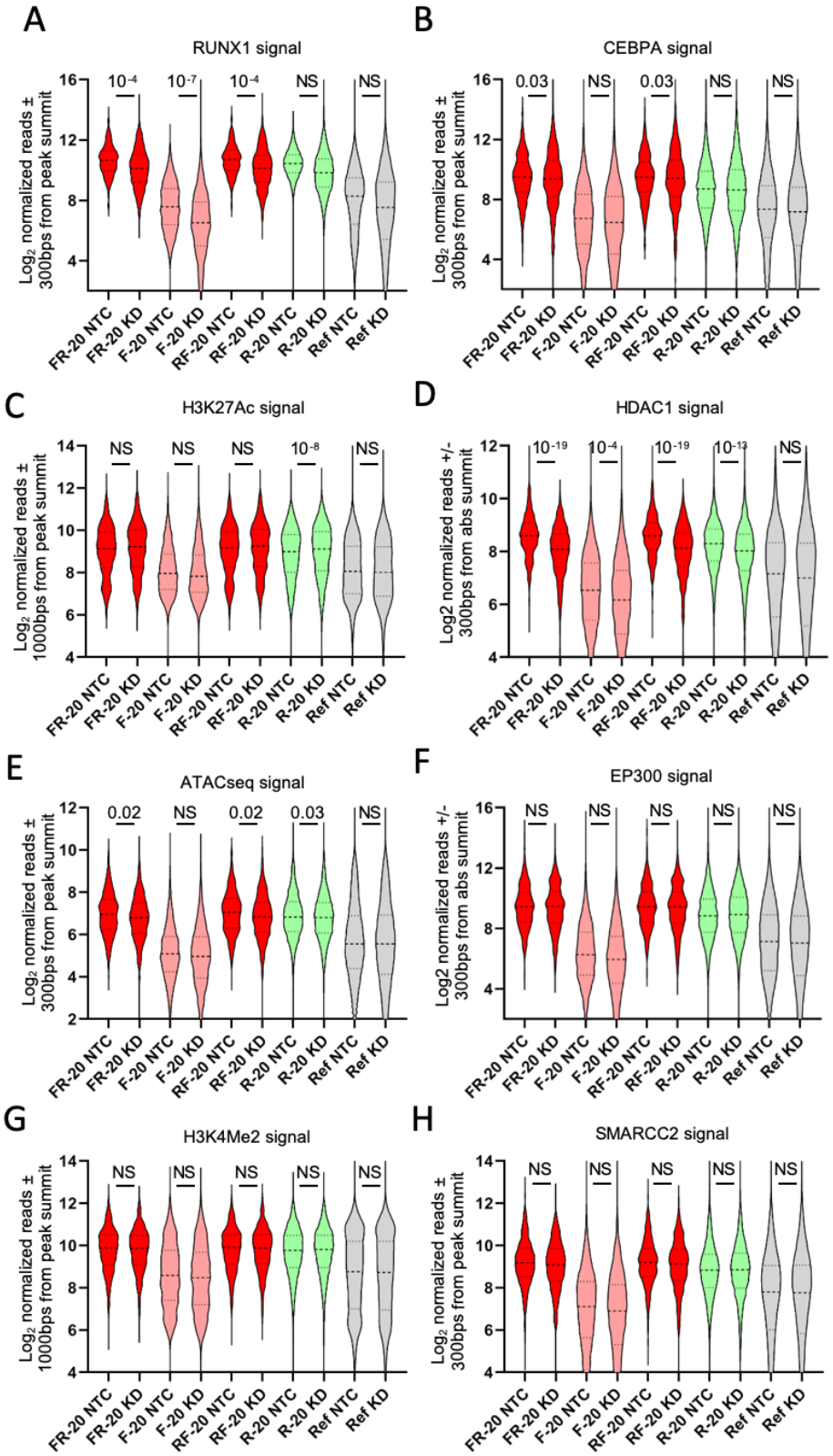


Figure 29. FOXC1-associated TFs consequences upon its depletion.

(A-H) Violin plots show distribution, median (thick dotted line) and interquartile range (light dotted lines) for ChIP signal for the indicated proteins at sites with strong FOXC1 and RUNX1 binding (FR-20, FOXC1 centered; RF-20 RUNX1 centered), FOXC1 binding (F-20) or RUNX1 binding (R-20) in Fujioka AML cells in control cells (NTC) or following FOXC1 KD. Ref, reference cohort used for normalization between experiments; NS, not significant. P values, unpaired t-test.

The most notable changes were in HDAC1 ChIP signal. There was a highly significant relative loss of HDAC1 ChIP signal at RF-20 enhancer sites (mean±SEM 446±12 versus 316±8 reads/600 base pairs; t-test $p=10^{-19}$; Figure 30A) and R-20 enhancer sites (mean±SEM 373±6 versus 325±5 reads/600 base pairs; t-test $p=10^{-13}$; Figure 30A). There was a lesser but still significant relative loss of HDAC1 ChIP signal at F-20 sites (mean±SEM 144±4 versus 123±4 reads/600 base pairs; t- test $p=10^{-4}$; Figure 30A). In addition to there being a greater proportional loss of HDAC1 signal at RF-20 versus R-20 enhancer sites following *FOXC1* knockdown (reduction of mean by 29.1% versus 12.9%), I also noted that in control cells there was significantly greater HDAC1 ChIP signal at RF-20 versus R-20 enhancer sites (mean±SEM 440±12 versus 373±6 reads/600 base pairs; t-test $p=10^{-9}$; Figure 30B). In contrast, in *FOXC1* KD cells there was no significant difference between HDAC1 ChIP signal at FR-20 versus R-20 enhancer sites (mean±SEM 316±8 versus 325±5 reads/600 base pairs; t-test p=not significant; Figure 30B-C).

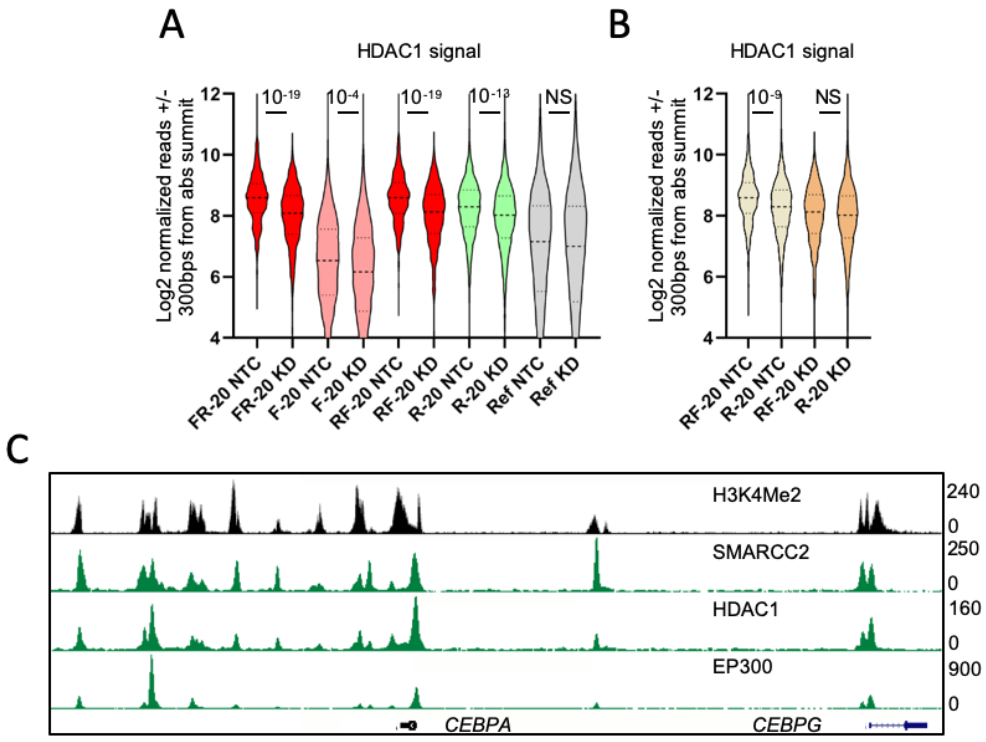


Figure 30. HDAC1 consequences upon FOXC1 depletion.

(A-B) Violin plots show distribution, median (thick dotted line) and interquartile range (light dotted lines) for ChIP signal for the indicated proteins at sites with strong FOXC1 and RUNX1 binding (FR-20, FOXC1 centered; RF-20 RUNX1 centered), FOXC1 binding (F-20) or RUNX1 binding (R-20) in Fujioka AML cells in control cells (NTC) or following *FOXC1* KD. Ref, reference cohort used for normalization between experiments; NS, not significant. P values, unpaired t-test. (C) Exemplar ChIP-seq tracks.

All together these data suggest that FOXC1 serves to stabilise the interaction of RUNX1 and HDAC1 with chromatin at a limited set of primed and active enhancers in myeloid leukemia cells.

4.7 FOXC1 acts as a repressor at a subset of primed and active enhancers

To evaluate the influence of FR-20 enhancer sites on gene expression I performed RNA sequencing in control and *FOXC1* KD Fujioka cells and identified 9,910 expressed protein coding genes (i.e. expressed at ≥2 FPKM in at least one sample). After *FOXC1* KD, 349 genes were upregulated by at least two-fold and 804 downregulated (Figure 31A). Upregulated genes included transcription factor genes with roles in monocyte/macrophage differentiation. Downregulated genes included those with roles in leukemic stem cell potential (e.g. *MYB*, *MYC*; Figure 31A) (Somervaille et al., 2009).
To further highlight the differentiation program induced by *FOXC1* KD in Fujioka cells, I used gene set enrichment analysis (GSEA) to compare the transcriptional changes with, as an example, those observed during phorbol ester-mediated terminal differentiation of THP1 AML cells into macrophages (Gazova et al., 2020) (Figure 31B); there was a highly significant overlap. Of note, quantitative PCR for key genes in primary patient cells following *FOXC1* KD gave similar results (Figure 31C).

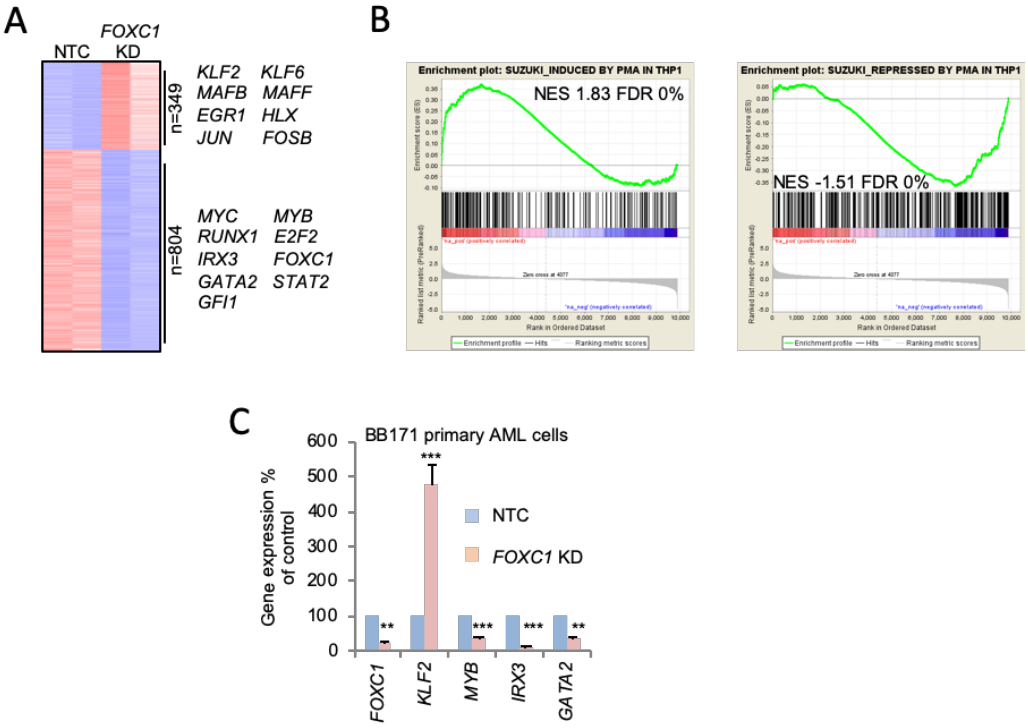


Figure 31. *FOXC1* KD induces upregulation of a myeloid gene set program.

(A-B) Human Fujioka AML cells were infected with a lentivirus targeting *FOXC1* for KD or a non-targeting control (NTC). (A) Heat map shows differentially expressed genes on Day 4 after initiation of KD, with transcription factor genes highlighted. (B) GSEA plots. NES, normalized enrichment score: FDR, false discovery rate. (C) Bar chart shows mean+SEM relative expression of the indicated genes in BB171 primary AML cells on Day 5 after initiation of *FOXC1* KD. * indicates p<0.05, ** indicates p<0.01 and *** indicates p<0.0001 for the indicated comparisons by an unpaired t test.

To facilitate additional analysis, I explored transcriptional changes using the Molecular Signatures Database Hallmark gene set collection, each of which conveys a specific biological state or process and displays coherent expression (Liberzon et al., 2011). Gene sets characteristic of cycling, metabolically active cells (i.e. MYC targets) were enriched among down regulated genes while using cell type specific gene “fingerprints” generated from microarray data from highly purified haematopoietic populations, I demonstrated that the monocyte fingerprint is enriched in up regulated genes (Figure 32A) (Chambers et al., 2007). In addition to specifically place the downregulated genes in an AML context I used gene sets generated from siRNA interference in the AML cell line THP1 and found the transcriptional signatures associated with *MYB* knockdown in these cells were enriched in downregulated genes (Figure 32B) (Gazova et al., 2020). Analysis of the most differentially expressed transcription factors revealed critical regulators of AML progression and/or differentiation as among the most strongly differential expressed TFs in Fujioka AML cells upon *FOXC1* depletion (Figure 32C).

Taken together these data demonstrate on a transcriptional level that the knockdown of *FOXC1* has important consequences for human Fujioka AML cells and results in a release of the repressed myeloid differentiation program.

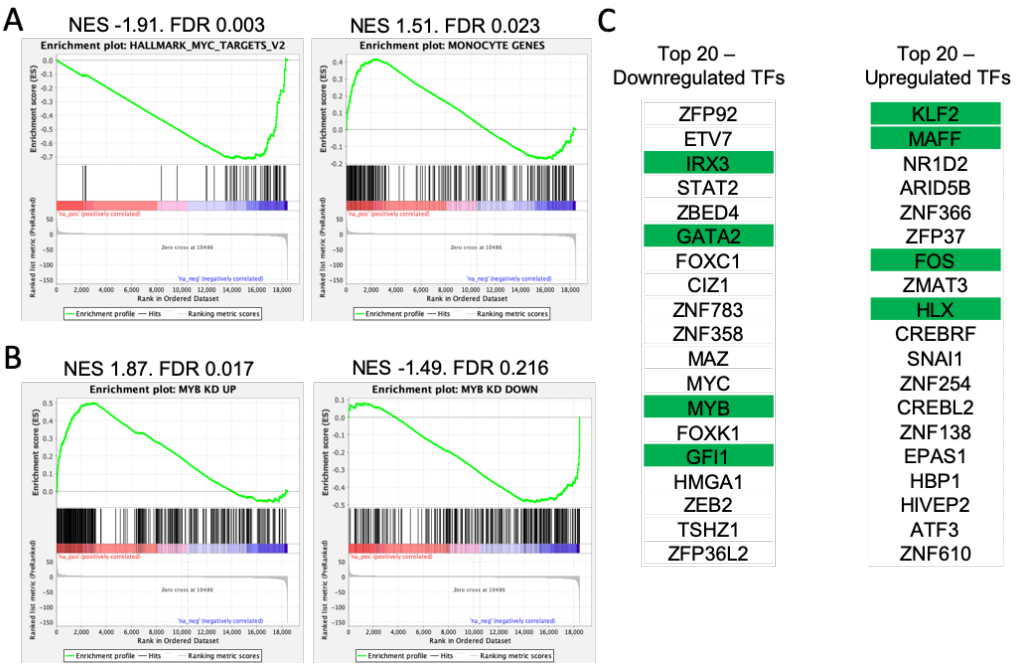


Figure 32. *FOXC1* Knockdown results in transcriptomic changes consistent with differentiation in Fujioka AML cells.

(A-C) Human Fujioka AML cells were infected with a lentivirus targeting *FOXC1* for KD or a non-targeting control (NTC). (A-B) GSEA plots. NES, normalized enrichment score: FDR, false discovery rate. (C) Tables show most downregulated and upregulated transcription factors. Genes highlighted in green are important for AML proliferation or differentiation in AML.

Given the overall relative loss of RUNX1 ChIP signal and gain of CEBPA ChIP signal from the 581 FR-20 enhancer sites (Figures 29A-B), I evaluated the proportional change in ChIP signal for both these factors at each enhancer. This facilitated grouping of FR-20 enhancer sites into four categories according to whether relative RUNX1 and CEBPA ChIP signal exhibited a relative increase or decrease following *FOXC1* KD (Figure 33A). Figure 33B shows one example of an enhancer for each group. I used Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010) to map genomic coordinates to the basal regulatory regions of nearby genes and then performed GSEA (Table 20-23). Within the limitations of this approach (regulatory elements do not necessarily control expression of the closest genes), I nevertheless observed strong enrichment of genes positioned close to Group 1 enhancers (reduced RUNX1 signal, increased CEBPA signal; Figures 33B, 33C, Table 20) among those upregulated following *FOXC1* KD. This contrasted with GSEA for the other three groups (Figure 33C) where no significant enrichment was observed among either upregulated or downregulated genes.

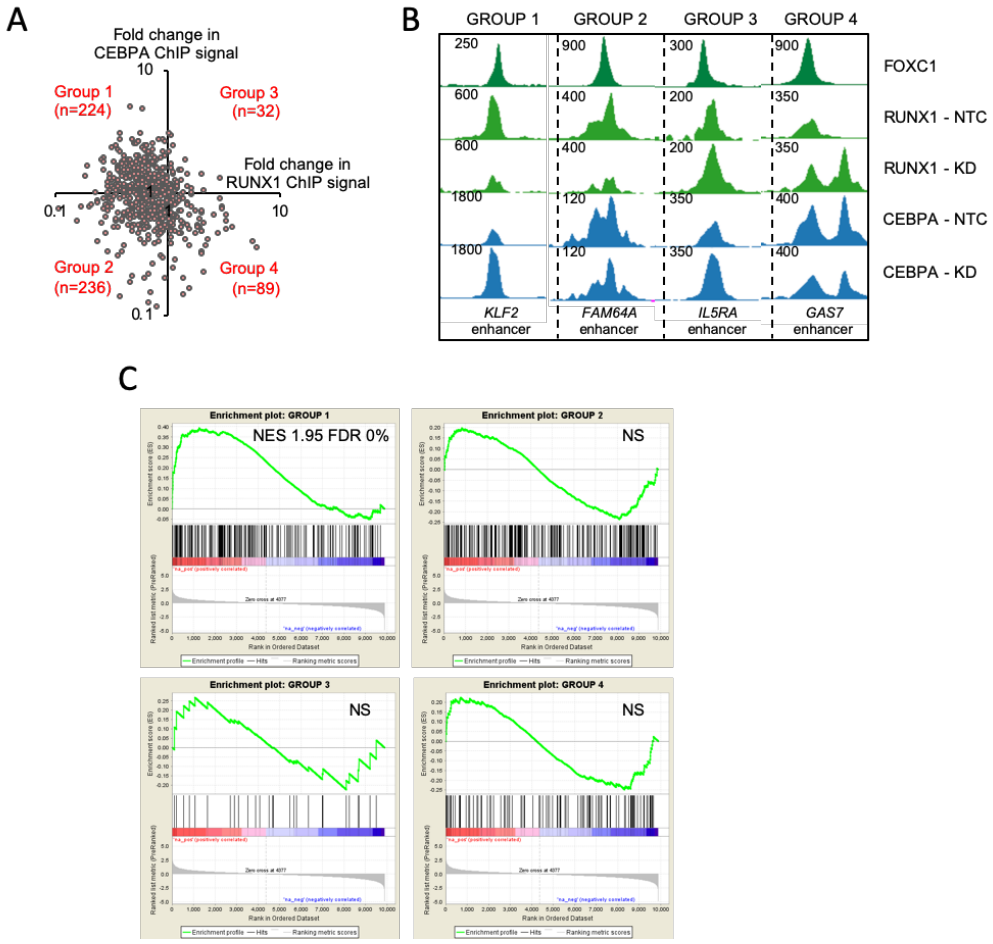


Figure 33. Reduced RUNX1 and increased CEBPA ChIP signal at enhancers controlling differentiation genes after *FOXC1* KD.

(A-C) Human Fujioka AML cells were infected with a lentivirus targeting *FOXC1* for KD or a non-targeting control (NTC). (A) Dot plot shows fold change in relative ChIP signal at 581 FR-20 enhancer sites and definition of four sub-groups. (B) Exemplar ChIP-seq tracks. (C) GSEA plots. NES, normalized enrichment score; FDR, false discovery rate.

Group 1							
ACTN1	CDNF	F5	IDNK	MEPE	PLEK	SIGLECL1	TPRG1
ADAMTS17	CEACAM21	FABP7	IL1R2	METTL9	PLEKHG3	SLAMF1	TRIB2
ADCY8	CEACAM4	FAM136A	IL1RL1	MINPP1	POLG	SLC15A1	TRIM36
ADGB	CELF2	FAM170A	IL1RL2	MKNK1	POLI	SLC34A2	TRPC1
ADM	CEP128	FAM208A	IMMP2L	MTBP	POU4F1	SLC37A3	TRPM2
ADSS	CEP350	FAM49B	INPP4A	MTFR2	PPARD	SLC45A1	TTC30A
AKNA	CERS3	FAM72B	IRF2BP2	MYC	PPIAL4C	SLC6A5	TTLL4
ALOX5AP	CITED2	FANCE	IRS2	MYCBP2	PPIAL4G	SLC7A1	TXLNB
ALX1	CLMP	FASLG	IVNS1ABP	MYH9	PPM1A	SLC9A8	TXN2
AMPD3	COA5	FBXL22	KCNH7	MYO16	PPP1R36	SLITRK5	UBAC2
ANTXR2	COG3	FBXW2	KCNH8	MYO1B	PRKCH	SMIM20	UBASH3B
APLF	COTL1	FCGR1A	KCNIP1	MYOM2	PRR16	SNTB1	UBE2J1
ARHGAP15	COX6C	FCGR1B	KCNN2	NAA20	PSAT1	SORL1	UBL3
ARHGAP22	CPN1	FFAR2	KCTD12	NABP1	PSEN1	SOX4	UGGT2
ARHGAP26	CPSF6	FGF9	KCTD6	NBPF11	PSIP1	SOX5	USP3
ARHGAP32	CRB2	FGGY	KDM7A	NBPF6	PTOV1	SP3	USP6NL
ARHGEF26	CRTAM	FIG4	KLF2	NDRG1	PTPRJ	SPATA13	VEGFA
ARHGEF3	CXCR4	FIGN	KLHL1	NELL1	QSOX1	SPATA8	VPS13B
ARMC2	CXCR6	FNBP1L	KLRC4-KLRK1	NPIPB3	RAB3C	SPERT	WDR27
ARRDC4	CYP27A1	FNDC3A	KLRK1	NPIPB5	RAP2B	SPP1	WISP1
ASAP1	CYSLTR2	FOSL2	KNCN	NR3C1	RASSF9	SPRY1	XIRP2
ATP2B1	DAAM1	FRMD3	KYNU	NRIP1	RBPJ	SRC	ZBTB18
ATP6V1G3	DACT1	FRMD4A	LAMB3	NUB1	RCVRN	SRI	ZCCHC2
ATP8A2	DAOA	FYCO1	LCP1	NUDT4	RERE	ST3GAL1	ZFAT
B3GALT1	DCAF5	G0S2	LILRB3	OLA1	RFK	ST8SIA6	ZNF107
B4GALT5	DCN	GABRP	LMO2	OR4B1	RGS2	STEAP4	ZNF117
BARX2	DCT	GABRR2	LPIN1	OR56A1	RHCG	STK17B	ZNF138
BBC3	DENND1A	GAS7	LPP	OR56A4	RHEB	STK24	ZNF175
BBX	DHRS7	GK2	LRCH1	ORM2	RHOBTB2	STXBP5	ZNF217
BCAR3	DIO2	GNG2	LRGUK	OSM	RILPL1	SUCNR1	
BCAS1	DMAC1	GOT2	LRMDA	PAPLN	RIN2	TANK	
BCAT1	DNAJC3	GPC5	LRRC3	PAPSS2	RIPOR2	TBC1D7	
BCKDHB	DOCK4	GPR183	LRRC8C	PCNX1	RPLP1	TBL2	
BCLAF1	DOCK9	GPR42	LRRC8D	PCOLCE2	RPS20	TEK	
BICC1	DUSP4	GPR6	LYN	PCSK1	RPS9	TFEC	
BICD1	EDNRB	GPR85	LYZ	PCSK5	RUFY3	TGFA	
BMT2	EEA1	GRAMD1B	MAGEF1	PDE11A	S100A12	THBS2	
BNC2	EFR3A	GRM8	MANBAL	PDE4D	S100A8	THSD7B	
BTG1	ELL2	GRSF1	MAP1LC3A	PDE7B	SAE1	TLDC1	
C14orf166	ENSA	H3F3C	MAP3K9	PDHB	SAMSN1	TLE3	
C1QTNF9	ENSG00000258989	HACD1	MAP4K4	PEBP4	SARAF	TM9SF2	
C3orf58	EPHA7	HBEGF	MAP7	PFDN1	SATB1	TMED2	
CACNA2D1	EPHB3	HECW2	MBD2	PGGT1B	SC5D	TMEM132E	
CAPRIN1	EPS15L1	HENMT1	MBNL1	PHACTR1	SCEL	TMEM63B	
CARMIL1	EQTN	HGF	MCL1	PHYHIPL	SEL1L	TNFRSF10B	
CBLB	ERLIN1	HMCN1	MDFIC	PIGU	SEL1L3	TNFRSF10C	
CCL1	ERRFI1	HOOK1	MED25	PKIB	SELP	TNFRSF10D	
CD33	ERV3-1	HS3ST1	MEDAG	PLB1	SESN1	TNFRSF11A	
CD48	ETS1	HS3ST2	MEF2C	PLCB1	SHISA2	TNFSF18	
CDKAL1	EXOC4	ID1	MEGF9	PLCB4	SIGLEC5	TOMM20	

Table 20. List of genes in GROUP 1.

Group 2								
ABCC4	C1GALT1	DAD1	FGGY	ITGA6	MLLT1	PSAPL1	ST3GAL6	VWA8
ABI2	C1QL2	DAOA	FILIP1L	ITPR1	MLXIP	PTAFR	STAMBPL1	WASHC3
ABLIM1	C8orf37	DBI	FLT1	JCAD	MPRIP	PTGDR	STMND1	WDR5
ABLIM3	C9orf47	DCBLD2	FOS	JPH2	MRPL36	PTPN22	STRIP2	WDR61
ACAD8	CA8	DDI1	FOXF1	JUN	MRPS33	PXDN	STX1A	WDR70
ACSBG2	CACNA2D1	DDX20	FOXI1	KCND3	MSMB	QKI	SUCO	WNT16
ACTA2	CAMKMT	DENND3	FOXO3	KCNRG	MTFR2	RAB11FIP1	SUMF1	YWHAQ
ACTL7B	CAPN12	DLX2	FYCO1	KCNS3	MTRNR2L6	RAB4A	SUSD6	YWHAZ
ACTN4	CASP2	DMXL1	FZD7	KDM8	MUS81	RAPH1	SV2B	ZAP70
ADAM17	CASP8	DNAJB12	GABRG1	KHDRBS1	MUSK	RASGRP3	SVIL	ZFAT
ADD3	CBX6	DNAJC5B	GCDH	KHDRBS3	MXI1	RBM26	SYCE2	ZNF280D
ADGRE1	CCZ1B	DPF1	GDF6	KL	MYO16	RBMS3	SYK	ZNF608
ADGRV1	CD1C	DRAM1	GDNF	KLF12	MYT1L	RDH14	TAB2	ZNF648
AFAP1	CD200	DTHD1	GLB1L3	KLF4	NEGR1	RGCC	TAC1	ZNF703
AFAP1L1	CD276	DUSP1	GLI2	KLF7	NFAM1	RHOH	TAPT1	ZNF706
AFAP1L2	CD28	DUSP16	GLI3	KLHL25	NKX6-2	RNASEH2B	TBC1D21	ZRANB2
AIM2	CD44	DYNC2H1	GLUL	KRTDAP	NOG	RNF19B	TBC1D5	
AK2	CD47	DYNLRB2	GNAI1	LAPTM5	NOTCH1	RNF219	TCF12	
AKAP13	CD5L	EBF1	GOT2	LCP2	NRF1	ROBO1	TCF20	
ANKS1A	CD82	EDNRB	GPC5	LHPP	NSRP1	RPL15	TCP11	
ANO6	CDA	EED	GPC6	LMAN1L	NTNG1	RPS6KA4	TESC	
ANTXRL	CDH10	EFCAB5	GPR37	LNX1	NXNL2	RRAGD	TGFBR1	
APCDD1	CDK15	EFEMP2	HAND2	LPAR1	OR10H1	RSBN1	THRAP3	
APOBEC3A	CDYL2	EFNB2	HDDC2	LPCAT1	OR4E2	RXFP2	TLE3	
ARHGAP24	CEACAM4	EGFL7	HEG1	LPIN1	OTUD5	RXRA	TLNRD1	
ARHGAP25	CEACAM7	EIF4A3	HEY2	LPP	P2RY1	SATB1	TMEM131	
ARID2	CENPV	EMX2	HGF	LRRK2	PAN3	SDHAF3	TMEM132C	
ARL5C	CEP44	ENSG00000283563	HHEX	MAFF	PCBP3	SEPT2	TMEM235	
ARL9	CHIC2	ERG	HIPK2	MAGI2	PCOTH	SFMBT2	TMEM37	
ARMC2	CHRNA9	ERN1	HOPX	MAP3K5	PDE7A	SHROOM3	TMEM39B	
ARRDC3	CLCN1	ERRFI1	HPS3	MAP3K7CL	PDE7B	SIGLEC7	TMEM67	
ATXN1	CLDN10	ETS2	HUS1B	MAP4K4	PDP1	SIGLEC9	TNFAIP8	
AUH	CLEC16A	ETV6	IBTK	MAP7	PDS5B	SIM1	TNFRSF19	
AZIN2	CLEC2L	EXOC2	ICAM2	MAP7D1	PDZD8	SIPA1L3	TNFRSF1B	
B3GLCT	CLINT1	EXOC6	ID2	MAPK10	PHLPP1	SIX3	TNFRSF8	
BACH1	CMSS1	EYA3	IFI16	MAPKAPK2	PICALM	SLC12A8	TOMM20	
BBS12	CNTN1	FAM124A	IFT57	MARCO	PIGC	SLC1A2	TOX	
BBX	CNTNAP2	FAM170A	IGF2R	MARF1	PIGL	SLC22A11	TOX2	
BCL11A	COL12A1	FAM46A	IL10	MATN1	PIM2	SLC35F1	TPK1	
BCL2	COL15A1	FAM53B	IL16	MB21D2	PINK1	SLC36A1	TRIB2	
BCL2L14	COL4A1	FAM98A	IL1R2	MBD3L5	PLA2G6	SLC45A1	TRIM15	
BCL6	COL6A1	FANCM	IL31	MBNL1	PLD6	SLC45A4	TRIM26	
BMP10	CP	FAT2	INHBA	MCHR2	PLN	SLCO3A1	TSPAN18	
BORCS5	CPED1	FBXO33	INHBB	MCMBP	PLPP2	SLITRK5	UACA	
BRAF	CREB1	FBXW8	INPP5A	MEF2C	POT1	SOCS1	UBE2E1	
BRF2	CSK	FCRL1	INPP5F	MICAL2	POU4F1	SOCS3	UBE2J1	
BTLA	CSMD3	FDFT1	IRF2BP2	MICALCL	PRKCQ	SPHAR	UST	
BUD23	CTSB	FFAR2	IRF8	MICU1	PROM1	SPRY2	VAPA	
C16orf45	CXCR6	FGF12	IRS2	MIER2	PRR16	SRSF5	VAV3	
C17orf67	CYP4F2	FGF2	ISL1	MIS18BP1	PRSS58	ST3GAL1	VPS37C	

Table 21. List of genes in GROUP 2.

Group 3			
ADGRE2	EGLN2	PEX11G	TMCC2
ADORA2B	EQTN	PHYKPL	TMEM106A
ARHGAP27	FKBP4	PRKCD	UBE2E3
ARHGEF18	FNDC3A	RCBTB2	VPS53
ARL4D	GAS7	RCVRN	
CACNA1C	HIPK2	REG1B	
CDH1	HLX	REG3G	
CDH3	HPCAL1	RFLNB	
CDYL	IL5RA	RRM2	
CLEC2L	IRS2	RUNX1	
CLIC6	ITGA4	SBNO2	
CNTN4	LINC00282	SLC15A1	
COL23A1	MAP3K14	SOHLH2	
CYP2A6	MLNR	SPART	
CYSLTR2	MYO16	STK24	
DHRS12	NUAK2	TBC1D26	
DUSP10	OR10K2	TEK	
ECI2	OR7C1	TKT	

Table 22. List of genes in GROUP 3.

Group 4			
AKT3	IRS2	PPP1R12B	WDR4
ALDH8A1	ITIH5	PRKCD	WDR45B
AMOTL1	IVD	PRKCQ	WDR5
APOD	KATNAL1	PSTPIP1	WNK1
ARHGAP26	KCTD12	RAB3D	WNT5A
BCL2L14	KCTD4	RABL2A	XYLT1
C2orf40	KLF2	RAP1B	ZFHX3
CARS	KMT2E	RAP1GAP2	ZFP36L2
CASR	KNSTRN	RCVRN	ZFYVE28
CD86	LEF1	RIN2	ZMIZ1
CEACAM1	LGALSL	RPL34	ZNF414
CHST4	LINC00282	RPS24	ZNF638
CMIP	LMO7	RRAGC	ZSCAN20
CNIH3	MACC1	RXRA	
COG2	MARCH10	SARAF	
CXCL17	MB21D1	SATB1	
CYTH1	MDM1	SDCCAG8	
DAD1	ME1	SFMBT2	
DAPK2	MOSPD1	SGK1	
DDX43	MPEG1	SLC39A13	
DHRS12	MRC2	SLC44A3	
DMRT1	MTMR12	SPAG6	
DMRT3	MXD4	SPI1	
DNAH17	MYBBP1A	SPINK2	
DUSP4	MYH11	SPNS2	
DYSF	MYO16	SPTLC1	
EPS15L1	NAA20	SRPK2	
ERC2	NCEH1	SUGP2	
ETV6	NCK2	SULT1A1	
EXT1	NDE1	SULT1A2	
F3	NINJ2	SUSD1	
FAM163A	NPIPA7	SV2B	
FAM71D	NPIPA8	SYT2	
FANCL	NR3C1	TARBP1	
FNBP1	OR1D5	TAT	
FOXD4L1	OR4E2	TBC1D5	
FOXK2	OR5AN1	TDRD5	
GAS7	OSBPL5	TES	
GGACT	OXGR1	TEX51	
GOLPH3	PCCA	TKT	
GPHN	PDE9A	TNFRSF11B	
GYPC	PELI1	TNFSF10	
HAAO	PFKFB3	TPT1	
HERC1	PGBD5	TSPAN16	
HOMER3	PHC2	TSPAN3	
HOPX	PIP4K2A	UBL3	
HS6ST3	PIWIL4	UGCG	
IARS	PLAC1	USP20	
ID2	PLCG2	VPS33B	
IRF2BP2	POU3F1	WDR26	

Table 23. List of genes in GROUP 4.

Analysis of transcription factor binding motifs surrounding the absolute summit of FOXC1 ChIPseq peaks in the four different enhancer groups confirmed strong, significant enrichment for FOXC1 motifs in all of those, with no major differences between the groups (Figure 34). In particular, Group 1 and Group 2 show a very similar overall motif binding for FOXC1 and the other transcription factors studied in this thesis (RUNX1, CEBPA and PU.1).

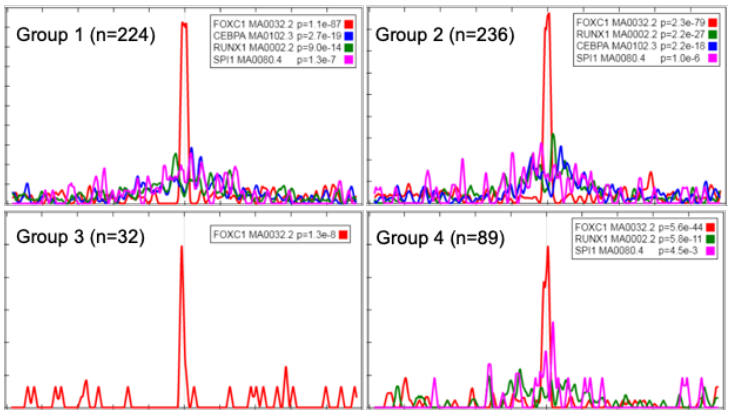


Figure 34. Motif analysis of FOXC1/RUNX1 bound enhancers.
MEME-ChIP motif enrichment plots.

Leading edge analysis revealed that the enrichment of Group 1 genes in GSEA was driven by those upregulated during normal monocytic lineage expression, such as *KLF2*, *MBD2*, *ID1*, *S100A12* and *RGS2*, among others (Figure 35A) (Bagger et al., 2019). Eighty-eight percent of Group 1 FR-20 enhancers exhibited primed or active chromatin configuration in basal conditions (Figure 35B).

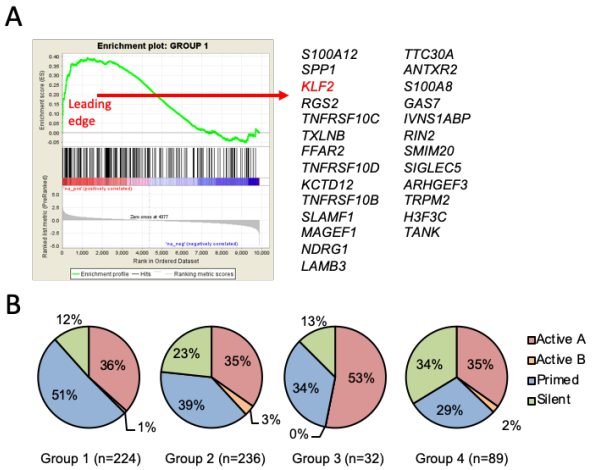


Figure 35. Gene expression changes after *FOXC1* knockdown.
(A-C) Human Fujioka AML cells were infected with a lentivirus targeting *FOXC1* for KD or a non-targeting control (NTC). (A) GSEA plots. NES, normalized enrichment score: FDR, false discovery rate. (B) Pie charts show chromatin categories of the four FR-20 enhancer groups.

Group 1 enhancers exhibited significantly greater baseline RUNX1 binding compared with Group 2 and 4 enhancers, and intermediate levels of CEBPA binding compared with Group 2 (lower) and Group 4 (higher) (Figure 36A). There was no difference in H3K4me2 signal between the groups but differences in the levels of ATAC, EP300 and SMARCC2 binding (Figure 36B). There was no change in H3K4Me2, SMARCC2 or EP300 ChIP signal, or ATAC seq signal, for any of the enhancer groupings following

FOXC1 knockdown (data not shown). However, there was a significant reduction in HDAC1 ChIP signal at Group 1 and Group 2 enhancers (i.e. those where there was a reduction in RUNX1 signal) and a significant relative increase in H3K27Ac signal at Group 1 sites (Figures 36C-E). As for RUNX1, there was significantly greater ChIP signal for HDAC1 at Group 1 versus Group 2 enhancers (Figure 36C).

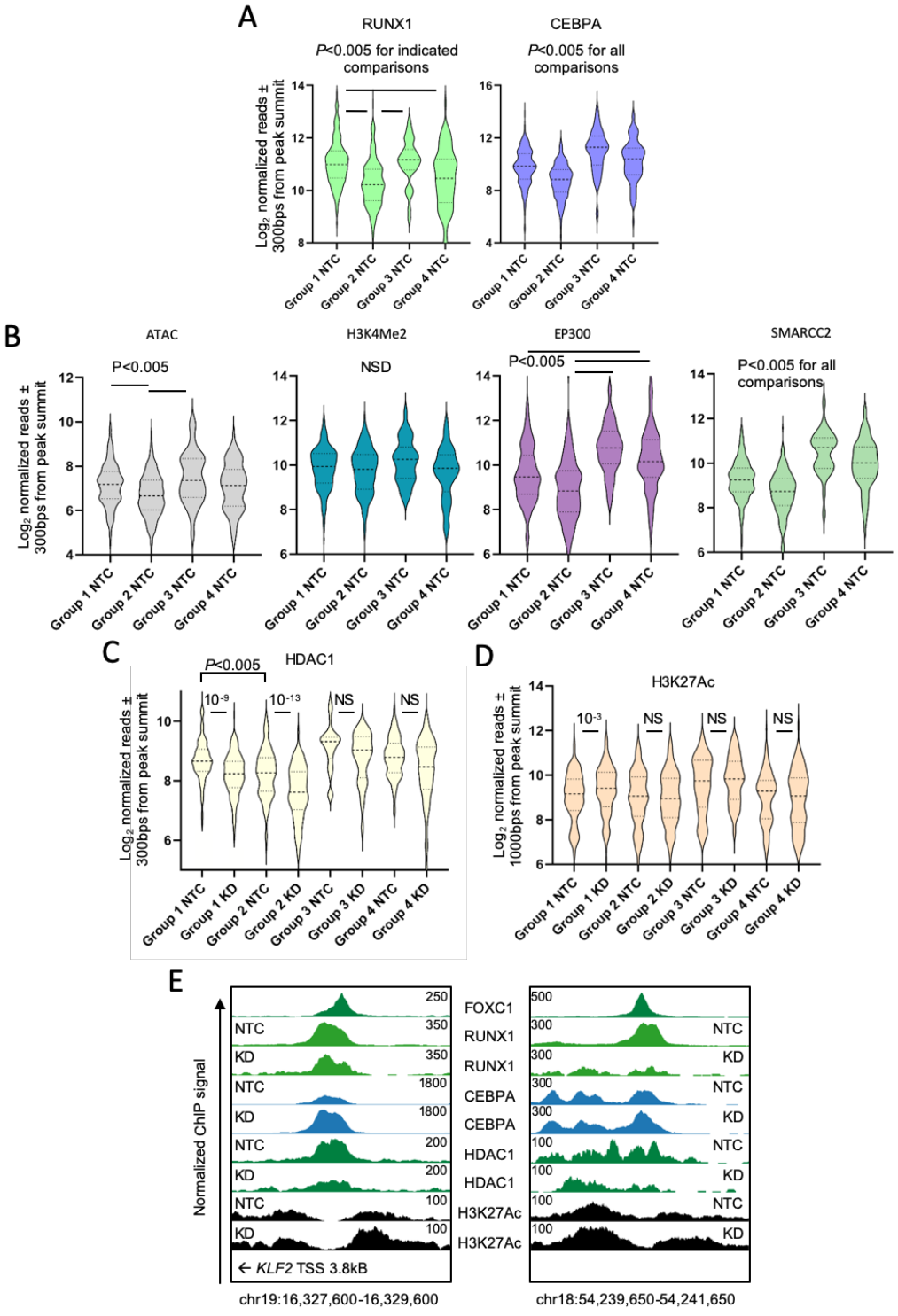


Figure 36. Reduced RUNX1 and increased CEBPA ChIP signal at enhancers controlling differentiation genes after *FOXC1* KD.

(A-C) Human Fujioka AML cells were infected with a lentivirus targeting *FOXC1* for KD or a non-targeting control (NTC). (A-D) Violin plots show distribution, median (thick dotted line) and interquartile range (light dotted lines) for ChIP signal for the indicated proteins and the indicated Groups of FR-20 enhancer sites in control (NTC) or *FOXC1* KD cells on Day 5. P value, one way ANOVA with Tukey post hoc; or unpaired t-test. (E) Exemplar ChIP-seq tracks.

To summarise, Group 1 enhancers lose RUNX1, HDAC1 and gain H3K27Ac signal upon *FOXC1* KD. Plus, these enhancers are the ones associated with upregulated genes in the RNAseq experiments.

One of the Group 1 FR-20 enhancers was positioned 4kB downstream of the TSS for *KLF2*, a gene involved in monocytic lineage differentiation and upregulated following *FOXC1* knockdown. KLF2 belongs to the Kruppel-like family of transcription factors (KLFs), which are critical regulators of a broad range of important cellular functions including development, growth, and differentiation and play an essential role along the monocyte/macrophage cellular differentiation pathway (Andreoli et al., 2010). Thus, KLF transcription factors are generally inhibited in acute myeloid leukaemia. In particular, *KLF2*, *KLF3*, *KLF5* and *KLF6* are generally expressed at lower levels in AML blast and progenitor cells than normal myeloid differentiated cells (Humbert et al., 2011).

To confirm its ability to promote differentiation of AML cells, I generated a conditional construct for full-length *KLF2*. Expression was induced in Fujioka cells using a doxycycline-regulated system (Figure 37A). I noted both upregulation of CD86 (Figure 37B) and reduction of clonogenic activity, with both fewer and smaller colonies in the presence of increased KLF2 expression (Figure 37C).

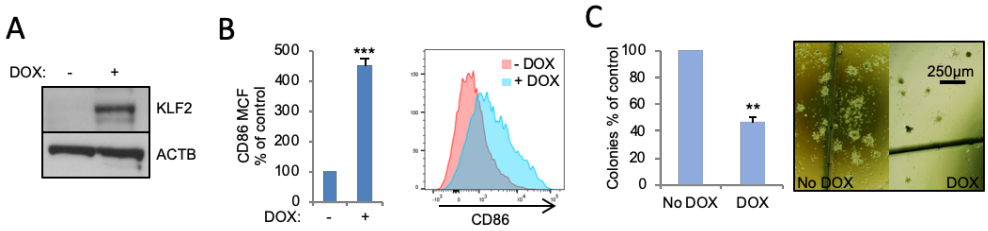


Figure 37. KLF2 overxpression induces differentiation in Fujioka AML cells.
(A-C) Fujioka cells were infected with lentiviruses expressing *KLF2* under the control of a doxycycline-regulated promoter. (A) Western blot shows induced expression of KLF2. (B) Bar chart (left panel) shows mean+SEM mean cell fluorescence (MCF) for CD86 as determined by flow cytometry in the indicated conditions (n=3). Right panel: representative flow cytometry plots. (C) Bar chart (left panel) shows means+SEM colony-forming cell (CFC) frequencies of KLF2-expressing Fujioka cells relative to control cells after 10 days in semi-solid culture (n=3). Right panel: representative colonies. * indicates $p<0.05$, ** indicates $p<0.01$ and *** indicates $p<0.0001$ for the indicated comparisons by an unpaired t test.

Interestingly, other genes involved in differentiation program were also upregulated (e.g. *CD86*, *JUN*, *KLF6*, *FOS* and *MEF2C*).

To determine whether KLF6 also participates directly in monocytic differentiation and has a key role in FOXC1 driven human leukemias, I generated a conditional construct

for full-length *KLF6*. Expression was induced in Fujioka AML cells using a doxycycline-regulated system. Exogenous expression of KLF6 was verified by Western analyses (Figure 38A). Furthermore, *KLF6* forced expression induced induction of the myeloid markers CD86 (Figure 38B). Taken together, these data indicate that KLF2-KLF6 are a critical regulator of Fujioka monocytic differentiation.

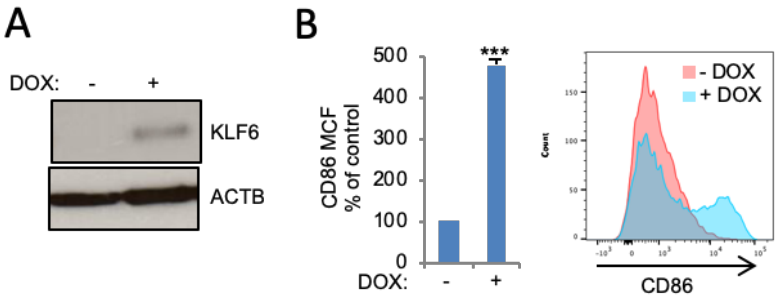


Figure 38. KLF6 overxpression induces differentiation in Fujioka AML cells.
(A-C) Fujioka cells were infected with lentiviruses expressing *KLF6* under the control of a doxycycline-regulated promoter. (A) Western blot shows induced expression of KLF2. (B) Bar chart (left panel) shows mean+SEM mean cell fluorescence (MCF) for CD86 as determined by flow cytometry in the indicated conditions (n=3). Right panel: representative flow cytometry plots. * indicates $p<0.05$, ** indicates $p<0.01$ and *** indicates $p<0.0001$ for the indicated comparisons by an unpaired t test.

Thus, at a discrete set of regulatory elements distributed close to genes upregulated during myelomonocytic differentiation, and which exhibit high RUNX1 and intermediate CEBPA binding (i.e. Group 1 FR-20 enhancers), FOXC1 serves as a transcription repressor through stabilising RUNX1 and HDAC1 binding, thus limiting enhancer activity.

4.8 Forced recruitment or displacement of RUNX1 from FOXC1 binding sites regulates expression of differentiation genes

To provide functional evidence that the protein:protein interaction of FOXC1 with RUNX1 is critical in conferring a differentiation block in FOXC1^{high} AML cells, I generated a conditional construct in which the DNA-binding domain of FOXC1 was fused directly to RUNX1 (FKD-RUNX1) so that RUNX1 remained bound at FOXC1 sites after *FOXC1* KD. Construct expression was induced in Fujioka AML cells using a doxycycline-regulated system (Figure 39A). As expected, *FOXC1* KD in Fujioka AML cells promoted differentiation, as evidenced by increased expression of the monocyte/macrophage lineage differentiation marker CD86 (used as a surrogate for differentiation in the experiments that follow) (Figure 39B), while the FKD-RUNX1 fusion protein completely blocked upregulation of CD86 expression in response to *FOXC1* KD (Figures 39B).

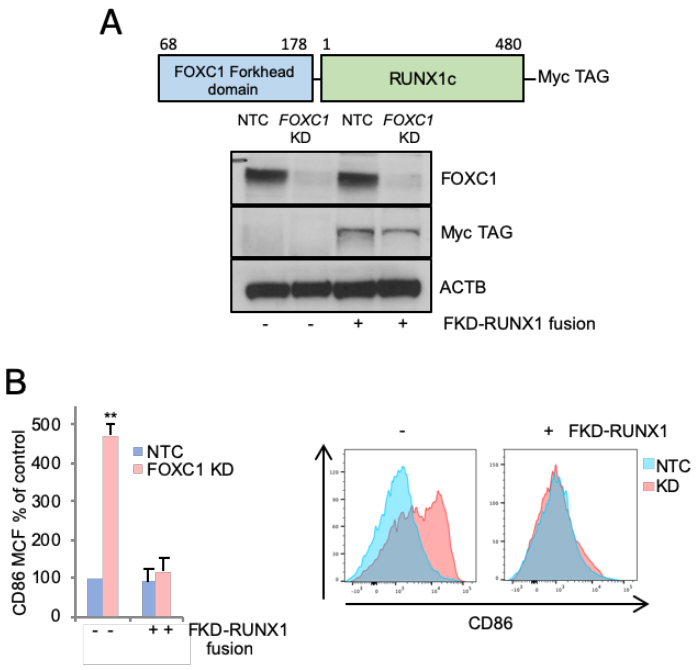


Figure 39. Forced recruitment of RUNX1 from FOXC1 binding sites blocks differentiation.

(A-B) Fujioka cells were infected with lentiviruses expressing a FOXC1 Forkhead domain–RUNX1c fusion protein under the control of a doxycycline-regulated promoter. (A) Western blots show expression of the indicated protein and transcription factors constructs. (B) Bar chart shows mean+SEM CD86 mean cell fluorescence (MCF) as determined by flow cytometry in the indicated conditions (n=3). Right panel: representative flow cytometry plots. * indicates $p<0.05$, ** indicates $p<0.01$ and *** indicates $p<0.0001$ for the indicated comparisons by an unpaired t test.

In contrast, induced expression in Fujioka cells of FOXC1 mutants G165R and F112S (Figure 40A), which exhibit reduced interaction with RUNX1 (Figure 20D), promoted immunophenotypic differentiation (Figure 40B). I performed Q-PCR in Fujioka cells expressing the FOXC1 G165R mutant and observed similar gene expression changes to those observed in Fujioka cells following *FOXC1* KD (Figure 40C).

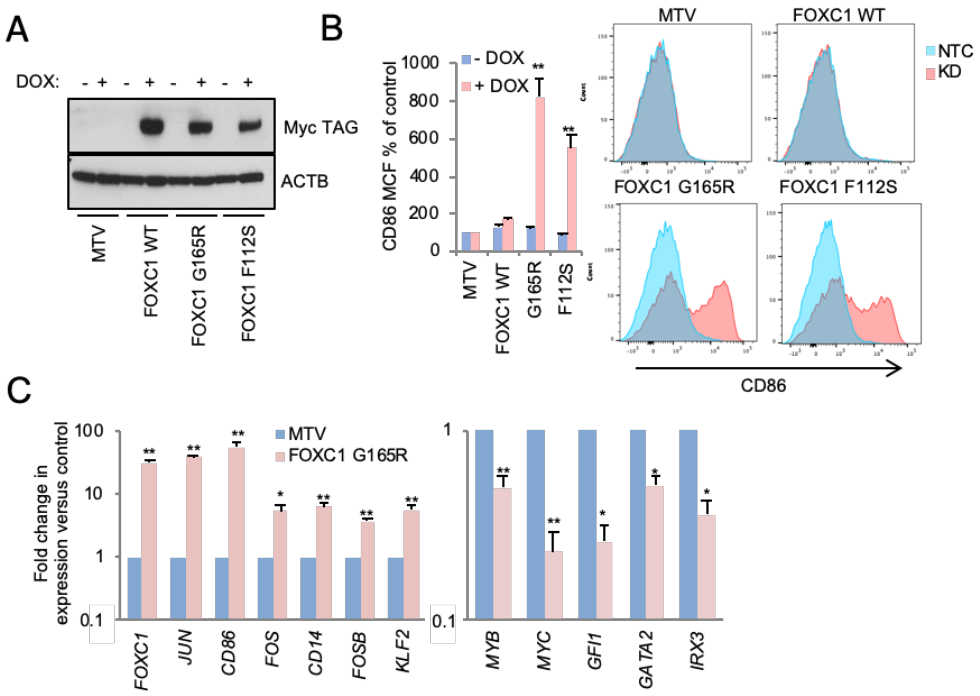


Figure 40. Forced displacement of RUNX1 from FOXC1 binding sites regulates expression of differentiation genes.
(A-C) Fujioka cells were infected with lentiviruses expressing sequences coding for FOXC1 WT, FOXC1 G165R and FOXC1 F112S under the control of a doxycycline-regulated promoter. (A) Western blots show expression of the indicated proteins. (B) Bar chart (left panel) shows mean+SEM CD86 mean cell fluorescence as determined by flow cytometry in the indicated conditions (n=3). Right panel: representative flow cytometry plots. (C) Bar chart shows mean+SEM relative expression of the indicated genes in Fujioka cells of Day 4 following induced expression of FOXC1 G165R. * indicates p<0.05, ** indicates p<0.01 and *** indicates p<0.0001 for the indicated comparisons by an unpaired t test.

4.9 *FOXC1* knockdown triggers redistribution of RUNX1 from enhancers to promoters

Evaluation of RUNX1 ChIP signal at FR-20 enhancer sites hinted at a redistribution of RUNX1 binding following *FOXC1* KD. MACS2 identified 17,589 RUNX1 binding peaks in *FOXC1* KD Fujioka cells but only 67% of these overlapped with RUNX1 peaks in control Fujioka cells (i.e. absolute peak summits within 200 base pairs of each other; Figure 41A). When the strongest 20% of RUNX1 peaks in KD cells (pileup value ≥168) were considered, only 37.2% overlapped with a strong RUNX1 peak in control cells (Figure 41B). I grouped strong RUNX1 peaks into three categories as shown in Figure 41B. Group A peaks were predominantly enhancer bound, with only 9% bound to 5'UTR or promoter regions. By contrast, 58% of Group C peaks were 5'UTR or promoter bound (Figure 41C). The shared set of Group B peaks exhibited an intermediate pattern.

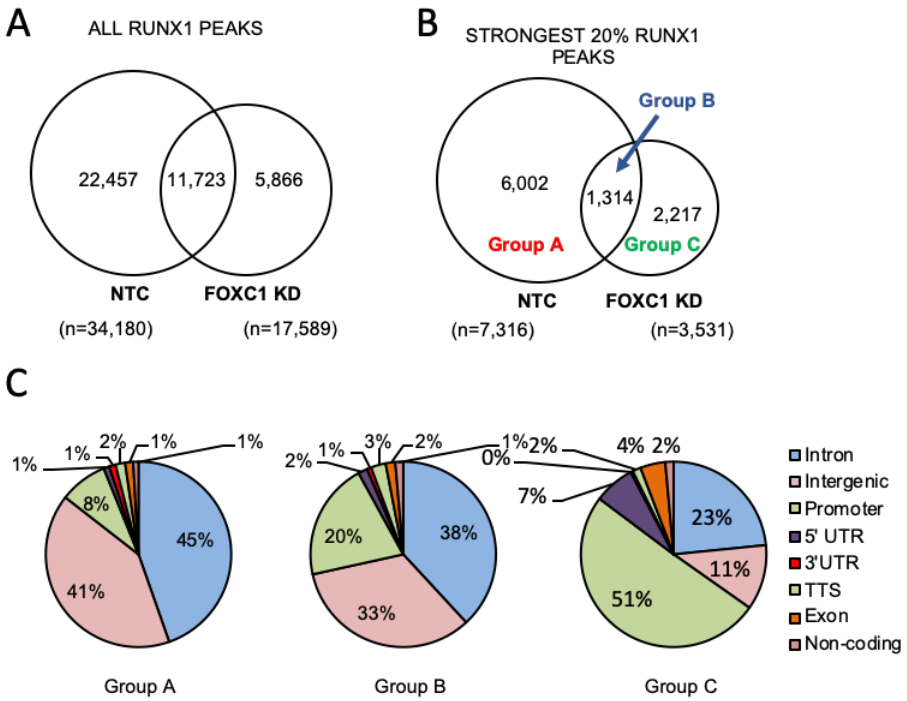


Figure 41. *FOXC1* knockdown triggers redistribution of RUNX1 binding
(A) Venn diagram shows intersection of all RUNX1 binding peaks in control (NTC) or *FOXC1* KD Fujioka AML cells. (B) Venn diagram shows intersection of the strongest 20% of RUNX1 binding peaks in control (NTC) or FOXC1 KD Fujioka AML cells & classification of groups. (C) Pie charts show genome annotations for RUNX1 binding peaks in Groups A-C.

There was a significant relative decrease in ChIP signal at Group A peaks for RUNX1 (mean±SEM 1425±10 versus 1056±50 reads/600 base pairs; t-test $p=10^{-13}$) and HDAC1 (mean±SEM 338±4 versus 290±3 reads/600 base pairs; t-test $p=10^{-23}$); and a significant increase in CEBPA (mean±SEM 707±18 versus 806±27 reads/600 base pairs; t-test $p=10^{-3}$) and H3K27Ac (mean±SEM 635±6 versus 679±7 reads/600 base pairs; t-test $p=10^{-6}$) ChIP signal (Figures 42A-D). Group C peaks displayed the opposite pattern. There was a significant relative increase in ChIP signal for RUNX1 (mean±SEM 543±10 versus 2523±52 reads/600 base pairs; t-test $p<10^{-50}$) and HDAC1 (mean±SEM 1649±26 versus 1871±28 reads/600 base pairs; t-test $p=10^{-9}$); and a significant decrease in CEBPA (mean±SEM 1220±53 versus 821±43 reads/600 base pairs; t-test $p=10^{-9}$) and H3K27Ac (mean±SEM 975±14 versus 896±13 reads/600 base pairs; t-test $p=10^{-5}$) ChIP signal (Figures 42A-D). Group B peaks displayed an intermediate pattern between Groups A and C.

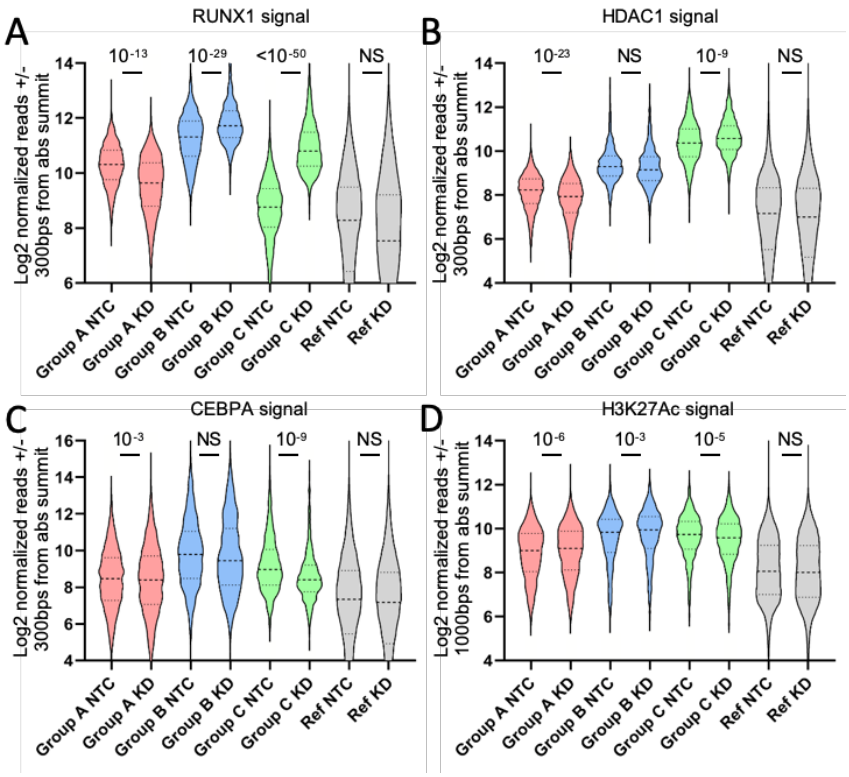


Figure 42. *FOXC1* knockdown triggers changes of TFs and epigenetic regulators binding on chromatin.

(A-D) Violin plots show distribution, median (thick dotted line) and interquartile range (light dotted lines) for ChIP signal for the indicated proteins and groups in control (NTC) or *FOXC1* KD Fujioka AML cells. Ref, reference cohort used for normalization between experiments; NS, not significant. P values, unpaired t-test

I confirmed the decrease and increase in RUNX1 ChIP signal by ChIP PCR for a number of Group A and Group C peaks (Figures 43A-B). MEME-ChIP confirmed that genomic sequences at the center of Group A and Group B RUNX1 binding peaks were strongly enriched for RUNX1 motifs (Figure 44A). By contrast, there was no enrichment for RUNX1 motifs in genomic sequences surrounding Group C RUNX1 peaks. Rather, there was enrichment for several short ungapped GC-rich motifs identified by Discriminative Regular Expression Motif Elicitation (DREME), as well as for SP1 a member of the SP1/Kruppel-like family of transcription factors. Interestingly, using ChIP PCR at three exemplar Group C promoters of genes whose expression was downregulated, in addition to an increase in RUNX1 ChIP signal we observed an increase in KLF2 ChIP signal and a concomitant decrease in SP1 ChIP signal (Figure 44B). I speculated that KLF2 might interact with RUNX1, recruiting it to Group C promoters through KLF2's interaction with GC-rich sequences (Figure 44C), in the process displacing related factors with transcription activating potential. In the same way that loss of RUNX1 from Group 1 FR-20 enhancers was associated with an increase in expression of nearby genes (Figure 33C), consistent with RUNX1's repressor activity, gain of RUNX1 at gene promoters was associated with downregulated expression.

Following *FOXC1* knockdown by GSEA there was strong enrichment among downregulated genes of those whose promoters or nearby enhancers were subject to the strongest accumulation of RUNX1 (Figure 44D).

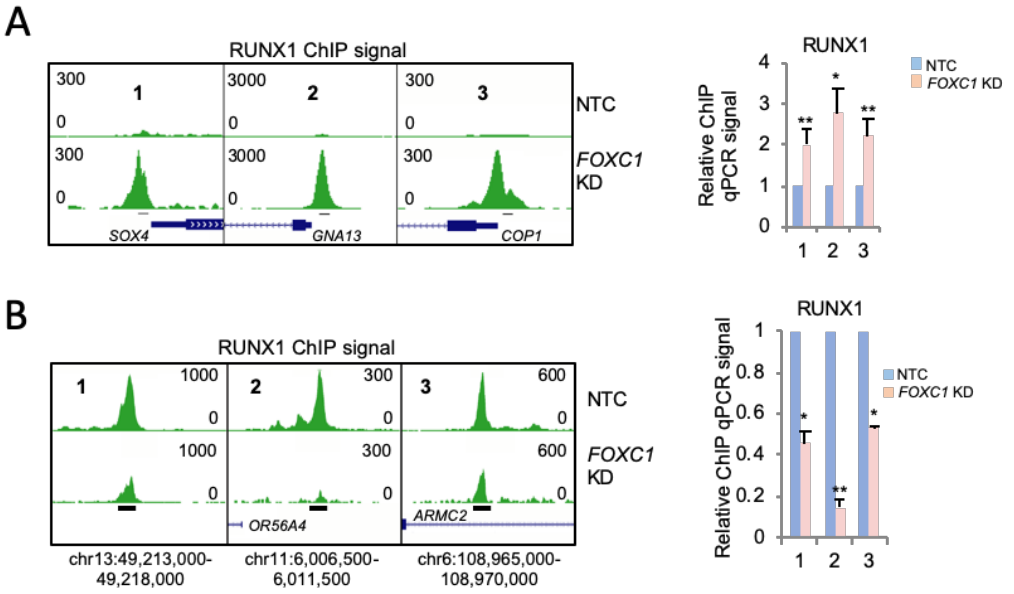


Figure 43. Validation of redistribution of RUNX1 binding
(A-B) Human Fujioka AML cells were infected with a lentivirus targeting *FOXC1* for KD or a non-targeting control (NTC). (A) Exemplar ChIPseq tracks at three genomic locations (left panel) with confirmatory ChIP-PCR for the indicated proteins (right panels); mean+SEM relative ChIP signal is shown (n=3). (B) Exemplar ChIPseq tracks at three genomic locations (left panel) with confirmatory ChIP-PCR for the indicated proteins (right panel); mean+SEM relative ChIP signal is shown (n=3). * indicates p<0.05, ** indicates p<0.01 and *** indicates p<0.0001 for the indicated comparisons by an unpaired t test.

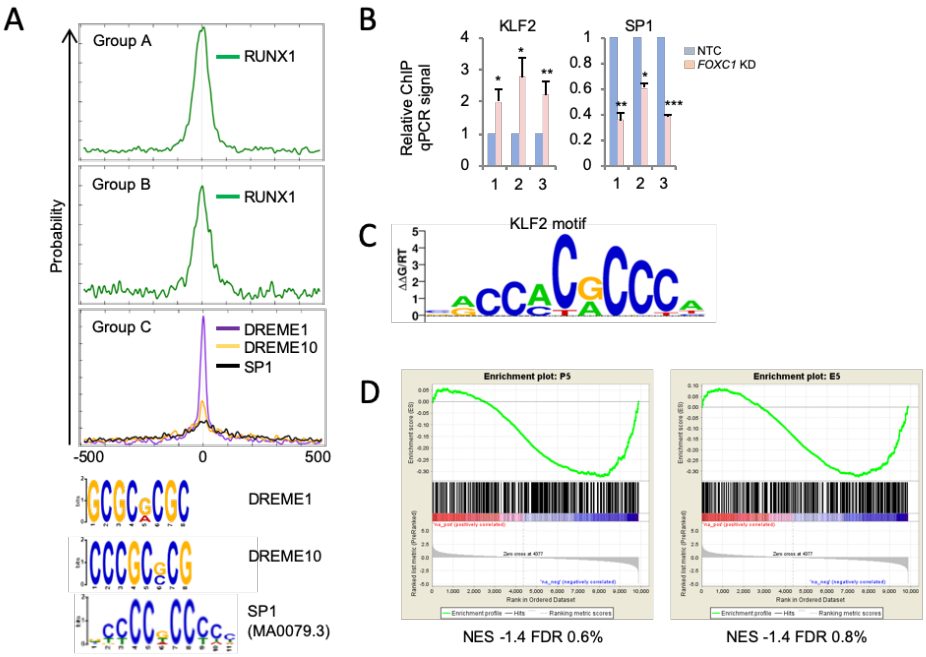


Figure 44. KLF2 increases its binding to the new RUNX1 peaks
(A-D) Human Fujioka AML cells were infected with a lentivirus targeting *FOXC1* for KD or a non-targeting control (NTC). (A) MEME-ChIP and DREME motif enrichment plots for the indicated groups. (B) Confirmatory ChIP-PCR for the indicated proteins; mean+SEM relative ChIP signal is shown (n=3). (C) KLF2 binding motif. (D) GSEA plots. P5, genes whose promoters exhibit ≥5-fold increase in ChIP signal for RUNX1; E5, enhancers which exhibit ≥5-fold increase in ChIP signal for RUNX1. NES, normalized enrichment score; FDR, false discovery rate. * indicates p<0.05, ** indicates p<0.01 and *** indicates p<0.0001 for the indicated comparisons by an unpaired t test.

Together these data demonstrate that *FOXC1* knockdown triggers a differentiation process which involves the redistribution of the transcription repressive activity of RUNX1 from enhancers to promoters.

4.10 *FOXC1* knockdown triggers redistribution of RUNX1 on *MYB*, *IRX3* and *GATA2* promoters

Analysis of the most downregulated transcription factors upon *FOXC1* KD revealed that *MYB*, *IRX3* and *GATA2* were amongst the most strongly downregulated gene in Fujioka KD cells by comparison with NTC (Figure 32C). Interestingly, RUNX1 ChIP-seq analysis using NTC or *FOXC1* KD infected cells revealed that a de novo RUNX1 redistribution on promoters of critical growth and renewal genes includes *MYB*, *IRX3* and *GATA2* promoters. *Myb* was first identified as an oncogene in retroviruses that induced leukemias but it has been shown to play a central role in the development and progression of AML driven by several different oncogenes, including mixed lineage leukemia (MLL)-fusion genes (Pattabiraman and Gonda, 2013). *IRX3* is overexpressed in about 30% of AML cases whereas its expression is almost undetectable during normal haematopoiesis. Moreover, functional experiments demonstrated its contribution to the block of differentiation (Somerville et al., 2018). *GATA2* expression is critical

throughout various stages of haematopoiesis, and its levels partly determine the fate along different myeloid cell lineages. Moreover, *GATA2* forced expression leads to erythroblast proliferation and it blocks erythroid terminal differentiation (Vicente et al., 2012). Taken together, these data raised a question as to whether the differentiation block release seen in *FOXC1* KD cells may be mediated at least in part by downregulation of expression of the *MYB*, *IRX3* and *GATA2.*

Figure 45A illustrates RUNX1 new peaks on *MYB*, *IRX3* and *GATA2* promoters upon *FOXC1* depletion. Downregulation of these transcription factors was confirmed by qPCR (Figure 45B) and western blotting (Figure 45C). Unfortunately, no GATA2 commercial antibody was found to work in Fujioka AML cells to check GATA2 protein levels. Of note, quantitative PCR for these key genes in primary patient BB171 cells following *FOXC1* KD gave similar results (Figure 31C) and same analysis in primary patient BB475 cells also revealed a strong downregulation of *MYB*, *IRX3* and *GATA2* genes (Figure 45D).

To investigate whether downregulation of these transcription factors in AML make a functional contribution to transformation, knockdown (KD) experiments were firstly performed in human Fujioka AML cells. First, KD efficiency was confirmed by qPCR (Figure 46A). Following *MYB*, *IRX3* and *GATA2* initiation of KD there was up regulation of the myeloid differentiation marker CD86 (Figures 46B) and loss of clonogenic potential (Figure 46B).

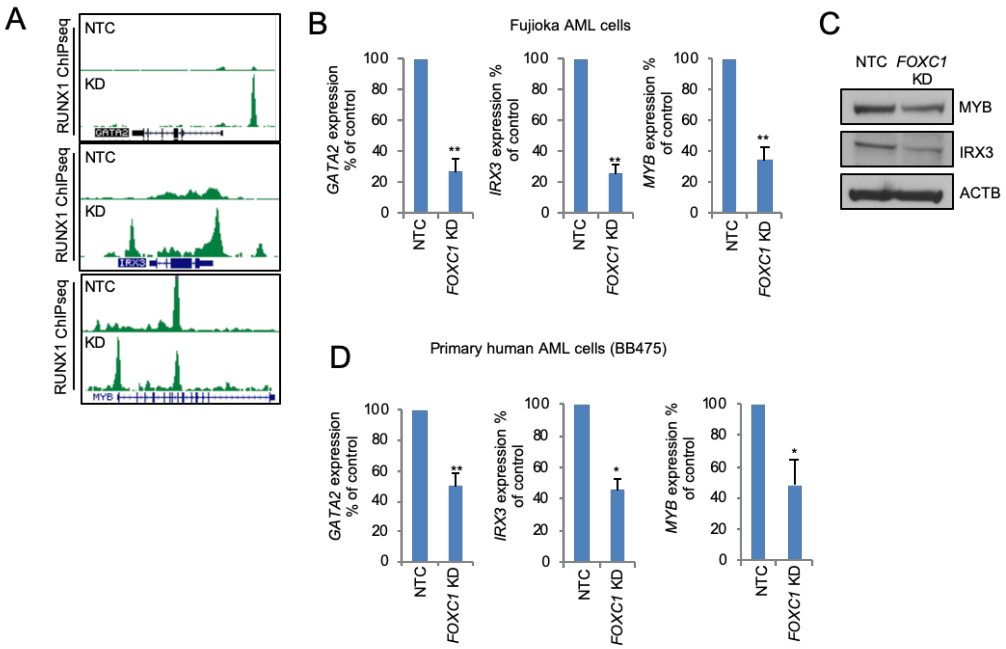


Figure 45. *FOXC1* kockdown leads to downregulation of *MYB*, *IRX3* and *GATA2*
(A-C) Human Fujioka AML cells were infected with a lentivirus targeting *FOXC1* for KD or a non-targeting control (NTC). (A) Exemplar ChIPseq tracks. (B) RPKM fold change is shown. (C) Western blots show expression of the indicated proteins. (D) Primary patient AML cells (BB475) were infected with a lentivirus targeting *FOXC1* for KD or a NTC with puromycin drug resistance as selectable marker (n=3). Bar chart shows mean+SEM relative transcript expression in KD versus control cells (n=3). * indicates p<0.05, ** indicates p<0.01 and *** indicates p<0.0001 for the indicated comparisons by an unpaired t test.

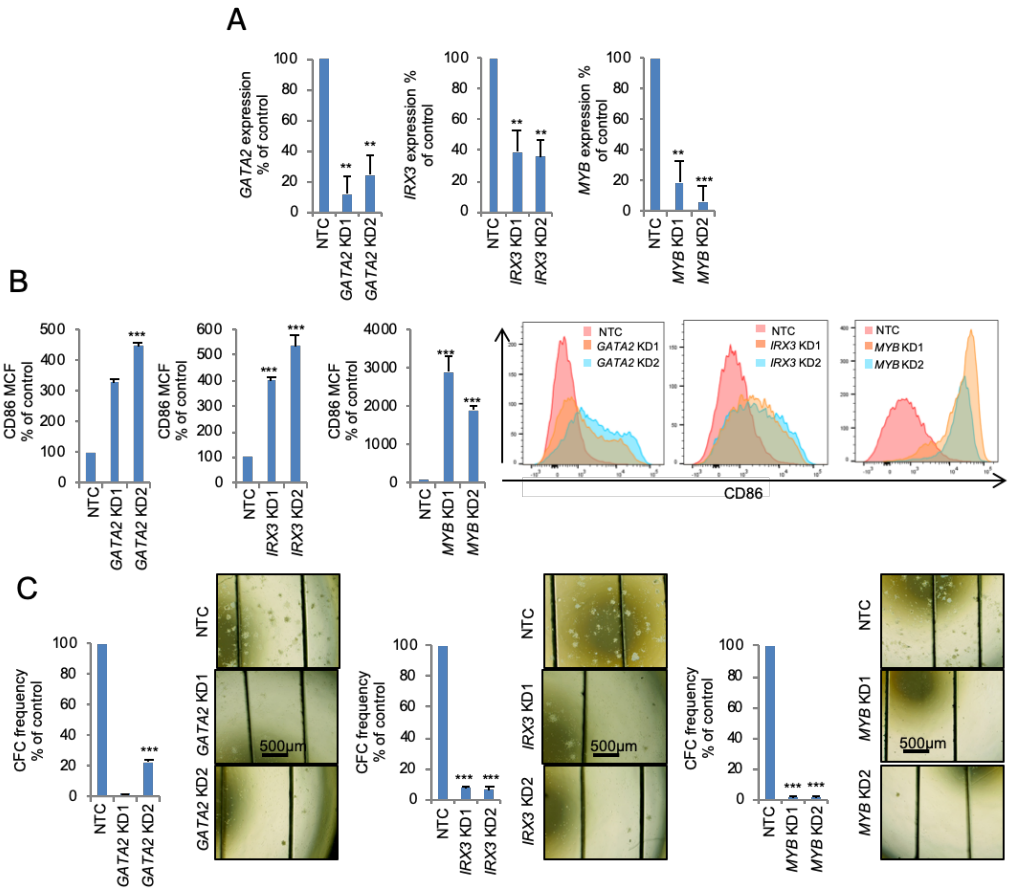


Figure 46. MYB, IRX3 and GATA2 knockdown induce differentiation of Fujioka AML cells.
(A-C) Human Fujioka AML cells were infected with a lentivirus targeting *MYB, IRX3 and GATA2* for KD or a non- targeting control (NTC). (A) Bar chart shows mean+SEM relative expression of the indicated gens in KD versus control cells (n=3) 72hrs after initiation of KD. (B) Bar chart (left panel) shows mean+SEM CD86 mean cell fluorescence (MCF) as determined by flow cytometry in the indicated conditions on Day 5 (n=3). Representative flow cytometry plots (right panel) are also shown. (C) Bar chart (left panel) shows mean+SEM colony-forming cell (CFC) frequencies in the indicated conditions relative to control cells enumerated after twelve days in semi-solid culture (n=3). Image (right panel) shows representative images of colonies. * indicates $p<0.05$, ** indicates $p<0.01$ and *** indicates $p<0.0001$ for the indicated comparisons by an unpaired t test.

Taken together, these data show that RUNX1 redistribution on promoters of key transcription factors is a hallmark of releasing the differentiation block imposed by FOXC1. Plus, RUNX1-mediated downregulations of these essential proliferation genes is key for induction of differentiation and loss of clonogenic potential in Fujioka and primary AML cells.

4.11 RUNX1/CBFB inhibitor Ro5-3335 induces differentiation in FOXC1 high AML cells

Given the importance of RUNX1 transcription factor in FOXC1^{high} AML cells, I hypothesised that RUNX1 inhibition may have utility in this type of leukaemia. RUNX1

forms a heterodimer with CBFB for DNA binding and previous studies suggest that physical interaction between the two proteins is essential for RUNX1 transcriptional activity (Cunningham et al., 2012).

I, therefore, screened a RUNX1/CBFB inhibitor, Ro5-335, in FOXC1^{high} leukaemia versus FOXC1^{low} leukaemia AML cells. I first treated Fujioka AML (FOXC1^{high}) cells with increasing doses of the compound, and found that it is able to strongly induce CD86, used as a marker of differentiation (Figure 47A). I did not observe any effect in inducing differentiation in K562 cells (FOXC1^{low}) even when really high doses of Ro5-335 were used (20 μM, Figure 47B). Moreover, I observed significant growth inhibition of Fujioka cells when treated with Ro5-3335 while no effect was observed in K562 cells (Figure 47C). Primary AML blast cells (FOXC1 high) were also treated with the RUNX1 inhibitor and differentiation was observed in all cases, as reported in Figure 47D.

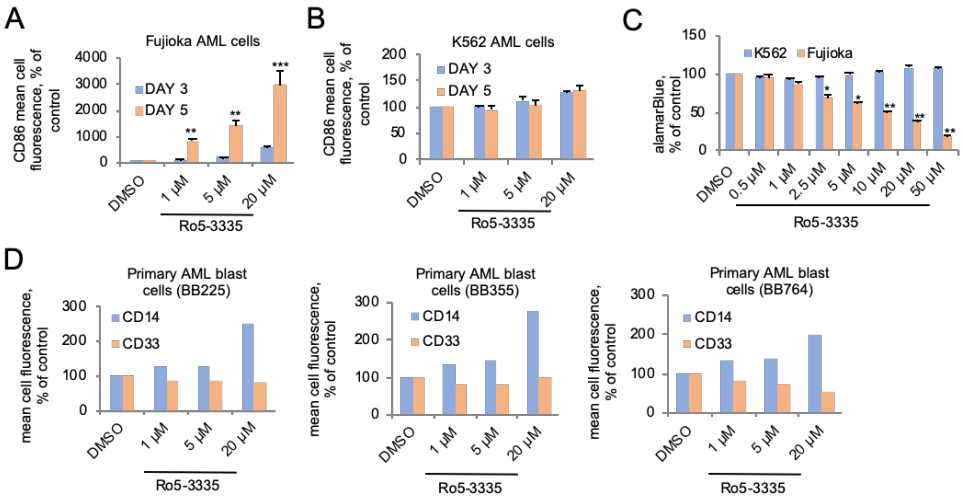


Figure 47. RUNX1/CBFB inhibitor Ro5-3335 induces differentiation in FOXC1^{high} AML cells.

(A-C) Fujioka and K562 AML cells were treated with Ro5-335 or DMSO with the indicated doses. (A-B) Bar charts indicate means+SEM for CD86 mean cell fluorescence, as determined by flow cytometry, in the indicated conditions (n=3). (C) Cell growth of Fujioka and K562 AML cells was assessed using alamarBlue assay. Data represents mean ± SEM in the indicated conditions. (D) Primary AML blast cells (BB225, BB355 and BB764) were treated with Ro5-335 or DMSO with the indicated doses. Bar charts indicate mean cell fluorescence for CD33 and CD14, as determined by flow cytometry, in the indicated conditions (n=1). * indicates $p<0.05$, ** indicates $p<0.01$ and *** indicates $p<0.0001$ for the indicated comparisons by an unpaired t test.

4.12 Enhanced recruitment of Groucho repressor TLE3 to RF-20 enhancer sites

Taken together all the data presented above, it is clear that RUNX1 is acting as a repressor in my experiments. RUNX1 is known to interact with three repressor groups: mSIN3A, HDACs and Groucho (Koh et al., 2013). Among the set of high confidence FOXC1 interacting proteins, I identified the Groucho co-repressor family member TLE3 (Table 19). The other two are instead missing from the list of FOXC1 pulldowns whereas TLE3 is present in all three FOXC1 RIME experiments as a strongly interacting FOXC1

protein. TLE proteins bind via their WD40 beta propeller domain to a range of transcription factors via either a C-terminal WRPW/Y motif or an internal Engrailed homology motif (FxIxxIL) to confer transcription repression through mechanisms that remain incompletely understood (Jennings and Ish-Horowicz, 2008). Drosophila Runt and Groucho interact genetically and the interaction can be mapped to a C-terminal VWRPY sequence present in all RUNX proteins (Aronson et al., 1997). I noted that the Eukaryotic Linear Motif (ELM) resource (Kumar et al., 2020) predicted a putative Engrailed homology motif in the inhibitory domain of FOXC1 (GFSVDNIMT; amino acids 307-315) (Figure 48A). Interestingly, an earlier study also suggested it might interact with residues in the FHD of FOXG1 (aa151-160 in FOXC1 by homology - YWTLDPSYN) (Marcal et al., 2005). Experiments performed in Fujioka AML cells show confirmed physical interactions of TLE3 with endogenous FOXC1 or RUNX1 (Figures 48B-C), suggesting that the interaction between the proteins may underlie the mechanisms behind FOXC1-induced differentiation block.

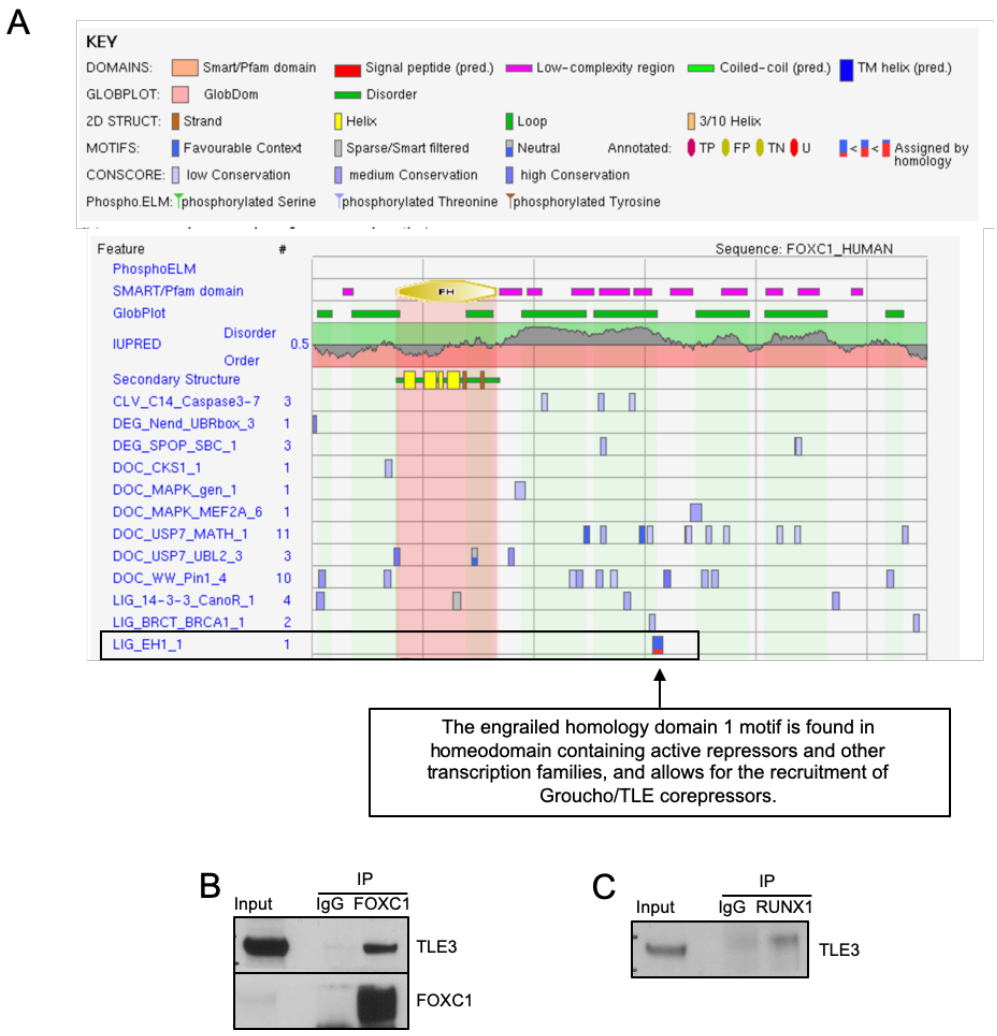


Figure 48. TLE3 interacts with FOXC1 and RUNX1.
(A) Eukaryotic Linear Motif (ELM) resource output (Kumar et al., 2020). (B) Anti-FOXC1 immunoprecipitation in Fujioka AML cells (representative of n=3 experiments). IP, immunoprecipitation. (C) Anti-RUNX1 immunoprecipitation in Fujioka AML cells (representative of n=3 experiments). IP, immunoprecipitation.

To confirm that TLE3 contributed to the differentiation block exhibited by Fujioka cells, knockdown (KD) experiments were firstly performed in human Fujioka cells. Mean transcript levels of *TLE3*, as determined by q-PCR, were reduced to 20% of control upon shRNA KD (Figure 49A). Western blotting analysis confirm KD efficiency (Figure 49B). Following the initiation of knockdown there was an increase of the myeloid differentiation marker CD86 (Figure 49C) and loss of clonogenic potential (Figure 49D). Moreover, *FOXC1* KD led to morphological differentiation (Figure 49E); I observed similar findings in primary patient AML cells (Figures 50A-C). Taken together, these findings provide strong evidence that TLE3 expression is functionally important to maintaining the hallmark properties of AML Fujioka and primary cells, proliferation and differentiation block.

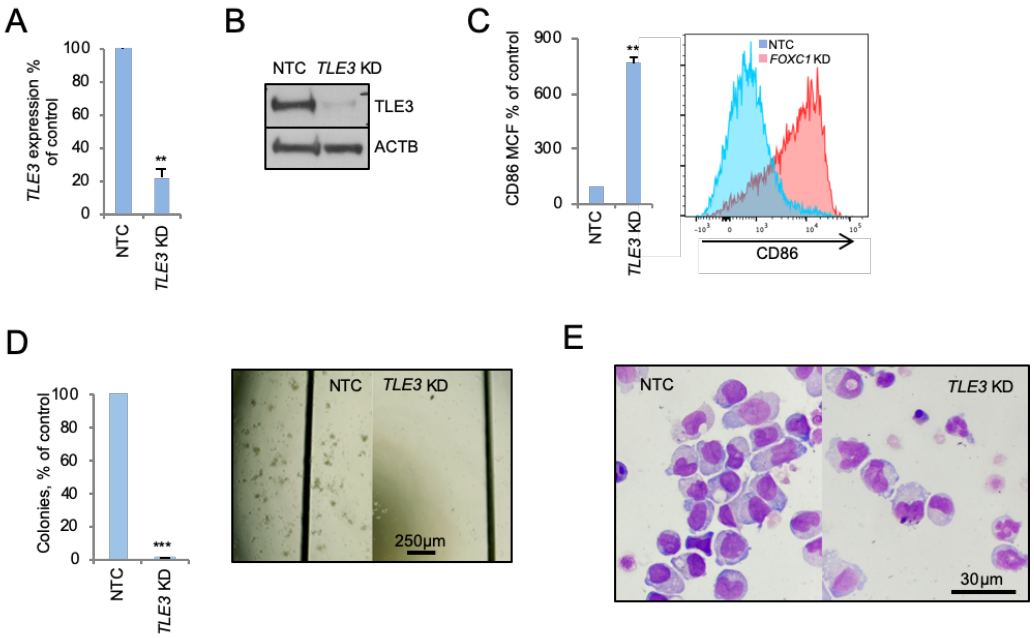


Figure 49. TLE3 sustains the differentiation block of Fujioka cells.
(A-D) Human Fujioka cells were infected with a lentivirus targeting *TLE3* for KD or a NTC with puromycin drug resistance as selectable marker (n=3). (A) Bar chart shows mean+SEM relative transcript expression in KD versus control cells (n=3). (B) Western blot shows *TLE3* KD in Fujioka cells 72hrs following initiation of KD. (C) Bar chart (left panel) shows mean+SEM percentage of cells positive for CD86 surface marker as determined by flow cytometry analysis in the indicated conditions (n=3). Representative flow cytometry plots (right panel) are also shown. (D) Bar chart (left panel) shows mean+SEM colony-forming cell (CFC) frequencies in the indicated conditions relative to control cells enumerated after twelve days in semi-solid culture (n=3). Image (right panel) shows representative images of colonies. (E) Representative images of cytospins of cells on Day 7. * indicates $p<0.05$, ** indicates $p<0.01$ and *** indicates $p<0.0001$ for the indicated comparisons by an unpaired t test.

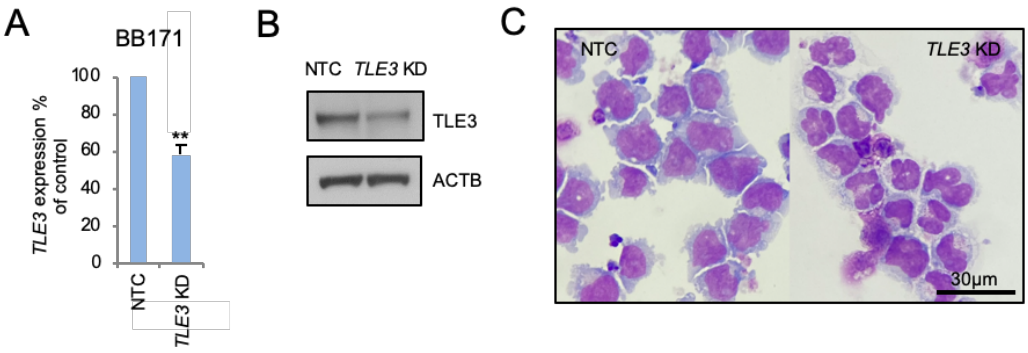


Figure 50. TLE3 sustains the differentiation block of primary AML cells.
(A-C) Human primary BB171 cells were infected with a lentivirus targeting *TLE3* for KD or a NTC with puromycin drug resistance as selectable marker (n=3). (A) Bar chart shows mean+SEM relative transcript expression in KD versus control cells (n=3). (B) Western blot shows *TLE3* KD in Fujioka cells 72hrs following initiation of KD. (E) Representative images of cytospins of cells on Day 7. * indicates $p<0.05$, ** indicates $p<0.01$ and *** indicates $p<0.0001$ for the indicated comparisons by an unpaired t test.

I, then, performed ChIP sequencing for TLE3 and observed a strong positive correlation genome-wide of ChIP signal for TLE3 and RUNX1 (Figures 51A-B). While there was no evidence in genome-wide analysis that FOXC1 alone was able to recruit TLE3 to chromatin (Figure 51C), there was nevertheless significantly more TLE3 ChIP signal at RF- 20 enhancer sites by comparison with R-20 enhancer sites (mean±SEM 2110±123 versus 1653±26 reads/600 base pairs; t-test $p=10^{-4}$) (Figure 52A). This suggests that FOXC1 enhances or stabilises the recruitment of the Groucho repressor TLE3 to chromatin by RUNX1. Following *FOXC1* KD there was a significant loss of TLE3 ChIP signal from Group 1 and Group 2 FR-20 enhancers (Figure 52B); indeed the fold changes in RUNX1 and TLE3 ChIP signal at FR-20 enhancer sites following *FOXC1* KD were highly correlated (Figure 52C). In the same way in which RUNX1 ChIP signal was redistributed from enhancers to promoters following *FOXC1* KD, so too was TLE3 ChIP signal (Figures 53A-B), with a similar pattern of motif enrichment (Figure 53C). Thus, FOXC1 stabilises association of RUNX1, HDAC1 and the Groucho family repressor protein TLE3 at a discrete set of enhancers to limit locoregional transcription. Depletion of FOXC1 triggers a redistribution of RUNX1, TLE3 and HDAC1 to promoters of critical growth and renewal genes, including for example *MYB* (Figure 53B) and *IRS2* (Figure 53D).

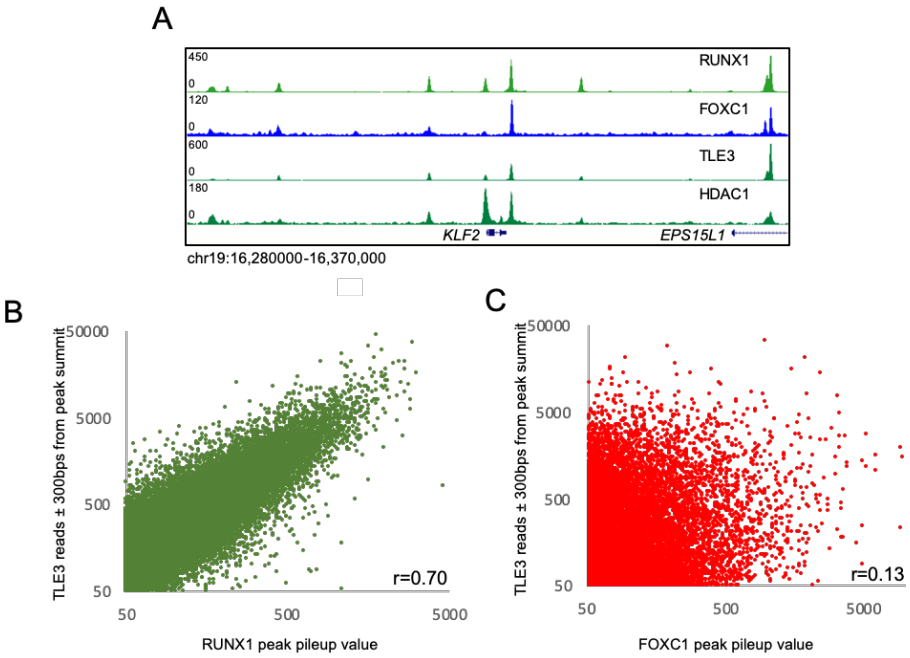


Figure 51. TLE3 ChIPseq in Fujioka AML cells.
(A) Exemplar ChIPseq tracks. (B) Dot plots show TLE3 ChIP signal at RUNX1 binding peaks in control Fujioka AML cells. (C) Dot plots show TLE3 ChIP signal at FOXC1 binding peaks in control Fujioka AML cells.

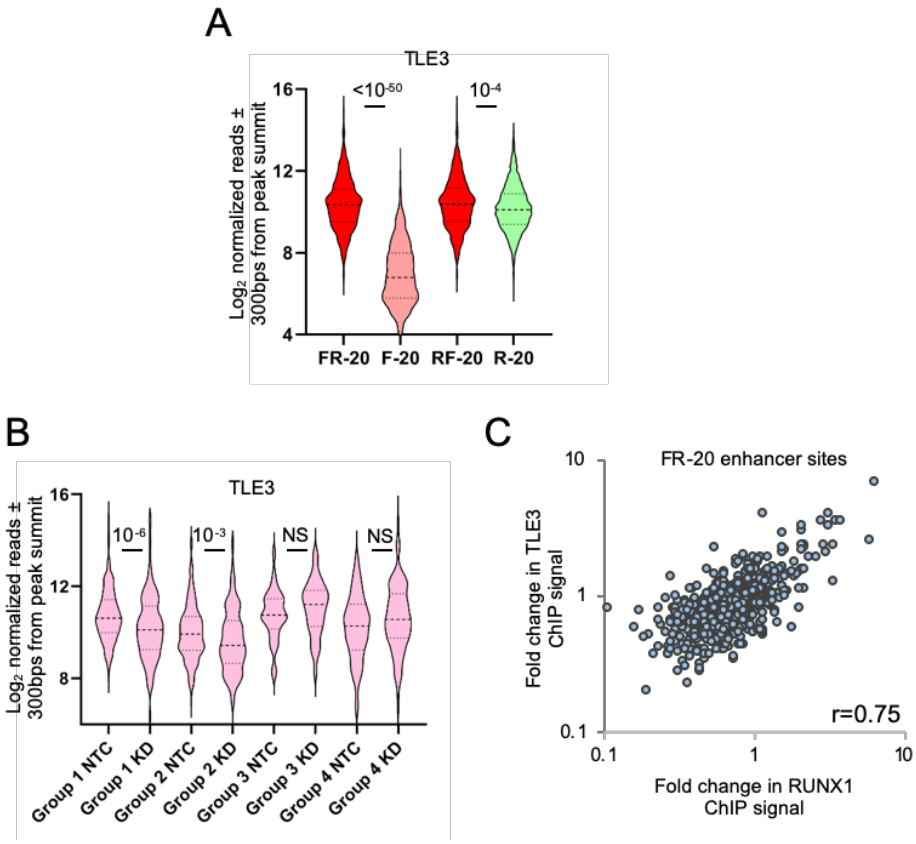


Figure 52. FOXC1 stabilizes TLE3 & RUNX1 binding at enhancers controlling differentiation.

(A) Violin plot show distribution, median (thick dotted line) and interquartile range (light dotted lines) for TLE3 ChIP signal at the indicated sites in control (NTC) Fujioka AML cells. P value, unpaired t-test. (B) Violin plot shows TLE3 ChIP signal at the indicated FR-20 enhancer sites in control and FOXC1 KD Fujioka cells. NS, not significant. P values, unpaired t-test. (C) Dot plot shows fold change in relative TLE3 and RUNX1 ChIP signal at each of 581 FR-20 enhancer sites.

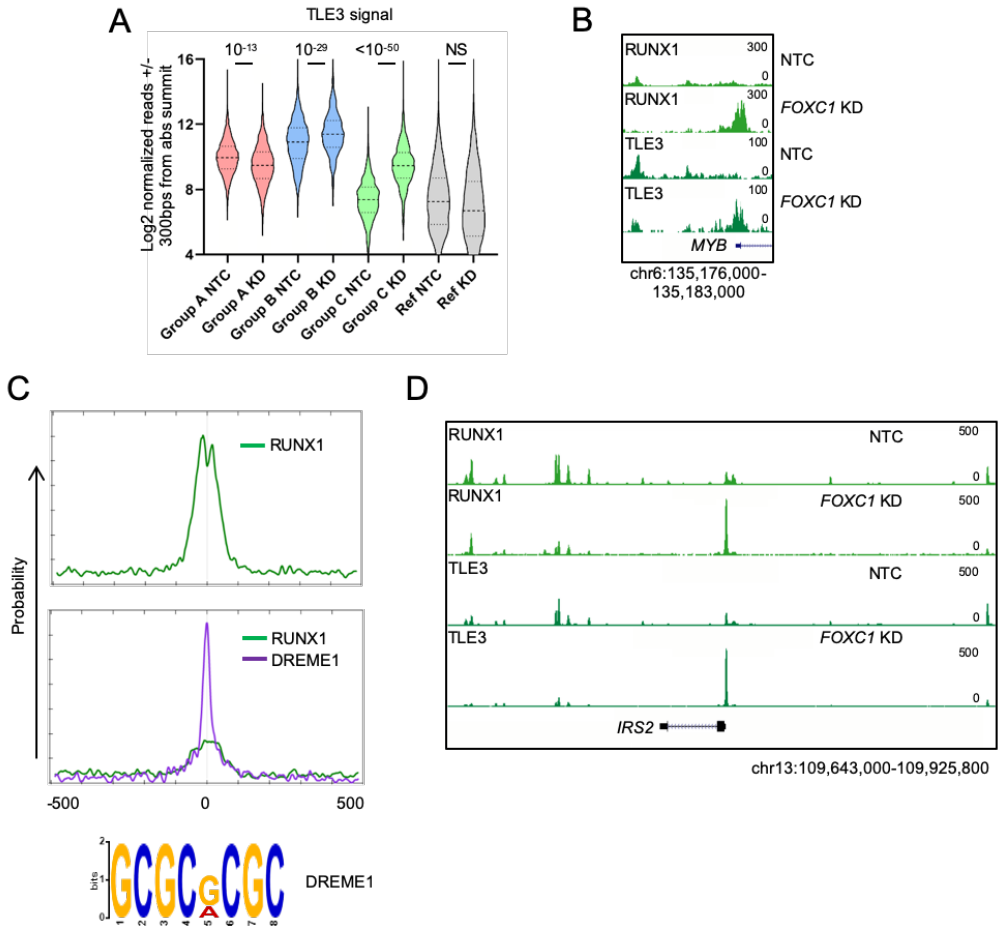


Figure 53. *FOXC1* knockdown triggers redistribution of TLE3 binding.
(A) Violin plot shows TLE3 ChIP signal at the indicated RUNX1 group binding sites in control and *FOXC1* KD Fujioka cells. Ref, reference cohort used for normalization between experiments; NS, not significant. P values, unpaired t-test. (B) Exemplar ChIPseq tracks. (C) MEME-ChIP motif enrichment plots at the strongest 20% of TLE3 binding peaks in control (upper panel) and *FOXC1* KD Fujioka AML cells (lower panel). The DREME1 motif is also shown. (D) Exemplar ChIP-seq tracks.

4.13 Summary

To conclude, Somerville et al. 2015 previously showed that *FOXC1* is over-expressed in a subset of HOXA/B+ AML leukaemias and blocks monocytic differentiation. I then globally identified FOXC1-interacting proteins in AML cells and identified 12 interacting transcription factors, including RUNX1 and C/EBPA which were confirmed by co-IP. Interaction with RUNX1 and CEBPA mapped to the FOXC1 Forkhead DNA-binding domain, with mutations G165R and F112S in the non-DNA contact face preventing interaction. shRNA-mediated KD of *FOXC1*, *RUNX1*, *CEBPA* each induced monocytic differentiation of Fujioka (FOXC1^{high}) AML cells.

I then performed FOXC1, RUNX1, CEBPA, PU.1, H3K27Ac, and H3K4me1 ChIP-Seq on Fujioka cells, with PU.1 predominating on K27Ac^{low}K4me1^{low} inactive sites and the other TFs distributed over the other three categories. In addition, 28% of FOXC1 peaks

corresponded to RUNX1 peaks, and 26.5% of FOXC1 peaks corresponded to CEBPA peaks. FOXC1/RUNX1 (FR) peaks predominated at primed or active enhancers, with stronger FOXC1 and RUNX1 binding at these sites compared to FOXC1 only or RUNX1 only sites. *FOXC1* KD led to reduced RUNX1 and increased CEBPA binding to FR enhancers.

I also performed ChIP-Seq for EP300, SMARCC2, HDAC1, and K3K4Me2, as well as ATAC-Seq. *FOXC1* KD led to significant reduction of HDAC1 at RF and RUNX1-only binding enhancers. RNA-seq identified 349 upregulated and 804 down-regulated genes after *FOXC1* KD. Upregulated genes, including those associated with monocytic differentiation (e.g. *KLF2*), correlated with reduced RUNX1, reduced HDAC1, increased CEBPA binding and increased H3K27Ac upon *FOXC1* KD. Exogenous *KLF2* expression induced monocytic differentiation.

Exogenous FOXC1 DBD-RUNX1 fusion protein prevented differentiation induction by *FOXC1* KD, while FOXC1(G165R) or (112S) which do not bind RUNX1 favored differentiation.

RUNX1 relocated from enhancers to promoters upon *FOXC1* KD along with HDAC1, with loss of CEBPA and H3K27Ac from promoters. Enhancer regions but not promoter-binding sites had underlying RUNX1 motifs (promoters had SP1/KLF2 motif). In three cases tested, such promoters lost SP1 and gained KLF2. And promoters that gained RUNX1 correlated with decreased RNA expression, including *MYB*.

The global protein interaction assay identified TLE3 as binding FOXC1, which was confirmed by co-IP. TLE3 is also able to bind to RUNX1 protein and its ablation in Fujioka AML cells induced differentiation. TLE3 ChIP-Seq showed enrichment at RF vs RUNX1-only binding enhancers, with loss of TLE3 at RF enhancers upon *FOXC1* KD, and increase TLE3 at promoters.

Chapter 5: Discussion

Despite the genetic heterogeneity, the differentiation block is the cardinal feature of AML and emerging evidence reveals that mutations cluster in particular categories of genes. Indeed, the great majority of patients with AML have one or more mutations targeting a transcription factor, chromatin modifier or regulator of DNA methylation and this emphasises the absolute centrality of epigenetic and transcription factor dysfunction to the disease (Chen et al., 2013a). Derepression of *FOXC1* is a novel oncogenic phenomenon in human AML, with approximately 20% of cases exhibiting high levels 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 (Goardon et al., 2011; Kikushige et al., 2010; Saito et al., 2010). 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*. Expression of *FOXC1*, which is not present in normal hematopoietic lineages, contributes to blocked myeloid differentiation in molecular subtypes of AML with concomitant high *HOX* gene expression. In particular, 95% of patients with high *FOXC1* expression also exhibit high *HOXA9* expression (Somerville et al., 2015). While the role of HOXA9 in AML has been intensively studied and many authors have identified and characterised its mechanism of action (Collins and Hess, 2016), derepression of *FOXC1* in human AML is a relatively unexplored mechanism and, given the inferior prognosis for patients with FOXC1 positive AML, understanding the molecular consequences of its aberrant expression may be a fundamental requirement for development of targeted therapies for *FOXC1*^{high} AML patients.

Although the mechanisms by which *FOXC1* is de-repressed in a lineage-inappropriate manner remain to be determined, it should be emphasised that, in addition to its defined genetic lesions, AML is also marked by widespread epigenetic changes, for example in DNA methylation (Cancer Genome Atlas Research et al., 2013). Genetic lesions may confer epigenetic plasticity and thus be permissive for outgrowth of clones driven or sustained by lineage-inappropriate transcriptional networks. Such networks would be attractive potential targets for therapy, given their lack of expression or importance in normal hematopoietic cells.

In the present study, I demonstrate that the derepression of the Forkhead transcription factor FOXC1 is functionally relevant to the differentiation block in primary AML. Here, I show that *FOXC1* KD significantly decreases the viability and growth of human AML cells, inducing differentiation, apoptosis, and reducing their clonogenic potential. This has important consequences in clinical settings as FOXC1 is not required nor expressed in any haematopoietic cells, making it a good therapeutic target. Interestingly, an indirect role for FOXC1 in blood development was recently reported. *FOXC1* is highly expressed

in bone marrow mesenchymal cells and selective downregulation of *Foxc1* leads to a depletion of the bone marrow niche and the number of HSCs (Omatsu et al., 2014).

To elucidate the molecular mechanisms of FOXC1 in leukemogenesis, I performed integrated ChIPseq analysis of FOXC1 in patient and Fujioka AML cells together with the FOXC1 proteome. I revealed FOXC1 binding peaks to be mainly distributed over intronic and intergenic regions, consistent with putative roles at enhancers. FOXC1 shares the FOX domain with other family members but the specificity and differential binding to target genes are thought to be influenced by the flanking regions of the domain and, importantly, by the complex interactions with other transcription factors at the chromatin level (Chen et al., 2016). Further to this, RIME-mass spectrometry analysis revealed a strong physical interaction with another two transcription factors with a well-established role in the disease: RUNX1 and CEBPA. Interestingly, both findings were highly conserved in Fujioka and primary AML cells, highlighting how the former can be considered a good model to investigate the molecular mechanisms mediated by *FOXC1* derepression.

The ability of FOXC1 to bind to and interfere with the activity of these two factors suggests an unexpected, additional mechanism in myeloid leukemogenesis. RUNX1 is a master regulator of developmental hematopoiesis, controlling the emergence of hematopoietic stem cells from hemogenic endothelium (Lancrin et al., 2009); in adulthood it is required for proper megakaryocyte and lymphoid development, and suppression of a myeloproliferative phenotype (Growney et al., 2005). Genetic lesions of *RUNX1* in AML, whether by somatic mutation or chromosomal translocation, are frequent although the mechanisms by which they promote leukemic transformation are incompletely understood. RUNX1- RUNX1T1 (AML1-ETO) recruits a multitude of co-repressors to its binding sites whereas somatic mutations of *RUNX1*, which often target sequences coding for the Runt Homology Domain, are inactivating or confer dominant negative activity (Sood et al., 2017). Functionally, RUNX1 may be sequestered away from chromatin by CBFB-SMMHC, or have its activity modified by interaction with CBFB-SMMHC which also recruits co-repressors to sites of RUNX1 binding (Beghini, 2019). More recent studies also suggest that wild-type RUNX1 is required for growth and survival of certain types of leukemia cells. Biallelic mutations in *CEBPA* which block CEBP factor homo- or heterodimerization, or DNA binding, are also frequent in AML (Wilhelmson and Porse, 2020).

I have discovered that the Forkhead domain of FOXC1 interacts with RUNX1 and together these factors co-occupy hundreds of primed and active enhancers, including many which are distributed close to genes upregulated in monocyte/macrophage differentiation. Sites of strong co-localized FOXC1 and RUNX1 binding exhibit higher levels of RUNX1, TLE3 and HDAC1 binding by comparison with strong RUNX1 binding

sites that do not have co-localised FOXC1. I demonstrate that FOXC1 acts mostly by repressing this set of myeloid-critical enhancers (e.g. *KLF2*) by stabilizing RUNX1, HDAC1 and TLE3 at these sites. My investigations indicate that, in Fujioka AML cells, the strongest FOXC1 binding peaks are co-localized with the strongest RUNX1 binding peaks on chromatin and that, following *FOXC1* KD, RUNX1 is displaced from these enhancer sites. Loss of RUNX1 leads to decreased binding of HDAC1 and TLE3 and increased acetylation and gene expression, activating a myeloid gene set program that pushes cells to differentiate. This is coupled with increase binding of CEBPA at these sites, whose role in myeloid differentiation is well defined (Tawana et al., 2017). Consistent with its role as an oncogene, FOXC1 represses important regulators of myeloid differentiation (such as *JUN*, *KLF2*, *KLF6*, *MEF2C*, *FOS*, *CD86*). All of these seem to be modulated by FOXC1 in a direct way since integration of chromatin occupancy data and gene expression profiling confirmed these findings. However, loss of FOXC1 in human AML cells revealed a much broader change towards monocyte differentiation (GSEA data) where master regulators of monocyte differentiation pathway (i.e. JUN, ERG1, MAFB, IRF8, MYC, GATA2, GFI1, CEBPA) are all modulated. In particular, my studies show that restoration of *KLF2* or *KLF6* levels in Fujioka AML cells is sufficient to induce myeloid differentiation and loss of clonogenic potential.

Therefore, in my studies, FOXC1 and RUNX1 serve as transcription repressors: in the genome-wide redistribution of RUNX1 binding which follows *FOXC1* KD, loss of RUNX1 from enhancers associates with increased expression of nearby genes, whereas the opposite is the case for both enhancers and promoters which gain RUNX1. This is further emphasised by my observation of an extremely strong genome-wide correlation in both control and *FOXC1* KD cells of binding peaks for RUNX1, HDAC1 and the Groucho repressor TLE3. As RUNX1 is redistributed with differentiation, so too is HDAC1 and TLE3 (Figure 54).

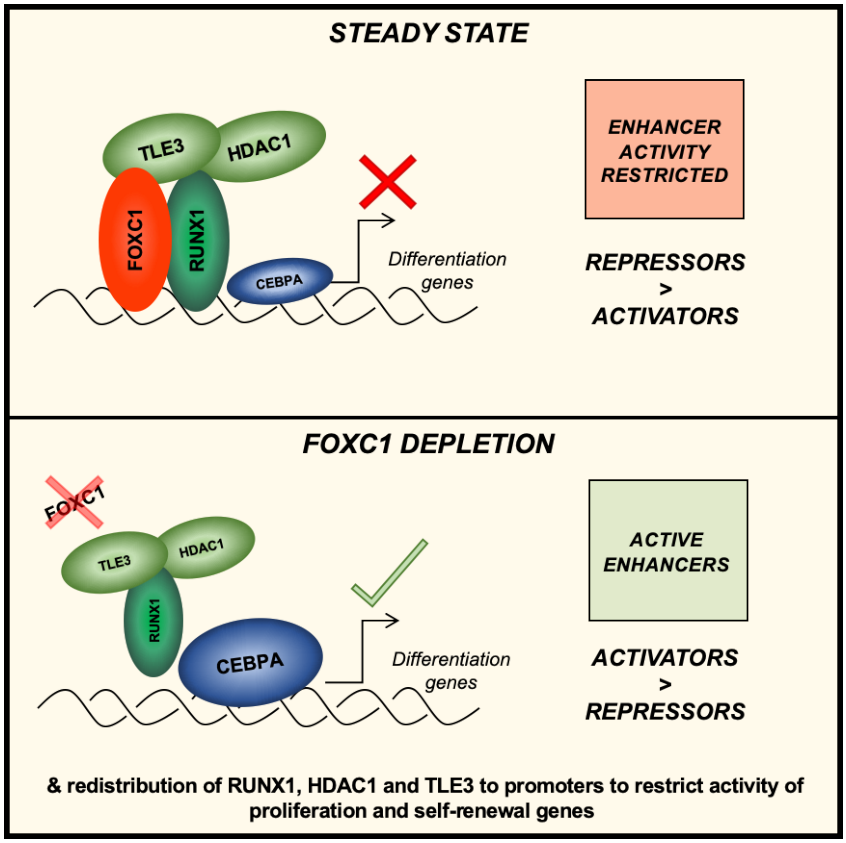


Figure 54. Graphical abstract
Schematic representation of the main finding of my PhD project.

Recruitment of TLE3 on the new promoter sites is mediated by RUNX1 but its mechanisms of repression are still not fully understood. The Groucho/TLE family of corepressors interacts with at least five families of transcription factor and plays critical roles in development. The ability to interact with many different transcription factors suggests the function of the Gro/TLEs will probably vary, depending on the cell type and the timing of expression during development, offering new therapeutic insights for cancer treatment (Jennings and Ish-Horowicz, 2008).

The mechanisms by which Groucho family proteins confer transcription repression are poorly understood, but may include reduction of chromatin accessibility and recruitment of deacetylase activity (Jennings and Ish-Horowicz, 2008). TLE3 has not previously been reported to have a significant role in AML, although other Groucho homologs including TLE1 and TLE4, whose genes are located within the commonly deleted region of del(9q), have interestingly been shown in KD studies to restrain Kasumi-1 AML cell growth (Dayyani et al., 2008).

My studies also highlight the challenges associated with determining which among many thousands of genome-wide transcription factor binding sites are functionally the most important, and the complexities of enhancer biology; it cannot be presumed that each binding site has equal biological significance. It is notable that those FOXC1 sites controlling expression of differentiation genes following *FOXC1* KD (Group 1 enhancers) account for fewer than 1.5% of the total. These sites were generally marked by primed

or active histone modifications, accessible chromatin, strong RUNX1 and FOXC1 binding, and intermediate levels of CEBPA binding. These findings shed a light on to the complexity of enhancer biology and how chromatin state, levels of binding of TFs, chromatin accessibility and other factors (e.g. interactions with other proteins) are essential contributors of a transcription factor binding site and how the presence of a TF on a particular genomic location does not always mediate a biological effect.

In fact, the consequences of FOXC1 depletion at any one RUNX1-bound enhancer were nevertheless variable, and likely dependent upon the presence or absence of many dozens of additional co-located factors. Consistent with FOXC1 having pioneer activity, I also observed widespread and strong binding at sites of silent chromatin, but found no evidence that cellular depletion of FOXC1 at these sites contributed acutely to cellular differentiation. Presumably, the inappropriate occupation of this discrete subset of primed and active enhancers by FOXC1 inhibits their normal activity by preventing, through RUNX1/TLE3/HDAC1 recruitment, the normal upregulation of critical genes required for differentiation.

RUNX1 and CEBPA have been shown to interact through the C-terminal basic leucine zipper domain of CEBPA and the Runt domain of RUNX1 (Zhang et al., 1996), and at many FOXC1/RUNX1 co-occupied enhancers in my study there was significant co-localized CEBPA binding. Nevertheless, overall I noted that loss of RUNX1 from genomic binding sites generally correlated with gain of CEBPA. Following *FOXC1* KD CEBPA was recruited to enhancers near to upregulated myeloid differentiation genes as RUNX1 was lost, and vice versa at enhancers and promoters close to genes downregulated with differentiation. I speculate that obstruction of RUNX1 and CEBPA transcription factor switching at enhancers and promoters may make a significant contribution to blockade of myeloid differentiation, a concept supported by the finding that targeting RUNX1 to sites of FOXC1 binding using a FOXC1 FKD-RUNX1c fusion blocked upregulation of differentiation genes.

FOXC1-RUNX1 physical interaction is a novel oncogenic mechanism, not explored to date, and therefore this work could serve as the basis for further investigations in the solid tumour area where FOXC1 is known to be a critical contributor to tumorigenesis. For example, *FOXC1* overexpression in MCF-7 breast cancer cells significantly enhances their proliferation and contributes to the stemness of the cancer cells (Gilding and Somervaille, 2019). Whether in solid tumours this interaction is still functional and still contributes to the disease is completely unknown.

The differentiation-associated redistribution of RUNX1 binding, which has also been observed following knockdown of *RUNX1-RUNX1T1* in Kasumi-1 cells (Ptasinska et al., 2012), is to short ungapped GC-rich DNA motifs rather than RUNX1 motifs. Currently, I do not know how the RUNX1/TLE3/HDAC1 repressor complex is directed to new sites, but it is likely that this involves the interaction of RUNX1 with other transcription factors

whose activity is altered by *FOXC1* KD, such as KLF2 or KLF4 as suggested by our ChIP-qPCR data. Previous studies demonstrated that FOXC1 represses expression of the Krüppel-like factor family, *KLF4*, and showed that restoration of *KLF4* expression in Hoxa9/FOXC1 AML cells inhibits proliferation and clonogenic potential (Somerville et al., 2015). Therefore, FOXC1 could generally have a role in repressing TFs belonging to the Krüppel-like factor family and so, blocking differentiation. It remains unclear whether additional mechanisms are involved in this switch, including post-translational modifications of RUNX1, for example.

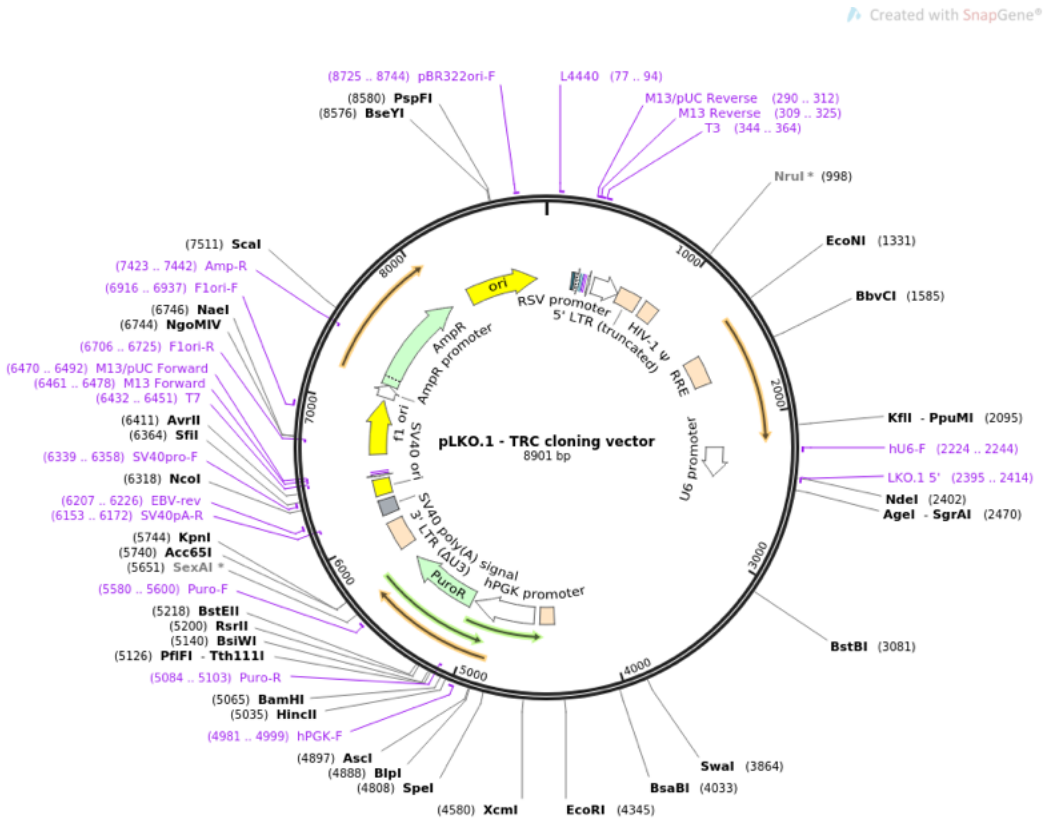
The induction of myeloid differentiation upon FOXC1 depletion therefore involves not only activation of a subset of myeloid enhancers but also increased recruitment of transcriptional repressors to additional sites (e.g. *MYB*, *MYC* promoters).

Finally, my work suggests novel targets for therapeutic intervention, and additionally provides insights that may be of value in furthering understanding of FOXC1^{high} solid malignancies, and the Axenfeld-Rieger Syndrome. Compounds that target the interaction of the Forkhead domain of FOXC1 with RUNX1 would be predicted to destabilise enhancer-bound RUNX1/TLE3/HDAC1 to promote differentiation and may be beneficial alone, or more likely in combination, in FOXC1^{high} HOX^{high} AMLs. Further, my studies of *TLE3* knockdown in primary patient AML cells suggest that compounds that target the interaction of the C-terminal VWRPY domain of RUNX1 with the WD40 beta propeller domain of TLE3, or the domains by which TLE3 undergoes oligomerization, may also be beneficial. In conclusion, I demonstrate that FOXC1 cooperates with RUNX1, HDAC1 and TLE3 to block myeloid differentiation in AML cells. Further understanding of the mechanism of these interactions and of the downstream pathways affected could have therapeutic implications.

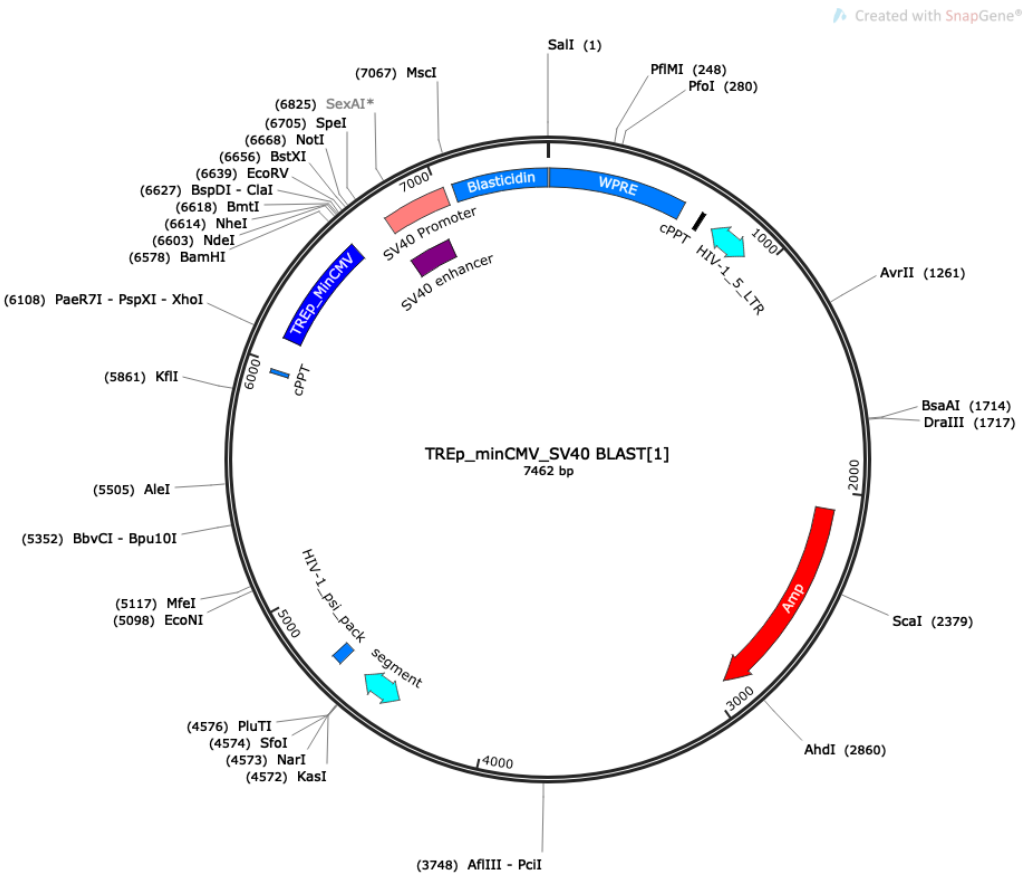
Chapter 6: Supplemental

6.1 Plasmid Maps

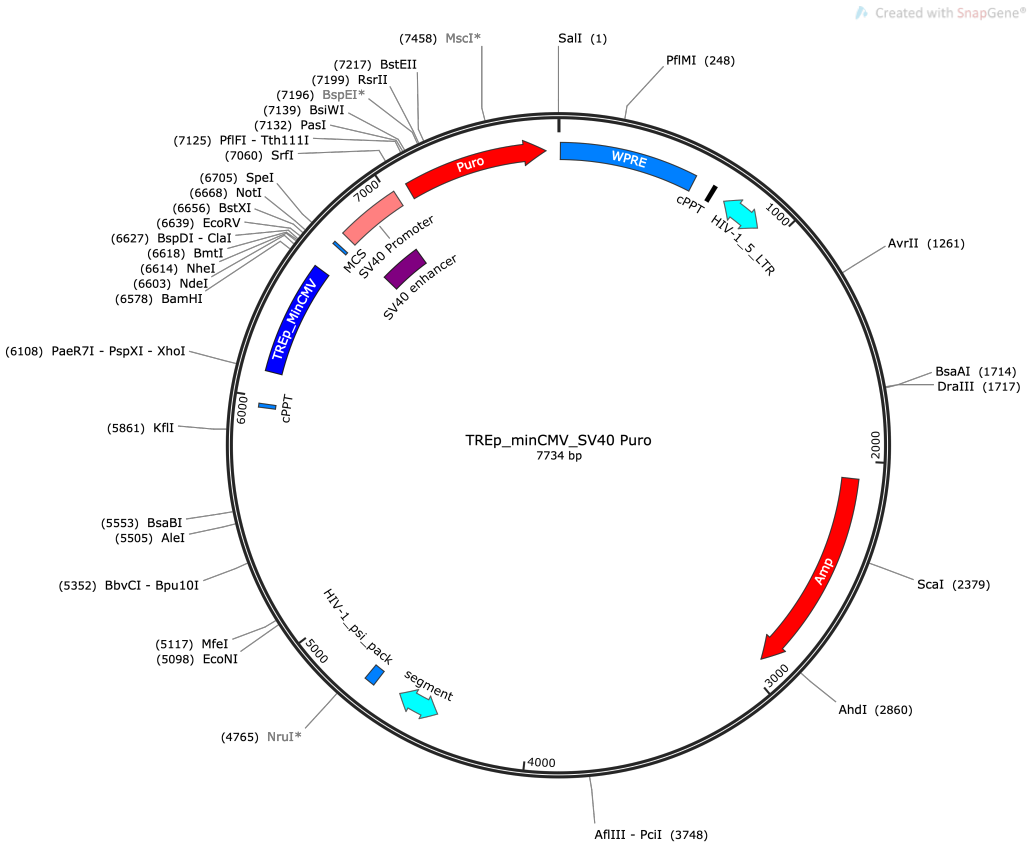
pLKO.1



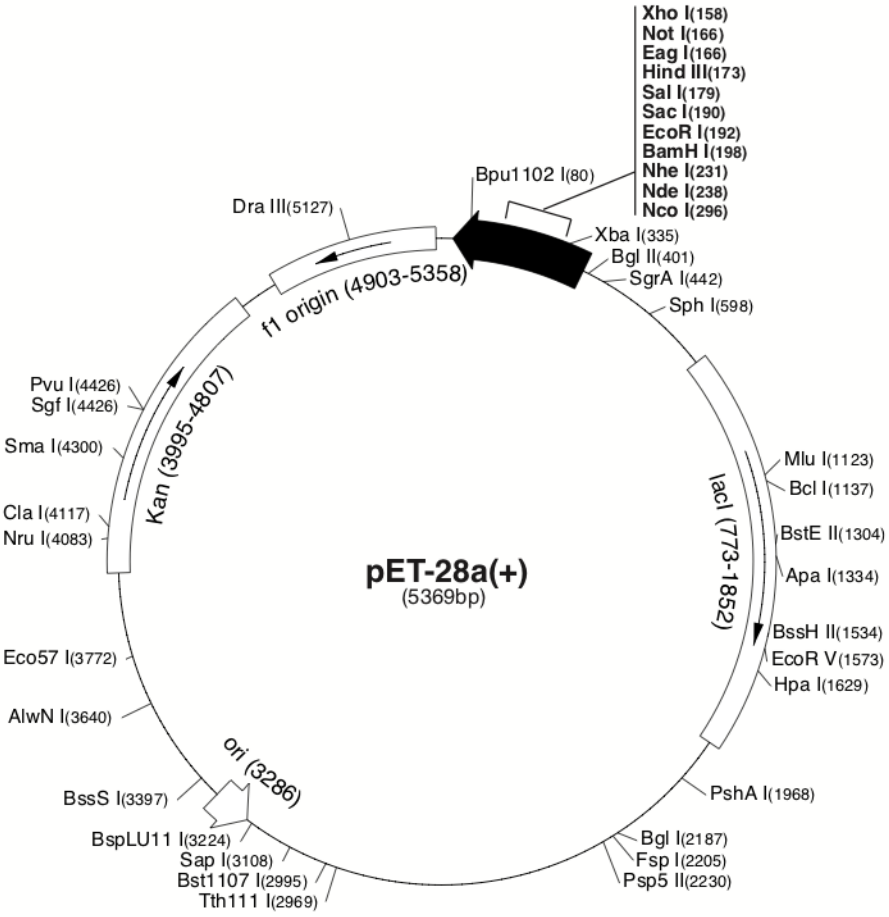
pLentiGS - blasticidn



pLentiGS – puromycin



Pet28


Xho I(158)
Not I(166)
Eag I(166)
Hind III(173)
Sal I(179)
Sac I(190)
EcoR I(192)
BamH I(198)
Nhe I(231)
Nde I(238)
Nco I(296)
Bpu1102 I(80)
Dra III(5127)
Xba I(335)
Bgl II(401)
SgrA I(442)
Sph I(598)
f1 origin (4903-5358)
Pvu I(4426)
Sgf I(4426)
Sma I(4300)
Kan (3995-4807)
Cla I(4117)
Nru I(4083)
pET-28a(+)
(5369bp)
Mlu I(1123)
Bcl I(1137)
lacI (773-1852)
BstE II(1304)
Apa I(1334)
BssH II(1534)
EcoR V(1573)
Hpa I(1629)
Eco57 I(3772)
AlwN I(3640)
ori (3286)
BssS I(3397)
BspLU11 I(3224)
Sap I(3108)
Bst1107 I(2995)
Tth111 I(2969)
PshA I(1968)
Bgl I(2187)
Fsp I(2205)
Psp5 II(2230)

References

Adolfsson, J., Borge, O.J., Bryder, D., Theilgaard-Monch, K., Astrand-Grundstrom, I., Sitnicka, E., Sasaki, Y., and Jacobsen, S.E.W. (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.

Alharbi, R.A., Pettengell, R., Pandha, H.S., and Morgan, R. (2013). The role of HOX genes in normal hematopoiesis and acute leukemia. Leukemia *27*, 1000-1008.

Andreoli, V., Gehrau, R.C., and Bocco, J.L. (2010). Biology of Kruppel-like factor 6 transcriptional regulator in cell life and death. IUBMB Life *62*, 896-905.

Andrews, R.G., Singer, J.W., and Bernstein, I.D. (1989). Precursors of Colony-Forming Cells in Humans Can Be Distinguished from Colony-Forming Cells by Expression of the Cd33 and Cd34 Antigens and Light Scatter Properties. J Exp Med *169*, 1721-1731.

Arber, D.A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M.J., Le Beau, M.M., Bloomfield, C.D., Cazzola, M., and Vardiman, J.W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood *127*, 2391-2405.

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

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

Assi, S.A., Imperato, M.R., Coleman, D.J.L., Pickin, A., Potluri, S., Ptasinska, A., Chin, P.S., Blair, H., Cauchy, P., James, S.R.*, et al.* (2019). Subtype-specific regulatory network rewiring in acute myeloid leukemia. Nat Genet *51*, 151-162.

Bagger, F.O., Kinalis, S., and Rapin, N. (2019). BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. Nucleic Acids Res *47*, D881-D885.

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

Becker, A.J., McCulloch, E.A., and Till, J.E. (2014). Pillars article: Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. Nature. 1963. 197:452-454. J Immunol *192*, 4945-4947.

Beghini, A. (2019). Core Binding Factor Leukemia: Chromatin Remodeling Moves Towards Oncogenic Transcription. Cancers (Basel) *11*.

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.

Bereshchenko, O., Mancini, E., Moore, S., Bilbao, D., Mansson, R., Luc, S., Grover, A., Jacobsen, S.E., Bryder, D., and Nerlov, C. (2009). Hematopoietic stem cell expansion precedes the generation of committed myeloid leukemia-initiating cells in C/EBPalpha mutant AML. Cancer Cell *16*, 390-400.

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

Berry, F.B., Saleem, R.A., and Walter, M.A. (2002). FOXC1 transcriptional regulation is mediated by N- and C-terminal activation domains and contains a phosphorylated transcriptional inhibitory domain. J Biol Chem *277*, 10292-10297.

Bhatia, M., Bonnet, D., Kapp, U., Wang, J.C.Y., Murdoch, B., and Dick, J.E. (1997). Quantitative analysis reveals expansion of human hematopoietic repopulating cells after short-term ex vivo culture. J Exp Med *186*, 619-624.

Bickmore, W.A. (2013). The spatial organization of the human genome. Annu Rev Genomics Hum Genet *14*, 67-84.

Bonnet, D., and Dick, J.E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nature Medicine *3*, 730-737.

Bowman, R.L., Busque, L., and Levine, R.L. (2018). Clonal Hematopoiesis and Evolution to Hematopoietic Malignancies. Cell Stem Cell *22*, 157-170.

Bowman, R.L., and Levine, R.L. (2017). TET2 in Normal and Malignant Hematopoiesis. Csh Perspect Med *7*.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.
Burnett, A., Wetzler, M., and Lowenberg, B. (2011). Therapeutic advances in acute myeloid leukemia. J Clin Oncol *29*, 487-494.

Butler, J.M., Nolan, D.J., Vertes, E.L., Varnum-Finney, B., Kobayashi, H., Hooper, A.T., Seandel, M., Shido, K., White, I.A., Kobayashi, M.*, et al.* (2010). Endothelial cells are essential for the self-renewal and repopulation of Notch-dependent hematopoietic stem cells. Cell Stem Cell *6*, 251-264.

Calkhoven, C.F., Muller, C., and Leutz, A. (2000). Translational control of C/EBPalpha and C/EBPbeta isoform expression. Genes Dev *14*, 1920-1932.

Cancer Genome Atlas Research, N., Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet *45*, 1113-1120.

Castilla, L.H., Wijmenga, C., Wang, Q., Stacy, T., Speck, N.A., Eckhaus, M., Marin-Padilla, M., Collins, F.S., Wynshaw-Boris, A., and Liu, P.P. (1996). Failure of embryonic hematopoiesis and lethal hemorrhages in mouse embryos heterozygous for a knocked-in leukemia gene CBFB-MYH11. Cell *87*, 687-696.

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.

Chalamalasetty, R.B., Garriock, R.J., Dunty, W.C., Kennedy, M.W., Jailwala, P., Si, H., and Yamaguchi, T.P. (2014). Mesogenin 1 is a master regulator of paraxial presomitic mesoderm differentiation. Development *141*, 4285-4297.

Challen, G.A., Sun, D.Q., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H.C., Xi, Y.X.*, et al.* (2012). Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet *44*, 23-U43.

Chambers, S.M., Boles, N.C., Lin, K.Y., Tierney, M.P., Bowman, T.V., Bradfute, S.B., Chen, A.J., Merchant, A.A., Sirin, O., Weksberg, D.C.*, et al.* (2007). Hematopoietic fingerprints: an expression database of stem cells and their progeny. Cell Stem Cell *1*, 578-591.

Chen, S.J., Shen, Y., and Chen, Z. (2013a). A panoramic view of acute myeloid leukemia. Nat Genet *45*, 586-587.

Chen, X., Muller, G.A., Quaas, M., Fischer, M., Han, N., Stutchbury, B., Sharrocks, A.D., and Engeland, K. (2013b). The forkhead transcription factor FOXM1 controls cell cycle-dependent gene expression through an atypical chromatin binding mechanism. Mol Cell Biol *33*, 227-236.

Chen, Y., Li, Y., Xue, J., Gong, A., Yu, G., Zhou, A., Lin, K., Zhang, S., Zhang, N., Gottardi, C.J.*, et al.* (2016). Wnt-induced deubiquitination FoxM1 ensures nucleus beta-catenin transactivation. Embo J *35*, 668-684.

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.

Collins, C.T., and Hess, J.L. (2016). Role of HOXA9 in leukemia: dysregulation, cofactors and essential targets. Oncogene *35*, 1090-1098.

Conway O'Brien, E., Prideaux, S., and Chevassut, T. (2014). The epigenetic landscape of acute myeloid leukemia. Adv Hematol *2014*, 103175.

Cunningham, L., Finckbeiner, S., Hyde, R.K., Southall, N., Marugan, J., Yedavalli, V.R., Dehdashti, S.J., Reinhold, W.C., Alemu, L., Zhao, L.*, et al.* (2012). Identification of benzodiazepine Ro5-3335 as an inhibitor of CBF leukemia through quantitative high throughput screen against RUNX1-CBFbeta interaction. Proc Natl Acad Sci U S A *109*, 14592-14597.

Dayyani, F., Wang, J., Yeh, J.R., Ahn, E.Y., Tobey, E., Zhang, D.E., Bernstein, I.D., Peterson, R.T., and Sweetser, D.A. (2008). Loss of TLE1 and TLE4 from the del(9q) commonly deleted region in AML cooperates with AML1-ETO to affect myeloid cell proliferation and survival. Blood *111*, 4338-4347.

de Rooij, J.D., Zwaan, C.M., and van den Heuvel-Eibrink, M. (2015). Pediatric AML: From Biology to Clinical Management. J Clin Med *4*, 127-149.

Decker, M., Martinez-Morentin, L., Wang, G.N., Lee, Y.J., Liu, Q.X., Leslie, J., and Ding, L. (2017). Leptin-receptor-expressing bone marrow stromal cells are myofibroblasts in primary myelofibrosis. Nat Cell Biol *19*, 677-+.

Desterke, C., Martinaud, C., Ruzehaji, N., and Le Bousse-Kerdiles, M.C. (2015). Inflammation as a Keystone of Bone Marrow Stroma Alterations in Primary Myelofibrosis. Mediators Inflamm *2015*, 415024.

Dick, J.E. (2009). Looking ahead in cancer stem cell research. Nat Biotechnol *27*, 44-46.
DiNardo, C.D., and Cortes, J.E. (2016). Mutations in AML: prognostic and therapeutic implications. Hematology Am Soc Hematol Educ Program *2016*, 348-355.

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

Dohner, H. (2007). Implication of the molecular characterization of acute myeloid leukemia. Hematology Am Soc Hematol Educ Program, 412-419.

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

Falini, B., Mecucci, C., Tiacci, E., Alcalay, M., Rosati, R., Pasqualucci, L., La Starza, R., Diverio, D., Colombo, E., Santucci, A.*, et al.* (2005). Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. N Engl J Med *352*, 254-266.

Falini, B., Nicoletti, I., Martelli, M.F., and Mecucci, C. (2007). Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. Blood *109*, 874-885.

Ferrara, F., and Schiffer, C.A. (2013). Acute myeloid leukaemia in adults. Lancet *381*, 484-495.

Ferreira, H.J., Heyn, H., Vizoso, M., Moutinho, C., Vidal, E., Gomez, A., Martinez-Cardus, A., Simo-Riudalbas, L., Moran, S., Jost, E.*, et al.* (2016). DNMT3A mutations mediate the epigenetic reactivation of the leukemogenic factor MEIS1 in acute myeloid leukemia. Oncogene *35*, 3079-3082.

Follows, G.A., Tagoh, H., Lefevre, P., Hodge, D., Morgan, G.J., and Bonifer, C. (2003). Epigenetic consequences of AML1-ETO action at the human c-FMS locus. Embo J *22*, 2798-2809.

Forsberg, E.C., Serwold, T., Kogan, S., Weissman, I.L., and Passegue, E. (2006). New evidence supporting megakaryocyte-erythrocyte potential of flk2/flt3+ multipotent hematopoietic progenitors. Cell *126*, 415-426.

Gaidzik, V.I., Teleanu, V., Papaemmanuil, E., Weber, D., Paschka, P., Hahn, J., Wallrabenstein, T., Kolbinger, B., Kohne, C.H., Horst, H.A.*, et al.* (2016). RUNX1 mutations in acute myeloid leukemia are associated with distinct clinico-pathologic and genetic features (vol 30, pg 2160, 2016). Leukemia *30*, 2282-2282.

Garcia-Fruitos, E. (2010). Inclusion bodies: a new concept. Microb Cell Fact *9*, 80.
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.

Gazova, I., Lefevre, L., Bush, S.J., Clohisey, S., Arner, E., de Hoon, M., Severin, J., van Duin, L., Andersson, R., Lengeling, A.*, et al.* (2020). The Transcriptional Network That Controls Growth Arrest and Macrophage Differentiation in the Human Myeloid Leukemia Cell Line THP-1. Front Cell Dev Biol *8*, 498.

Gilding, L.N., and Somervaille, T.C.P. (2019). The Diverse Consequences of FOXC1 Deregulation in Cancer. Cancers (Basel) *11*.

Gilliland, D.G., and Griffin, J.D. (2002). Role of FLT3 in leukemia. Curr Opin Hematol *9*, 274-281.

Glont, S.E., Papachristou, E.K., Sawle, A., Holmes, K.A., Carroll, J.S., and Siersbaek, R. (2019). Identification of ChIP-seq and RIME grade antibodies for Estrogen Receptor alpha. PLoS One *14*, e0215340.

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.

Gombart, A.F., Hofmann, W.K., Kawano, S., Takeuchi, S., Krug, U., Kwok, S.H., Larsen, R.J., Asou, H., Miller, C.W., Hoelzer, D.*, et al.* (2002). Mutations in the gene encoding the transcription factor CCAAT/enhancer binding protein alpha in myelodysplastic syndromes and acute myeloid leukemias. Blood *99*, 1332-1340.

Gowher, H., Loutchanwoot, P., Vorobjeva, G., Handa, V., Jurkowska, R.Z., Jurkowski, T.P., and Jeltsch, A. (2006). Mutational analysis of the catalytic domain of the murine

Dnmt3a DNA-(cytosine C5)-methyltransferase. Journal of Molecular Biology *357*, 928-941.

Grier, D.G., Thompson, A., Kwasniewska, A., McGonigle, G.J., Halliday, H.L., and Lappin, T.R. (2005). The pathophysiology of HOX genes and their role in cancer. J Pathol *205*, 154-171.

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.

Growney, J.D., Shigematsu, H., Li, Z., Lee, B.H., Adelsperger, J., Rowan, R., Curley, D.P., Kutok, J.L., Akashi, K., Williams, I.R.*, et al.* (2005). Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. Blood *106*, 494-504.

Han, B., Bhowmick, N., Qu, Y., Chung, S., Giuliano, A.E., and Cui, X. (2017). FOXC1: an emerging marker and therapeutic target for cancer. Oncogene *36*, 3957-3963.
Han, B., Qu, Y., Jin, Y., Yu, Y., Deng, N., Wawrowsky, K., Zhang, X., Li, N., Bose, S., Wang, Q.*, et al.* (2015). FOXC1 Activates Smoothened-Independent Hedgehog Signaling in Basal-like Breast Cancer. Cell Rep *13*, 1046-1058.

Hanekamp, D., Cloos, J., and Schuurhuis, G.J. (2017). Leukemic stem cells: identification and clinical application. Int J Hematol *105*, 549-557.

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.

Hart, S.M., and Foroni, L. (2002). Core binding factor genes and human leukemia. Haematologica *87*, 1307-1323.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

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.

Holtschke, T., Lohler, J., Kanno, Y., Fehr, T., Giese, N., Rosenbauer, F., Lou, J., Knobeloch, K.P., Gabriele, L., Waring, J.F.*, et al.* (1996). Immunodeficiency and chronic myelogenous leukemia-like syndrome in mice with a targeted mutation of the ICSBP gene. Cell *87*, 307-317.

Hooper, A.T., Butler, J.M., Nolan, D.J., Kranz, A., Iida, K., Kobayashi, M., Kopp, H.G., Shido, K., Petit, I., Yanger, K.*, et al.* (2009). Engraftment and Reconstitution of Hematopoiesis Is Dependent on VEGFR2-Mediated Regeneration of Sinusoidal Endothelial Cells. Cell Stem Cell *4*, 263-274.

Hossain, D.M., Dos Santos, C., Zhang, Q., Kozlowska, A., Liu, H., Gao, C., Moreira, D., Swiderski, P., Jozwiak, A., Kline, J.*, et al.* (2014). Leukemia cell-targeted STAT3 silencing and TLR9 triggering generate systemic antitumor immunity. Blood *123*, 15-25.

Huang, L., Chi, J., Berry, F.B., Footz, T.K., Sharp, M.W., and Walter, M.A. (2008). Human p32 is a novel FOXC1-interacting protein that regulates FOXC1 transcriptional activity in ocular cells. Invest Ophthalmol Vis Sci *49*, 5243-5249.

Humbert, M., Halter, V., Shan, D., Laedrach, J., Leibundgut, E.O., Baerlocher, G.M., Tobler, A., Fey, M.F., and Tschan, M.P. (2011). Deregulated expression of Kruppel-like factors in acute myeloid leukemia. Leuk Res *35*, 909-913.

Ichikawa, M., Asai, T., Saito, T., Seo, S., Yamazaki, I., Yamagata, T., Mitani, K., Chiba, S., Ogawa, S., Kurokawa, M.*, et al.* (2004). AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. Nat Med *10*, 299-304.

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.
I
wasaki, H., Somoza, C., Shigematsu, H., Duprez, E.A., Iwasaki-Arai, J., Mizuno, S., Arinobu, Y., Geary, K., Zhang, P., Dayaram, T.*, et al.* (2005). Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. Blood *106*, 1590-1600.

Jackson, B.C., Carpenter, C., Nebert, D.W., and Vasiliou, V. (2010). Update of human and mouse forkhead box (FOX) gene families. Hum Genomics *4*, 345-352.
Jagannathan-Bogdan, M., and Zon, L.I. (2013). Hematopoiesis. Development *140*, 2463-2467.

Jennings, B.H., and Ish-Horowicz, D. (2008). The Groucho/TLE/Grg family of transcriptional co-repressors. Genome Biol *9*, 205.

Johansen, L.M., Iwama, A., Golub, T.R., and Tenen, D.G. (1999). c-myc is a critical target for C/EBP alpha in granulopoiesis. Blood *94*, 59a-59a.

Kaestner, K.H., Lee, K.H., Schlondorff, J., Hiemisch, H., Monaghan, A.P., and Schutz, G. (1993). Six members of the mouse forkhead gene family are developmentally regulated. Proc Natl Acad Sci U S A *90*, 7628-7631.

Kato, N., Kitaura, J., Doki, N., Komeno, Y., Watanabe-Okochi, N., Togami, K., Nakahara, F., Oki, T., Enomoto, Y., Fukuchi, Y.*, et al.* (2011). Two types of C/EBPalpha mutations play distinct but collaborative roles in leukemogenesis: lessons from clinical data and BMT models. Blood *117*, 221-233.

Kelly, L.M., and Gilliland, D.G. (2002). Genetics of myeloid leukemias. Annu Rev Genomics Hum Genet *3*, 179-198.

Khan, I., Halasi, M., Patel, A., Schultz, R., Kalakota, N., Chen, Y.H., Aardsma, N., Liu, L., Crispino, J.D., Mahmud, N.*, et al.* (2018). FOXM1 contributes to treatment failure in acute myeloid leukemia. Jci Insight *3*.

Khan, S.N., Jankowska, A.M., Mahfouz, R., Dunbar, A.J., Sugimoto, Y., Hosono, N., Hu, Z., Cheriyath, V., Vatolin, S., Przychodzen, B.*, et al.* (2013). Multiple mechanisms deregulate EZH2 and histone H3 lysine 27 epigenetic changes in myeloid malignancies. Leukemia *27*, 1301-1309.

Kiel, M.J., He, S., Ashkenazi, R., Gentry, S.N., Teta, M., Kushner, J.A., Jackson, T.L., and Morrison, S.J. (2007). Haematopoietic stem cells do not asymmetrically segregate chromosomes or retain BrdU. Nature *449*, 238-242.

Kiel, M.J., Yilmaz, O.H., Iwashita, T., Yilmaz, O.H., 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.*, et al.* (2010). TIM-3 is a promising target to selectively kill acute myeloid leukemia stem cells. Cell Stem Cell *7*, 708-717.

Kirstetter, P., Schuster, M.B., Bereshchenko, O., Moore, S., Dvinge, H., Kurz, E., Theilgaard-Monch, K., Mansson, R., Pedersen, T.A., Pabst, T.*, et al.* (2008). Modeling of C/EBPalpha mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells. Cancer Cell *13*, 299-310.

Klemsz, M.J., Mckercher, S.R., Celada, A., Vanbeveren, C., and Maki, R.A. (1990). The Macrophage and B-Cell Specific Transcription Factor Pu.1 Is Related to the Ets Oncogene. Cell *61*, 113-124.

Koh, C.P., Wang, C.Q., Ng, C.E., Ito, Y., Araki, M., Tergaonkar, V., Huang, G., and Osato, M. (2013). RUNX1 meets MLL: epigenetic regulation of hematopoiesis by two leukemia genes. Leukemia *27*, 1793-1802.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Krumlauf, R. (1994). Hox genes in vertebrate development. Cell *78*, 191-201.

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

Kumar, M., Gouw, M., Michael, S., Samano-Sanchez, H., Pancsa, R., Glavina, J., Diakogianni, A., Valverde, J.A., Bukirova, D., Calyseva, J.*, et al.* (2020). ELM-the eukaryotic linear motif resource in 2020. Nucleic Acids Res *48*, D296-D306.

Kume, T., Jiang, H.Y., Topczewska, J.M., and Hogan, B.L.M. (2001). The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. Gene Dev *15*, 2470-2482.

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.

Lancrin, C., Sroczynska, P., Stephenson, C., Allen, T., Kouskoff, V., and Lacaud, G. (2009). The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. Nature *457*, 892-895.

Lawrence, H.J., Helgason, C.D., Sauvageau, G., Fong, S., Izon, D.J., Humphries, R.K., and Largman, C. (1997). Mice bearing a targeted interruption of the homeobox gene HOXA9 have defects in myeloid, erythroid, and lymphoid hematopoiesis. Blood *89*, 1922-1930.

Lawrence, H.J., Rozenfeld, S., Cruz, C., Matsukuma, K., Kwong, A., Komuves, L., Buchberg, A.M., and Largman, C. (1999). Frequent co-expression of the HOXA9 and MEIS1 homeobox genes in human myeloid leukemias. Leukemia *13*, 1993-1999.

Levine, R.L. (2013). Molecular pathogenesis of AML: translating insights to the clinic. Best Pract Res Clin Haematol *26*, 245-248.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739-1740.

Lidonnici, M.R., Audia, A., Soliera, A.R., Prisco, M., Ferrari-Amorotti, G., Waldron, T., Donato, N., Zhang, Y., Martinez, R.V., Holyoake, T.L.*, et al.* (2010). Expression of the Transcriptional Repressor Gfi-1 Is Regulated by C/EBP alpha and Is Involved in Its Proliferation and Colony Formation-Inhibitory Effects in p210BCR/ABL-Expressing Cells. Cancer Res *70*, 7949-7959.

Lin, F.T., Macdougald, O.A., Diehl, A.M., and Lane, M.D. (1993). A 30-Kda Alternative Translation Product of the Ccaat Enhancer-Binding Protein-Alpha Message - Transcriptional Activator Lacking Antimitotic Activity. P Natl Acad Sci USA *90*, 9606-9610.

Lin, S., Ptasinska, A., Chen, X.T.T., Shrestha, M., Assi, S.A., Chin, P.S., Imperato, M.R., Aronow, B.J., Zhang, J.S., Weirauch, M.T.*, et al.* (2017). A FOXO1-induced oncogenic network defines the AML1-ETO preleukemic program. Blood *130*, 1213-1222.

Liu, L.L., Zhang, D.H., Mao, X., Zhang, X.H., and Zhang, B. (2014). Over-expression of FoxM1 is associated with adverse prognosis and FLT3-ITD in acute myeloid leukemia. Biochem Bioph Res Co *446*, 280-285.

Liu, P., and Han, Z.C. (2003). Treatment of acute promyelocytic leukemia and other hematologic malignancies with arsenic trioxide: review of clinical and basic studies. Int J Hematol *78*, 32-39.

Lorsbach, R.B., Moore, J., Ang, S.O., Sun, W.L., Lenny, N., and Downing, J.R. (2004). Role of RUNX1 in adult hematopoiesis: analysis of RUNX1-IRES-GFP knock-in mice reveals differential lineage expression. Blood *103*, 2522-2529.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696-1697.

Maiques-Diaz, A., Spencer, G.J., Lynch, J.T., Ciceri, F., Williams, E.L., Amaral, F.M.R., Wiseman, D.H., Harris, W.J., Li, Y., Sahoo, S.*, et al.* (2018). Enhancer Activation by Pharmacologic Displacement of LSD1 from GFI1 Induces Differentiation in Acute Myeloid Leukemia. Cell Rep *22*, 3641-3659.

Majeti, R. (2012). Clonal evolution of pre-leukemic hematopoietic stem cells precedes human acute myeloid leukemia. Onkologie *35*, 14-14.

Marcal, N., Patel, H., Dong, Z., Belanger-Jasmin, S., Hoffman, B., Helgason, C.D., Dang, J., and Stifani, S. (2005). Antagonistic effects of Grg6 and Groucho/TLE on the transcription repression activity of brain factor 1/FoxG1 and cortical neuron differentiation. Mol Cell Biol *25*, 10916-10929.

Martens, J.H.A., and Stunnenberg, H.G. (2010). The molecular signature of oncofusion proteins in acute myeloid leukemia. Febs Lett *584*, 2662-2669.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.

Metzeler, K.H., Maharry, K., Radmacher, M.D., Mrozek, K., Margeson, D., Becker, H., Curfman, J., Holland, K.B., Schwind, S., Whitman, S.P.*, et al.* (2011). TET2 Mutations Improve the New European LeukemiaNet Risk Classification of Acute Myeloid

Leukemia: A Cancer and Leukemia Group B Study. Journal of Clinical Oncology *29*, 1373-1381.

Miranda, M.B., and Johnson, D.E. (2007). Signal transduction pathways that contribute to myeloid differentiation. Leukemia *21*, 1363-1377.

Miyamoto, T., Weissman, I.L., and Akashi, K. (2000). AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. Proc Natl Acad Sci U S A *97*, 7521-7526.

Miyoshi, H., Shimizu, K., Kozu, T., Maseki, N., Kaneko, Y., and Ohki, M. (1991). t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. Proc Natl Acad Sci U S A *88*, 10431-10434.

Mohammed, H., Taylor, C., Brown, G.D., Papachristou, E.K., Carroll, J.S., and D'Santos, C.S. (2016). Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. Nat Protoc *11*, 316-326.

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., and Spradling, A.C. (2008). Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. Cell *132*, 598-611.

Mrozek, 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.

Mullighan, C.G., Kennedy, A., Zhou, X., Radtke, I., Phillips, L.A., Shurtleff, S.A., and Downing, J.R. (2007). Pediatric acute myeloid leukemia with NPM1 mutations is characterized by a gene expression profile with dysregulated HOX gene expression distinct from MLL-rearranged leukemias. Leukemia *21*, 2000-2009.

Murphy, T.C., Saleem, R.A., Footz, T., Ritch, R., McGillivray, B., and Walter, M.A. (2004). The wing 2 region of the FOXC1 forkhead domain is necessary for normal DNA-binding and transactivation functions. Invest Ophthalmol Vis Sci *45*, 2531-2538.

Myatt, S.S., and Lam, E.W. (2007). The emerging roles of forkhead box (Fox) proteins in cancer. Nat Rev Cancer *7*, 847-859.

Nakamura, S., Hirano, I., Okinaka, K., Takemura, T., Yokota, D., Ono, T., Shigeno, K., Shibata, K., Fujisawa, S., and Ohnishi, K. (2010). The FOXM1 transcriptional factor promotes the proliferation of leukemia cells through modulation of cell cycle progression in acute myeloid leukemia. Carcinogenesis *31*, 2012-2021.

Nerlov, C., and Ziff, E.B. (1995). Ccaat Enhancer-Binding Protein-Alpha Amino-Acid Motifs with Dual Tbp and Tfiib Binding Ability Cooperate to Activate Transcription in Both Yeast and Mammalian-Cells. Embo J *14*, 4318-4328.

Ng, R.K., Kong, C.T., So, C.C., Lui, W.C., Chan, Y.F., Leung, K.C., So, K.C., Tsang, H.M., Chan, L.C., and Sham, M.H. (2014). Epigenetic dysregulation of leukaemic HOX code in MLL-rearranged leukaemia mouse model. J Pathol *232*, 65-74.

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.

Nucifora, G., Begy, C.R., Kobayashi, H., Roulston, D., Claxton, D., Pedersen-Bjergaard, J., Parganas, E., Ihle, J.N., and Rowley, J.D. (1994). Consistent intergenic splicing and production of multiple transcripts between AML1 at 21q22 and unrelated genes at 3q26 in (3;21)(q26;q22) translocations. Proc Natl Acad Sci U S A *91*, 4004-4008.

Nutt, S.L., Metcalf, D., D' Amico, A., Polli, M., and Wu, L. (2005). Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors. J Exp Med *201*, 221-231.
Okuda, T., vanDeursen, J., Hiebert, S.W., Grosveld, G., and Downing, J.R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell *84*, 321-330.

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-+.

Omatsu, Y., Sugiyama, T., Kohara, H., Kondoh, G., Fujii, N., Kohno, K., and Nagasawa, T. (2010). The Essential Functions of Adipo-osteogenic Progenitors as the Hematopoietic Stem and Progenitor Cell Niche. Immunity *33*, 387-399.

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

Pabst, T., Eyholzer, M., Fos, J., and Mueller, B.U. (2009). Heterogeneity within AML with CEBPA mutations; only CEBPA double mutations, but not single CEBPA mutations are associated with favourable prognosis. Brit J Cancer *100*, 1343-1346.

Pabst, T., Mueller, B.U., Harakawa, N., Schoch, C., Haferlach, T., Behre, G., Hiddemann, W., Zhang, D.E., and Tenen, D.G. (2001). AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. Nat Med *7*, 444-451.

Papachristou, E.K., Kishore, K., Holding, A.N., Harvey, K., Roumeliotis, T.I., Chilamakuri, C.S.R., Omarjee, S., Chia, K.M., Swarbrick, A., Lim, E.*, et al.* (2018). A quantitative mass spectrometry-based approach to monitor the dynamics of endogenous chromatin-associated protein complexes. Nat Commun *9*, 2311.

Passegue, E., Jochum, W., Schorpp-Kistner, M., Mohle-Steinlein, U., and Wagner, E.F. (2001). Chronic myeloid leukemia with increased granulocyte progenitors in mice lacking junB expression in the myeloid lineage. Cell *104*, 21-32.

Pattabiraman, D.R., and Gonda, T.J. (2013). Role and potential for therapeutic targeting of MYB in leukemia. Leukemia *27*, 269-277.

Porse, B.T., Bryder, D., Theilgaard-Monch, K., Hasemann, M.S., Anderson, K., Damgaard, I., Jacobsen, S.E., and Nerlov, C. (2005). Loss of C/EBP alpha cell cycle control increases myeloid progenitor proliferation and transforms the neutrophil granulocyte lineage. J Exp Med *202*, 85-96.

Ptasinska, A., Assi, S., Martinez-Soria, N., Imperato, M.R., Piper, J., Cauchy, P., Pickin, A., James, S., Hoogenkamp, M., Tenen, D.G.*, et al.* (2014). Identification of a Dynamic Core Transcriptional Network in t(8;21) AML Regulating Differentiation Block and Self-Renewal. Blood *124*.

Ptasinska, A., Assi, S.A., Mannari, D., James, S.R., Williamson, D., Dunne, J., Hoogenkamp, M., Wu, M., Care, M., McNeill, H.*, et al.* (2012). Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. Leukemia *26*, 1829-1841.

Pundhir, S., Bratt Lauridsen, F.K., Schuster, M.B., Jakobsen, J.S., Ge, Y., Schoof, E.M., Rapin, N., Waage, J., Hasemann, M.S., and Porse, B.T. (2018). Enhancer and

Transcription Factor Dynamics during Myeloid Differentiation Reveal an Early Differentiation Block in Cebpa null Progenitors. Cell Rep *23*, 2744-2757.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Radomska, H.S., Huettner, C.S., Zhang, P., Cheng, T., Scadden, D.T., and Tenen, D.G. (1998). CCAAT enhancer binding protein alpha is a regulatory switch sufficient for induction of granulocytic development from bipotential myeloid progenitors. Mol Cell Biol *18*, 4301-4314.

Ran, D., Lam, K., Shia, W.J., Lo, M.C., Fan, J.B., Knorr, D.A., Ferrell, P.I., Ye, Z.H., Yan, M., Cheng, L.Z*., et al.* (2013). The role of RUNX1 isoforms in hematopoietic commitment of human pluripotent stem cells Response. Blood *121*, 5252-5253.

Rasmussen, K.D., and Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. Gene Dev *30*, 733-750.

Reilly, J.T. (2005). Pathogenesis of acute myeloid leukaemia and inv(16)(p13;q22): a paradigm for understanding leukaemogenesis? Br J Haematol *128*, 18-34.

Rezsohazy, R., Saurin, A.J., Maurel-Zaffran, C., and Graba, Y. (2015). Cellular and molecular insights into Hox protein action. Development *142*, 1212-1227.

Rosenbauer, F., Owens, B.M., Yu, L., Tumang, J.R., Steidl, U., Kutok, J.L., Clayton, L.K., Wagner, K., Scheller, M., Iwasaki, H*., et al.* (2006). Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU.1. Nat Genet *38*, 27-37.

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.

Sakoe, Y., Sakoe, K., Kirito, K., Ozawa, K., and Komatsu, N. (2010). FOXO3A as a key molecule for all-trans retinoic acid-induced granulocytic differentiation and apoptosis in acute promyelocytic leukemia. Blood *115*, 3787-3795.

Saleh, M., Rambaldi, I., Yang, X.J., and Featherstone, M.S. (2000). Cell signaling switches HOX-PBX complexes from repressors to activators of transcription mediated by histone deacetylases and histone acetyltransferases. Mol Cell Biol *20*, 8623-8633.
Scheijen, B., Ngo, H.T., Kang, H., and Griffin, J.D. (2004). FLT3 receptors with internal tandem duplications promote cell viability and proliferation by signaling through Foxo proteins. Oncogene *23*, 3338-3349.

Schwieger, M., Lohler, J., Fischer, M., Herwig, U., Tenen, D.G., and Stocking, C. (2004). A dominant-negative mutant of C/EBPalpha, associated with acute myeloid leukemias, inhibits differentiation of myeloid and erythroid progenitors of man but not mouse. Blood *103*, 2744-2752.

Scott, E.W., Simon, M.C., Anastasi, J., and Singh, H. (1994). Requirement of Transcription Factor Pu.1 in the Development of Multiple Hematopoietic Lineages. Science *265*, 1573-1577.

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

Seo, S., Chen, L.S., Liu, W.Z., Zhao, D.M., Schultz, K.M., Sasman, A., Liu, T., Zhang, H.F., Gage, P.J., and Kume, T. (2017). Foxc1 and Foxc2 in the Neural Crest Are

Required for Ocular Anterior Segment Development. Invest Ophth Vis Sci *58*, 1368-1377.

Shafarenko, M., Amanullah, A., Gregory, B., Liebermann, D.A., and Hoffman, B. (2004). Fos modulates myeloid cell survival and differentiation and partially abrogates the c-Myc block in terminal myeloid differentiation. Blood *103*, 4259-4267.

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.

Shigesada, K., van de Sluis, B., and Liu, P.P. (2004). Mechanism of leukemogenesis by the inv(16) chimeric gene CBFB/PEBP2B-MHY11. Oncogene *23*, 4297-4307.

Shih, A.H., Abdel-Wahab, O., Patel, J.P., and Levine, R.L. (2012). The role of mutations in epigenetic regulators in myeloid malignancies. Nat Rev Cancer *12*, 599-612.

Sive, J.I., Basilico, S., Hannah, R., Kinston, S.J., Calero-Nieto, F.J., and Gottgens, B. (2016). Genome-scale definition of the transcriptional programme associated with compromised PU.1 activity in acute myeloid leukaemia. Leukemia *30*, 14-23.
Slany, R.K. (2009). The molecular biology of mixed lineage leukemia. Haematol-Hematol J *94*, 984-993.

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.

Somerville, T.D., Wiseman, D.H., Spencer, G.J., Huang, X., Lynch, J.T., Leong, H.S., Williams, E.L., Cheesman, E., and Somervaille, T.C. (2015). Frequent Derepression of the Mesenchymal Transcription Factor Gene FOXC1 in Acute Myeloid Leukemia. Cancer Cell *28*, 329-342.

Somerville, T.D.D., Simeoni, F., Chadwick, J.A., Williams, E.L., Spencer, G.J., Boros, K., Wirth, C., Tholouli, E., Byers, R.J., and Somervaille, T.C.P. (2018). Derepression of the Iroquois Homeodomain Transcription Factor Gene IRX3 Confers Differentiation Block in Acute Leukemia. Cell Rep *22*, 638-652.

Sood, R., Kamikubo, Y., and Liu, P. (2017). Role of RUNX1 in hematological malignancies. Blood *129*, 2070-2082.

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. Nature Reviews Cancer *2*, 502-513.

Sugiyama, T., and Nagasawa, T. (2012). Bone marrow niches for hematopoietic stem cells and immune cells. Inflamm Allergy Drug Targets *11*, 201-206.
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.

Sun, J.J., Ishii, M., Ting, M.C., and Maxson, R. (2013). Foxc1 controls the growth of the murine frontal bone rudiment by direct regulation of a Bmp response threshold of Msx2. Development *140*, 1034-1044.

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

Tamura, T., Kurotaki, D., and Koizumi, S. (2015). Regulation of myelopoiesis by the transcription factor IRF8. Int J Hematol *101*, 342-351.

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.

Tawana, K., Rio-Machin, A., Preudhomme, C., and Fitzgibbon, J. (2017). Familial CEBPA-mutated acute myeloid leukemia. Semin Hematol *54*, 87-93.

Tenen, D.G. (2001). Abnormalities of the CEBP alpha transcription factor: a major target in acute myeloid leukemia. Leukemia *15*, 688-689.

Tenen, D.G., Hromas, R., Licht, J.D., and Zhang, D.E. (1997). Transcription factors, normal myeloid development, and leukemia. Blood *90*, 489-519.

Testa, U., Labbaye, C., Castelli, G., and Pelosi, E. (2016). Oxidative stress and hypoxia in normal and leukemic stem cells. Exp Hematol *44*, 540-560.

Thol, F., Damm, F., Ludeking, A., Winschel, C., Wagner, K., Morgan, M., Yun, H., Gohring, G., Schlegelberger, B., Hoelzer, D.*, et al.* (2011). Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia. J Clin Oncol *29*, 2889-2896.
Tupler, R., Perini, G., and Green, M.R. (2001). Expressing the human genome. Nature *409*, 832-833.

Vangala, R.K., Heiss-Neumann, M.S., Rangatia, J.S., Singh, S.M., Schoch, C., Tenen, D.G., Hiddemann, W., and Behre, G. (2003). The myeloid master regulator transcription factor PU.1 is inactivated by AML1-ETO in t(8;21) myeloid leukemia. Blood *101*, 270-277.

Vardiman, J.W., Thiele, J., Arber, D.A., Brunning, R.D., Borowitz, M.J., Porwit, A., Harris, N.L., Le Beau, M.M., Hellstrom-Lindberg, E., Tefferi, A.*, et al.* (2009). The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood *114*, 937-951.

Verhaak, R.G., Goudswaard, C.S., van Putten, W., Bijl, M.A., Sanders, M.A., Hugens, W., Uitterlinden, A.G., Erpelinck, C.A., Delwel, R., Lowenberg, B.*, et al.* (2005). Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. Blood *106*, 3747-3754.

Vicente, C., Conchillo, A., Garcia-Sanchez, M.A., and Odero, M.D. (2012). The role of the GATA2 transcription factor in normal and malignant hematopoiesis. Crit Rev Oncol Hematol *82*, 1-17.

Visnjic, D., Kalajzic, Z., Rowe, D.W., Katavic, V., Lorenzo, J., and Aguila, H.L. (2004). Hematopoiesis is severely altered in mice with an induced osteoblast deficiency. Blood *103*, 3258-3264.

Wang, J., Hoshino, T., Redner, R.L., Kajigaya, S., and Liu, J.M. (1998). ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. Proc Natl Acad Sci U S A *95*, 10860-10865.

Wang, J.C., and Dick, J.E. (2005). Cancer stem cells: lessons from leukemia. Trends Cell Biol *15*, 494-501.

Wang, J.H., Xu, Y.L., Li, L., Wang, L., Yao, R., Sun, Q., and Du, G.H. (2017). FOXC1 is associated with estrogen receptor alpha and affects sensitivity of tamoxifen treatment in breast cancer. Cancer Med-Us *6*, 275-287.

Wang, Z.Y., and Chen, Z. (2008). Acute promyelocytic leukemia: from highly fatal to highly curable. Blood *111*, 2505-2515.

Weissman, I. (2005). Stem cell research - Paths to cancer therapies and regenerative medicine. Jama-J Am Med Assoc *294*, 1359-1366.

Wilhelmson, A.S., and Porse, B.T. (2020). CCAAT enhancer binding protein alpha (CEBPA) biallelic acute myeloid leukaemia: cooperating lesions, molecular mechanisms and clinical relevance. Br J Haematol.

Wilson, A., Laurenti, E., Oser, G., van der Wath, R.C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C.F., Eshkind, L., Bockamp, E.*, et al.* (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell *135*, 1118-1129.

Wingelhofer, B., and Somervaille, T.C.P. (2019). Emerging Epigenetic Therapeutic Targets in Acute Myeloid Leukemia. Front Oncol *9*.

Winters, A.C., and Bernt, K.M. (2017). MLL-Rearranged Leukemias-An Update on Science and Clinical Approaches. Front Pediatr *5*, 4.

Wiseman, D.H., Greystoke, B.F., and Somervaille, T.C.P. (2014). The variety of leukemic stem cells in myeloid malignancy. Oncogene *33*, 3091-3098.

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.

Yang, L., Rau, R., and Goodell, M.A. (2015). DNMT3A in haematological malignancies. Nat Rev Cancer *15*, 152-165.

Zhang, D.E., Hetherington, C.J., Meyers, S., Rhoades, K.L., Larson, C.J., Chen, H.M., Hiebert, S.W., and Tenen, D.G. (1996). CCAAT enhancer-binding protein (C/EBP) and AML1 (CBF alpha2) synergistically activate the macrophage colony-stimulating factor receptor promoter. Mol Cell Biol *16*, 1231-1240.

Zhang, D.E., Zhang, P., Wang, N.D., Hetherington, C.J., Darlington, G.J., and Tenen, D.G. (1997). Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. P Natl Acad Sci USA *94*, 569-574.

Zhang, J., Niu, C., Ye, L., Huang, H., He, X., Tong, W.G., Ross, J., Haug, J., Johnson, T., Feng, J.Q.*, et al.* (2003). Identification of the haematopoietic stem cell niche and control of the niche size. Nature *425*, 836-841.

Zhang, P., Iwasaki-Arai, J., Iwasaki, H., Fenyus, M.L., Dayaram, T., Owens, B.M., Shigematsu, H., Levantini, E., Huettner, C.S., Lekstrom-Himes, J.A.*, et al.* (2004). Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. Immunity *21*, 853-863.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhou, J., and Chng, W.J. (2014). Identification and targeting leukemia stem cells: The path to the cure for acute myeloid leukemia. World J Stem Cells *6*, 473-484.

Zou, Y.R., Kottmann, A.H., Kuroda, M., Taniuchi, I., and Littman, D.R. (1998). Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development. Nature *393*, 595-599.

Appendix 1

Revert the SIRT: Normalizing SIRT1 Activity in Myelodysplastic Stem Cells

Fabrizio Simeoni, Tim C.P. Somervaille. (2018)

Cell Stem Cell 23, 315-317.

A preview article written by myself and Professor Somervaille.

caused in part by capricious gods sending mixed signals: strategically employing reprogramming factors that do not inadvertently recruit corepressors, or engineering them to this end, could expedite the odyssey, perhaps broadly availing reprogramming into diverse lineages.

REFERENCES

dos Santos, R.L., Tosti, L., Radzisheuskaya, A., Caballero, I.M., Kaji, K., Hendrich, B., and Silva, J.C.R. (2014). MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. Cell Stem Cell *15*, 102–110.

Kloet, S.L., Baymaz, H.I., Makowski, M., Groenewold, V., Jansen, P.W.T.C., Berendsen, M., Niazi, H., Kops, G.J., and Vermeulen, M. (2015). Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. FEBS J. *282*, 1774–1785.

Loh, K.M., and Lim, B. (2013). Stem cells: Close encounters with full potential. Nature *502*, 41–42.

Low, J.K.K., Webb, S.R., Silva, A.P.G., Saathoff, H., Ryan, D.P., Torrado, M., Brofelth, M., Parker, B.L., Shepherd, N.E., and Mackay, J.P. (2016). CHD4 Is a Peripheral Component of the Nucleosome Remodeling and Deacetylase Complex. J. Biol. Chem. *291*, 15853–15866.

Mor, N., Rais, Y., Sheban, D., Peles, S., Aguilera-Castrejon, A., Zviran, A., Elinger, D., Viukov, S., Geula, S., Krupalnik, V., et al. (2018). Neutralizing Gatad2a-Chd4-Mbd3/NuRD Complex Facilitates Deterministic Induction of Naive Pluripotency. Cell Stem Cell *23*, this issue, 412–425.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., et al. (2013). Deterministic direct reprogramming of somatic cells to pluripotency. Nature *502*, 65–70.

Sakai, H., Urano, T., Ookata, K., Kim, M.-H., Hirai, Y., Saito, M., Nojima, Y., and Ishikawa, F. (2002). MBD3 and HDAC1, two components of the NuRD complex, are localized at Aurora-A-positive centrosomes in M phase. J. Biol. Chem. *277*, 48714–48723.

Torrado, M., Low, J.K.K., Silva, A.P.G., Schmidberger, J.W., Sana, M., Sharifi Tabar, M., Isilak, M.E., Winning, C.S., Kwong, C., Bedward, M.J., et al. (2017). Refinement of the subunit interaction network within the nucleosome remodelling and deacetylase (NuRD) complex. FEBS J. *284*, 4216–4232.

Zhang, W., Aubert, A., Gomez de Segura, J.M., Karuppasamy, M., Basu, S., Murthy, A.S., Diamante, A., Drury, T.A., Balmer, J., Cramard, J., et al. (2016). The Nucleosome Remodeling and Deacetylase Complex NuRD Is Built from Preformed Catalytically Active Sub-modules. J. Mol. Biol. *428*, 2931–2942.

Zhuang, Q., Li, W., Benda, C., Huang, Z., Ahmed, T., Liu, P., Guo, X., Ibañez, D.P., Luo, Z., Zhang, M., et al. (2018). NCoR/SMRT co-repressors cooperate with c-MYC to create an epigenetic barrier to somatic cell reprogramming. Nat. Cell Biol. *20*, 400–412.

Revert the SIRT: Normalizing SIRT1 Activity in Myelodysplastic Stem Cells

Fabrizio Simeoni[1] **and Tim C.P. Somervaille**[1,*]
[1]Leukaemia Biology Laboratory, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, M20 4JG, UK
*Correspondence: tim.somervaille@cruk.manchester.ac.uk
https://doi.org/10.1016/j.stem.2018.08.003

Myelodysplastic syndromes are hematologic malignancies with few treatment options and a propensity to transform to acute myeloid leukemia. In this issue of *Cell Stem Cell*, Sun et al. (2018) report that low SIRT1 levels in myelodysplastic stem cells contribute to aberrant self-renewal through enabling hyperacetylation and reduced activity of TET2.

The seven mammalian sirtuins form a highly conserved family of nicotinamide adenine dinucleotide $(NAD)^+$-dependent histone deacetylases with wide-ranging activities and subcellular localizations (Anderson et al., 2014). SIRT1, the best-studied mammalian sirtuin, serves to deacetylate many targets in the nucleus and cytoplasm including histones, metabolic enzymes, transcription factors, and DNA repair proteins. Its activity counters the development of inflammation and insulin resistance, as well as age-associated conditions such as heart disease, diabetes, and neurodegeneration, at least in part through sensing nutritional stress. In cancer, SIRT1 has variably been reported to have both tumor suppressing and tumor promoting activities in a context-dependent manner (Chalkiadaki and Guarente, 2015).

In this issue of *Cell Stem Cell*, Sun and colleagues (Sun et al., 2018) report that in human myelodysplastic syndrome (MDS) stem and progenitor cells, SIRT1 protein levels are very low and that genetic and pharmacological SIRT1 activation counters their abnormal self-renewal. SIRT1 is therefore a novel therapeutic target in this difficult-to-treat disease. The human myelodysplastic syndromes are a diverse group of clonal myeloid malignancies characterized clinically by ineffective hematopoiesis leading to anemia, leucopenia, and/or thrombocytopenia. Patients often require blood product and/or growth factor support, and there is a frequent propensity for evolution to acute myeloid leukemia (AML). Currently available therapies including hypomethylating agents (e.g., azacitidine) or allogeneic transplantation have only limited efficacy or applicability.

Interestingly, previous studies have found SIRT1 overexpression in other myeloid malignancies compared with its presence in normal hematopoietic stem and progenitor cells (HSPCs), and they have done so to functional effect. In primary human chronic myeloid leukemia (CML) HSPCs, genetic or pharmacological inhibition of SIRT1 impaired growth and enhanced apoptosis through increasing the acetylation and activity of

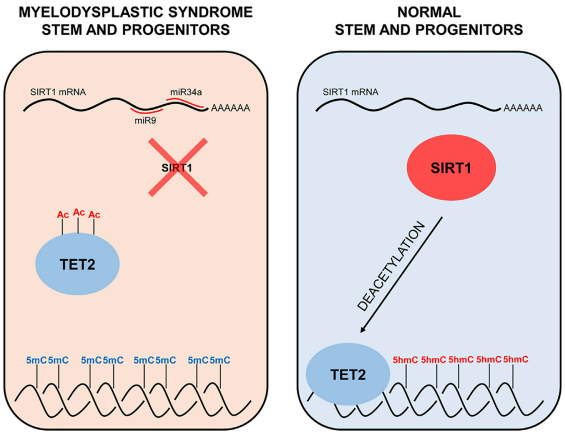


Figure 1. Low SIRT1 Levels in Myelodysplastic Stem and Progenitor Cells Enable Hyperacetylation and Reduced Activity of TET2

TP53 (Li et al., 2012). Likewise, in human AML associated with internal tandem duplications of FLT3 (FLT3-ITD), HSPCs also expressed high levels of SIRT1. This was due to enhanced SIRT1 protein stability resulting from expression of the deubiquitinase USP22, which removes conjugated polyubiquitin chains. Again, genetic or pharmacological inhibition of SIRT1 impaired growth and enhanced apoptosis through increasing acetylation and activity of TP53 (Li et al., 2014). By contrast, *Sirt1* knockout mice show largely unperturbed hematopoiesis during development and early adulthood (Leko et al., 2012), but expansion of the HSPC compartment is observed under conditions of hematopoietic stress (Singh et al., 2013).

In the present study, Sun and colleagues found that SIRT1 protein levels were significantly lower in a cohort of primary human MDS HSPC samples relative to their normal counterparts, likely due to a block in translation caused by high expression of miR9 and miR34a (Figure 1). Integrated analysis of human MDS-L cells, NUP98-HOXD13 transgenic MDS mice, and primary patient samples indicated an inverse relationship between SIRT1 levels or activity and MDS HSPC self-renewal: increasing SIRT1 activity suppressed growth and dysplastic phenotypes and promoted differentiation. To identify the most important target of SIRT1's deacetylase activity in MDS-L cells, the authors adopted an acetyllysine immunoprecipitation mass spectrometry approach followed by functional screening of differentially acetylated candidates. The authors identified TET2 (Tet methylcytosine dioxygenase 2), a dioxygenase that regulates turnover of DNA methylation through oxidation of methylated cytosine (5mC) to 5-hydroxymethylcytosine (5hmC). This is remarkable, given the well-established importance of TET2 in MDS. Heterozygous, loss-of-function somatic mutations of TET2 leading to haploinsufficiency are found in 30%–35% of patients (Haferlach et al., 2014) and are associated with DNA hypermethylation.

Proteomic and mutational analyses demonstrated that deacetylation of K1472, K1473, and K1478 of TET2 by SIRT1 enhanced the catalytic activity of TET2, likely by facilitating its interaction with DNA. These residues are conserved and located within the catalytic domain. Consistent with this, forced expression of SIRT1 in MDS-L cells led to increased levels of 5hmC at CpG islands and enhancers (Figure 1).

An important question arising from these data is whether pharmacologic activation of SIRT1 might prove to be an effective therapy for patients with MDS. To address this, the authors made use of synthetic SIRT1 activators, SRT1720 and SRT2104, and observed that the clonogenic activity of $CD34^+$ MDS HSPCs, but not that of their normal counterparts, was dose-dependently inhibited *in vitro* (whether a TET2 mutation was present or not), and reduced engraftment of treated cells *in vivo*. Importantly, SRT1720-treated primary MDS samples exhibited reduced TET2 acetylation and increased TET2 catalytic activity, as evidenced by increased 5hmC levels. Results were confirmed using the NUP98-HOXD13 transgenic MDS mouse model: 120 days post-transplant, secondary recipients of NUP98-HOXD13 transgenic bone marrow were treated for 12 weeks with SRT1720 or vehicle. Mice exposed to pharmacologic SIRT1 activation had better blood counts, reduced levels of chimerism with MDS cells, and fewer MDS-initiating cells. Once more, there was decreased TET2 acetylation and increased TET2 activity, as evidenced by increased 5hmC.

Taken together, these data indicate that the authors have uncovered a previously unappreciated mechanism by which TET2 activity is reduced, which is of potential clinical significance. Low SIRT1 levels in MDS HSPCs result in hyperacetylated TET2 with reduced catalytic activity, which in turn results in higher levels of DNA methylation. This observation may explain why many MDS patient samples that lack a TET2, IDH1, or IDH2 mutation nevertheless exhibit very low 5hmC levels (Liu et al., 2013) and suggests that reduced cellular TET2 activity can also be achieved through a non-mutational post-translational mechanism.

The selective and potent synthetic SIRT1 activator SRT2104 has already been evaluated in diverse clinical scenarios with some positive results including satisfactory safety, improved lipid profiles, reduction in biomarkers of inflammatory bowel disease activity, and improvement in psoriasis (Dai et al., 2018), although poor and variable pharmacokinetics were a feature. Clinical trial evaluation of SIRT1 activators in MDS, aiming to restore TET2 activity, is now

of clear interest. One note of caution is that, given the substantial cell context differences in SIRT1's deacetylation targets and functions, choice of patient and careful monitoring would be important to ensure that positive selection of leukemic sub-clones does not take place. For example, based on the authors' prior discoveries (Li et al., 2012, 2014), pharmacologic activation of SIRT1 in myeloid malignancies such as FLT3-ITD AML or CML, and potentially other genetic subtypes, might be predicted to exacerbate disease through TP53 inactivation.

REFERENCES

Anderson, K.A., Green, M.F., Huynh, F.K., Wagner, G.R., and Hirschey, M.D. (2014). SnapShot: Mammalian sirtuins. Cell *159*, 956e1.

Chalkiadaki, A., and Guarente, L. (2015). The multifaceted functions of sirtuins in cancer. Nat. Rev. Cancer *15*, 608–624.

Dai, H., Sinclair, D.A., Ellis, J.L., and Steegborn, C. (2018). Sirtuin activators and inhibitors: Promises, achievements, and challenges. Pharmacol. Ther. *188*, 140–154.

Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., Schnittger, S., Sanada, M., Kon, A., Alpermann, T., et al. (2014). Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia *28*, 241–247.

Leko, V., Varnum-finney, B., Li, H., Gu, Y., Flowers, D., Nourigat, C., Bernstein, I.D., and Bedalov, A. (2012). SIRT1 is dispensable for function of hematopoietic stem cells in adult mice. Blood *119*, 1856–1860.

Li, L., Wang, L., Li, 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, L., Osdal, T., Ho, Y., Chun, S., McDonald, T., Agarwal, P., Lin, A., Chu, S., Qi, J., Li, L., et al. (2014). SIRT1 activation by a c-MYC oncogenic network promotes the maintenance and drug resistance of human FLT3-ITD acute myeloid leukemia stem cells. Cell Stem Cell *15*, 431–446.

Liu, X., Zhang, G., Yi, Y., Xiao, L., Pei, M., Liu, S., Luo, Y., Zhong, H., Xu, Y., Zheng, W., and Shen, J. (2013). Decreased 5-hydroxymethylcytosine levels are associated with *TET2* mutation and unfavorable overall survival in myelodysplastic syndromes. Leuk. Lymphoma *54*, 2466–2473.

Singh, S.K., Williams, C.A., Klarmann, K., Burkett, S.S., Keller, J.R., and Oberdoerffer, P. (2013). Sirt1 ablation promotes stress-induced loss of epigenetic and genomic hematopoietic stem and progenitor cell maintenance. J. Exp. Med. *210*, 987–1001.

Sun, J., He, X., Zhu, Y., Ding, Z., Dong, H., Feng, Y., Du, J., Wang, H., Wu, X., Zhang, L., et al. (2018). SIRT1 Activation Disrupts Maintenance of Myelodysplastic Syndrome Stem and Progenitor Cells by Restoring TET2 Function. Cell Stem Cell *23*, this issue, 355–369.

Appendix 2

A stress-responsive enhancer induces dynamic drug resistance in acute myeloid leukemia

Mark S. Williams, Fabio M. Amaral, Fabrizio Simeoni, Tim C.P. Somervaille. (2020) The Journal of Clinical Investigation 130, 1217-1232.

A research article written by Mark S. Williams, where I am a contributing author.

The Journal of Clinical Investigation RESEARCH ARTICLE

A stress-responsive enhancer induces dynamic drug resistance in acute myeloid leukemia

Mark S. Williams, Fabio M.R. Amaral, Fabrizio Simeoni, and Tim C.P. Somervaille

Leukaemia Biology Laboratory, Cancer Research UK Manchester Institute, University of Manchester, Manchester, United Kingdom.

The drug efflux pump ABCB1 is a key driver of chemoresistance, and high expression predicts treatment failure in acute myeloid leukemia (AML). In this study, we identified and functionally validated the network of enhancers that controls expression of *ABCB1*. We show that exposure of leukemia cells to daunorubicin activated an integrated stress response–like transcriptional program to induce *ABCB1* through remodeling and activation of an ATF4-bound, stress-responsive enhancer. Protracted stress primed enhancers for rapid increases in activity following re-exposure of cells to daunorubicin, providing an epigenetic memory of prior drug treatment. In primary human AML, exposure of fresh blast cells to daunorubicin activated the stress-responsive enhancer and led to dose-dependent induction of *ABCB1*. Dynamic induction of *ABCB1* by diverse stressors, including chemotherapy, facilitated escape of leukemia cells from targeted third-generation ABCB1 inhibition, providing an explanation for the failure of ABCB1 inhibitors in clinical trials. Stress-induced upregulation of *ABCB1* was mitigated by combined use of the pharmacologic inhibitors U0126 and ISRIB, which inhibit stress signaling and have potential for use as adjuvants to enhance the activity of ABCB1 inhibitors.

Introduction

Resistance of leukemia cells, including leukemia stem cells with disease-reconstituting activity, to the chemotherapy drugs used in standard induction and consolidation regimens is the most common cause of treatment failure in acute myeloid leukemia (AML). Primary drug resistance is certainly linked to the genetic lesions driving AML; for example, leukemias with an $NPM1^{mut}$ FLT3-ITD genotype are substantially more difficult to cure with chemotherapy alone than $NPM1^{mut}$ AMLs that lack FLT3-ITD, although the reasons for such differential sensitivity remain obscure. Levels of expression of drug-detoxifying enzymes, topoisomerase II, microRNAs, and the propensity of cells to undergo autophagy have all been suggested to contribute to intrinsic drug resistance (1).

Most significantly, high expression of the ABCB1 drug efflux pump (also known as MDR1 or P-glycoprotein), which actively exports anthracyclines, predicts treatment failure in AML (2, 3). More generally, ABCB1 is highly expressed in many poor-risk malignancies as well as in normal gut, liver, and kidney and the blood-brain barrier (4). Inhibitors of ABCB1 have been tested in clinical trials in AML but with limited success. Nevertheless, in view of its significant role in the disease, the rationale for targeting ABCB1 remains strong (3). Furthermore, given the abundance of preclinical evidence supporting a role for ABCB1 in drug resistance, the failure in clinical trials of inhibitors of ABCB1 has not been adequately explained.

A greater understanding of the cancer-specific regulation of ABCB1 and its role in drug resistance is required to facilitate the design of new therapeutic strategies. Specifically, it is unclear how *ABCB1* expression is established and maintained in human AML. Whether expression is constitutive or dynamic is of critical relevance to the clinical application of ABCB1 inhibitors, where previous trials have assumed constant expression (5). Advances in enhancer biology have established that these distal regulatory elements govern cell type–specific gene expression and frequently respond to environmental conditions and homeostatic perturbations (6, 7). Critically, the enhancer landscape of *ABCB1* has yet to be defined.

Results

Resistance to daunorubicin due to stereotypical induction of ABCB1. We initially set out to evaluate mechanistic heterogeneity in the acquisition of resistance to daunorubicin, which is the mainstay drug of AML induction chemotherapy regimens. To do this we generated multiple daunorubicin-resistant K562 leukemia cell lines in parallel. K562 cells are derived from the pleural effusion of a patient with chronic myeloid leukemia in terminal myeloid blast crisis (8), and, unmanipulated, they undergo apoptosis in response to daunorubicin with an IC_{50} of approximately 40 nM. We selected this line in view of its extensive use as a model system by the ENCODE Consortium.

Three separate vials of early-passage K562 cells were thawed and cultured separately for 2 weeks. The 3 drug-sensitive lines were designated K562_S1–3, and aliquots were cryopreserved for later use. Each line was then exposed to escalating doses of daunorubicin in continuing culture until they were able to expand in 500 nM (Figure 1A). Resistant lines were designated K562_R1–3, and the time taken to acquire this level of resistance was 106 days (K562_R1 and R3) or 117 days (K562_R2). The daunorubicin IC_{50}

Conflict of interest: The authors have declared that no conflict of interest exists.

Submitted: June 3, 2019; **Accepted:** November 21, 2019; **Published:** February 4, 2020.
Reference information: *J Clin Invest*. 2020;130(3):1217–1232.
https://doi.org/10.1172/JCI130809.

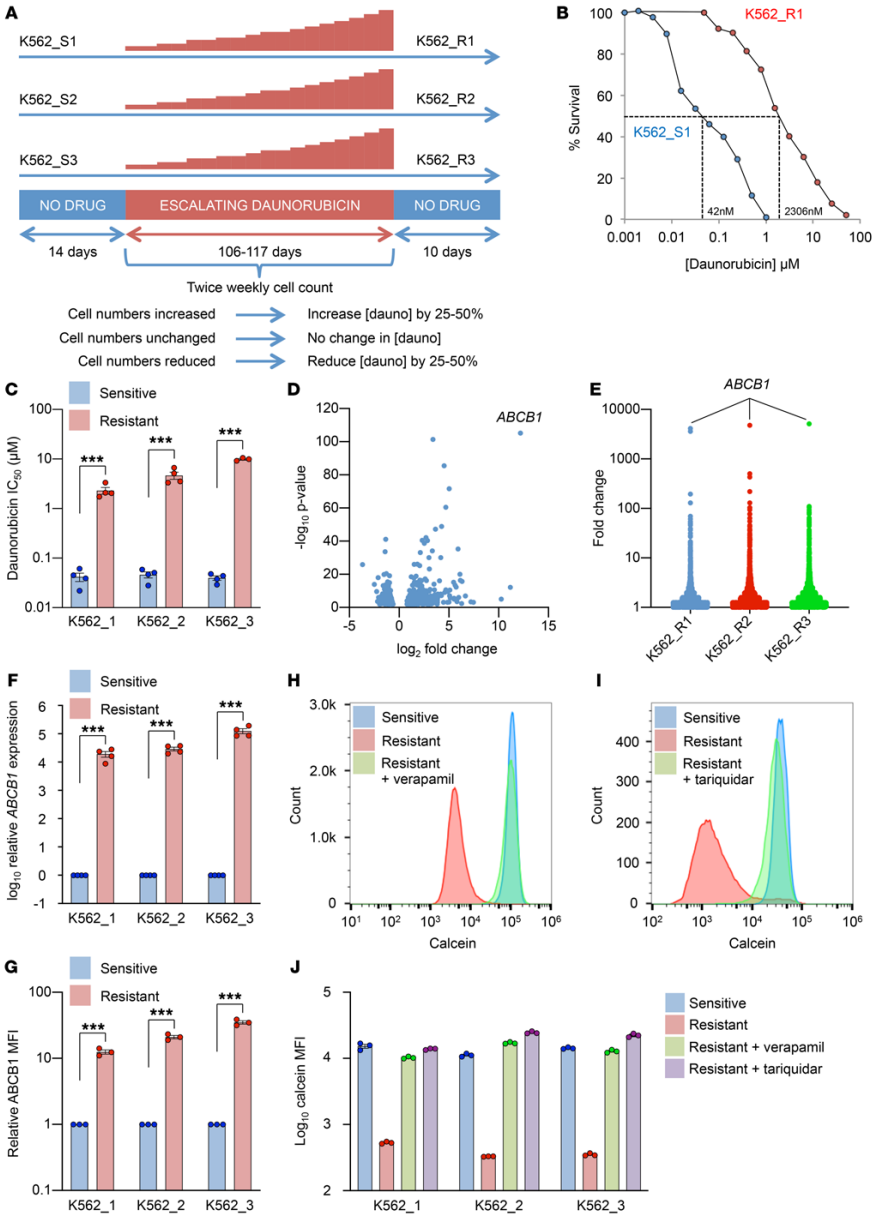

A
K562_S1
K562_R1
K562_S2
K562_R2
K562_S3
K562_R3
NO DRUG
ESCALATING DAUNORUBICIN
NO DRUG
14 days
106-117 days
10 days
Twice weekly cell count
Cell numbers increased
Increase [dauno] by 25-50%
Cell numbers unchanged
No change in [dauno]
Cell numbers reduced
Reduce [dauno] by 25-50%
B
K562_R1
K562_S1
% Survival
42nM
2306nM
[Daunorubicin] μM
C
Sensitive
Resistant
Daunorubicin IC50 (μM)
K562_1
K562_2
K562_3
D
ABCB1
-log10 p-value
log2 fold change
E
ABCB1
Fold change
K562_R1
K562_R2
K562_R3
F
Sensitive
Resistant
log10 relative ABCB1 expression
K562_1
K562_2
K562_3
H
Sensitive
Resistant
Resistant + verapamil
Count
Calcein
I
Sensitive
Resistant
Resistant + tariquidar
Count
Calcein
G
Sensitive
Resistant
Relative ABCB1 MFI
K562_1
K562_2
K562_3
J
Log10 calcein MFI
K562_1
K562_2
K562_3
Sensitive
Resistant
Resistant + verapamil
Resistant + tariquidar

Figure 1. Resistance to daunorubicin due to stereotypical induction of *ABCB1*. (**A**) Outline of experiment. (**B**) Dose-response curves for sensitive and resistant lines following 72 hours of treatment with the indicated doses of daunorubicin. (**C**) Bar chart shows mean ± SEM IC_{50} values for daunorubicin for all lines (n = 4). ***P < 0.001 by unpaired t test. (**D**) Volcano plot shows differential gene expression between sensitive (K562_S1–3) and resistant (K562_R1–3) cell lines. (**E**) *ABCB1* is the most highly upregulated gene in each resistant line compared with its sensitive parental line. (**F**) Mean ± SEM fold increase in *ABCB1* expression, as determined by quantitative PCR (n = 4). ***P < 0.001 by unpaired t test. (**G**) Mean ± SEM fold increase in ABCB1 median fluorescence intensity (MFI), as determined by flow cytometry (n = 3). ***P < 0.001 by unpaired t test. (**H** and **I**) Representative flow histograms show calcein AM retention in the indicated lines in the presence or absence of verapamil 40 μM (**H**) or tariquidar 50 nM (**I**). (**J**) Summary of calcein AM retention data for all 3 line pairs for verapamil and tariquidar (n = 3).

values were 2.3 μM, 4.7 μM, and 9.9 μM, respectively, with 55-fold, 101-fold, and 249-fold increases versus drug-sensitive lines K562_S1–3, respectively (Figure 1, B and C).

To evaluate changes in gene expression, we performed RNA sequencing. To avoid detecting transient changes in gene expression associated with recent daunorubicin exposure or contamination with apoptotic cells, each line was propagated for a further 10 days without daunorubicin prior to RNA extraction. RNA sequencing was performed using a single replicate for each sensitive line (K562_S1–3) and 2 replicates for each resistant line (K562_R1–3). When each drug-resistant line was compared with the sensitive lines, the most highly upregulated protein coding gene in each case was *ABCB1* (mean 4700-fold) even though the lines had been cultured separately from one another for at least 4 months (Figure 1, D and E, and Supplemental Table 1; supplemental material available online with this article; https://doi.org/10.1172/JCI130809DS1). Increased *ABCB1* expression was confirmed by quantitative PCR, and this correlated well with increased cell surface ABCB1 protein (Figure 1, F and G). To confirm that the upregulated protein expression was functional, we performed fluorescent dye efflux experiments. Drug-sensitive K562_S lines did not efflux calcein acetoxymethyl (calcein AM), whereas drug-resistant K562_R lines exhibited robust drug efflux (Figure 1, H and I). Efflux was completely reversed by either verapamil (a nonspecific ABC transporter substrate) or tariquidar (a highly specific inhibitor of ABCB1) (5). This confirmed that all drug efflux was due to ABCB1 (Figure 1J). No other ABC transporter gene was upregulated more than 2.5-fold in resistant cells (Supplemental Table 1). Thus even when chemoresistance is induced in separate lines, the mechanism of acquisition (i.e., *ABCB1* upregulation) is stereotypical.

Daunorubicin-resistant leukemia cells express a common integrated stress response–like gene signature. Unsupervised hierarchical clustering analysis, using cosine distance and average linkage, of 5953 expressed protein coding genes revealed that transcriptomes of sensitive and resistant lines differed substantially from one another (Figure 2A). Interestingly, principal component analysis revealed differences in the transcriptome of K562_S3 compared with both K562_S1 and K562_S2 (PC2), which were preserved as cells developed resistance (Figure 2B). PC1 accounted for 50% of the variance and defined the transition from sensitive to resistant in each case. Differential gene expression analysis identified 223 and 154 genes as significantly upregulated or downregulated, respectively (t test, P < 0.01, fold change >2 or <0.5) (Figure 2C and Supplemental Table 2). Among the upregulated gene set there was significant enrichment for Gene Ontology Biological Process terms, reflecting cellular stress including "response to endoplasmic reticulum stress" and "endoplasmic reticulum unfolded protein response"; among the downregulated gene set there was enrichment for "rRNA processing" and "mRNA splicing, via spliceosome" (9). Gene set enrichment analysis (GSEA) revealed that, of the 2414 curated gene sets from the Molecular Signatures Database tested (version 6.2) (10), those reflecting the response of HL-60 promyelocytic leukemia cells to the aminopeptidase inhibitor tosedostat (11), and arterial endothelial cells to hypoxia (12), were the most significantly enriched among both up- and downregulated genes in daunorubicin-resistant versus sensitive K562 cells (Figure 2D and Supplemental Figure 1A). Tosedostat is an aminopeptidase inhibitor that induces intracellular amino acid deprivation and consequent activation of the integrated stress response (ISR). Likewise, hypoxia activates the ISR by impairing disulfide bond formation, causing protein misfolding and endoplasmic reticulum stress (13).

To identify candidate regulators of high-level *ABCB1* transcription, and more generally the associated ISR-like transcriptional program, we identified transcription factor genes upregulated in resistant versus sensitive cells (Table 1 and Supplemental Tables 1 and 2). The most highly expressed was *ATF4*. Others included ATF4-bound transcriptional targets such as *ATF3*, *XBP1*, and *CEBPB* (Supplemental Table 2) or ATF4 binding partners including *JUN*, *JUNB*, *CEBPB*, *CEBPG*, *DDIT3*, and *ATF3* (14, 15). Consistent with ATF4 being a core driver of the upregulated ISR-like transcriptional program, GSEA demonstrated highly significant enrichment for ATF4 target genes among genes upregulated in daunorubicin-resistant versus sensitive K562 cells (Figure 2E). In this analysis, ATF4 target genes were those identified as genes closest to the strongest 500 ATF4 binding peaks identified by ChIP-Seq in K562 cells (Supplemental Table 2 and refs. 14, 16). Similar analyses using sets of genes located closest to the 500 strongest CEBPB, CEBPG, ATF3, JUN, or JUNB binding peaks in K562 cells also revealed significant enrichment in daunorubicin-resistant K562 cells (Supplemental Table 2 and refs. 14, 16). Notably, however, enrichment scores were lower than for the analysis using ATF4 target genes (Supplemental Figure 1B). All together these data demonstrate that the acquisition of an ABCB1-dependent daunorubicin-resistant cellular state in myeloid leukemia cells is associated with sustained upregulation of an ISR-like transcriptional program, with the transcription factor ATF4 at its core.

Expression of ABCB1 is regulated by a stress-responsive enhancer. Despite its clinical significance as a critical regulator of chemoresistance, knowledge of the transcriptional control of *ABCB1* is incomplete. Constitutive expression of its promoter requires motifs within 250 bp of the transcription start site that facilitate binding of nuclear factor-Y and SP1, and promoter binding sites for EGR1, WT1, HIF1A, CEBPB, FOXO factors, and TCF7 have been reported (17). TP53 may repress or activate the *ABCB1* promoter depending on whether it is wild-type or mutant; promoter DNA methylation represses *ABCB1* expression; and genetic translocations may activate *ABCB1* expression through juxtaposition

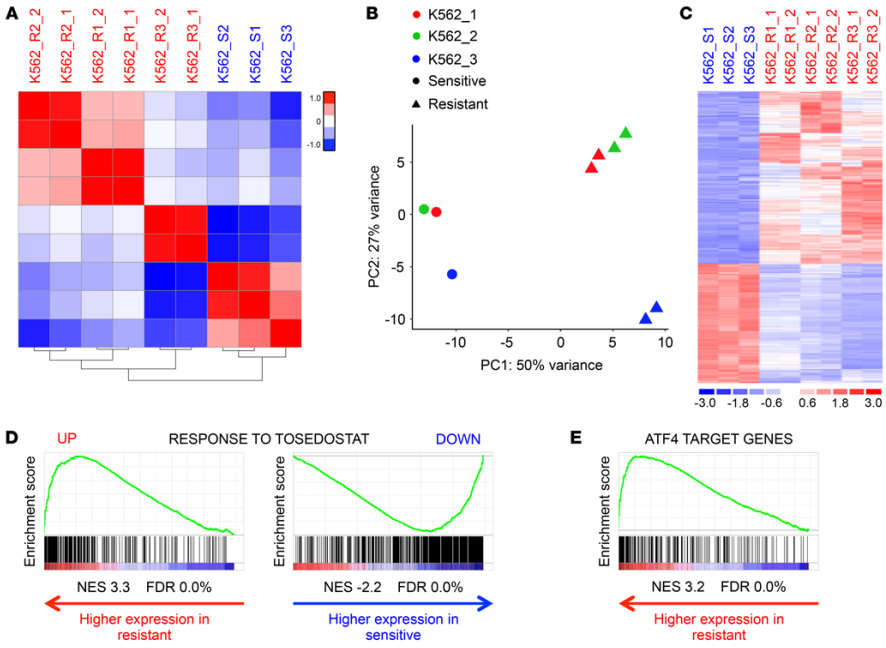


Figure 2. Daunorubicin-resistant leukemia cells express a common ISR-like gene signature. (**A**) Similarity matrix and hierarchical clustering of samples by differential gene expression. (**B**) Principal component (PC) analysis of gene expression from all sensitive and resistant cell lines. (**C**) Heatmap shows differentially expressed genes (223 upregulated and 154 genes downregulated; *t* test, $P < 0.01$, fold change >2 or <0.5). (**D** and **E**) Gene set enrichment analysis plots.

of the native promoter to that of more active but unrelated genes (4). *ABCB1* expression may also be induced by stressful stimuli, and roles for the AP-1 transcription factor family and nuclear factor-κB have been suggested, but supporting data are indirect (4, 17). There is no knowledge as to whether *ABCB1* is regulated by enhancer elements, and, if so, which factors control these.

To identify candidate *ABCB1* enhancer elements, we performed ChIP-Seq for H3K27 acetylation, a histone modification that marks active enhancers (18), using sensitive (K562_S1–3) and resistant (K562_R1–3) lines. We searched a 2-Mb region centered on *ABCB1* for differentially acetylated regions in resistant versus sensitive lines; the great majority of *cis*-regulatory elements lie within 1 Mb of target genes (19). Consistent with the dramatic increase in transcription, there was strong promoter acetylation in drug-resistant lines, which was not observed in sensitive lines. In addition, we identified 4 acetylation peaks, designated E1–E4, in intronic sequences of *ABCB1* (E1–E3) or upstream of the promoter (E4) in resistant but not sensitive cell lines (Figure 3A).

Using H3K27Ac ChIP-Seq data from ENCODE (14), we also searched for candidate enhancer elements in normal liver and adrenal gland, the tissues with the highest constitutive levels of *ABCB1* expression (Figure 3A). The pattern was tissue-specific, although putative *ABCB1* enhancers from K562_R1–3 lines were acetylated in liver (E1 and E3) or adrenal gland (E1, E2, and E3). Normal human $CD34^+$ hematopoietic stem and progenitor cells (HSPCs) express intermediate levels of *ABCB1*, and H3K27 acetylation of E3 was observed (Figure 3A). Interestingly, E3 and four additional sites were marked by H3K4 monomethylation in $CD34^+$ HSPCs, a histone modification that marks poised as well as active enhancers.

To determine the nature of candidate regulatory element contacts, we next performed 4C sequencing in drug-resistant cells with a viewpoint centered on the *ABCB1* promoter. There were particularly strong interactions between E3 and E4 and the promoter, and lower-level interactions between E1 and E2 and the promoter (Figure 3A and Supplemental Figure 2A). Strong contact was also observed between 3 additional regions, termed C1, C2, and C3, and the promoter. C1 is H3K4-monomethylated and weakly H3K27-acetylated in $CD34^+$ HSPCs, and strongly acetylated in liver; C2 is H3K4-monomethylated in $CD34^+$ HSPCs; and C3 is acetylated in the adrenal gland. These observations suggest that C1–C3 may exhibit tissue-specific enhancer activity, although

Table 1. The most significantly upregulated transcription factors in resistant versus sensitive lines

Gene	FPKM	Fold change	*P* value
ATF4	178.0	2.7	5.9×10^{-31}
JUN	93.5	4.8	1.9×10^{-11}
JUNB	69.7	2.1	4.7×10^{-4}
XBP1	29.7	2.4	1.7×10^{-10}
KLF6	25.8	3.5	1.2×10^{-24}
DDIT3	25.4	4.8	4.6×10^{-25}
CEBPB	23.9	3.5	3.1×10^{-15}
KLF10	21.3	2.3	1.6×10^{-13}
FOSB	18.7	60.9	3.7×10^{-40}
CEBPG	15.7	3.9	4.8×10^{-22}
CSRNP1	15.6	5.0	7.8×10^{-64}
ATF3	6.1	12.7	3.2×10^{-61}

Genes are ranked by mean expression in resistant lines. FPKM, fragments per kilobase of transcript per million mapped reads.

the presence of constitutive contact with the promoter in K562_R cells may be explained by C1 being bound by CTCF and cohesin (Supplemental Figure 2B). The reason for contacts between C2, C3, and the promoter was not apparent. Thus, the *ABCB1* promoter exhibits a network of physical contacts with nearby enhancers in drug-resistant K562 leukemia cells.

To confirm that putative enhancers were functional, we next performed targeted silencing using a CRISPR-dCas9-KRAB system. We designed multiple sgRNAs for each region (Supplemental Figure 3, A–D) and screened them in K562_R3 cells, using loss of cell surface ABCB1 expression or increased calcein AM retention as a measure of activity (Supplemental Figure 3, E and F). The most active guides were then selected for use in all resistant cell lines. K562_R1–3 cells were dual-infected with pHR-SFFV-dCas9-BFP-KRAB and pLKO5.sgRNA.EFS.tRFP657, the latter expressing an sgRNA targeting an enhancer or the promoter, or a nontargeting control (Figure 3B). We used ChIP-Seq for H3K9 trimethylation to confirm that silencing was discrete and accurate: induced regions of heterochromatin ranged in size from 3 to 8 kb, were centered on the target sequence for each guide, and did not target the promoter, even where 4C-Seq had shown the enhancer region to be in close physical proximity (Figure 3C). Quantitative PCR and flow cytometry assessment of the effect of *ABCB1* promoter silencing revealed substantial repression of transcription (Figure 4, A and B). The enhancer silencing experiments revealed either modest or no significant contribution to ABCB1 expression from E1, E2, and E4. The most extensive reductions in expression of *ABCB1* transcripts and protein were observed following silencing of E3, demonstrating that this was the most active enhancer, consistent with its high level of H3K27 acetylation and close contact with the promoter.

Within E3 is a DNase I–hypersensitive site (Figure 4C and ref. 14). Motif analysis of the 30-bp sequence revealed consensus binding sites for several of the transcription factors upregulated in drug-resistant cells, including ATF4, JUN, and CEBPB (Tables 1 and 2 and Figure 4C). We used ENCODE ChIP-Seq data from unmanipulated K562 cells to characterize binding of those factors to each enhancer (Figure 4C and Supplemental Figure 4). Data sets were available for 6 of the 12 factors upregulated in resistant cells (Supplemental Table 3), all of which were bound to the E3 enhancer, suggesting it to be stress-responsive (Figure 4C). There was some modest ATF3 and ATF4 binding at E1 and adjacent to E4. Critically, binding of AP-1 transcription factors to the promoter was absent (Supplemental Figure 4). Interestingly, E2 exhibited binding of TAL1 and GATA2, key hematopoietic stem cell (HSC) transcription factors that are active in AML and associated with poor clinical outcome (20). To confirm the ENCODE data and to determine whether there was increased binding of stress-responsive transcription factors at E3 in drug-resistant cells, we performed ChIP PCR. We observed significant increases in the binding of ATF4, ATF3, CEBPB, JUND, and JUN to E3 in K562_R1 compared with K562_S1 cells (Figure 4D). These data together demonstrate that acquisition of daunorubicin resistance is associated with activation of a stress-responsive, AP-1–bound enhancer element in intron 4 of *ABCB1*.

Dynamic induction of ABCB1 by diverse cellular stressors. To explore further the relationship between cell stress and expression of *ABCB1*, but over a shorter time scale, we induced intracellular amino acid depletion using the aminopeptidase inhibitor tosedostat (11). Tosedostat is able to induce cellular stress in both sensitive and resistant cells because it is not an ABCB1 substrate subject to cellular extrusion in $ABCB1^{hi}$ cells (Supplemental Figure 5A). There was significant upregulation of *ABCB1* expression in all K562 lines after 48 hours, although the absolute level of increase was far greater in drug-resistant lines (Figure 5A). Activation of the ISR upregulates ATF4 through a translational mechanism (15), so it was unsurprising that changes in *ATF4* transcript levels were modest (Supplemental Figure 5B). Instead, as a surrogate measure of ATF4 activity, we quantified expression of 3 genes that are known direct targets of ATF4: *DDIT3*, *DDIT4*, and *CEBPB* (21). Expression of all three was robustly induced by tosedostat, again with the absolute level of increase being greater in drug-resistant lines (Figure 5A). Tosedostat also induced expression of the AP-1 transcription factor *JUN* in all lines (Supplemental Figure 5C). Similar observations were made following treatment of unmanipulated early-passage K562 cells with alternate stressors: thapsigargin, which activates the ISR through blockade of the endoplasmic reticulum Ca^{2+} ATPase (Supplemental Figure 5D and ref. 22), and high-density culture (cell density of $>10^6$/mL for 48 hours; Supplemental Figure 5E). Thus, diverse cellular stressors induce dynamic upregulation of *ABCB1* and other direct targets of ATF4.

Exposure of sensitive and resistant K562 lines to 100 nM and 500 nM daunorubicin, respectively, for 72 hours also induced *ABCB1* (Figure 5B). As for tosedostat, the greatest absolute levels of increase were observed in drug-resistant lines, and they correlated with significant increases in *DDIT3*, *DDIT4*, and *CEBPB* (Figure 5B). By contrast with tosedostat, the fold-change increases in *DDIT3*, *DDIT4*, and *CEBPB* induced by 500 nM daunorubicin were lower, and increased expression of *ATF4* was not observed, suggesting that daunorubicin may be a somewhat less efficient activator of the ISR pathway (Supplemental Figure 5F).

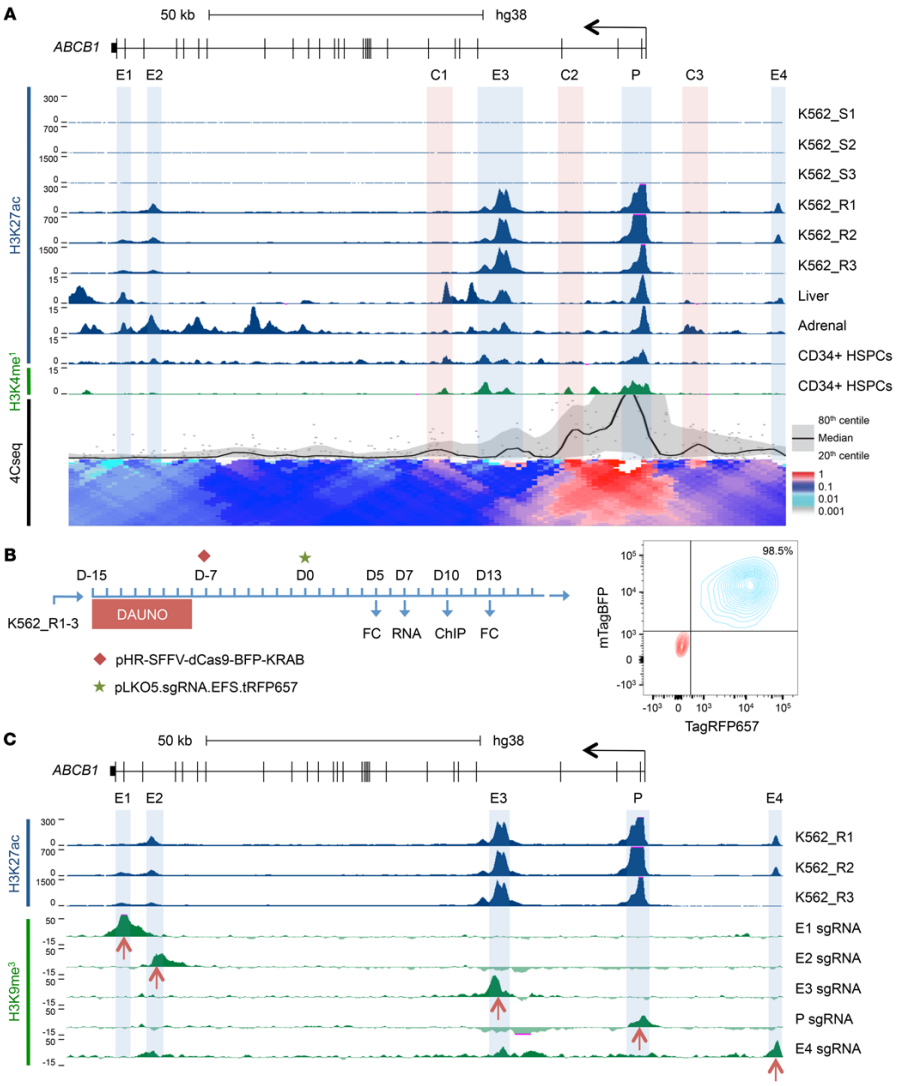


Figure 3. Regulatory element landscape of *ABCB1*. (**A**) ChIP-Seq tracks for H3K27Ac and H3K4me1 surrounding *ABCB1* (chr7:87,495,508–87,626,404; hg38) in the indicated human cells and tissues, including CD34$^+$ hematopoietic stem and progenitor cells. Putative enhancers (E1–E4) are highlighted in blue. Lower track shows a local contact profile generated from 4C sequencing of K562_R1 using a viewpoint centered on the *ABCB1* promoter. Regions of contact that do not contain an active enhancer in K562_R1–3 are highlighted in red (C1–C3). (**B**) Experimental outline (left); and representative flow cytometry plot (right) showing double-positive population (blue; K562_R1 BFP$^+$RFP$^+$) and negative control population (red). FC, flow cytometry. (**C**) H3K9me3 ChIP-Seq tracks for each sgRNA. Signal from empty vector was subtracted to show only histone methylation resulting from presence of the guide. Red arrows indicate the position of the target sequence. H3K27Ac ChIP-Seq tracks from **A** are included for reference.

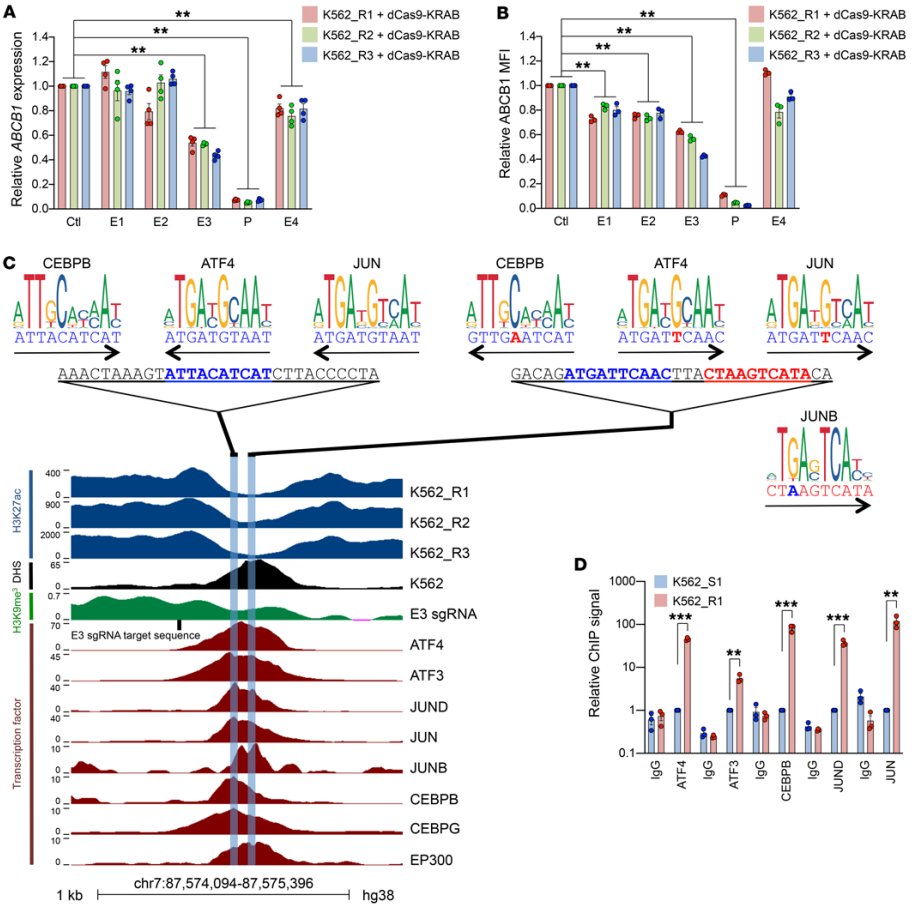


Figure 4. Expression of *ABCB1* is regulated by a stress-responsive enhancer. (**A**) Mean ± SEM *ABCB1* expression by quantitative PCR in dCas9-KRAB$^+$ resistant cell lines (K562_R1–3) expressing sgRNAs targeting the indicated putative enhancer elements (E1–E4) or the promoter (P), relative to control cells expressing a nontargeting guide (Ctl). $^{**}P < 0.01$ by 1-way ANOVA with Tukey's post hoc test ($n = 4$). (**B**) As for **A**, but with mean ± SEM ABCB1 median fluorescence intensity (MFI) by flow cytometry. $^{**}P < 0.01$ by 1-way ANOVA with Tukey's post hoc test ($n = 3$). (**C**) ChIP-Seq tracks for H3K27Ac, H3K9me3 (our data), and the indicated transcription factors in K562 cells (ENCODE); and DNase-Seq (ENCODE) at the E3 enhancer. Sites of AP-1 binding motifs are indicated. (**D**) Mean ± SEM relative ChIP PCR signal for the indicated transcription factors for K562_R1 and K562_S1 using primers for the E3 enhancer. $^{**}P < 0.01$, $^{***}P < 0.001$ by unpaired *t* test ($n = 3$).

Daunorubicin also induced expression of *JUN* in resistant lines (Supplemental Figure 5G). The differences in response to both tosedostat and daunorubicin of drug-sensitive versus drug-resistant K562 lines are in keeping with the observed enhancer remodeling at the *ABCB1* locus induced by prolonged (>100 days) daunorubicin exposure.

These data demonstrate that brief daunorubicin exposure also induces ATF4 target gene expression, including *ABCB1*. Importantly, ABCB1 expression in daunorubicin-resistant K562 lines was dynamic and diminished over time if cells were not continuously exposed to drug (Figure 5C). Loss of ABCB1 expression was more pronounced when cells were propagated at low densi-

Table 2. Transcription factor binding motifs identified in a 30-bp sequence taken from the H3K27 acetylation nadir at the center of enhancer E3

Motif	Score	Strand
JUN	0.99	–
JUND(var.2)	0.98	–
CEBPB	0.95	+
ATF4	0.95	–
HLF	0.94	+
GATA2	0.94	–
DBP	0.93	–
FOXC1	0.93	+
FOSL2::JUN(var.2)	0.92	–
CEBPA	0.92	+

ty (<200,000/mL), emphasizing the need for rigorous control of cell culture conditions when performing stress experiments. Even modest elevations of cell density (>200,000/mL) were sufficient to cause significant increases in ABCB1 in comparison with low-density controls (Figure 5, C and D). Re-exposure of K562_R1–3 cells to daunorubicin (100 or 500 nM for 7 days) following a 24-day daunorubicin-free period of culture led to a dose-dependent reestablishment of *ABCB1* expression, an effect that was dependent on the activity of the ATF4-bound E3 enhancer, since it was attenuated when E3 was silenced with dCas9-KRAB (Figure 5E). All together these data demonstrate that expression of the daunorubicin drug export pump ABCB1 is dynamically regulated in leukemia cells though the ATF4-bound E3 enhancer.

Daunorubicin activates a stress-responsive ABCB1 enhancer in primary AML cells. We next examined *ABCB1* enhancer accessibility and usage in primary AML. We identified cases of relapsed or refractory AML from Manchester Cancer Research Centre's Tissue Biobank with high *ABCB1* expression by quantitative PCR (Figure 6A and Supplemental Table 4) and performed ChIP sequencing for H3K27Ac in high-expressing cases where sufficient cryopreserved bulk blast cells were available (red bars in Figure 6A). We also made use of a recently published DNase-Seq primary AML data set (23). Quantitative PCR analysis revealed that *ATF4* expression correlated significantly with *ABCB1* (Figure 6B; $r = 0.53$, $P = 0.005$). Considering the genomic region encompassing the coding sequence of *ABCB1* and sequences 20 kb upstream and 10 kb downstream, we identified 5 DNase I–hypersensitive sites (DHSs) (in addition to the DHS observed at the promoter) in multiple cases of AML (Figure 6, C and D). These included E1 and E3 (accessible in 12 of 36 and 13 of 36 primary AML cases, respectively), which became strongly acetylated in drug-resistant K562 cells, and the CTCF binding site C1 (accessible in 32 of 36 primary AML cases) (Figure 3A and Figure 6, C and D). Regions E2 and E4 (Figure 6D and data not shown) were not accessible. Two additional sites (A and B; accessible in 13 of 36 and 14 of 36 cases, respectively), which were not acetylated in drug-resistant K562 cells, were also DNase I–hypersensitive. DHS site B was adjacent to other confirmed *ABCB1* enhancers (E1 and E2; Figure 6D) and was acetylated in ABCB1-expressing adrenal tissue. Importantly, this site also contains binding motifs for ATF4, JUN, and CEBPB, suggesting that it too may serve as a stress-responsive enhancer (Figure 6E). Across the totality of primary AML samples profiled by Assi et al. (23), 6 of 36 exhibited DHS at both B and E3 sites, 8 of 36 at B only, 7 of 36 at E3 only, and 15 of 36 at neither B nor E3. Together these data show that stress-responsive regulatory elements are accessible in bulk primary AML cells. Our own H3K27Ac ChIP-Seq analyses further demonstrated that *ABCB1*-expressing samples exhibited peaks of acetylation surrounding these sites: of the 10 samples analyzed, 4 had discernible H3K27Ac peaks at B only, 2 at E3 only, 1 at both, and 3 at neither. In 1 case there was a peak of acetylation at A. Thus, in a substantial proportion of primary AML cases, stress-responsive *ABCB1* regulatory elements are accessible and active.

To determine whether primary AML cells respond to stress in a similar manner to drug-resistant K562 cells, we exposed fresh bulk primary AML blasts from bone marrow or blood (Supplemental Table 4) to daunorubicin (10 nM, 100 nM, and 1000 nM) for 18 hours. We observed dose-dependent induction of ATF4 target genes *ABCB1*, *DDIT3*, *DDIT4*, *CEBPB*, and *JUN*, although, as before, changes in *ATF4* transcripts were modest or absent (Figure 6F and Supplemental Figure 6A). It was of note that this response was not observed when similar analyses were performed using cryopreserved AML samples following a freeze-thaw cycle (Supplemental Figure 6B). Vehicle-treated freeze-thawed samples exhibited substantially higher levels of *ATF4* and *DDIT3* compared with vehicle-treated fresh samples (Supplemental Figure 7A), suggesting that the freeze-thaw process activates cellular stress pathways consequently obscuring the response to daunorubicin exposure. Two additional fresh primary AML samples were treated with 1000 nM daunorubicin or vehicle for 18 hours (Supplemental Figure 7B) and subjected to ChIP PCR for H3K27Ac surrounding E3 (Figure 6G). Significant increases in acetylation were observed, confirming acute stress–induced regulation of E3. By contrast, daunorubicin had no effect on the acetylation of the CTCF binding site C1 (Figure 6G).

We also assessed the effect of daunorubicin exposure on 2 other ABC transporter genes previously associated with chemoresistance in AML (3). *ABCG2* expression increased significantly in 4 of 5 fresh samples following daunorubicin exposure, but absolute levels of expression were very low as judged by cycle threshold (Supplemental Figure 7C). Induction was not observed in 3 of 4 freeze-thawed samples (Supplemental Figure 7D). *ABCC1* was more highly expressed and its expression increased significantly in all fresh samples (Supplemental Figure 7E), with responses again smaller or absent in freeze-thawed samples (Supplemental Figure 7F). The change in expression of these efflux pumps in response to daunorubicin mirrors that of *ABCB1*, suggesting regulation by similar mechanisms. Interestingly, ENCODE data in unmodified K562 cells shows intronic binding of CEBPB, CEBPG, JUND, JUN, ATF4, and ATF3 within *ABCC1*, suggesting that this efflux pump may also be stress-responsive (Supplemental Figure 8 and ref. 14).

ABCB1 is also expressed in normal HSCs and downregulated during differentiation. Indeed, extrusion of rhodamine 123 or Hoechst 33342 by ABCB1 has been used to identify long-term repopulating HSCs (24). HSCs also make use of the ISR and

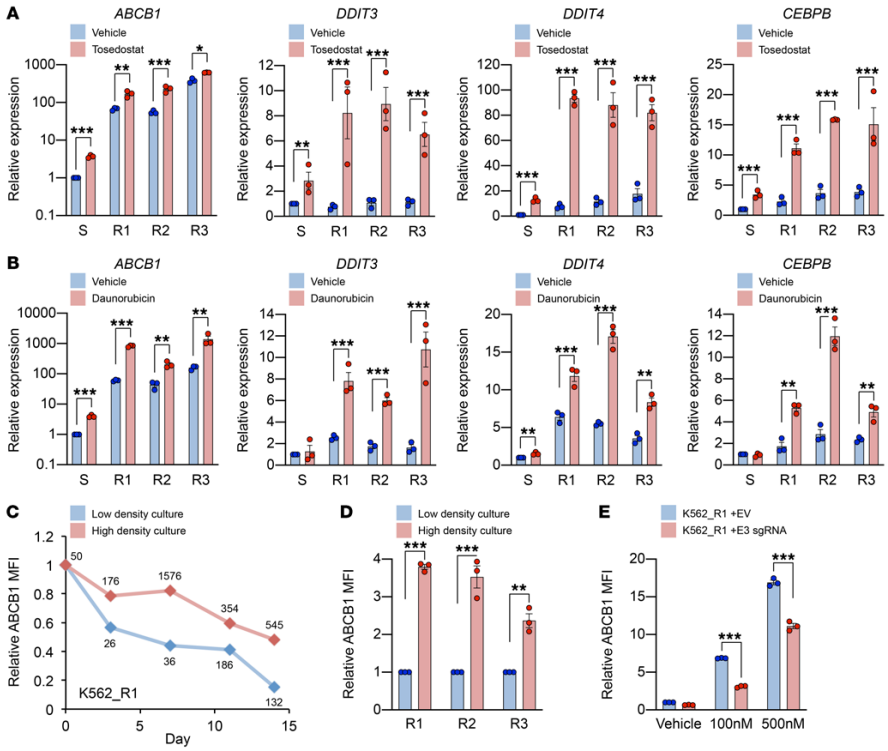


Figure 5. Dynamic induction of *ABCB1* by diverse cellular stressors. (**A** and **B**) Mean ± SEM expression of the indicated genes by quantitative PCR relative to a fresh aliquot of unmanipulated drug-sensitive K562 cells ($n = 3$) following exposure to tosedostat (50 μM) for 48 hours (**A**) or daunorubicin (100 nM for sensitive or 500 nM for resistant lines) for 72 hours (**B**). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by unpaired *t* test. (**C**) ABCB1 mean fluorescence intensity (MFI) over time in K562_R1 cells maintained in high- or low-density culture. Numbers indicate cell density (K/mL). (**D**) ABCB1 MFI in K562_R1–3 following 14 days of high- or low-density culture ($n = 3$). $^{**}P < 0.01$, $^{***}P < 0.001$ by unpaired *t* test. (**E**) Mean ± SEM ABCB1 MFI in dCas9-KRAB$^+$ resistant cells (K562_R1) expressing either an E3-targeting sgRNA or a nontargeting sgRNA (EV) following 7 days of exposure to the indicated dose of daunorubicin ($n = 3$). $^{***}P < 0.001$ by unpaired *t* test.

ATF4 to protect against homeostatic cellular stress and to preserve the integrity of the stem cell pool (25). Given this account of an adaptive, prosurvival ISR signature in HSCs, we examined the expression of K562 resistance–associated transcription factor genes across normal hematopoiesis (26). As previously described, *ABCB1* expression diminished as cells differentiated, with the highest expression seen in early HSCs (Supplemental Figure 9A). *ATF4* expression followed a similar pattern and was highly correlated with *ABCB1* (Supplemental Figure 9, A–C; $r = 0.91$, $P < 0.001$). Given the predominantly translational regulation of ATF4, we also studied its transcriptional target *DDIT3*; changes in expression correlated even more closely with that of *ABCB1* (Supplemental Figure 9, A–C; $r = 0.95$, $P < 0.001$). Indeed, all of the transcription factorsthat were upregulated in daunorubicin-resistant K562 cells were significantly correlated with *ABCB1* expression across normal hematopoiesis (Supplemental Figure 9B). Reflecting this observation, GSEA revealed highly significant enrichment of expression of the 223 genes upregulated in drug-resistant versus sensitive K562 cells in normal hematopoietic stem cell/multipotent progenitor versus downstream myeloid progenitor populations (Supplemental Table 2 and Supplemental Figure 9D), suggesting a common gene expression program driven by adaptive prosurvival stress signaling. Analysis of H3K27Ac and H3K4me1 ChIP-Seq and DNase-Seq data (ENCODE) from normal CD34$^+$ HSPCs confirmed that regulatory elements A, E1, E3, and, to a lesser extent, B were accessible in CD34$^+$ HSPCs and marked by H3K4 monomethylation and, in the case of E3, by H3K27 acetylation (Supplemental Figure 9E).

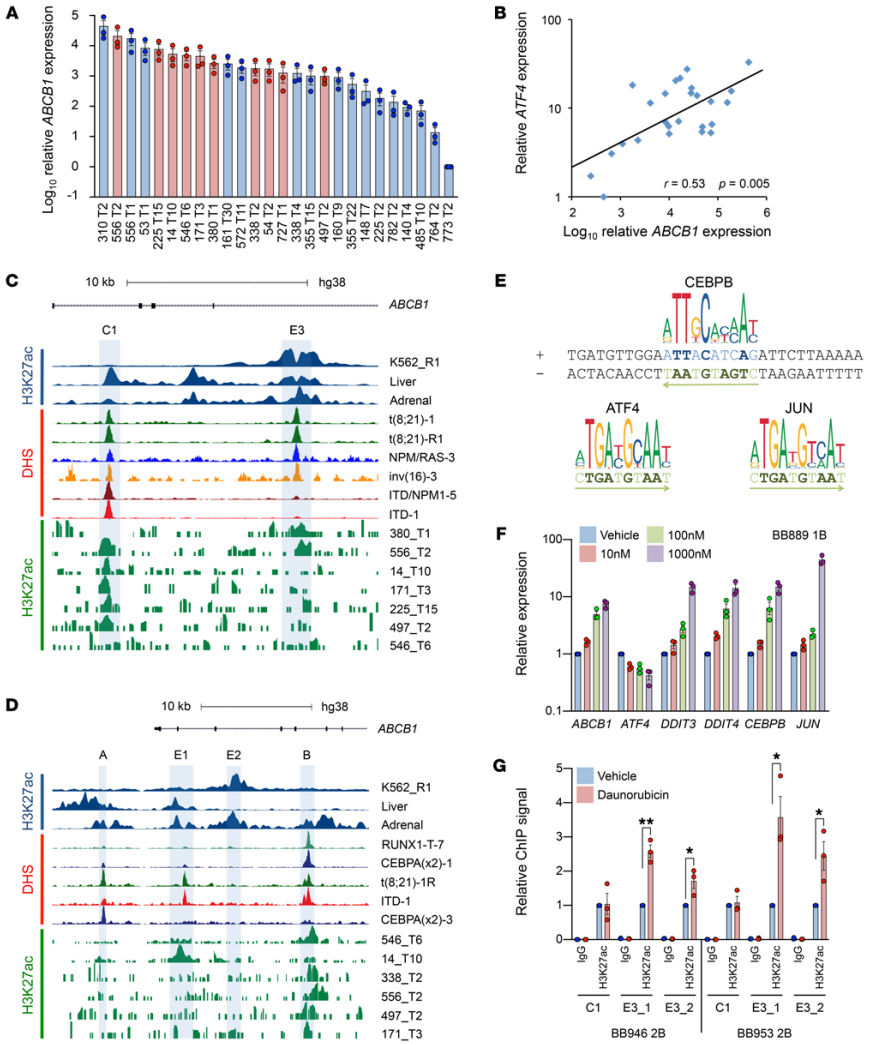


Figure 6. Daunorubicin activates a stress-responsive *ABCB1* enhancer in primary AML cells. (**A**) *ABCB1* expression by quantitative PCR in primary AML samples ($n = 3$). H3K27Ac ChIP-Seq was performed on the samples highlighted in red. (**B**) Correlation of *ATF4* and *ABCB1* expression; r = Pearson product-moment correlation coefficient. (**C** and **D**) ChIP-Seq (our data) and DNase-Seq tracks (23) surrounding C1 and E3 (chr7:87,561,371–87,579,610; hg38) (**C**) and A, B, E1, and E2 (chr7:87,494,187–87,522,854; hg38) (**D**) from the indicated cell lines, human tissues (ENCODE), and primary AML samples. (**E**) Transcription factor binding motifs identified at the center of site B. (**F**) Mean ± SEM relative expression of the indicated genes following exposure of fresh primary AML blast cells to the indicated doses of daunorubicin for 18 hours ($n = 3$). BB numbers indicate Biobank identifier. (**G**) Mean ± SEM relative ChIP PCR signal for H3K27Ac using fresh primary AML blast cells exposed to 1000 nM daunorubicin or vehicle for 18 hours ($n = 3$). Data from 2 patients (BB946 and BB953) are shown. PCR was performed using 2 primer sets for the E3 enhancer (E3_1 and E3_2) and 1 for the CTCF binding site C1. $^*P < 0.05$, $^{**}P < 0.01$ by unpaired t test. BB numbers indicate Biobank identifier.

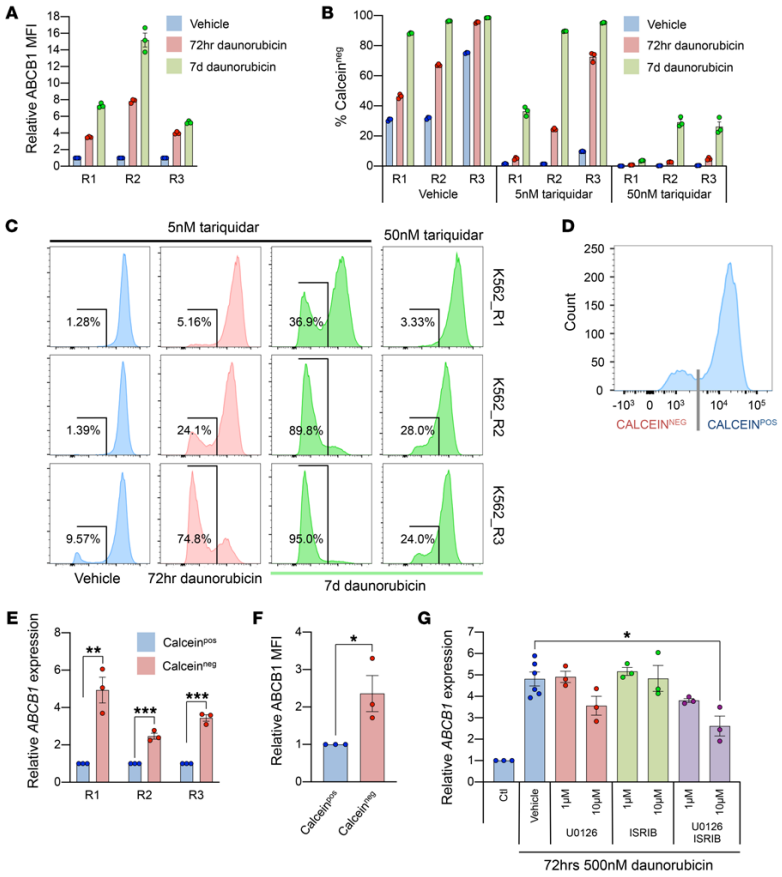


Figure 7. Activation of an ISR-like response facilitates escape from ABCB1 inhibition. (**A**) Mean ± SEM relative ABCB1 median fluorescence intensity (MFI) in K562_R1–3 following exposure to 500 nM daunorubicin or vehicle for 72 hours or 7 days (n = 3). (**B**) Proportion of cells that are calcein AM$^-$ following exposure of K562_R1–3 to the indicated conditions as determined by flow cytometry (n = 3). (**C**) As for **B**, but showing individual flow histograms for each of the indicated conditions. (**D**) Experimental outline depicting FACS of calcein AM$^-$ and calcein AM$^+$ populations. (**E**) Mean ± SEM *ABCB1* expression by quantitative PCR of calcein AM$^-$ and calcein AM$^+$ populations (n = 3). $^{**}P < 0.01$, $^{***}P < 0.001$ by unpaired t test. (**F**) As for **E**, but with relative ABCB1 MFI by flow cytometry (n = 3). (**G**) Mean ± SEM *ABCB1* expression by quantitative PCR in K562_R1 following exposure to 500 nM daunorubicin or vehicle for 72 hours with the indicated inhibitors (n = 3–6). $^*P < 0.05$ by 1-way ANOVA with Tukey's post hoc test (n = 3).

These data demonstrate the close link between expression of a stress-responsive genetic program and resistance to daunorubicin through upregulation of *ABCB1*; they further demonstrate that chemotherapy treatment with daunorubicin activates a stress-responsive enhancer and induces upregulation of a drug resistance mechanism in AML blast cells that may contribute to therapeutic failure and disease relapse.

Activation of an ISR-like response facilitates escape from ABCB1 inhibition. Pharmacologic inhibitors of ABCB1 have been tested in clinical trials as adjuncts to AML therapy, but without significant success (3). Trials of the third-generation inhibitor tariquidar used doses of 2 mg/kg (resulting in plasma concentrations of ~4 nM), based on maximal inhibition of rhodamine 123 efflux in CD56$^+$ NK cells, which exhibit relatively high, stable levels of ABCB1 expres-

sion (5). Our observation of dynamic, stress-responsive *ABCB1* expression raised a question of whether the dose of ABCB1 inhibitors used to inhibit steady-state cells might be ineffective under conditions of cellular stress.

We re-exposed drug-resistant K562 cells (lines R1–3) that had been cultured without daunorubicin for 24 days to 500 nM daunorubicin or vehicle for 72 hours and assessed the ability of 5 nM tariquidar to inhibit efflux of calcein AM. As expected, re-exposure of cells to daunorubicin further induced ABCB1 expression (Figure 7A). Concomitant treatment of cells with 5 nM tariquidar abolished calcein AM efflux in vehicle-treated cells, but in daunorubicin–re-exposed cells, where ABCB1 had been further induced, in each case a population of cells was observed that failed to retain calcein AM. This demonstrates continued activity of ABCB1 drug efflux in a subpopulation of cells despite exposure of the cell population to levels of tariquidar approximating those achieved in clinical trials (Figure 7, B and C). This effect became yet more apparent when daunorubicin exposure was extended to 7 days but could be overcome by increasing of the concentration of tariquidar (Figure 7, A–C), indicating that the effect was due to differential expression of ABCB1. We confirmed this by flow-sorting tariquidar-treated K562_R1–3 cells into calcein AM^- and calcein AM^+ populations and evaluating ABCB1 expression (Figure 7, D–F). Similar observations were made when K562_R1–3 cells were exposed to tosedostat, demonstrating that this effect was not specific to daunorubicin and was likely consequent upon activation of an ISR-like program (Supplemental Figure 10, A–C). To identify an approach to overcome the phenomenon of daunorubicin- and stress-induced escape from ABCB1 inhibition, we evaluated stress pathway inhibitors. U0126 antagonizes AP-1 target gene transcription via inhibition of MEK1/2, and ISRIB (integrated stress response inhibitor) antagonizes the consequences of eIF2a phosphorylation through a mechanism involving binding of eIF2B to restore normal translation of factors including ATF4 (27, 28). Treatment with 10 μM of U0126 was suggestive of reduced *ABCB1* induction in K562_R1 exposed to 500 nM daunorubicin for 72 hours compared with vehicle (Figure 7G). While ISRIB alone did not have an effect, combined treatment with U0126 led to significant dose-dependent suppression of *ABCB1* induction (Figure 7G).

Thus daunorubicin- and stress-induced acute induction of *ABCB1* can overcome pharmacologic inhibition of ABCB1, leading to leukemia cell survival, and this can, at least in part, be mitigated by concomitant treatment of cells with inhibitors of stress signaling.

Discussion

Efflux of chemotherapeutic agents by ABCB1 is an important cause of treatment failure in human cancer. High expression levels may be an intrinsic feature of the cell type, or due to promoter translocations between *ABCB1* and genes with strong constitutive expression, such as those found in patients with breast or ovarian cancer who relapse after prior therapy (29, 30). We found that primary AML cells displayed dynamic expression of *ABCB1*, suggesting physiological regulation rather than control by constitutively active captured promoters. Our analyses reveal the network of enhancers that controls intrinsic expression of *ABCB1* in human leukemia cells; the gene is a direct target of the transcription factor ATF4, which is activated through a chemotherapy-induced cellular stress response.

Drug resistance in cancer can arise through multiple mechanisms, including genetic events and stochastic transcriptional changes in rare cells (31). We observed significant transcriptional differences in one of our cell lines (K562_3) that persisted as the line acquired resistance. These differences are likely due to the genetic and transcriptional divergence that frequently accompanies the propagation of cancer cell lines (32). Indeed, this was our primary motivation for creating 3 independent resistant cell lines, and it is significant that in spite of these differences all lines developed daunorubicin resistance through induction of *ABCB1*.

For cancer cells to survive they must adapt to stressful stimuli. The ISR is activated by endoplasmic reticulum stress, hypoxia, amino acid deprivation, and oxidative stress, common consequences of uncontrolled proliferation and outgrowth of the vascular supply. The transcription factor ATF4 is a critical effector of the ISR and is highly expressed in many cancers as a result of extrinsic stress or direct activation by constitutive oncogene expression (33). During stress it is efficiently translated as a result of eIF2α phosphorylation, permitting heterodimer formation with transcription factors such as JUN, FOS, and CEBPB, and binding of transcriptional targets (15, 34). We found that adaptation of leukemic cells to prolonged daunorubicin exposure (>100 days) involved expression of an ATF4-centered, ISR-like transcriptional program that led to sustained upregulation of *ABCB1*. ATF4 and its interaction partners bind a stress-responsive enhancer in intron 4, suggesting that this element responds specifically to stress signaling with an adaptive, prosurvival output. The dramatic differences in histone acetylation surrounding, in particular, enhancer E3 are indicative of enhancer remodeling, although the molecular mechanisms underlying this process remain unclear. It is also unclear how the various active enhancers cooperate to regulate transcriptional activity at the promoter. Uncovering these mechanisms would be of great interest, not least because the adaptive molecular changes surrounding E3 appear to serve as the basis for the epigenetic memory of prior cellular stress, at least as far as expression of *ABCB1* is concerned.

Previous reports of *ABCB1* responses to cellular stress have been contradictory, demonstrating induction or repression, even after exposure to the same stressor (35). These conflicting results might be consistent with the pleiotropic function of ATF4, which is able to orchestrate adaptation and survival or apoptosis depending on cellular context and the severity of the insult. Indeed, downregulation of *ABCB1* appears to precede cell death, suggesting that the gene is negatively regulated by ISR signaling where apoptosis is the result (36).

Even in the era of targeted therapies, tumor bulk continues to predict treatment failure for many cancers (37), and the total white cell count in blood at presentation is strongly predictive of outcome in AML (38). Our observation that prolonged daunorubicin exposure elicited a transcriptional response that was shared by cells exposed to amino acid deprivation or hypoxia suggests that extrinsic stress applied experimentally has similar consequences to environmental stresses experienced by cancer cells in vivo. We speculate that protracted cellular stress primes stress-responsive *ABCB1* enhancers for both strong constitutive activity and augmented responses following exposure to additional stressors, such as chemotherapy, leading to increased drug efflux: chemo-

therapy may induce its own chemoresistance mechanism. While steady-state expression of ABC transporters is seen only in a subset of resistant AML cases, we found that dynamic upregulation of *ABCB1* following daunorubicin exposure occurred in all fresh primary samples tested (39). Rapid adaptation to therapy may therefore represent a more common mechanism of resistance than previously appreciated, especially considering that the biopsies that provide primary material for research are seldom taken during treatment.

The importance of ABCB1 in hematopoiesis is well established. Its expression is a hallmark of hematopoietic stem cells (HSCs) and accounts for their reduced staining with Hoechst 33342. HSCs display high levels of prosurvival ISR activity, which we found to correlate with the expression of *ABCB1* and the transcription factor combination expressed in our resistant cells (25). Leukemic stem cells (LSCs) can also be identified by their capacity for ABCB1-mediated dye efflux (40, 41); LSCs occupy hypoxic bone marrow niches that may contribute to *ABCB1* expression and chemoresistance (42). Given the abundance of evidence supporting a role for ABCB1 in drug resistance, the lack of success of clinical trials of ABCB1 inhibitors is puzzling. A potential explanation is suggested by our observation that exposure of leukemia cells with primed *ABCB1* enhancers to daunorubicin leads to rapid and substantial upregulation of ABCB1, with escape of a leukemia cell subpopulation from the effects of drug efflux pump inhibition.

The emerging role of the ISR as driver of adaptation and survival in cancer has led to interest in pharmacologic manipulation of this pathway. We found that stress-induced upregulation of *ABCB1* could be mitigated by use of the MEK inhibitor U0126 alone or in combination with ISRIB, suggesting a possible therapeutic strategy for testing in early-phase trials. Given that the output of the ISR is dependent on the precise state of each cell, there is a risk that a therapy designed to promote apoptosis may inadvertently drive adaptation and survival in a subset of cells. The precise function of ABCB1 as an effector of adaptive stress signaling also needs to be defined. We also found that *ABCC1* expression was induced by daunorubicin exposure in fresh primary AML cells and that intronic regions likely bind the same transcription factors that drive *ABCB1* expression. The evolution of the ABC superfamily has involved gene duplication, and members presumably share previously unrecognized regulatory features (43). ABC transporters are also highly evolutionarily conserved, contributing to both nutrient import and multidrug resistance in bacteria (44). These pumps efflux a wide range of endogenous compounds and have been shown to influence paracrine signaling, membrane lipid composition, and cellular redox state (45). It is therefore likely that expression of *ABCB1* has physiological effects that mitigate certain forms of stress. In fact, the removal of chemotherapy from leukemia cells may simply be an unfortunate by-product of its primary function.

In summary, we show that cellular stress can drive chemoresistance through *ABCB1* enhancers, providing an explanation for the failure of clinical trials of ABCB1 inhibitors and suggesting an approach to overcome drug resistance. This study has implications for the study of resistance mechanisms more generally, as these data demonstrate that the behavior of cancer cells is highly dependent on cell context and environmental factors. Studies of cells in steady state alone may be potentially misleading.

Methods

Cell culture. K562 cells were from DSMZ and were cultured in RPMI 1640 medium supplemented with 2 mM L-glutamine (Life Technologies) and 10% FBS (Sigma-Aldrich). While under drug selection, cells were counted and replated every third day. Cell lines were confirmed mycoplasma-free and authenticated by short tandem repeat DNA profiling.

Primary AML samples. Primary human AML samples were from the Manchester Cancer Research Centre Tissue Biobank (approved by the South Manchester Research Ethics Committee). Their use was authorized by the Tissue Biobank's scientific subcommittee, with the informed consent of donors. For ChIP, selected samples were thawed or collected fresh and immediately cross-linked. For treatment with daunorubicin, fresh leukemic blast cells were obtained by density gradient centrifugation of bone marrow or peripheral blood. Cells were treated in α-MEM medium supplemented with 12.5% heat-inactivated FBS, 12.5% heat-inactivated horse serum, 2 mM L-glutamine, 57.2 μM β-mercaptoethanol, 1 μM hydrocortisone (Sigma-Aldrich), and IL-3, G-CSF, and TPO (all at 20 ng/mL; PeproTech).

Reagents. Daunorubicin, verapamil, and ISRIB were from Sigma-Aldrich; tosedostat and tariquidar were from Generon; and thapsigargin and U0126 were from Merck. Compounds were resuspended in DMSO (tosedostat, tariquidar, thapsigargin, ISRIB, U0126) or ddH20 (verapamil and daunorubicin), aliquoted, and stored at –20°C. Final DMSO concentration was less than 0.5% in all experiments.

Cell viability assays. 5×10^3 cells were plated in each well of a 96-well plate with media containing a serial dilution of daunorubicin. Plates were incubated for 72 hours at 37°C. Twenty microliters of 140 μg/mL resazurin (Sigma-Aldrich) was added to each well. Plates were then incubated for a further 4 hours and read using a POLARstar Omega plate reader (BMG Labtech).

RNA sequencing and data analysis. Total RNA was extracted from 5×10^5 cells using QIAshredder spin columns and an RNeasy Plus Micro kit (Qiagen). Before sequencing, RNA integrity was checked using a 2100 Bioanalyzer (Agilent Technologies). Poly(A) libraries were prepared using a SureSelect Poly(A) kit (Agilent Technologies); samples were then barcoded and pooled. Sequencing was performed using a NextSeq desktop sequencing system (Illumina). A single run (400 million reads) of 75 bp paired-end sequencing produced a mean of 45.7 million reads per sample. Reads were aligned to the human genome (hg38) using STAR version 2.4.2a (46). DEseq2 was used to perform differential gene expression analysis and calculate FPKM (fragments per kilobase of transcript per million mapped reads) values for each transcript (47). Hierarchical clustering, similarity matrix, and heatmap visualizations were created using clustergrammer (48). Principal component analysis was performed using ggplot2 (49). Gene set enrichment analyses (10) were performed with GSEA version 2.0 software (http://www.broad.mit.edu/gsea) using signal-to-noise for gene ranking and 1000 data permutations. To identify ATF4, CEBPB, CEBPG, ATF3, JUN, or JUNB target genes, K562 ChIP-Seq data sets were downloaded from the ENCODE Consortium (14). The strongest peaks by pileup value were identified by Model-based Analysis of ChIP-Seq version 2 (MACS2) using default parameters (Supplemental Table 2 and ref. 50). Gene expression and *ABCB1* correlations in sorted cord blood populations were analyzed using data from Laurenti et al. (26). Raw data files for RNA sequencing are available at the Gene Expression Omnibus with the accession number GSE131825.

Quantitative PCR. cDNA was generated using a High Capacity Reverse Transcription kit (Applied Biosystems). Quantitative PCR reactions were performed in MicroAmp optical 384-well reaction plates and analyzed using a QuantStudio 5 PCR system (Applied Biosystems). Reactions were performed in triplicate or quadruplicate and included primers for β-actin (ACTB) as a housekeeping gene. Primers were designed using the Universal Probe Library (UPL) Assay Design Center (Roche) and purchased from Integrated DNA Technologies. Raw fluorescence data were converted to Ct values using the Thermo Fisher Cloud facility and normalized to ACTB. For primer sequences and associated probes see Supplemental Table 5.

FACS, flow cytometry, and assessment of calcein-AM retention. Flow cytometry was performed using an LSR II flow cytometer (BD Biosciences). A FACSAria II (BD Biosciences) was used for cell sorting experiments. FlowJo version 10.1 (BD Biosciences) was used to analyze data. To assess calcein AM retention, 5×10^5 cells were resuspended in PBS containing 10 nM freshly prepared calcein AM (BioLegend) with 40 μM verapamil, 5 or 50 nM tariquidar, or vehicle. Samples were incubated for 20 minutes at 37°C, then resuspended in prewarmed culture medium and incubated for a further 10 minutes to ensure optimal retention. Calcein AM accumulation was assessed by flow cytometry. ABCB1 expression was assessed using CD243-PE or CD243-APC (clone UIC2, eBioscience), the latter being used when cells were treated with daunorubicin, which has similar excitation and emission spectra to PE.

ChIP and next-generation sequencing. Chromatin immunoprecipitation (ChIP) was performed using anti-H3K27Ac (ab4729) and anti-H3K9me3 (ab8898, Abcam). 10^8 cells were used for each precipitation using the method described by Lee et al. (51). Briefly, cells were cross-linked with 1% formaldehyde for 10 minutes at room temperature before the reaction was quenched with 0.125 M glycine. Cell pellets were washed twice with PBS and nuclear lysates sonicated for 6 cycles using a Bioruptor Pico (Diagenode). Antibody (10 μg) bound to 100 μL of magnetic beads (Dynabeads Protein G, Invitrogen) was added to each sample and immunoprecipitation performed overnight on a rotator at 4°C and 20 rpm. After 5 washes with RIPA buffer (50 mM HEPES [pH 7.6], 1 mM EDTA, 0.7% Na deoxycholate, 1% NP-40, 0.5 M LiCl), ChIP-bound fractions were extracted by incubation for 15 minutes at 65°C with elution buffer (50 mM Tris-HCl [pH 8], 10 mM EDTA, 1% SDS). Cross-linking was then reversed by incubation at 65°C for 6 hours. RNase A (1 mg/mL) and proteinase K (20 mg/mL) were added to eliminate RNA and protein from the samples. DNA was extracted using phenol/chloroform/isoamyl alcohol and precipitated by addition of 2 volumes of ice-cold 100% ethanol, glycogen (20 μg/μL), 200 mM NaCl and freezing at –80°C for at least 1 hour. Pellets were washed with 70% ethanol and eluted in 50 μL 10 mM Tris-HCl (pH 8.0).

Libraries were prepared for sequencing using a Microplex Library Preparation Kit (Diagenode). Fragments of 200–800 bp were selected using AMPure beads (Beckman Coulter) and quantified by quantitative PCR with a KAPA Library Quantification Kit (Kapa Biosystems). Sequencing was performed using a NextSeq desktop sequencing system (Illumina) with 75-bp, paired-end high output generating 40–65 million reads per sample. Reads were aligned to the human genome (hg38) using BWA-MEM version 0.7.15 (52). Read duplicates were removed using Picard version 2.1.0. Reads were further filtered using Bedtools version 2.25.0 to keep only paired reads that mapped to standard chromosomes and to remove reads with a mapping quality of less than 10. Reads mapped to blacklisted regions defined by ENCODE were then removed using Bedtools (http://mitra.stanford.edu/kundaje). To define H3K9 trimethylation caused by dCas9-KRAB, we subtracted nontargeting control reads from each sgRNA track using the BAMcompare function from deepTools2 (53). Results were correlated with ChIP-Seq from ENCODE (Supplemental Table 3) and publicly available DNase I–hypersensitivity site (DHS) data (14, 23). Motif analysis was performed using JASPAR (http://jaspar.genereg.net). Raw data files for ChIP sequencing are available at the Gene Expression Omnibus with the accession number GSE131825.

ChIP PCR. ChIP was performed using anti-H3K27Ac (ab4729, Abcam), anti-ATF4 (ab23760, Abcam), anti-ATF3 (D2Y5W, Cell Signaling Technology), anti–c-JUN (60A8, Cell Signaling Technology), anti-JUND (D17G2, Cell Signaling Technology), and anti-CEBPB (ab322588, Abcam). Cells were cross-linked using ChIP Cross-link Gold (C01019027; Diagenode) for 30 minutes in PBS with 1 mM $MgCl_2$ and then with 1% formaldehyde for 10 minutes. The reaction was then quenched with 0.125 M glycine. Cell pellets were washed twice with cold PBS containing protease inhibitors (Complete EDTA-free tablets, Roche). Ten million cells were used per ChIP, as described in the protocol reported by Lee et al. (51). Nuclear lysates were sonicated using a Bioruptor Pico (Diagenode) for either 10 (K562) or 8 (BB953 and BB946) cycles. Immunoprecipitation was performed overnight at 20 rpm and 4°C, with 10 μL magnetic beads (Dynabeads Protein G, Invitrogen) per 1 μg antibody. Washing and DNA extraction were performed as for ChIP sequencing. For ChIP, quantitative PCR assays were performed in 384-well MicroAmp optical reaction plates using TaqMan Fast Universal PCR Mastermix (Life Technologies), and with probes and primers designed using the Universal Probe Library System (Roche). Signal was detected using an ABI PRISM 7900HT Sequence Detection System (Life Technologies). For primer sequences and associated probes, see Supplemental Table 6.

4C sequencing. 4C primer sequences and enzyme combinations were selected using the University of Chicago online tool (http://mnlab.uchicago.edu/4Cpd) with coordinates from the ABCB1 promoter active in K562_R1–3 cells (hg38, chr7:87,598,302–87,601,399). 4C sequencing was performed according to the protocol developed by Splinter et al. (54). Briefly, 10^7 cells were cross-linked with 2% formaldehyde for 10 minutes at room temperature before the reaction was quenched with 0.125 M glycine. Cells were lysed with buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% TX-100, and 1× complete protease inhibitors (11245200, Roche). The cross-linked nuclear preparation was then incubated with DpnII. Digestion was confirmed by reversing cross-linking for an aliquot and running it on a 0.6% agarose gel. Samples were then ligated overnight at 16°C using T4 DNA ligase (799009, Roche). Ligation efficiency was again confirmed with 0.6% agarose gel. Cross-linking was reversed and DNA extracted using phenol-chloroform; samples were then subjected to a second digestion using Csp6I. Ligation was again performed overnight at 16°C using T4 DNA ligase; DNA was then extracted using phenol-chloroform and purified with a QIAquick PCR purification kit (28104, Qiagen). PCR primers were designed to incorporate 4C primers with a barcode and

Illumina adapter sequences: reading primer, 5′ P5-Barcode-Primer 3′; nonreading primer, 5′ P7-Primer 3′; reading, GAGATACCAGGTCTGATC; nonreading, AGGGTAGGTATTCCACTTTT; barcode, CTTGTA; illumina adapter sequence P5, AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT; P7, CAAGCAGAAGACGGCATACGAGAT; nonreading primer, CAAGCAGAAGACGGCATACGAGATAGGGTAGGTATTCCACTTTT; and reading primer, AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTGTAGAGATACCAGGTCTGATC.

PCR was performed with Expand Long Template Polymerase (11759060001, Roche) using 3.2 μg of 4C template product, and then purified using a High Pure PCR Product Purification Kit (11732676001, Roche). Library quality was assessed using a 2100 Bioanalyzer (Agilent Technologies). Samples were sequenced with 10% phiX using a MiSeq desktop sequencing system (Illumina) with 75-bp, single-end settings generating a mean of 1.3 million reads per sample. Sequencing data were deconvoluted using cutadapt version 1.18. Reads were mapped and analysis performed using 4Cseqpipe (55).

CRISPR-dCas9-KRAB enhancer silencing. CRISPR guides were designed with Off-Spotter (https://cm.jefferson.edu/Off-Spotter/), using putative enhancer sequences from K562_R1–3 H3K27Ac ChIP-Seq data. Several guides were selected for each enhancer and the promoter to allow preliminary screening of sgRNA activity. Guides were chosen to provide relatively even coverage across each enhancer and targeting of both DNA strands (Supplemental Figure 3, A–D). Primers incorporating the sgRNA sequence were designed as follows (primer sequences are shown in Supplemental Table 7; 4 nucleotides [green] were added to the guide sequence to permit ligation to the cut vector and if the guide sequence did not start with a guanine then one was added [yellow] to allow efficient transcription by the U6 promoter):

5′ CACCGXXXXXXXXXXXXXXXXXXX 3′
3′ CXXXXXXXXXXXXXXXXXXXCAAA 5′

Primers were annealed by heating reagents A (Supplemental Table 8) to 98°C for 5 minutes, then allowing slow cooling by removing the heat block from the heating element until equilibrated to room temperature. Annealed primers were then ligated into pLKO5.sgRNA.EFS.tRFP657 (57824, Addgene) using combined digestion-ligation with *BsmBI* and T4 DNA Ligase (M180A, Promega). Reagents B (Supplemental Table 8) were heated to 55°C for 2 hours; reagents C (Supplemental Table 8) were then added and the temperature reduced to 37°C for 1 hour. Lentivirus was produced using 293FT packaging cells (Life Technologies) cultured in DMEM (Sigma-Aldrich) with 10% FBS. Four micrograms of vector was added to 1 mL DMEM with 21 μL polyethylenimine (Polysciences), 2 μg pCMVd8.91, and 1 μg pMD2.G. The mixture was incubated for 30 minutes at room temperature, then added dropwise to a 10-cm dish containing 75% confluent 293FT cells; medium was replaced after 24 hours. Conditioned medium containing lentivirus was collected at 48 and 72 hours after transfection; packaging cells were removed using a 0.45-μm filter. K562_R1–3 were reselected for 7 days with 500 nM daunorubicin to ensure high-level ABCB1 expression prior to lentiviral transduction with pHR-SFFV-dCas9-BFP-KRAB (46911, Addgene). Transduction was performed by resuspension of 2×10^6 K562 cells in fresh viral supernatant containing 8 μg/mL Polybrene. After 24 hours the medium was exchanged to remove the virus. Seven days later a second transduction was performed using lentivirus containing ligated pLKO5.sgRNA.EFS.tRFP657. After 5 days, expression of mTagBFP, tRFP657, and ABCB1 and calcein AM retention were assessed by flow cytometry. All sgRNAs were screened for activity using K562_R3 (Supplemental Figure 3, E and F). The most active guide for each enhancer was then used to transduce dCas9-KRAB$^+$ K562_R1–3. Flow cytometry, RNA extraction, and ChIP were then performed on days 5, 7, and 10 after transduction, respectively. A further assessment of mTagBFP, tRFP657, and ABCB1 expression was made on day 13 to confirm stable expression (Figure 3B).

Assessment of ABCB1 ATPase activity. The Pgp-Glo Assay System (V3601, Promega) was used to assess the ability of tosedostat to induce ABCB1 ATPase activity. The assay was performed as described in the product literature. Briefly, Na_3VO_4 (0.1 mM), verapamil (0.2 mM), or tosedostat (0.2 mM) was incubated for 40 minutes at 37°C with 5 mM ATP and membranes containing recombinant ABCB1. Residual ATP was then assessed by addition of Ultra-Glo luciferase and incubation at room temperature for 20 minutes. Luminescence was quantified using a GloMax-Multi detection system (Promega). Na_3VO_4 inhibits ABCB1 ATPase activity, providing a negative control. Verapamil is a known ABCB1 substrate, inducing ATPase activity and providing a positive control.

Statistics. For flow cytometry, quantitative PCR, ChIP PCR, and luciferase assays, statistical significance was determined using the unpaired, 2-tailed Student's *t* test when comparing 2 experimental groups, or with 1-way ANOVA with Tukey's correction when comparing 3 or more groups. All tests were performed in Prism 8 (GraphPad). *P* values less than 0.05 were considered statistically significant. The statistical methods used to analyze next-generation sequencing data are detailed in the relevant sections of Methods.

Study approval. Primary human AML samples were from the Manchester Cancer Research Centre Tissue Biobank (approved by the South Manchester Research Ethics Committee). Their use was authorized by the Tissue Biobank's scientific subcommittee, with the informed consent of donors.

Author contributions

MSW and TCPS designed the study, and MSW and FS performed experiments. MSW, FMRA, and TCPS performed bioinformatics analyses. MSW and TCPS wrote the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgments

This work was supported by Cancer Research UK grant C5759/A20971 and a Kay Kendall Leukaemia Fund Clinical Training Fellowship to MSW.

Address correspondence to: Tim C.P. Somervaille, Leukaemia Biology Laboratory, Cancer Research UK Manchester Institute, The University of Manchester, Oglesby Cancer Research Centre Building, 555 Wilmslow Road, Manchester M20 4GJ, United Kingdom. Phone: 44.161.306.3240; Email: tim.somervaille@cruk.manchester.ac.uk.

1. Zhang J, Gu Y, Chen B. Mechanisms of drug resistance in acute myeloid leukemia. *Onco Targets Ther*. 2019;12:1937–1945.
2. Leith CP, et al. Acute myeloid leukemia in the elderly: assessment of multidrug resistance (MDR1) and cytogenetics distinguishes biologic subgroups with remarkably distinct responses to standard chemotherapy. A Southwest Oncology Group study. *Blood*. 1997;89(9):3323–3329.
3. Shaffer BC, Gillet JP, Patel C, Baer MR, Bates SE, Gottesman MM. Drug resistance: still a daunting challenge to the successful treatment of AML. *Drug Resist Updat*. 2012;15(1–2):62–69.
4. Chen KG, Sikic BI. Molecular pathways: regulation and therapeutic implications of multidrug resistance. *Clin Cancer Res*. 2012;18(7):1863–1869.
5. Stewart A, Steiner J, Mellows G, Laguda B, Norris D, Bevan P. Phase I trial of XR9576 in healthy volunteers demonstrates modulation of P-glycoprotein in CD56$^+$ lymphocytes after oral and intravenous administration. *Clin Cancer Res*. 2000;6(11):4186–4191.
6. Andersson R, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507(7493):455–461.
7. Schick S, Fournier D, Thakurela S, Sahu SK, Garding A, Tiwari VK. Dynamics of chromatin accessibility and epigenetic state in response to UV damage. *J Cell Sci*. 2015;128(23):4380–4394.
8. Lozzio CB, Lozzio BB. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood*. 1975;45(3):321–334.
9. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57.
10. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550.
11. Krige D, et al. CHR-2797: an antiproliferative aminopeptidase inhibitor that leads to amino acid deprivation in human leukemic cells. *Cancer Res*. 2008;68(16):6669–6679.
12. Manalo DJ, et al. Transcriptional regulation of vascular endothelial cell responses to hypoxia by HIF-1. *Blood*. 2005;105(2):659–669.
13. Rozpedek W, Pytel D, Mucha B, Leszczynska H, Diehl JA, Majsterek I. The role of the PERK/eIF2α/ATF4/CHOP signaling pathway in tumor progression during endoplasmic reticulum stress. *Curr Mol Med*. 2016;16(6):533–544.
14. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74.
15. Pakos-Zebrucka K, Koryga I, Mnich K, Ljujic M, Samali A, Gorman AM. The integrated stress response. *EMBO Rep*. 2016;17(10):1374–1395.
16. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010;28(5):495–501.
17. Scotto KW. Transcriptional regulation of ABC drug transporters. *Oncogene*. 2003;22(47):7496–7511.
18. Creyghton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*. 2010;107(50):21931–21936.
19. Williamson I, Hill RE, Bickmore WA. Enhancers: from developmental genetics to the genetics of common human disease. *Dev Cell*. 2011;21(1):17–19.
20. Diffner E, et al. Activity of a heptad of transcription factors is associated with stem cell programs and clinical outcome in acute myeloid leukemia. *Blood*. 2013;121(12):2289–2300.
21. Han J, et al. ER-stress-induced transcriptional regulation increases protein synthesis leading to cell death. *Nat Cell Biol*. 2013;15(5):481–490.
22. Quynh Doan NT, Christensen SB. Thapsigargin, origin, chemistry, structure-activity relationships and prodrug development. *Curr Pharm Des*. 2015;21(38):5501–5517.
23. Assi SA, et al. Subtype-specific regulatory network rewiring in acute myeloid leukemia. *Nat Genet*. 2019;51(1):151–162.
24. Chaudhary PM, Roninson IB. Expression and activity of P-glycoprotein, a multidrug efflux pump, in human hematopoietic stem cells. *Cell*. 1991;66(1):85–94.
25. van Galen P, et al. Integrated stress response activity marks stem cells in normal hematopoiesis and leukemia. *Cell Rep*. 2018;25(5):1109–1117.e5.
26. Laurenti E, et al. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat Immunol*. 2013;14(7):756–763.
27. Favata MF, et al. Identification of a novel inhibitor of mitogen-activated protein kinase kinase. *J Biol Chem*. 1998;273(29):18623–18632.
28. Zyryanova AF, et al. Binding of ISRIB reveals a regulatory site in the nucleotide exchange factor eIF2B. *Science*. 2018;359(6383):1533–1536.
29. Patch AM, et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature*. 2015;521(7553):489–494.
30. Christie EL, et al. Multiple ABCB1 transcriptional fusions in drug resistant high-grade serous ovarian and breast cancer. *Nat Commun*. 2019;10(1):1295.
31. Shaffer SM, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature*. 2017;546(7658):431–435.
32. Ben-David U, et al. Genetic and transcriptional evolution alters cancer cell line drug response. *Nature*. 2018;560(7718):325–330.
33. Denoyelle C, et al. Anti-oncogenic role of the endoplasmic reticulum differentially activated by mutations in the MAPK pathway. *Nat Cell Biol*. 2006;8(10):1053–1063.
34. Wortel IMN, van der Meer LT, Kilberg MS, van Leeuwen FN. Surviving stress: modulation of ATF4-mediated stress responses in normal and malignant cells. *Trends Endocrinol Metab*. 2017;28(11):794–806.
35. Hano M, Tomášová L, Šereš M, Pavlíková L, Breier A, Sulová Z. Interplay between P-glycoprotein expression and resistance to endoplasmic reticulum stressors. *Molecules*. 2018;23(2):E337.
36. Li XM, Liu J, Pan FF, Shi DD, Wen ZG, Yang PL. Quercetin and aconitine synergistically induces the human cervical carcinoma HeLa cell apoptosis via endoplasmic reticulum (ER) stress pathway. *PLoS One*. 2018;13(1):e0191062.
37. Kratz JD, et al. Metastatic bulk independently predicts outcomes for EGFR precision targeting in colorectal cancer. *J Natl Compr Canc Netw*. 2018;16(12):1442–1450.
38. Ossenkoppele GJ, Janssen JJ, van de Loosdrecht AA. Risk factors for relapse after allogeneic transplantation in acute myeloid leukemia. *Haematologica*. 2016;101(1):20–25.
39. Horibata S, Gui G, Lack J, DeStefano CB, Gottesman MM, Hourigan CS. Heterogeneity in refractory acute myeloid leukemia. *Proc Natl Acad Sci U S A*. 2019;116(21):10494–10503.
40. Wulf GG, et al. A leukemic stem cell with intrinsic drug efflux capacity in acute myeloid leukemia. *Blood*. 2001;98(4):1166–1173.
41. van Rhenen A, et al. High stem cell frequency in acute myeloid leukemia at diagnosis predicts high minimal residual disease and poor survival. *Clin Cancer Res*. 2005;11(18):6520–6527.
42. Ishikawa F, et al. Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. *Nat Biotechnol*. 2007;25(11):1315–1321.
43. Moitra K, Dean M. Evolution of ABC transporters by gene duplication and their role in human disease. *Biol Chem*. 2011;392(1–2):29–37.
44. Davidson AL, Chen J. ATP-binding cassette transporters in bacteria. *Annu Rev Biochem*. 2004;73:241–268.
45. Begicevic RR, Falasca M. ABC transporters in cancer stem cells: beyond chemoresistance. *Int J Mol Sci*. 2017;18(11):E2362.
46. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
47. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
48. Fernandez NF, et al. Clustergrammer, a web-based heatmap visualization and analysis tool for high-dimensional biological data. *Sci Data*. 2017;4:170151.
49. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, New York, USA: Springer-Verlag; 2016.
50. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
51. Lee TI, Johnstone SE, Young RA. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc*. 2006;1(2):729–748.
52. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
53. Ramírez F, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44(W1):W160–W165.
54. Splinter E, de Wit E, van de Werken HJ, Klous P, de Laat W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods*. 2012;58(3):221–230.
55. van de Werken HJ, et al. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods*. 2012;9(10):969–972.

Appendix 3

Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis.

Akihide Yoshimi, Kuan-Ting Lin, Daniel H Wiseman, Mohammad Alinoor Rahman, Alessandro Pastore, Bo Wang, Stanley Chun-Wei Lee, Jean-Baptiste Micol, Xiao Jing Zhang, Stephane de Botton, Virginie Penard-Lacronique, Eytan M Stein, Hana Cho, Rachel E Miles, Daichi Inoue, Todd R Albrecht, Tim C P Somervaille, Kiran Batta, Fabio Amaral, Fabrizio Simeoni, Deepti P Wilks, Catherine Cargo, Andrew M Intlekofer, Ross L Levine, Heidi Dvinge, Robert K Bradley, Eric J Wagner, Adrian R Krainer, Omar Abdel-Wahab. (2019)

Nature 574, 273-281.

A research article written by Akihide Yoshimi, Kuan-Ting Lin and Daniel H Wiseman, where I am a contributing author.

LETTER

https://doi.org/10.1038/s41586-019-1618-0

Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis

Akihide Yoshimi[1,12], Kuan-Ting Lin[2,12], Daniel H. Wiseman[3,4,12], Mohammad Alinoor Rahman[2], Alessandro Pastore[1], Bo Wang[1], Stanley Chun-Wei Lee[1], Jean-Baptiste Micol[5], Xiao Jing Zhang[1], Stephane de Botton[5], Virginie Penard-Lacronique[5], Eytan M. Stein[6], Hana Cho[1], Rachel E. Miles[1], Daichi Inoue[1], Todd R. Albrecht[7], Tim C. P. Somervaille[3], Kiran Batta[4], Fabio Amaral[3], Fabrizio Simeoni[3], Deepti P. Wilks[8], Catherine Cargo[9], Andrew M. Intlekofer[1], Ross L. Levine[1,6], Heidi Dvinge[10], Robert K. Bradley[11], Eric J. Wagner[7], Adrian R. Krainer[2] & Omar Abdel-Wahab[1,6]*

Transcription and pre-mRNA splicing are key steps in the control of gene expression and mutations in genes regulating each of these processes are common in leukaemia[1,2]. Despite the frequent overlap of mutations affecting epigenetic regulation and splicing in leukaemia, how these processes influence one another to promote leukaemogenesis is not understood and, to our knowledge, there is no functional evidence that mutations in RNA splicing factors initiate leukaemia. Here, through analyses of transcriptomes from 982 patients with acute myeloid leukaemia, we identified frequent overlap of mutations in *IDH2* and *SRSF2* that together promote leukaemogenesis through coordinated effects on the epigenome and RNA splicing. Whereas mutations in either *IDH2* or *SRSF2* imparted distinct splicing changes, co-expression of mutant IDH2 altered the splicing effects of mutant SRSF2 and resulted in more profound splicing changes than either mutation alone. Consistent with this, co-expression of mutant IDH2 and SRSF2 resulted in lethal myelodysplasia with proliferative features in vivo and enhanced self-renewal in a manner not observed with either mutation alone. *IDH2* and *SRSF2* double-mutant cells exhibited aberrant splicing and reduced expression of *INTS3*, a member of the integrator complex[3], concordant with increased stalling of RNA polymerase II (RNAPII). Aberrant *INTS3* splicing contributed to leukaemogenesis in concert with mutant IDH2 and was dependent on mutant SRSF2 binding to *cis* elements in *INTS3* mRNA and increased DNA methylation of *INTS3*. These data identify a pathogenic crosstalk between altered epigenetic state and splicing in a subset of leukaemias, provide functional evidence that mutations in splicing factors drive myeloid malignancy development, and identify spliceosomal changes as a mediator of *IDH2*-mutant leukaemogenesis.

Mutations in RNA splicing factors are common in cancer and impart specific changes to splicing that are identifiable by mRNA sequencing (RNA-seq)[4–6]. Somatic mutations involving the proline 95 residue of the spliceosome component SRSF2 are among the most recurrent in myeloid malignancies and alter SRSF2 binding to RNA in a sequence-specific manner[6,7]. We analysed RNA-seq data in The Cancer Genome Atlas (TCGA)[1] from 179 patients with acute myeloid leukaemia (AML), evaluating them for spliceosomal alterations. Aberrant splicing events characteristic of *SRSF2* mutations, including *EZH2*[6,7] poison exon inclusion, were observed in 19 patients ($P = 1.6 \times 10^{-12}$; Fisher's exact test; Fig. 1a, Extended Data Fig. 1a, b, Supplementary Table 1). Although only one patient with a mutation in *SRSF2* was reported in the TCGA AML publication[1], mutational analysis of RNA-seq data identified *SRSF2* hotspot mutations in each of these 19 patients (11% of the patients with AML). Therefore, these data retrospectively identify *SRSF2* as one of the most commonly mutated genes in the TCGA AML cohort.

Notably, 47% of patients with mutated *SRSF2* also had a mutation in *IDH2* and conversely, 56% of patients with mutated *IDH2* had a mutation in *SRSF2* ($P = 1.7 \times 10^{-6}$; Fisher's exact test; Fig. 1b, Extended Data Fig. 1c, d, Supplementary Table 2). Similar results were seen in RNA-seq data from 498 and 263 patients with AML from the Beat AML[8] and Leucegene[9] studies, respectively (Fig. 1c, d, Extended Data Fig. 1e–j, Supplementary Table 2). Across these datasets, variant allele frequencies of *IDH2* and *SRSF2* mutations were high and significantly correlated (Extended Data Fig. 1k), suggesting their common placement as early events in AML.

Beyond these datasets, combined *IDH2* and *SRSF2* mutations were identified in 5.2–6.2% of 1,643 unselected consecutive patients with AML in clinical practice (Supplementary Table 3). Although not statistically significant, *IDH2* and *SRSF2* double-mutant AML cases had the shortest overall survival across the four studied genotypes (Extended Data Fig. 2a). Whereas patients with *IDH2* and *SRSF2* double mutations had mostly intermediate cytogenetic risk, their prognosis was comparable to those with adverse cytogenetic risk (Extended Data Fig. 2b). The patients with *IDH2* and *SRSF2* double mutations were also significantly older than those with mutations in *IDH2* only, or with wild-type *IDH2* and *SRSF2* (Extended Data Fig. 2d; clinical and genetic features are summarized in Extended Data Fig. 2 and Supplementary Table 3).

Mutations in *IDH2* confer neomorphic enzymatic activity that results in the generation of 2-hydroxyglutarate (2HG)[10], which in turn induces DNA hypermethylation via the competitive inhibition of the α-ketoglutarate-dependent enzymes TET1–TET3. Unsupervised hierarchical clustering of DNA methylation data from the TCGA AML cohort revealed that *IDH2* and *SRSF2* double-mutant AML cases form a distinct cluster with higher DNA methylation than *IDH2* single-mutant AML cases (Extended Data Fig. 1l-o). Collectively, these data identify *IDH2* and *SRSF2* double-mutant leukaemia as a recurrent genetically defined AML subset with a distinct epigenomic profile.

We next sought to understand the basis for co-enrichment of *IDH2* and *SRSF2* mutations. Although mutations in splicing factors are frequently found in leukaemias, there is no functional evidence that they can transform cells in vivo. Overexpression of human $IDH2^{R140Q}$ or $IDH2^{R172K}$ in bone marrow (BM) cells from *Vav-cre* $Srsf2^{P95H/+}$ or *Vav-cre* $Srsf2^{+/+}$ mice revealed a clear collaborative effect between mutant *IDH2* and *Srsf2* (Extended Data Fig. 3a). Four weeks after transplantation, the peripheral blood of recipient mice transplanted with *IDH2* and

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. [3]Leukaemia Biology Laboratory, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, UK. [4]Division of Cancer Sciences, The University of Manchester, Manchester, UK. [5]Gustave Roussy, Université Paris-Saclay, Villejuif, France. [6]Leukemia Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [7]Department of Biochemistry & Molecular Biology, The University of Texas Medical Branch at Galveston, Galveston, Texas, USA. [8]Manchester Cancer Research Centre Biobank, The University of Manchester, Manchester, UK. [9]Haematological Malignancy Diagnostic Service, St James's University Hospital, Leeds, UK. [10]Department of Biomolecular Chemistry, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, USA. [11]Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [12]These authors contributed equally: Akihide Yoshimi, Kuan-Ting Lin, Daniel H. Wiseman. *e-mail: abdelwao@mskcc.org

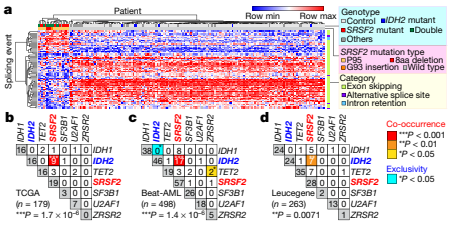


Fig. 1 | Frequently co-occurring *IDH2* and *SRSF2* mutations in AML. **a**, Heat map of per cent-spliced-in values for mutant *SRSF2*-specific splicing events in TCGA AML samples. **b**–**d**, Co-occurrence of mutations in *IDH1*, *IDH2*, *TET2* and RNA-splicing factors in the TCGA (**b**), Beat AML (**c**) and Leucegene (**d**) cohorts. Number of patients indicated; co-occurrence or exclusivity noted by colour coding; two-sided Fisher's exact test. Double refers to *SRSF2* and *IDH2* double mutant throughout.

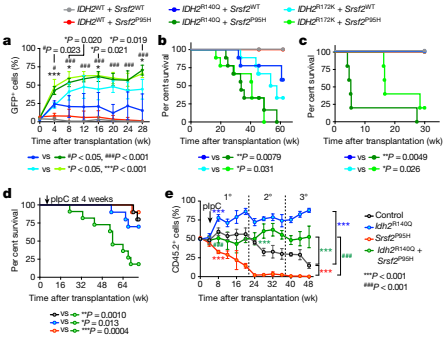


Fig. 2 | Mutant *IDH2* cooperates with mutant *Srsf2* to promote leukaemogenesis. **a**, Chimerism of GFP$^+$ cells in the blood of recipients transplanted with BM cells with indicated genotypes over time ($n = 5$ per group; data at 0 week represent transduction efficiency; mean percentage ± s.d.; two-way analysis of variance (ANOVA) with Tukey's multiple comparison test). **b**–**d**, Kaplan–Meier survival analysis of primary recipients (**b**) ($n = 10$ mice per genotype), recipients of serial transplant (**c**) ($n = 5$) and primary recipients transplanted non-competitively with BM cells from knock-in mice (**d**) ($n = 10$). log-rank Mantel–Cox test (two-sided). **e**, Chimerism of peripheral blood CD45.2$^+$ cells in competitive transplantation. $n = 10$ mice per group; mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test.

Srsf2 double-mutant cells had a substantially higher percentage of GFP$^+$ cells than in an *Srsf2* wild-type background (Fig. 2a, Extended Data Fig. 3b, c). Moreover, these mice exhibited significant myeloid skewing, macrocytic anaemia and thrombocytopaenia of greater magnitude than seen with mutant *IDH2* (Extended Data Fig. 3d–h). *IDH2* and *Srsf2* double mutants showed no difference in plasma 2HG levels from *IDH2* single mutants (Extended Data Fig. 3i, j). Serial replating of BM cells from leukaemic mice revealed markedly enhanced clonogenicity of *IDH2* and *Srsf2* double-mutant cells compared with other genotypes; the *IDH2* and *Srsf2* cells exhibited a blastic morphology and immature immunophenotype (Extended Data Fig. 3k–m). Consistent with these in vitro results, mice transplanted with *IDH2* and *Srsf2* double-mutant cells developed a lethal myelodysplastic syndrome (MDS) characterized by pancytopaenia, macrocytosis, myeloid dysplasia, expansion of immature BM progenitors and splenomegaly (Fig. 2b, Extended Data Fig. 3n–w). The *IDH2* and *Srsf2* double-mutant cells were also serially transplantable in sublethally irradiated recipients (Fig. 2c, Extended Data Fig. 3x). By contrast, *IDH2* single-mutant controls developed leukocytosis, myeloid skewing without clear dysplasia and less pronounced splenomegaly, whereas *Srsf2* single-mutant cells exhibited impaired repopulation capacity. These results provide evidence that spliceosomal gene mutations can promote leukaemogenesis in vivo.

We next sought to verify the effects of mutant *Idh2* and *Srsf2* using models in which both mutants were expressed from endogenous loci. *Mx1-cre Srsf2*$^{P95H/+}$ mice were crossed with *Idh2*$^{R140Q/+}$ mice to generate control, *Idh2*^{R140Q} single-mutant, *Srsf2*^{P95H} single-mutant and *Idh2* and *Srsf2* double knock-in (DKI) mice (Extended Data Fig. 4a). As expected, 2HG levels in peripheral blood mononuclear cells were increased and 5-hydroxymethylcytosine levels in KIT$^+$ BM cells were decreased in *Idh2* single-mutant and DKI primary mice compared with controls (Extended Data Fig. 4b, c). We next performed non-competitive transplantation, in which each mutation was induced alone or together following stable engraftment in recipients. DKI mice showed stable engraftment over time, similar to *Idh2* single-mutant or control mice (Extended Data Fig. 4d). However, DKI mice developed a lethal MDS with proliferative features and significantly shorter survival compared to controls (Fig. 2d). In competitive transplantation, expression of mutant *Idh2*^{R140Q} rescued the impaired self-renewal capacity of *Srsf2* single-mutant cells (Fig. 2e). These observations were supported by an increase in haematopoietic stem–progenitor cells in DKI mice compared with *Srsf2* single-mutant or control mice in primary and serial transplantation (Extended Data Fig. 4e–i). These results confirm cooperativity between mutant IDH2 and SRSF2 in promoting leukaemogenesis in vivo.

On the basis of data identifying 2HG-mediated inhibition of TET2 as a mechanism of *IDH2* mutant leukaemogenesis[11], we also evaluated whether loss of *TET2* might promote transformation of *SRSF2* mutant cells. However, deletion of *Tet2* in an *Srsf2* mutant background was insufficient to rescue the impaired self-renewal capacity of *Srsf2* single-mutant cells (Extended Data Fig. 4j–n). Similarly, restoration of TET2 function did not affect the self-renewal capacity of *Idh2* and *Srsf2* double-mutant cells in vivo (Extended Data Fig. 4o–q). These data indicated that the collaborative effects of mutant *Idh2* and *Srsf2* are not solely dependent on TET2. Consistent with this, combined silencing of *Tet2* and *Tet3* partially rescued the impaired replating capacity of *Srsf2* mutant cells in vitro (Extended Data Fig. 4r, s) and the impaired self-renewal of *Srsf2* mutant cells in vivo (Extended Data Fig. 4t–v). Because FTO and ALKBH5—which have roles in RNA processing as *N*6-methyladenosine (m^{6}A) RNA demethylases[12,13]—are also dependent on α-ketoglutarate, we investigated the effects of their loss on cooperativity with mutant *Srsf2*. However, collaborative effects were not observed between loss of *Fto* or *Alkbh5* and *Srsf2*^{P95H} (Extended Data Fig. 4w, x).

To understand the basis for cooperation between *IDH2* and *SRSF2* mutations, we next analysed RNA-seq data from the TCGA ($n = 179$ patients), Beat AML ($n = 498$ patients) and Leucegene ($n = 263$ patients) cohorts as well as two previously unpublished RNA-seq datasets targeting defined *IDH2* and *SRSF2* genotype combinations ($n = 42$ patients) and the knock-in mouse models. This revealed that cells with mutations in both *IDH2* and *SRSF2* consistently contained more aberrant splicing events than cells with mutations in *SRSF2* only. Moreover, *IDH2* mutations were associated with a small but reproducible change in RNA splicing (Fig. 3a, b, Extended Data Fig. 5a–g, Supplementary Tables 4–20). By contrast, AML cases in which both *TET2* and *SRSF2* were mutated had fewer changes in splicing than those in which *IDH2* and *SRSF2* were mutated (Extended Data Fig. 5h–m, Supplementary Tables 21, 22).

The majority of splicing changes associated with *SRSF2* mutations involved altered cassette-exon splicing, consistent with *SRSF2* mutations promoting inclusion of C-rich RNA sequences[6,7]. The sequence specificity of mutant SRSF2 on splicing was not influenced by concomitant *IDH2* mutations (Extended Data Fig. 5n–q) and a number of these events were validated by PCR with reverse transcription

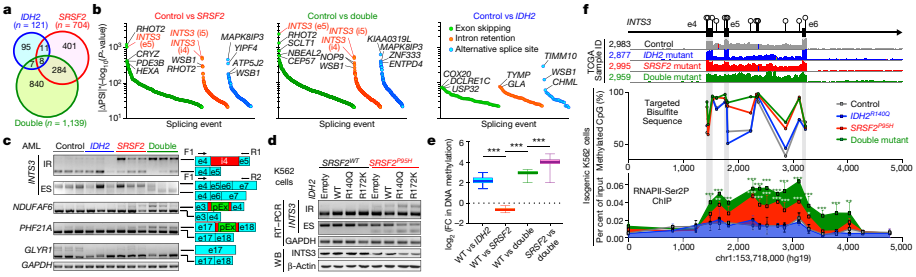


Fig. 3 | Collaborative effects of mutant *IDH2* and *SRSF2* on aberrant splicing. a, Venn diagram showing numbers of differentially spliced events from TCGA AML samples. **b**, Number of differentially spliced events ($|\Delta PSI| > 10\%$ and $P < 0.01$) in indicated genotypes are ranked by $((|\Delta PSI| \times |(-\log(P))$ according to class of event (e5, exon 5; i4 and i5, intron 4 and 5, respectively). PSI and *P* values adjusted for multiple comparisons were calculated using PSI-Sigma[25]. **c**, Representative RT–PCR results of aberrantly spliced transcripts in samples from patients with AML (pEx: exon with premature stop codon; $n = 3$ patients per genotype; three technical replicates with similar results). ES, exon skipping; IR, intron retention; F1, R1 and R2 represent primers used for RT–PCR. **d**, RT–PCR and western blots in isogenic K562 human leukaemia cells (representative images from three biologically independent experiments with similar results). **e**, Mean fold change (expressed in $\log_2$) in DNA cytosine methylation (*y* axis) at regions of genomic DNA encoding mRNA that undergo differential splicing (*x* axis). DNA methylation levels were determined by enhanced reduced representation bisulfite sequencing (eRRBS). $n = 3$ per genotype; the line represents mean, box edges represent the 25th and 75th percentiles and whiskers show 2.5th and 97.5th percentiles; one-way ANOVA with Tukey's multiple comparison test; $^{***}P < 2.2 \times 10^{-16}$. **f**, Genomic locus of *INTS3* around exons 4–6 with CpG dinucleotides indicated (top), representative RNA-seq from four patients with AML (top; $n = 1$ per genotype), results of targeted bisulfite sequencing (middle; $n = 1$ per genotype) and results of ChIP-walking experiments targeting RNAPII phosphorylated on Ser2 (Ser2P) (bottom; $n = 3$; mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test compared with control.). Double mutant, $IDH2^{R140Q}SRSF2^{P95H}$. $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$.

(RT–PCR) of primary AML samples from an independent cohort (Fig. 3c). Among the mis-splicing events in AML with mutations in both *IDH2* and *SRSF2* was a complex event in *INTS3* involving intron retention across two contiguous introns and skipping of the intervening exon (Fig. 3b, c, Extended Data Figs. 5e–g, r–y, 6a–c). Aberrant *INTS3* splicing was demonstrated in isogenic and non-isogenic leukaemia cells with or without *IDH2* and/or *SRSF2* mutations (Fig. 3d, Extended Data Fig. 6d–f), and *INTS3* transcripts with both intron retention and exon skipping resulted in nonsense-mediated decay (Extended Data Fig. 6g–j). Consistent with these observations, INTS3 protein expression was reduced in *SRSF2* mutant cells (Fig. 3d, Extended Data Fig. 6e, f, k–n, Supplementary Table 23). Moreover, silencing of INTS3 was associated with reduced protein levels of additional integrator subunits in *SRSF2* mutant AML compared to *SRSF2* wild-type AML. Consistent with these observations, steady-state protein expression levels of integrator subunits were correlated with one another (Extended Data Fig. 6o). Overall, these data indicate that aberrant splicing and consequent loss of INTS3 was a consistent feature of *IDH2* and *SRSF2* double-mutant cells and was associated with reduced expression of multiple integrator subunits.

We next sought to understand how *IDH2* mutations, which affect the epigenome, might influence splicing catalysis. Splice-site choice is influenced by *cis*-regulatory elements engaged by RNA-binding proteins as well as RNAPII elongation, which is regulated by DNA cytosine methylation and histone modifications[14]. We therefore generated a controlled system to dissect the contribution of RNA-binding elements and DNA methylation to *INTS3* intron retention. We constructed a minigene of *INTS3* spanning exons 4 and 5 and the intervening intron 4 (Extended Data Fig. 7a–c). Transfection of this minigene into leukaemia cells containing combinations of *IDH2* and *SRSF2* mutations revealed that retention of *INTS3* intron 4 is driven by mutant SRSF2 and further enhanced in the *IDH2* and *SRSF2* double-mutant setting (Extended Data Fig. 7d). SRSF2 normally binds C- or G-rich motif sequences in RNA equally well to promote splicing[15]. Leukaemia-associated mutations in *SRSF2* promote its avidity for C-rich sequences while reducing the ability to recognize G-rich sequences[6,7]. Of note, exon 4 of *INTS3* contains the highest number of predicted SRSF2-binding motifs over the entire *INTS3* genomic region (Extended Data Fig. 7c). We evaluated the role of putative SRSF2 motifs in regulating *INTS3* splicing by mutating all six CCNG motifs in exon 4 to G-rich sequences. In this G-rich version of the minigene, intron retention no longer occurred (*INTS3*-GGNG) (Extended Data Fig. 7e). Conversely, when all G-rich SRSF2 motifs were converted to C-rich sequences (*INTS3*-CCNG), intron retention became evident (Extended Data Fig. 7f). These results confirmed the sequence-specific activity of mutant SRSF2 in *INTS3* intron retention and identified a role for mutant IDH2 in regulating splicing.

Because *IDH2* mutations promote increased DNA methylation and DNA methylation can affect splicing[14], we generated genome-wide maps of DNA cytosine methylation from patients with AML across four genotypes (Supplementary Table 23). This revealed that differentially spliced events in *IDH2* single-mutant as well as *IDH2* and *SRSF2* double-mutant AML (compared to *IDH2* and *SRSF2* wild-type and *SRSF2* single-mutant AML) contained significant hypermethylation of DNA. Thus regions of differential DNA hypermethylation significantly overlapped with regions of differential RNA splicing (Fig. 3e, Extended Data Fig. 7j).

The above results suggest a strong link between increased DNA methylation mediated by mutant IDH2 and altered RNA splicing by mutant SRSF2. To evaluate this further, we next examined DNA methylation levels around endogenous *INTS3* exons 4–6 by targeted bisulfite sequencing. This revealed increased DNA methylation at all CpG dinucleotides in this region in *IDH2* and *SRSF2* double-mutant cells compared to control or single-mutant cells (Fig. 3f, Extended Data Fig. 7k). A functional role of DNA methylation at these sites was verified by evaluating splicing in versions of the *INTS3* minigene in which each CG dinucleotide was converted to an AT to prevent cytosine methylation. In these CG-to-AT versions of the minigene, *IDH2* mutations no longer promoted mutant-*SRSF2*-mediated intron retention (Extended Data Fig. 7g–i). As further confirmation of the influence of mutant *IDH2* on *INTS3* splicing, cell-permeable 2HG increased *INTS3* intron retention whereas treatment of *IDH2* and *SRSF2* double-mutant cells with the DNA methyltransferase inhibitor 5-aza-2′-deoxycytidine (5-AZA-CdR) inhibited *INTS3* intron retention (Extended Data Fig. 7l, m).

Given that changes in epigenetic state may affect splicing by influencing RNAPII stalling[14,16], we evaluated the abundance of RNAPII

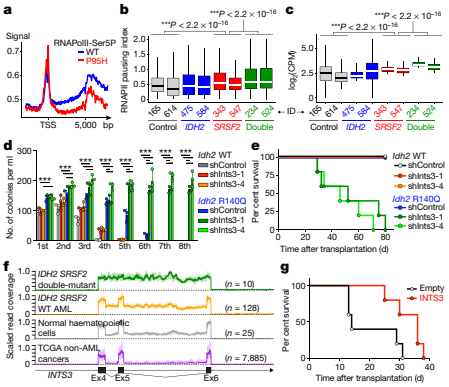


Fig. 4 | RNAPII stalling in *IDH2* and *SRSF2* double-mutant AML and contribution of INTS3 loss to leukaemogenesis. a, Metagene plot of genome-wide RNAPII Ser5 phosphorylation (Ser5P) in isogenic $SRSF2^{WT}$ or $SRSF2^{P95H}$ mutant K562 cells. **b**, **c**, RNAPII pausing index[20] in primary AML samples calculated as the ratio of normalized ChIP–seq reads of RNAPII Ser5P on transcription start sites (TSSs) (± 250 bp) over that of the corresponding bodies (+500 bp to +1000 bp from TSSs) (**b**) and RNAPII abundance over the differentially spliced regions between *SRSF2* single-mutant and *IDH2* and *SRSF2* double-mutant AML determined by RNAPII Ser2P ChIP–seq (*y* axis shows $\log_2$(counts per million (CPM))) (**c**). *x* axis shows patient ID. Box plots were generated from ChIP–seq data from an individual primary AML sample; the line represents mean, box edges represent 25th and 75th percentiles and whiskers show 2.5th and 97.5th percentiles. One-way ANOVA with Tukey's multiple comparison test. **d**, Colony numbers from serial replating assays of either *Mx1-cre* $Idh2^{+/+}$ *or* $Idh2^{R140Q/+}$ BM cells transduced with shRNA targeting *Ints3* (shInts3-1 or shInts3-4) or control shRNA (shControl). $n = 3$ biologically independent experiments; mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test. $^{***}P < 0.004$. **e**, **g**, Kaplan–Meier survival analysis of recipients. $n = 5$ per group; log-rank (Mantel–Cox) test (two-sided). In **e**: shInts3-1 versus shControl, $P = 0.0018$; shInts3-4 versus shControl, $P = 0.0018$. In **g**, $P = 0.034$ (INTS3 versus empty vector control). **f**, RNA-seq read coverage between exons 4–6 of *INTS3* and 1,000 bp either side of *INTS3* is scaled and shown as mean (line) ± s.d. (shaded region) (generated from TCGA datasets; sample list shown in legend for Extended Data Fig. 10e).

using chromatin precipitation with DNA sequencing (ChIP–seq) in isogenic $SRSF2^{WT}$ and $SRSF2^{P95H}$ cells as well as the primary samples from patients with AML. This revealed increased promoter-proximal transcriptional pausing and decreased RNAPII occupancy over gene bodies in *SRSF2* mutant cells, which was further enhanced in *IDH2* and *SRSF2* double-mutant cells (Fig. 4a, b, Extended Data Fig. 7n–q, Supplementary Table 23). Transcriptional pausing was also evident at *INTS5* and *INTS14* in *SRSF2* mutant cells (Extended Data Fig. 7r, s), which—in combination with aberrant splicing of several integrator subunits (Supplementary Table 24)—suggested impaired function of the entire integrator complex in *SRSF2* mutant cells. Similar to DNA cytosine methylation levels, RNAPII was more abundant over differentially spliced regions in *SRSF2* single-mutant AML than in *SRSF2* wild-type AML, and further enhanced over differentially spliced regions in *IDH2* and *SRSF2* double-mutant AML compared with those in *SRSF2* single-mutant AML (Fig. 4c, Extended Data Fig. 7t).

The above data provide further links between increased DNA cytosine methylation and RNAPII stalling with altered RNA splicing in *IDH2* and *SRSF2* double-mutant AML. To further evaluate this model, we performed ChIP for RNAPII across 4,766 bp of the *INTS3* locus in isogenic leukaemia cells (Fig. 3f). This revealed substantial accumulation of RNAPII across this locus in *IDH2* and *SRSF2* double-mutant cells. Treatment with 5-AZA-CdR significantly reduced RNAPII stalling, which was coupled with decreased aberrant *INTS3* splicing (Extended Data Fig. 7k–m). These data reveal that *IDH2* and *SRSF2* mutations coordinately dysregulate splicing through alterations in RNAPII stalling in addition to aberrant sequence recognition of *cis* elements in RNA.

INTS3 encodes a component of the integrator complex that participates in small nuclear RNA (snRNA) processing[3] in addition to RNAPII pause–release[17]. Consistent with this, *SRSF2* single-mutant cells had altered snRNA cleavage similar to those seen with direct INTS3 downregulation, which was exacerbated in *IDH2* and *SRSF2* double-mutant cells (Extended Data Fig. 8a–h). Attenuation of INTS3 expression in *SRSF2* mutant cells caused a blockade of myeloid differentiation, an effect further enhanced in an *IDH2* mutant background (Extended Data Fig. 8i–n). Notably, direct *Ints3* downregulation in the $Idh2^{R140Q/+}$ background resulted in enhanced clonogenic capacity of cells with an immature morphology and immunophenotype (Fig. 4d, Extended Data Fig. 8o–r) and promoted clonal dominance of *Idh2* mutant cells (Extended Data Fig. 9a–d). Moreover, mice transplanted with $Idh2^{R140Q/+}$ BM cells treated with short hairpin RNA (shRNA) targeting *Ints3* exhibited myeloid skewing, anaemia and thrombocytopaenia (Extended Data Fig. 9e–g) and developed a lethal MDS with proliferative features—phenotypes resembling those seen in *IDH2* and *Srsf2* double-mutant mice (Fig. 4e, Extended Data Fig. 9h, i).

The defects in snRNA processing in *SRSF2* single-mutant and *IDH2* and *SRSF2* double-mutant cells were partially rescued by INTS3 cDNA expression (Extended Data Fig. 8s–x). In addition, restoration of INTS3 expression released *SRSF2* single-mutant and *IDH2* and *SRSF2* double-mutant HL-60 cells from differentiation block (Extended Data Fig. 8y, z). Xenografts of *IDH2* and *SRSF2* double-mutant HL-60 cells demonstrated that forced expression of INTS3 induced myeloid differentiation and slowed leukaemia progression in vivo (Extended Data Fig. 9j-s). Collectively, these data suggest that INTS3 loss due to aberrant splicing by mutant IDH2 and SRSF2 contributes to leukaemogenesis.

Although loss of INTS3 resulted in measurable changes in snRNA processing, the degree of snRNA mis-processing did not have a substantial effect on splicing as determined by RNA-seq of $IDH2^{R140Q}$ mutant HL-60 cells with INTS3 silencing. By contrast, INTS3 depletion in these cells significantly affected transcriptional programs associated with myeloid differentiation, multiple oncogenic signalling pathways, RNAPII elongation-linked transcription and DNA repair (Extended Data Fig. 10a–d, Supplementary Table 25). This latter association of INTS3 loss with DNA repair is potentially consistent with previous reports that sensor of single-stranded DNA complexes containing INTS3 participate in DNA damage response[18,19].

These data uncover an important role for RNA splicing alterations in *IDH2* mutant tumorigenesis and identify perturbations in integrator as a driver of transformation of *IDH2* and *SRSF2* mutant cells. However, INTS3 is not known to be recurrently affected by coding-region alterations in leukaemias. We therefore evaluated *INTS3* splicing across 32 additional cancer types as well as normal blood cells to evaluate whether aberrant *INTS3* splicing might be a common mechanism in AML. This revealed that, whereas *INTS3* mis-splicing is most evident in *IDH2* and *SRSF2* double-mutant AML, *INTS3* mis-splicing is also prevalent across other molecular subtypes of AML but is not present in blood cells from healthy subjects or RNA-seq data from more than 7,000 samples from other cancer types (Fig. 4f, Extended Data Fig. 10e, f). To further evaluate the effects of enforced INTS3 expression in myeloid leukaemia with a wild-type splicing phenotype, we used *MLL-AF9*;$Nras^{G12D}$ mouse leukaemia (RN2) cells. INTS3 overexpression reduced colony-forming capacity (Extended Data Fig. 10g, h) and enhanced differentiation of RN2 cells, resulting in decelerated leukaemia progression in vivo (Fig. 4g, Extended Data Fig. 10i–s).

These data highlight a role for loss of INTS3 in broad genetic subtypes of AML. Further efforts to determine how integrator loss

promotes leukaemogenesis and other non-mutational mechanisms mediating *INTS3* aberrant splicing will be critical for understanding and targeting leukaemias with integrator loss. Previous studies have identified that integrator[17,20] and SRSF2[21] have direct roles in modulating transcriptional pause–release. The accumulation of RNAPII at certain mis-spliced loci in this study is consistent with recent data that suggest that mutant SRSF2 is defective in promoting RNAPII pause–release[22]. Identifying how aberrant splicing mediated by mutant SRSF2 is influenced by altered RNAPII pause–release may therefore be informative.

In addition to modifying splicing in *SRSF2* mutant cells, *IDH2* mutations were associated with reproducible changes in splicing in haematopoietic cells. There is a strong correlation between aberrant splicing in *IDH2* and *IDH1* mutant low-grade gliomas ($P = 2.2 \times 10^{-16}$ (binominal proportion test); Extended Data Fig. 10t–w, Supplementary Tables 26–28). A significant number of splicing events that were dysregulated in *IDH2* mutant AML from the TCGA and Leucegene cohorts were differentially spliced in *IDH2* mutants versus *IDH1* and *IDH2* wild-type low-grade gliomas ($P = 1.8 \times 10^{-9}$ and $P = 1.3 \times 10^{-8}$, respectively; binominal proportion test). These data suggest that *IDH1* and *IDH2* mutations impart a consistent effect on splicing regardless of tumour type. Finally, these results have important translational implications given the substantial efforts to pharmacologically inhibit mutant IDH1 and IDH2 as well as mutant splicing factors[23,24]. The frequent coexistence of *IDH2* and *SRSF2* mutations underscores the enormous therapeutic potential for modulation of splicing in the approximately 50% of patients with *IDH2* mutant leukaemia who also have a spliceosomal gene mutation.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1618-0.

Received: 24 May 2018; Accepted: 28 August 2019;
Published online 2 October 2019.

1. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
2. Papaemmanuil, E. et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616–3627, quiz 3699 (2013).
3. Wu, Y., Albrecht, T. R., Baillat, D., Wagner, E. J. & Tong, L. Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance. *Proc. Natl Acad. Sci. USA* **114**, 4394–4399 (2017).
4. Darman, R. B. et al. Cancer-associated SF3B1 hotspot mutations induce cryptic 3′ splice site selection through use of a different branch point. *Cell Rep.* **13**, 1033–1045 (2015).
5. Ilagan, J. O. et al. *U2AF1* mutations alter splice site recognition in hematological malignancies. *Genome Res.* **25**, 14–26 (2015).
6. Kim, E. et al. *SRSF2* mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell* **27**, 617–630 (2015).
7. Zhang, J. et al. Disease-associated mutation in *SRSF2* misregulates splicing by altering RNA-binding affinities. *Proc. Natl Acad. Sci. USA* **112**, E4726–E4734 (2015).
8. Tyner, J. W. et al. Functional genomic landscape of acute myeloid leukaemia. *Nature* **562**, 526–531 (2018).
9. Lavallée, V. P. et al. The transcriptomic landscape and directed chemical interrogation of *MLL*-rearranged acute myeloid leukemias. *Nat. Genet.* **47**, 1030–1037 (2015).
10. Dang, L. et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
11. Figueroa, M. E. et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).
12. Jia, G. et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).
13. Zheng, G. et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49**, 18–29 (2013).
14. Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.* **84**, 165–198 (2015).
15. Daubner, G. M., Cléry, A., Jayne, S., Stevenin, J. & Allain, F. H. A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *EMBO J.* **31**, 162–174 (2012).
16. Shukla, S. et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79 (2011).
17. Gardini, A. et al. Integrator regulates transcriptional initiation and pause release following activation. *Mol. Cell* **56**, 128–139 (2014).
18. Huang, J., Gong, Z., Ghosal, G. & Chen, J. SOSS complexes participate in the maintenance of genomic stability. *Mol. Cell* **35**, 384–393 (2009).
19. Li, Y. et al. HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J. Biol. Chem.* **284**, 23525–23531 (2009).
20. Stadelmayer, B. et al. Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat. Commun.* **5**, 5531 (2014).
21. Ji, X. et al. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* **153**, 855–868 (2013).
22. Chen, L. et al. The augmented R-loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations. *Mol. Cell* **69**, 412–425 (2018).
23. Seiler, M. et al. H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nat. Med.* **24**, 497–504 (2018).
24. Stein, E. M. et al. Enasidenib in mutant *IDH2* relapsed or refractory acute myeloid leukemia. *Blood* **130**, 722–731 (2017).

METHODS

Data reporting. The number of mice in each experiment was chosen to provide 90% statistical power with a 5% error level. Otherwise, no statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Mice. All mice were housed at Memorial Sloan Kettering Cancer Center (MSK). All mouse procedures were completed in accordance with the Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committees at MSK. Six- to eight-week-old week female CD45.1 C57BL/6 mice were purchased from The Jackson Laboratory (Stock No: 002014) . Male and female CD45.2 *Srsf2*$^{P95H/+}$ conditional knock-in mice, *Idh2*$^{R140Q/+}$ conditional knock-in mice, and *Tet2* conditional knockout mice (all on C57BL/6 background) were also analysed and used as bone marrow donors (generation of these mice were as described[6,26,27]). For BM transplantation assays with IDH2 overexpression, *Srsf2*$^{P95H/+}$ and littermate control mice were crossed to *Vav-cre* transgenic mice[28]. CBC analysis was performed on peripheral blood collected from submandibular bleeding, using a Procyte Dx Hematology Analyzer (IDEXX Veterinary Diagnostics). For all mouse experiments, the mice were monitored closely for signs of disease or morbidity daily and were euthanized for visible tumour formation at tumour volume >1 cm^3, failure to thrive, weight loss >10% total body weight, open skin lesions, bleeding, or any signs of infection. In none of the experiments were these limits exceeded.

BM transplantation assays. Freshly dissected femurs and tibias were isolated from *Mx1-cre*, *Mx1-cre/Idh2*$^{R140Q/+}$, *Mx1-cre Srsf2*$^{P95H/+}$, *Mx1-cre Idh2*$^{R140Q/+}$*Srsf2*$^{P95H/+}$, *Mx1-cre Tet2*$^{fl/fl}$, or *Mx1-cre Tet2*$^{fl/fl}$*Srsf2*$^{P95H/+}$ CD45.2$^+$ mice. BM was flushed with a 3-cm^3 insulin syringe into cold PBS supplemented with 2% bovine serum albumin to generate single-cell suspensions. BM cells were pelleted by centrifugation at 1,500 r.p.m. for 4 min and red blood cells (RBCs) were lysed in ammonium chloride-potassium bicarbonate lysis (ACK) buffer (150 mM NH_4Cl + 10 mM $KHCO_3$ + 0.1 mM EDTA; Thermo Fisher Scientific) for 3 min on ice. After centrifugation, cells were resuspended in PBS/2% BSA, passed through a 40-μm cell strainer, and counted. For competitive transplantation experiments, 0.5 × 10^6 BM cells from *Mx1-cre*, *Mx1-cre Idh2*$^{R140Q/+}$, *Mx1-cre Srsf2*$^{P95H/+}$, *Mx1-cre Idh2*$^{R140Q/+}$*Srsf2*$^{P95H/+}$, *Mx1-cre Tet2*$^{fl/fl}$ or *Mx1-cre Tet2*$^{fl/fl}$*Srsf2*$^{P95H/+}$ CD45.2$^+$ mice were mixed with 0.5 × 10^6 wild-type (WT) CD45.1$^+$ BM and transplanted via tail-vein injection into 8-week-old lethally irradiated (900 cGy) CD45.1$^+$ recipient mice. The CD45.1$^+$:CD45.2$^+$ ratio was confirmed to be approximately 1:1 by flow cytometry analysis pre-transplant. To activate the conditional alleles, mice were treated with 3 doses of polyinosinic:polycytidylic acid (pIpC; 12 mg per kg (body weight) per day; GE Healthcare) every second day via intraperitoneal injection. Peripheral blood chimerism was assessed every four weeks by flow cytometry. For noncompetitive transplantation experiments, 1 × 10^6 total BM cells from *Mx1-cre*, *Mx1-cre Idh2*$^{R140Q/+}$, *Mx1-cre Srsf2*$^{P95H/+}$, *Mx1-cre Idh2*$^{R140Q/+}$*Srsf2*$^{P95H/+}$, *Mx1-cre Tet2*$^{fl/fl}$, or *Mx1-cre Tet2*$^{fl/fl}$*Srsf2*$^{P95H/+}$ CD45.2$^+$ mice were injected into lethally irradiated (950 cGy) CD45.1$^+$ recipient mice. Peripheral blood chimerism was assessed as described for competitive transplantation experiments. Additionally, for each bleeding whole blood cell counts were measured on an automated blood analyser. Mice that were lost owing to pIpC toxicity were excluded from analysis.

Retroviral transduction and transplantation of primary haematopoietic cells. *Vav-cre Srsf2*$^{+/+}$ and *Vav-cre Srsf2*$^{P95H/+}$ mice were treated with a single dose of 5-fluoruracil (150 mg kg^{-1}) followed by BM collection from the femurs, tibias and pelvic bones 5 days later. RBCs were removed by ACK lysis buffer, and nucleated BM cells were transduced with viral supernatants containing MSCV-*IDH2*$^{WT/R140Q/R172K}$-IRES-GFP for 2 days in RPMI/20% FCS supplemented with mouse stem cell factor (mSCF, 25 ng ml^{-1}), mouse interleukin-3 (mIL3, 10 ng ml^{-1}) and mIL6 (10 ng ml^{-1}), followed by injection of about 0.5 × 10^6 cells per recipient mouse via tail vein injection into lethally irradiated (950 cGy) CD45.1$^+$ mice. Transplantation of primary BM cells with TET2 catalytic domain cDNA and anti-*Ints3* or *Tet3* shRNAs was similarly performed. For secondary transplantation experiments, 8-week old, lethally (900–950 cGy) or sub-lethally (450–700 cGy) irradiated C57/BL6 recipient mice were injected with unsorted 1 × 10^6 BM cells from the primary transplantation. *IDH2*^{WT} + *Srsf2*^{WT} and *IDH2*^{WT} + *Srsf2*^{P95H} mice were euthanized at day 315 post-transplant to collect BM for the serial transplantation. All cytokines were purchased from R&D Systems.

Flow cytometry analyses and antibodies. Surface-marker staining of haematopoietic cells was performed by first lysing cells with ACK lysis buffer and washing cells with ice-cold PBS. Cells were stained with antibodies in PBS/2% BSA for 30 min on ice. For haematopoietic stem/progenitor staining, cells were stained with the following antibodies: B220-APCCy7 (clone: RA3-6B2; purchased from BioLegend; catalogue no.: 103224; dilution: 1:200); B220-Bv711 (RA3-6B2; BioLegend; 103255; 1:200); CD3-PerCPCy5.5 (17A2; BioLegend; 100208; 1:200); CD3-APC (17A2; BioLegend; 100236; 1:200); CD3-APCCy7 (17A2; BioLegend; 100222; 1:200); Gr1-PECy7 (RB6-8C5; eBioscience; 25-5931-82; 1:500); CD11b-PE (M1/70; eBioscience; 12-0112-85; 1:500); CD11b-APCCy7 (M1/70; BioLegend; 101226; 1:200); CD11c-APCCy7 (N418; BioLegend; 117323; 1:200); NK1.1-APCCy7 (PK136; BioLegend; 108724; 1:200); Ter119-APCCy7 (BioLegend; 116223: 1:200); KIT-APC (2B8; BioLegend; 105812; 1:200); KIT-PerCPCy5.5 (2B8; BioLegend; 105824; 1:100); KIT-Bv605 (ACK2; BioLegend; 135120; 1:200); Sca1-PECy7 (D7; BioLegend; 108102; 1:200); CD16/CD32 (Fc-γRII/III)-Alexa700 (93; eBioscience; 56-0161-82; 1:200); CD34-FITC (RAM34; BD Biosciences; 553731; 1:200); CD45.1-FITC (A20; BioLegend; 110706; 1:200); CD45.1-PerCPCy5.5 (A20; BioLegend; 110728; 1:200); CD45.1-PE (A20; BioLegend; 110708; 1:200); CD45.1-APC (A20; BioLegend; 110714; 1:200); CD45.2-PE (104; eBioscience; 12-0454-82; 1:200); CD45.2-Alexa700 (104; BioLegend; 109822; 1:200); CD45.2-Bv605 (104; BioLegend; 109841; 1:200); CD48-Bv711 (HM48-1; BioLegend; 103439; 1:200); CD150 (9D1; eBioscience; 12-1501-82; 1:200). DAPI was used to exclude dead cells. For sorting human leukaemia cells, cells were stained with a lineage cocktail including CD34-PerCP (8G12; BD Biosciences; 345803; 1:200); CD117-PECy7 (104D2; eBioscience; 25-1178-42; 1:200); CD33-APC (P67.6; BioLegend; 366606; 1:200); HLA-DR-FITC (L243; BioLegend; 307604; 1:200); CD13-PE (L138; BD Biosciences; 347406; 1:200); CD45-APC-H7 (2D1; BD Biosciences; 560178; 1:200). The composition of mature haematopoietic cell lineages in the BM, spleen and peripheral blood was assessed using a combination of CD11b, Gr1, B220, and CD3. For the haematopoietic stem and progenitor analysis, a combination of CD11b, CD11c, Gr1, B220, CD3, NK1.1 and Ter119 was stained as lineage-positive cells. Fluorescence-activated cell sorting (FACS) was performed on a FACS Aria, and analysis was performed on an LSRII or LSR Fortessa (BD Biosciences). For western blotting, DNA dot blot assays, and ChIP assays, the following antibodies were used: INTS1 (purchased from Bethyl laboratories; catalogue no.: A300-361A; dilution: 1:1,000), INTS2 (Abcam; ab74982; 1:1,000), INTS3 (Bethyl laboratories; A300-427A; 1:1,000, Abcam; ab70451; 1:1,000), INTS4 (Bethyl laboratories; A301-296A; 1:1,000), INTS5 (Abcam; ab74405; 1:1,000), INTS6 (Abcam; ab57069; 1:1,000), INTS7 (Bethyl laboratories; A300-271A; 1:1,000), INTS8 (Bethyl laboratories; A300-269A; 1:1,000), INTS9 (Bethyl laboratories; A300-412A; 1:1,000), INTS11 (Abcam; ab84719; 1:1,000), Flag-M2 (Sigma-Aldrich; F-1084; 1:1,000), Myc-tag (Cell Signaling; 2276S; 1:1,000), β-actin (Sigma-Aldrich; A-5441; 1:2,000), 5-hydroxymehylcytosine (5hmC) (Active motif; 39769), RNAPII CTD repeat YSPTSPS (phospho S2) (Abcam; ab5095), RNAPII CTD repeat YSPTSPS (phospho S5) (Abcam; ab5408), and UPF1 (Abcam; ab109363; 1:1,000).

Minigene assay. We constructed *INTS3*-WT minigene spanning exons 4 to 5 of human *INTS3* into pcDNA3.1(+) vector (Invitrogen) using BamHI and XhoI sites, respectively. Artificial mutations were engineered into *INTS3*-WT minigene using the QuikChange Site-Directed Mutagenesis Kit (Agilent) to generate *INTS3*-GGNG, *INTS3*-CCNG, *INTS3*-WT_CG(−) *INTS3*-GGNG_CG(−), and *INTS3*-CCNG_CG(−) minigenes, respectively, and the sequences of inserts were verified by Sanger sequencing. Plasmids (1 μg) were transfected using Lipofectamine LTX reagent with PLUS reagent (Invitrogen) including 0.2 μg of eGFP and 0.8 μg of *INTS3* minigene, per well of a 6-well plate. Total RNA was extracted 48 h after transfection using TRIzol reagent (Ambion), followed by DNase I treatment (Qiagen). cDNA was synthesized with an oligo-dT primer using ImProm-II reverse transcriptase (Promega). Radioactive PCR was done with ^{32}P-α-dCTP, 1.25 units of AmpliTaq (Invitrogen) and 26 cycles using primer pairs 5′-GCTTGGTACCGAGCTCGGATC-3′ (vector-specific forward primer) and 5′-CAGTTCCCGTACCAACCACAC-3′ (reverse primer for *INTS3* versions of minigene), or 5′-CAGTTCCATTACCAACCACAC-3′ (reverse primer for *INTS3*_CG(-) versions of minigene). Products were run on a 5% PAGE and the bands were quantified using a Typhoon FLA 7000 (GE Healthcare). eGFP was used as a control for transfection efficiency and exogenous eGFP was amplified using a vector specific forward primer and reverse primer on eGFP. eGFP products were loaded after we ran the INTS3 products for 20–30 min. Percentages of intron 4 retention were normalized against exogenous eGFP.

Cell culture. K562 (human chronic myeloid/erythroleukaemia cell line) and HL-60 (human promyelocytic leukaemia cell line) leukaemia cells, K052 (human multilineage leukaemia cell line) leukaemia cells, TF1 (human erythroleukaemia cell line) leukaemia cells, *MLL-AF9/Nras*^{G12D} murine leukaemia (RN2) cells[29], and Ba/F3 (murine pro-B cell line) cells were cultured in RPMI/10% fetal calf serum (FCS, heat inactivated), RPMI/20% FCS, RPMI/10% FCS + human granulocyte-macrophage colony-stimulating factor (GM-CSF, R&D Systems; 5 ng ml^{-1}), and RPMI/10% FCS + mIL3 (R&D Systems; 1 ng ml^{-1}), respectively. None of the cell lines above were listed in the database of commonly misidentified cell lines maintained by ICLAC and NCBI Biosample.

MSCV-*IDH2*$^{WT/R140Q/R172K}$-IRES-GFP, MSCV-3×Flag-*INTS3*-puro, MSCV-IRES-3×Flag-*INTS3*-mCherry, MSCV-IRES-TET2 catalytic domain cDNA-mCherry ('TET2CD'), and empty vectors of these constructs were used for retroviral overexpression studies and pRRLSIN.cPPT.PGK-mCherry.WPRE-*SRSF2*$^{WT/P95H}$ constructs were used for lentiviral overexpression studies. TET2CD

pooling 96 samples with 5% PhiX onto a single NextSeq high output, 2 × 151-bp sequencing run; variant call format (VCF) files were analysed using Illumina's Variant Studio software; (b) a 40 gene panel (Oncomine Myeloid Research Assay; ThermoFisher), processing 8 samples per Ion 530 chip on the IonTorrent platform; data analysis performed using the Ion Reporter software; (c) a 27 gene custom panel (48 × 48 Access Array; Fluidigm) sequenced by Leeds HMDS on the MiSeq platform (300v2); or (d) MSK HemePACT[33] targeting all coding regions of 585 genes known to be recurrently mutated in leukaemias, lymphomas, and solid tumours. All panels provide sufficient coverage to detect minimum variant allele fraction 5% for all genes, except for the Access Array panel and *SRSF2*; all samples genotyped by this approach underwent manual Sanger sequencing of *SRSF2* exon 1 using the following primers (tagged with Fluidigm Access Array sequencing adaptors CS1/CS2): forward: acactgacgacatggttctacacccgtttacctgcggctc, reverse: tacggtagc agagacttggtctccttcgttcgctttcacgacaa.

Statistics and reproducibility. Statistical significance was determined by (1) unpaired two-sided Student's *t*-test after testing for normal distribution, (2) one-way or two-way ANOVA followed by Tukey's, Sidak's, or Dunnett's multiple comparison test, or (3) Kruskal–Wallis tests with uncorrected Dunn's test where multiple comparisons should be adjusted (unless otherwise indicated). Data were plotted using GraphPad Prism 7 software as mean values, with error bars representing standard deviation. For categorical variables, statistical analysis was done using Fisher's exact test or χ^2-test (two-sided). Representative western blot and PCR results are shown from three or more than three biologically independent experiments. Representative flow cytometry results and cytomorphology are shown from biological replicates ($n \geq 3$). $^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$, respectively, unless otherwise specified.

mRNA isolation, sequencing, and analysis. RNA was extracted as shown above. Poly(A)-selected, unstranded Illumina libraries were prepared with a modified TruSeq protocol. 0.5× AMPure XP beads were added to the sample library to select for fragments <400 bp, followed by 1× beads to select for fragments >100 bp. These fragments were then amplified with PCR (15 cycles) and separated by gel electrophoresis (2% agarose). DNA fragments 300 bp in length were isolated and sequenced on an Illumina HiSeq 2000 (about 100 × 10^6 101-bp reads per sample).

Primary samples from the Manchester Cancer Research Centre Haematological Malignancies Biobank with known *IDH2*/*SRSF2* mutation genotype were FACS-sorted to enrich for blasts on a FACS Aria III sorter using a panel including the following antibodies (all mouse anti-human): CD34-PerCP (8G12, BD); CD117-PECy7 (104D2, eBioscience); CD33-APC (P67.6, BioLegend); HLA-DR-FITC (L243, BioLegend); CD13-PE (L138, BD); CD45-APC-H7 (2D1, BD). RNA was extracted immediately using a Qiagen Micro RNeasy kit. All RNA samples had RIN values > 8. Poly(A)-selected, strand-specific SureSelect (Agilent) mRNA libraries were prepared using 200 ng RNA according to the manufacturer's protocol. Libraries were pooled and sequenced (2 × 101 bp paired end) to >100 million reads per sample on two HiSeq 2500 high throughput runs before retrospective merger of FASTQ files for downstream alignment and splicing analysis as described below. Transcriptional analysis was done using gene set enrichment analysis (GSEA)[34].

Publicly available RNA-seq data. Unprocessed RNA-seq reads of TCGA and Leucegene datasets (patients with AML) were downloaded from NCI's Genomic Data Commons Data Portal (GDC Legacy Archive; TCGA-LAML dataset) and NCBI's Sequence Read Archive (SRA; accession number SRP056295). The TCGA dataset consists of paired-end 2 × 50-bp libraries, with an average read count of 76.92 M. The Leucegene dataset consists of paired-end 2 × 100-bp libraries, with an average read count of 50.40 M per sample. The RNA-seq samples in the Leucegene dataset have 1–3 sequencing runs (about 50 M each run), and only one run was used to represent each RNA-seq sample.

Genome and splice junction annotations. Human assembly hg38 (GRCh38) and Ensembl database (human release 87) were used as the reference genome and gene annotation, respectively. RNA-seq reads were aligned by using 2-pass STAR 2.5.2a[35]. Known splice junctions from the gene annotation and new junctions identified from the alignments of the TCGA dataset were combined to create the database of alternative splicing events for splicing analysis.

Mutational analysis for the RNA-seq data. Samtools (1.3.1) was used to generate VCF files for seven target genes: *IDH1*, *IDH2*, *TET2*, *SF3B1*, *SRSF2*, *U2AF1*, and *ZRSR2* with mpileup parameters (-Bvu). The VCF files were further processed by our in-house scripts to filter out mutations with a variant allele frequency (VAF) lower than 15%. The filtered VCF files were used for variant effect predictor (v.89.4) to annotate the consequences of the mutations. We defined control patient samples as those without mutations in the seven target genes, *IDH2* mutated samples as those with only *IDH2* mutations but no mutations in the other six target genes, *SRSF2* mutated samples as those with only *SRSF2* mutations but no mutations in the other six target genes, Double-mutant samples as those with both *IDH2* and *SRSF2* mutations but no mutations in the other five target genes, and 'others' as those with mutations in *IDH1*, *TET2*, *SF3B1*, *U2AF1*, and *ZRSR2*.

Identification and quantification of differential splicing. The inclusion ratios of alternative exons or introns were estimated by using PSI-Sigma[25]. In brief, the new PSI index considers all isoforms in a specific gene region and can report the PSI value of individual exons in a multiple-exon-skipping or more complex splicing event. The database of splicing events was constructed based on both gene annotation and the alignments of RNA-seq reads. A new splicing event not known to the gene annotation is labelled as 'novel' and a splicing event with a reference transcript that is known to induce nonsense-mediated decay is labelled as 'NMD' in Supplementary Tables. The inclusion ratio of an intron retention isoform is estimated based on the median of 5 counts of intronic reads at the 1st, 25th, 50th, 75th and 99th percentiles in the intron. A splicing event is reported when both sample-size and statistical criteria are satisfied. The sample-size criterion requires a splicing event to have more than 20 supporting reads in more than 75% of the 2 populations in the comparison. For example, for a comparison of 130 control versus 6 *IDH2* mutant samples, a splicing event would be reported only when having more than 98 controls and 5 *IDH2* mutant samples with more than 20 supporting reads. In addition, a splicing event is reported only when it has more than 10% PSI change in the comparison and has a *P* value lower than 0.01.

To generate Fig. 4f, RNA-seq reads were mapped and PSI values were calculated using junction-spanning reads as previously described[36,37]. All reads mapping to the INTS3 introns (chr1:153,718,433-153,722,231; hg19) were extracted from the .bam files and the per-nucleotide coverage was calculated. Data from normal peripheral blood and BM mononuclear cells and $CD34^+$ cord blood cells are combined and shown as normal haematopoietic cells.

Motif enrichment and distribution. Motif analysis was done by using MEME SUITE[38]. In brief, the sequences of alternative exons of exon-skipping events were extracted from a given strand of the reference genome. The sequences were used as the input for MEME SUITE to search for motifs. One occurrence per sequence was set to be the expected site distribution. The width of motif was set to 5. The top 1 motif was selected on the basis of the ranking of *E*-value.

Heat map and sample clustering (differential splicing). The heat maps and sample clustering were done by using MORPHEUS (https://software.broadinstitute.org/morpheus/). The individual values in the matrix for the analysis were PSI values of a splicing event from a given RNA-seq sample. Splicing events were selected based on three criteria: (1) present in both TCGA and Leucegene datasets; (2) more than 15% PSI changes; and (3) false discovery rate smaller than 0.01. Unsupervised hierarchical clustering was based on one minus Pearson's correlation (complete linkage).

Correlation between global changes in splicing and DNA methylation. DNA methylation levels were determined by eRRBS and differentially spliced events were obtained from RNA-seq data. In Fig. 3e, Overlaps of differentially methylated regions of DNA with differential splicing was obtained by evaluating differential cytosine methylation in 500-bp segments of DNA at genomic coordinates at which differential RNA splicing were observed comparing AML with distinct *IDH2*/*SRSF2* genotypes shown (WT represents patients without mutations in *IDH1*/*IDH2*/spliceosomal genes).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

RNA-seq, ChIP–seq and eRRBS data have been deposited in the NCBI Sequence Read Archive under accession number SRP133673. Gel source data are shown in Supplementary Fig. 1. Other data that support the findings of this study are available from the authors upon reasonable request.

25. Lin, K. T. & Krainer, A. R. PSI-Sigma: a comprehensive splicing-detection method for short-read and long-read RNA-seq analysis. *Bioinformatics* btz438 (2019).
26. Moran-Crusio, K. et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20**, 11–24 (2011).
27. Shih, A. H. et al. Combination targeted therapy to disrupt aberrant oncogenic signaling and reverse epigenetic dysfunction in *IDH2*- and *TET2*-mutant acute myeloid leukemia. *Cancer Discov.* **7**, 494–505 (2017).
28. Georgiades, P. et al. *VavCre* transgenic mice: a tool for mutagenesis in hematopoietic and endothelial lineages. *Genesis* **34**, 251–256 (2002).
29. Zuber, J. et al. Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. *Nat. Biotechnol.* **29**, 79–83 (2011).
30. Lee, M. et al. Engineered split-TET2 enzyme for inducible epigenetic remodeling. *J. Am. Chem. Soc.* **139**, 4659–4662 (2017).
31. Kleppe, M. et al. Dual targeting of oncogenic activation and inflammatory signaling increases therapeutic efficacy in myeloproliferative neoplasms. *Cancer Cell* **33**, 29–43 (2018).
32. Maiques-Diaz, A. et al. Enhancer activation by pharmacologic displacement of LSD1 from GFI1 induces differentiation in acute myeloid leukemia. *Cell Rep.* **22**, 3641–3659 (2018).
33. Cheng, D. T. et al. Memorial Sloan Kettering–integrated mutation profiling of actionable cancer targets (MSK–IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn.* **17**, 251–264 (2015).

34. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
35. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
36. Dvinge, H. & Bradley, R. K. Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* **7**, 45 (2015).
37. Hubert, C. G. et al. Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. *Genes Dev.* **27**, 1032–1045 (2013).
38. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
39. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
40. Intlekofer, A. M. et al. Hypoxia induces production of l-2-hydroxyglutarate. *Cell Metab.* **22**, 304–311 (2015).
41. Dvinge, H. et al. Sample processing obscures cancer-specific alterations in leukemic transcriptomes. *Proc. Natl Acad. Sci. USA* **111**, 16802–16807 (2014).
42. Macrae, T. et al. RNA-seq reveals spliceosome and proteasome genes as most consistent transcripts in human cancer cells. *PLoS ONE* **8**, e72884 (2013).

Acknowledgements We thank D. L. Fei, Y. Huang, E. Wang, I. Aifantis, M. Patel, A. S. Shih, A. Penson, E. Kim, Y. R. Chung, B. H. Durham and H. Kunimoto for technical support, J. Wilusz for sharing recent data on integrator and B. J. Druker for sharing the Beat AML RNA-seq data. A.Y. is supported by grants from the Aplastic Anemia and MDS International Foundation (AA&MDSIF) and the Lauri Strauss Leukemia Foundation. A.Y. is a Special Fellow of The Leukemia and Lymphoma Society. A.Y., S.C.-W.L. and D.I. are supported by the Leukemia and Lymphoma Society Special Fellow Award. A.Y. and D.I. are supported by JSPS Overseas Research Fellowships. D.H.W. is supported by a Bloodwise Clinician Scientist Fellowship (15030). D.H.W. and K.B. are supported by fellowships from The Oglesby Charitable Trust. S.C.-W.L. is supported by the NIH/NCI (K99 CA218896) and the ASH Scholar Award. T.C.P.S. is supported by Cancer Research UK grant number C5759/A20971. E.J.W. is supported by grants from the CPRIT (RP140800) and the Welch Foundation (H-1889-20150801). R.K.B. and O.A.-W. are supported by grants from NIH/NHLBI (R01 HL128239) and the Department of Defense Bone Marrow Failure Research Program (W81XWH-16-1-0059). A.R.K. and O.A.-W. are supported by grants from the Starr Foundation (I8-A8-075) and the Henry & Marilyn Taub Foundation. O.A.-W. is supported by grants from the Edward P. Evans Foundation, the Josie Robertson Investigator Program, the Leukemia and Lymphoma Society and the Pershing Square Sohn Cancer Research Alliance.

Author contributions A.Y., K.-T.L., A.R.K. and O.A.-W. designed the study. A.Y., B.W., S.C.-W.L., J.-B.M., X.J.Z., H.C., R.E.M., D.I., T.R.A., K.B., F.S. and E.J.W. performed mouse experiments. K.-T.L. and M.A.R. performed RNA-seq analyses and minigene assays, respectively, under the supervision of A.R.K. A.P. performed DNA methylation and ChIP–seq analyses. T.R.A. and E.J.W. provided antibodies to detect integrator components and assays for snRNA cleavage. H.D., R.K.B. and F.A. performed RNA-seq analyses. D.H.W., T.C.P.S., D.P.W., S.d.B., V.P.-L., E.M.S. and R.L.L. provided clinical samples. D.H.W. and C.C. provided clinical correlative data for primary datasets. D.H.W. performed ChIP–seq experiments under the supervision of T.C.P.S. A.M.I. provided *Idh2*^{R140Q} knock-in mice. R.L.L. provided *Tet2* knockout mice. A.Y., K.-T.L., D.H.W. and O.A.-W. prepared the manuscript with help from all co-authors.

Competing interests A.M.I. has served as a consultant and advisory board member for Foundation Medicine. E.M.S. has served on advisory boards for Astellas Pharma, Daiichi Sankyo, Bayer, Novartis, Syros, Pfizer, PTC Therapeutics, AbbVie, Agios and Celgene and has received research support from Agios, Celgene, Syros and Bayer. R.L.L. is on the Supervisory Board of Qiagen and the Scientific Advisory Board of Loxo, reports receiving commercial research grants from Celgene, Roche and Prelude, has received honoraria from the speakers bureaus of Gilead and Lilly, has ownership interest (including stock and patents) in Qiagen and Loxo, and is a consultant and/or advisory board member for Novartis, Roche, Janssen, Celgene and Incyte. A.R.K. is a founder, director, advisor, stockholder and chair of the Scientific Advisory Board of Stoke Therapeutics and receives compensation from the company. A.R.K. is a paid consultant for Biogen; he is a member of the SABs of Skyhawk Therapeutics, Envisagenics BioAnalytics and Autoimmunity Biologic Solutions, and has received compensation from these companies in the form of stock. A.R.K. is a research collaborator of Ionis Pharmaceuticals and has received royalty income from Ionis through his employer, Cold Spring Harbor Laboratory. O.A.-W. has served as a consultant for H3 Biomedicine, Foundation Medicine, Merck and Janssen. O.A.-W. has received personal speaking fees from Daiichi Sankyo. O.A.-W. has received previous research funding from H3 Biomedicine unrelated to the current manuscript. D.I., R.K.B. and O.A.-W. are inventors on a provisional patent application (patent number FHCC.P0044US.P) applied for by Fred Hutchinson Cancer Research Center on the role of reactivating BRD9 expression in cancer by modulating aberrant *BRD9* splicing in *SF3B1* mutant cells.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1618-0.
Correspondence and requests for materials should be addressed to O.A.-W.
Peer review information *Nature* thanks Rotem Karni and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.

Appendix 4

Derepression of the Iroquois Homeodomain Transcription Factor Gene IRX3 Confers Differentiation Block in Acute Leukemia.

Tim D. D. Somerville, Fabrizio Simeoni, John A Chadwick, Emma L. Williams, Gary J. Spencer, Katalin Boros, Christopher Wirth, Eleni Tholouli, Richard J. Byers, Tim C. P. Somervaille. (2018)

Cell Reports 574, 273-281.

A research article written by Tim D. D. Somerville, where I am a contributing author.

OPEN ACCESS
CellPress

Cell Reports
Article

Derepression of the Iroquois Homeodomain Transcription Factor Gene *IRX3* Confers Differentiation Block in Acute Leukemia

Tim D.D. Somerville,[1] Fabrizio Simeoni,[1] John A. Chadwick,[1] Emma L. Williams,[1] Gary J. Spencer,[1] Katalin Boros,[2] Christopher Wirth,[3] Eleni Tholouli,[4] Richard J. Byers,[2] and Tim C.P. Somervaille[1,5,*]
[1]Leukaemia Biology Laboratory, Cancer Research UK Manchester Institute, The University of Manchester, Manchester M20 4BX, UK
[2]Department of Histopathology, Manchester University NHS Foundation Trust, Manchester M13 9WL, UK
[3]Applied Computational Biology and Bioinformatics Group, Cancer Research UK Manchester Institute, The University of Manchester, Manchester M20 4BX, UK
[4]Department of Haematology, Manchester University NHS Foundation Trust, Manchester M13 9WL, UK
[5]Lead Contact
*Correspondence: tim.somervaille@cruk.manchester.ac.uk
https://doi.org/10.1016/j.celrep.2017.12.063

SUMMARY

The Iroquois homeodomain transcription factor gene *IRX3* is expressed in the developing nervous system, limb buds, and heart, and transcript levels specify obesity risk in humans. We now report a functional role for *IRX3* in human acute leukemia. Although transcript levels are very low in normal human bone marrow cells, high *IRX3* expression is found in ~30% of patients with acute myeloid leukemia (AML), ~50% with T-acute lymphoblastic leukemia, and ~20% with B-acute lymphoblastic leukemia, frequently in association with high-level HOXA gene expression. Expression of *IRX3* alone was sufficient to immortalize hematopoietic stem and progenitor cells (HSPCs) in myeloid culture and induce lymphoid leukemias *in vivo*. *IRX3* knockdown induced terminal differentiation of AML cells. Combined *IRX3* and *Hoxa9* expression in murine HSPCs impeded normal T-progenitor differentiation in lymphoid culture and substantially enhanced the morphologic and phenotypic differentiation block of AML in myeloid leukemia transplantation experiments through suppression of a terminal myelomonocytic program. Likewise, in cases of primary human AML, high *IRX3* expression is strongly associated with reduced myelomonocytic differentiation. Thus, tissue-inappropriate derepression of *IRX3* contributes significantly to the block in differentiation, which is the pathognomonic feature of human acute leukemias.

INTRODUCTION

The cardinal pathologic feature of the acute leukemias is a block to normal blood cell differentiation that results in an accumulation in the bone marrow (BM) of incompletely differentiated blast cells and failure of normal hematopoiesis (Wiseman et al., 2014). Although the spectrum of mutations associated with these diseases is now well established, the biologic basis of how mutations interact with one another to establish the pathognomonic differentiation block is less well understood. We recently reported that the Forkhead box transcription factor FOXC1 is misexpressed in approximately 20% of patients with acute myeloid leukemia (AML), in particular in those cases exhibiting high HOXA/B gene expression (Somerville et al., 2015). *FOXC1* is neither required for nor expressed in normal hematopoietic cells but is essential for normal development of mesenchymal tissues such as the skeleton, heart, and eye and for the normal function of BM niche cells (Omatsu et al., 2014). Its misexpression in leukemic hematopoiesis contributes to a block in differentiation along both monocytic and B cell lineages and is associated with inferior survival. Given the lack of one-to-one correlation with any specific mutation, tissue-inappropriate derepression of *FOXC1* is paradigmatic for a non-mutational mechanism contributing to cellular transformation in myeloid malignancy.

Whether tissue-inappropriate misexpression of other transcription factors contributes to the differentiation block of leukemia is not known. One candidate is the Iroquois homeobox transcription factor gene *IRX3*, which, like *FOXC1*, is expressed in a significant proportion of patients with AML (Somerville et al., 2015) but minimally expressed in both normal hematopoietic stem and progenitor cells (HSPCs) and mature blood cells (ENCODE data; Zhou et al., 2011). IRX3 is a member of the three-amino-acid-loop-extension (TALE) superfamily of homeodomain transcription factors, which also includes MEIS1 and PBX1 (Mukherjee and Bürglin, 2007). In embryogenesis, it is strongly expressed in the developing nervous system, as well as in mesoderm-derived tissues such as the limb buds, kidney, and heart (Houweling et al., 2001). Of note, the developmental expression pattern of the *Irx3* paralog *Irx5*, which sits in the same 2MB topologically associating domain, is strikingly similar (Cohen et al., 2000; Claussnitzer et al., 2015). These genes exhibit functional redundancy because although *Irx3*-null and *Irx5*-null mice are viable and fertile, mice lacking both genes die in utero because of severe cardiac and skeletal defects (Zhang et al., 2011; Gaborit et al., 2012; Li et al., 2014; Smemo et al., 2014). Interestingly, non-coding variation in an enhancer

region 500 kb downstream of *IRX3* provides the strongest genetic association with risk for human obesity. Pertinent to this, adult *Irx3*-null mice display a 25%–30% reduction in body weight due to loss of fat mass, increased basal metabolic rate, and browning of white adipose tissue, attributable to loss of hypothalamic (Smemo et al., 2014) or preadipocyte (Claussnitzer et al., 2015) *Irx3* expression. The rs1421085 single-nucleotide variant present in the obesity risk region dictates the extent of local recruitment of ARID5B to the *IRX3* enhancer, with consequent regulation of *IRX3* expression (Claussnitzer et al., 2015).

Whether *IRX3* has a role in human malignancy is unclear. One study reported that *IRX3* is strongly expressed in colorectal adenomas in comparison with normal mucosa and negatively regulates TGF-β signaling in colorectal cancer cell lines (Martorell et al., 2014). However, little else is known. Given this, and the observation that *IRX3* is highly expressed in a subset of AML patients, we evaluated whether *IRX3* has a functional role in acute leukemia.

RESULTS

IRX3 Is Frequently Co-expressed with HOX Genes in Human AML

To ascertain the frequency and extent of *IRX3* expression in AML and in flow-sorted normal human BM cell populations, we performed both qPCR and analyses of published datasets. In a Dutch cohort of AML patients treated intensively with anthracycline-based chemotherapy on the Hemato-Oncologie voor Volwassenen Nederland (HOVON) protocols, *IRX3* transcripts were detected at high level (i.e., with a probeset [229638_at] value of $\log_2 > 7.1$, approximating to a value among the top 25% of array probeset values) in 159 of 461 bulk presentation samples (34%) (Wouters et al., 2009) (Figure 1A). Likewise, in The Cancer Genome Atlas Research Network series, 49 of 163 cases (30%) expressed *IRX3* at high level (Cerami et al., 2012; Ley et al., 2013) (Figure S1A). In flow-sorted populations of AML cells with immature immunophenotypes, *IRX3* was highly expressed (i.e., among the top 25% of array probeset values) in 33% (Saito et al., 2010) (Figure S1B), 58% (Kikushige et al., 2010) (Figure S1C), and 19% (Goardon et al., 2011) (Figure S1D) of samples. Concomitant microarray profiling of normal human immunophenotypic HSPCs suggested low or absent *IRX3* expression (Figures S1B–S1D). In keeping with this, our qPCR analyses revealed very low levels of *IRX3* transcripts in all normal BM populations tested, but in 10 of 29 AML samples (34%), *IRX3* expression was increased at least 250-fold over levels observed in the lowest expressing AML sample (Figure 1B; Table S1). *IRX3* transcript levels were higher in normal human $CD45^{neg}$ BM stromal cells than in normal BM cell populations (Figure 1B) but not as high as those observed in many AML samples. Given that expression of *IRX3* and *IRX5* are co-regulated in development, we also performed qPCR for *IRX5* transcripts. We did not detect *IRX5* expression in normal human BM cell populations (data not shown) but did detect low-level expression in 11 of 28 AML samples (39%), typically in cases with high *IRX3* expression (Figure S1E). In the Dutch AML cohort, 44 of 461 cases (10%) expressed *IRX5* at high level (i.e., probeset [210239_at] value of $\log_2 > 7.1$), and in every case there was also high IRX3 expression (data not shown). Thus, the set of $IRX5^{high}$ AML cases is a subset of the group of $IRX3^{high}$ cases.

To confirm that IRX3 protein was also expressed in AML, we performed immunohistochemical staining of a trephine biopsy tissue array. The array included 58 samples from patients with AML and 9 samples from patients with non-malignant BM conditions (Table S2). Immune staining was H-scored by blind evaluation. Strong or moderate nuclear immune staining (i.e., H score $\geq$ 80) was observed in 20 of 58 cases (34%). Weak or absent immune staining was observed in 38 of 58 AML cases (66%) and in all non-malignant cases (Figures 1C and 1D).

High *IRX3* expression in AML was strongly and positively associated with the presence of an *NPM1* mutation or a *FLT3* internal tandem duplication (Tables S3 and S4) (Wouters et al., 2009; Ley et al., 2013). A strong positive association with acute promyelocytic leukemia (APML) was also noted (Table S3). There were weaker positive associations with intermediate cytogenetic risk, normal karyotype and the presence of an *IDH1* mutation, t(6;9), or an *MLL* gene rearrangement. 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 mutations in *NRAS*, *TP53*, or *RUNX1*, or a double *CEBPA* mutation and the presence of high *MECOM* expression (Tables S3 and S4). Although detailed genotyping of the tissue array samples was not performed, there was nevertheless a significant association of normal karyotype with strong or moderate IRX3 expression (14 of 17 available karyotypes were normal where strong or moderate nuclear immune staining was present versus 17 of 34 with weak or absent staining, $p = 0.035$, Fisher's exact test) (Table S2).

To identify genes co-expressed with *IRX3* in human AML, we next compared $IRX3^{high}$ AMLs (probeset 229638_at value of $\log_2 > 7.1$) with $IRX3^{low}$ AMLs (probeset 229638_at value of $\log_2 < 6.1$) (Wouters et al., 2009) and found *HOXA9* and *HOXA5* to be the most differentially expressed transcription factor genes in the $IRX3^{high}$ group whether (data not shown) or not (Figure 1E) cases of APML (which do not express HOX genes) were included. Of the non-APML cases, 133 of 138 $IRX3^{high}$ samples (96%) exhibited high *HOXA9* expression (i.e., probeset value of $\log_2 > 7.1$), and of the $HOXA9^{high}$ samples, 133 of 319 (42%) expressed high levels of *IRX3* (Figure 1F). Of the five $IRX3^{high}$ $HOXA9^{low}$ samples, three expressed one or more alternative HOX genes at significant level (i.e., $\log_2 > 7.1$), indicating that overall, 98.6% of $IRX3^{high}$ non-APML samples exhibited HOX gene co-expression. Similar results were observed in analyses of The Cancer Genome Atlas (TCGA) dataset (Figure S1F) (Ley et al., 2013) and qPCR analysis of our own samples (Figure 1G). Taken together, these data demonstrate that in human AML, the Iroquois homeodomain transcription factor gene *IRX3*, which is minimally expressed in normal hematopoiesis, is often misexpressed in conjunction with high HOX gene expression, as well as in APML. The former association suggests an explanation for the statistically significant associations of *IRX3* expression with the presence of an *NPM1* mutation, *MLL* gene rearrangements, and a t(6;9) because these molecular subtypes are associated with high HOX gene expression.

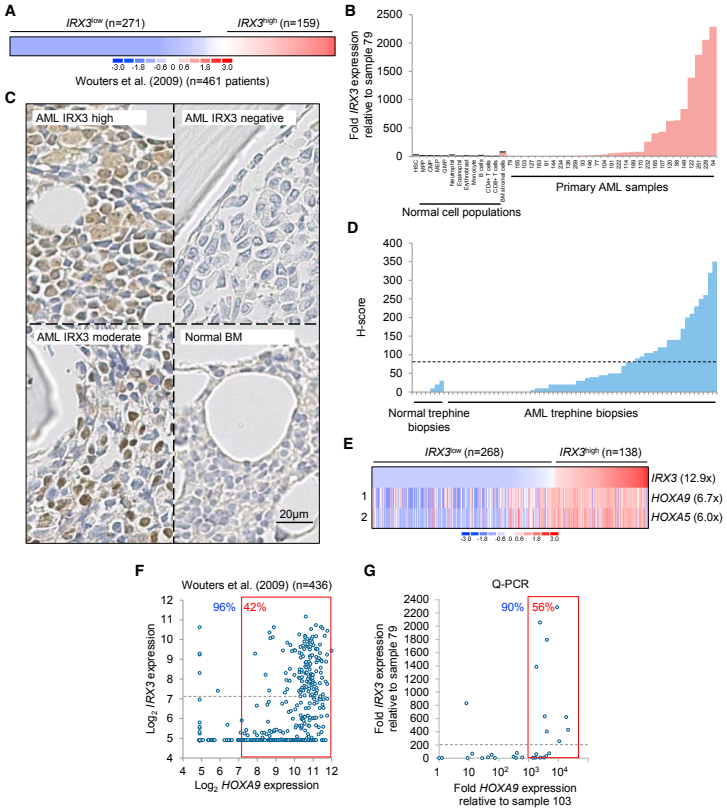


Figure 1. *IRX3* Expression in Human AML

(A) *IRX3* expression in bulk human AML samples.

(B) Relative *IRX3* expression in bulk primary human AML samples (n = 29) and normal human BM cell populations (n = 4 separate individuals per cell type; error bars = SEM). BM, bone marrow; CMP, common myeloid progenitor; Eosin, eosinophils; EryB, erythroblast; GMP, granulocyte-macrophage progenitor; HSC, hematopoietic stem cell ($CD34^+38^-90^+45RA^-Lin^-$); MEP, megakaryocyte-erythrocyte progenitor; Mono, monocytes; MPP, multipotent progenitor; Neut, neutrophils. AML sample numbers refer to Biobank identifier.

(C) Representative images of IRX3 immune staining of human BM trephine biopsies.

(D) H scores for IRX3 immune staining. Dashed line indicates cutoff value for moderate/strong immune staining.

(E) Heatmap shows $IRX3^{high}$ versus $IRX3^{low}$ AML cases (excluding APML) (Wouters et al., 2009) and the most differentially expressed transcription factor genes with mean fold change.

(F and G) Scatterplots show *IRX3* versus *HOXA9* expression in primary AML samples as determined by (F) array values or (G) qPCR. For scatterplots, 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 boxes) additionally exhibiting high *IRX3* expression (above the dotted gray line).

See also Figure S1 and Tables S1–S4.

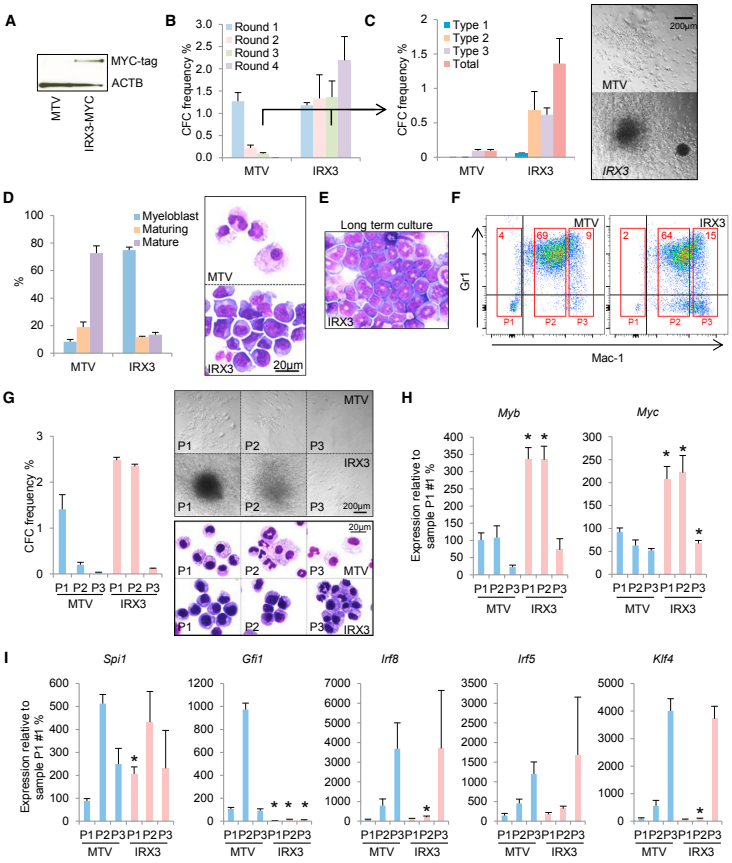


Figure 2. IRX3 Enhances Clonogenic Potential in Normal HSPCs

(A) MYC-tagged IRX3 expressed in murine KIT$^+$ BM HSPCs 72 hr after retroviral infection. MTV, empty vector.

(B) Mean + SEM colony-forming cell (CFC) frequencies during serial replating (n = 3).

(C) Bar chart (left) shows mean + SEM colony types in round 3 of culture (n = 3). Type 1, tightly packed colonies, contain blast cells only; type 2, contain mixed population of blasts and mature cells; type 3, contain mature cells only. Image (right) shows representative colonies.

(D) Bar chart (left) shows mean + SEM percentage of the indicated cell types in cytospin preparations after 7 days of culture of retrovirally infected murine KIT$^+$ BM cells in liquid conditions supporting myeloid development (n = 3). Representative images (right) are shown.

(E) Representative image of *IRX3*-expressing mouse BM cells after 5 weeks in liquid culture.

(F) Representative flow cytometry plots for the indicated cell surface markers after 7 days in liquid culture; red boxes indicate cell sorting gates. Numbers within boxes indicate percentage of cells with the indicated phenotype.

(legend continued on next page)

IRX3 Promotes Serial Replating of Normal BM Stem and Progenitor Cells

To evaluate whether misexpressed *IRX3* has functional consequences, we performed expression experiments in murine KIT$^+$ BM HSPCs (Figure 2A). Serial replating assays in conditions supporting myeloid lineage differentiation demonstrated that in comparison with control cells, *IRX3*$^+$ cells exhibited enhanced clonogenic activity (Figures 2A and 2B). In the third round of culture, control cells exclusively formed colonies with type 3 morphology (i.e., diffuse colonies containing mature macrophages), whereas *IRX3*$^+$ cells in addition formed type 1 (i.e., tightly packed colonies containing blast cells) and type 2 colonies (i.e., mixed colonies containing blast cells and mature cells) (Figure 2C). Reflecting these observations, *IRX3*$^+$ BM HSPCs cultured in liquid conditions (with interleukin-3 [IL-3], IL-6, granulocyte-macrophage colony-stimulating factor [GM-CSF], and stem cell factor [SCF]) for 7 days following retroviral transduction exhibited significantly impaired morphologic differentiation in comparison with control cells (Figure 2D) and could readily be grown on for at least 5 weeks (Figure 2E). The great majority of day 7 *IRX3*$^+$ and control cells in liquid culture exhibited a Mac1$^+$Gr1$^+$ immunophenotype, although there was a modest but significant reduction in the percentage of Gr1$^+$ cells in the *IRX3*$^+$ condition (Figure S1G). Flow-sorting analyses confirmed, as expected, that the clonogenic activity of control day 7 cells was near exclusively confined to cells with a Mac-1^{neg} immunophenotype. In contrast, IRX3$^+$ cells with a Mac-1^{intermediate} immunophenotype also exhibited strong clonogenic activity (Figures 2F and 2G). qPCR analysis of genes associated with self-renewal in myeloid leukemia (Somervaille et al., 2009) demonstrated significant upregulation of *Myc* and *Myb* in *IRX3*$^+$ populations with significant clonogenic potential (P1 and P2). Furthermore, there was significantly reduced expression in the aberrantly clonogenic *IRX3*$^+$ P2 population of transcription factor genes such as *Gfi1*, *Irf8*, and *Klf4*, which are associated with myelomonocytic lineage differentiation (Figure 2I). Thus, expression of *IRX3* in normal BM HSPCs in conditions supporting myeloid lineage differentiation confers a morphologic and functional differentiation block resulting in sustained clonogenic activity, elevated expression of self-renewal genes *Myc* and *Myb*, and reduced expression of myelomonocytic differentiation genes *Gfi1*, *Irf8*, and *Klf4*.

IRX3 Cooperates with HOXA9 to Enhance Differentiation Block in AML

Given the strong association of *IRX3* and *HOXA9* expression in human AML, we next evaluated the consequences of *IRX3* co-expression in *Hoxa9*-expressing murine BM HSPCs. Cells were infected in pairwise combinations with retroviral vectors expressing *Hoxa9*, *IRX3*, or a control vector (to generate *Hoxa9*/*IRX3* and *Hoxa9*/MTV cells, respectively) and serially replated in semisolid culture. Consistent with *IRX3* co-expression conferring an enhanced differentiation block to *Hoxa9*-immortalized HSPCs, by the fourth round of culture, *Hoxa9*/*IRX3* cells formed significantly more colonies than *Hoxa9*/MTV cells (Figure 3A), and these were significantly more likely to exhibit type 1 blast-like morphology (Figure 3B). Of note, *Hoxa9*/*IRX3* type 1 colonies were on average 25% smaller in cross-sectional area than *Hoxa9*/MTV type 1 colonies (Figure 3C); this was explained by there being a significantly reduced proportion of *Hoxa9*/*IRX3* cells in the SG_2M phase of the cell cycle (Figure S1H). The immunophenotype of *Hoxa9*/*IRX3* and *Hoxa9*/MTV cells was similar (Figure S1I).

Transplantation of *Hoxa9*-expressing murine BM HSPCs into irradiated syngeneic recipients results in short latency AMLs, which exhibit significant myelomonocytic maturation (Kroon et al., 1998; Somerville et al., 2015). To determine whether co-expression of IRX3 influences leukemia cell differentiation *in vivo*, we transplanted *Hoxa9*/*IRX3* and *Hoxa9*/MTV double-transduced BM HSPCs into irradiated congenic recipients. Consistent with the cell-cycle status of *in vitro* transformed *Hoxa9*/*IRX3* cells, analysis of blood at 4, 8, and 12 weeks post-transplantation demonstrated reduced donor:recipient chimerism in blood in *Hoxa9*/*IRX3* versus *Hoxa9*/MTV recipients (Figure 3D), and *Hoxa9*/*IRX3* recipient mice exhibited delayed onset of donor-derived AML in comparison with *Hoxa9*/MTV recipients (median 125 versus 270 days) (Figure 3E). Mice from both cohorts presented with substantially elevated blood leucocyte counts (Figure S2A), hepatosplenomegaly (Figure S2B), and effacement of BM due to infiltration by leukemia cells (Figures S2C–S2F).

In keeping with a model whereby *IRX3* misexpression *in vivo* blocks myeloid lineage differentiation, evaluation of the lineage composition of donor-derived cells in blood prior to development of AML demonstrated a significant proportional reduction in myeloid and a proportional increase in B-lineage differentiation at 4 and 12 weeks post-transplantation in *Hoxa9*/*IRX3* versus *Hoxa9*/MTV recipients (Figures 3F and S3A). Once mice developed full-fledged AML, this was further confirmed by analysis of blood smear and BM cytospin morphology: *Hoxa9*/*IRX3* recipients developed leukemias exhibiting significantly greater differentiation block in comparison with *Hoxa9*/MTV controls, as demonstrated by the proportion of blast cells being on average twice as high in the former versus the latter and the proportion of more differentiated leukemia cells being approximately half as much (Figures 3G and 3H). Flow cytometry analysis of leukemic BM cells confirmed donor origin and myeloid lineage (Figures 3I, 3J, and S3B) and also revealed a distinctive immunophenotype: *Hoxa9*/*IRX3* AMLs expressed significantly lower levels of both CD45.1 (Figures 3I and S3C) and the monocyte/macrophage differentiation marker F4/80 versus *Hoxa9*/MTV controls. There were also significantly fewer Mac1$^+$Gr1$^-$ leukemia cells in *Hoxa9*/*IRX3* recipients and, on average, double the percentage of cells positive for the stem and progenitor

(G) Bar chart (left) shows mean + SEM CFC frequencies for the indicated flow-sorted cell populations enumerated after 5 days (n = 3). Images (right) show representative colonies (top) and cytospins (bottom).
(H and I) Bar charts show mean + SEM relative transcript expression for the indicated genes in the indicated cell populations: (H) *Myb* and *Myc*; (I) *Spi1*, *Gfi1*, *Irf8*, *Irf5*, and *Klf4*. *p $\leq$ 0.05 for the indicated *IRX3*$^+$ population versus the equivalent control population (unpaired t test).
See also Figure S1.

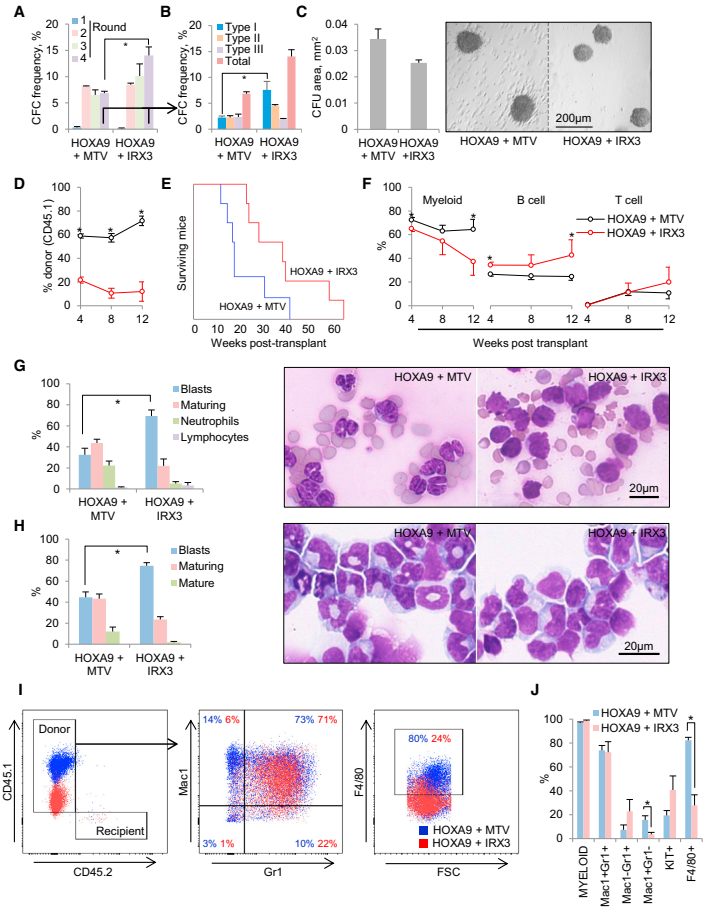


Figure 3. IRX3 Expression Confers an Enhanced Myeloid Differentiation Block in *Hoxa9*⁺ AML

(A) Bar chart shows mean + SEM colony-forming cell (CFC) frequencies during serial replating of murine KIT⁺ BM cells co-transduced with the indicated retroviral or control (MTV) expression vectors (n = 3). *p < 0.05 by unpaired t test.

(B) Bar chart shows mean + SEM colony types in round 4 (n = 3). *p < 0.05 by unpaired t test.

(C) Bar chart (left) shows mean + SEM area of type 1 colonies (n = 26–40) from (B) and representative images (right).

(D–J) Murine CD45.1⁺ KIT⁺ BM cells were infected in pairwise combination with retroviral vectors, and 96 hr later 10^6 drug-resistant cells were transplanted into CD45.2⁺ irradiated congenic recipients. (D) Line graph shows mean + SEM percentage donor-derived CD45.1⁺ cells in blood at the indicated times post-transplantation. (E) Survival curves of transplanted mice (n = 7 per cohort). (F) Line graphs show the mean + SEM percentage contribution of donor-derived cells in blood to the indicated lineages at the indicated times post-transplantation. (G) Bar chart (left) shows mean + SEM percentage leucocyte type in blood at death, as determined by morphologic analysis of blood smears (n = 4 or 5 per cohort). Representative images (right) are shown. (H) Bar chart (left) shows mean + SEM

(legend continued on next page)

marker KIT (Figure 3J), although in view of heterogeneous levels of expression, this difference did not reach statistical significance. *Hoxa9/IRX3* AML cells induced leukemias in secondarily transplanted recipients (Figure S3D), and high-level *IRX3* expression was readily detected in *Hoxa9/IRX3* AML BM cells (Figure S3E). Interestingly, in contrast to observations *in vitro*, cell-cycle analysis of leukemia cells from BM and spleen of leukemic mice demonstrated no difference in the fraction of cycling cells between the cohorts (Figure S3F), although in both cases, the SG2M fraction was substantially lower in the *in vivo* setting in comparison with the *in vitro* setting (i.e., 10%–20% versus 40%–50%).

Altogether our observations demonstrate that co-expression of *IRX3* modulates the phenotypic consequences of *Hoxa9* expression both *in vitro* and *in vivo*, conferring a significantly enhanced myeloid lineage differentiation block.

Functional Contribution of IRX3 to Differentiation Block in AML Cells

To further confirm that tissue-inappropriate expression of *IRX3* contributes functionally to the differentiation block in AML, we performed knockdown (KD) experiments in human THP1 AML cells, which exhibit the highest levels of *IRX3* expression among AML cell lines we tested (Figure S4A). *IRX3* KD led to loss of clonogenic potential (Figures 4A–4C), which was due to induction of differentiation, as evidenced by upregulation of the myeloid differentiation markers CD11b, CD14, and CD86 (Figures 4D and 4E) and morphologic analysis (Figure 4F). The proportion of apoptotic cells was unaffected (Figure S4B). To confirm that the observed phenotype was an on-target consequence of *IRX3* KD, similar experiments were performed using KD construct #2, which targets the 3′UTR region of *IRX3*, in a THP1 line engineered to express an *IRX3* cDNA that lacks it. Sustained *IRX3* expression in THP1 cells infected with lentiviral vectors expressing KD construct #2 was confirmed (Figures 4G and 4H), as was rescue of the differentiation phenotype (i.e., loss of clonogenic potential and upregulation of the differentiation marker CD86) (Figures 4I, S4C, and S4D). Similar experiments in murine MLL-AF9 AML cells, which also express *Irx3*, gave similar results: KD cells exhibited loss of clonogenic potential and terminal monocyte/macrophage lineage differentiation (Figures S4E–S4G). *IRX3* KD in a range of additional human AML cell lines also led to loss of clonogenic potential in many cases, but not all (Figure S4H). Of note, KD construct #2 had no effect on the clonogenic potential of K562 cells, which do not express *IRX3* (Figure S4A). Furthermore, the formation of colony-forming unit granulocyte monocyte (CFU-GM) and colony-forming unit monocyte/macrophage (CFU-M) from normal human $CD34^+$ cells (which express *IRX3* at very low levels) was unaffected by *IRX3* KD construct #2 (Figures S4I and S4J), although, for unclear reasons, there was a reduction in the formation of erythroid burst-forming units.

Thus, in keeping with *in vitro* and *in vivo* murine experiments, misexpressed *IRX3* contributes to the differentiation block in AML cells and sustains clonogenic potential.

IRX3 Expression in AML Represses a Myelomonocytic Differentiation Program

To evaluate the consequences for the transcriptome of co-expression of *IRX3* with *Hoxa9*, we performed RNA sequencing of flow-sorted KIT^+Gr1^+ leukemia cells recovered from the BM of sick mice, three from each cohort. Cells with this immunophenotype are enriched for leukemia-initiating activity in *Hoxa9/Meis1* murine leukemias (Gibbs et al., 2012). Analysis of 8,954 protein-coding genes that passed threshold criteria (i.e., expression >0.5 reads per kilobase per million mapped reads [RPKM] in at least one sample) revealed that 197 were upregulated and 403 downregulated by at least 2-fold in *Hoxa9/IRX3* versus *Hoxa9/MTV* leukemias (Figure 5A). Gene Ontology analysis revealed significant enrichment within the *Hoxa9/IRX3* downregulated gene set of biological process terms such as "immune response," "leucocyte activation," and "defense response" (Table S5), suggesting that IRX3-repressed genes are associated with mature myeloid cells. At similar levels of significance, there were no enriched terms among the *Hoxa9/IRX3* upregulated gene set. Gene set enrichment analysis (GSEA) confirmed that co-expression of *IRX3* with *Hoxa9* led to repression of a mature myeloid lineage program in KIT^+Gr1^+ leukemia cells, in keeping with morphologic analysis (Figures 3G and 3H). Among genes downregulated in *Hoxa9/IRX3* versus *Hoxa9/MTV* AML cells, there was significant enrichment for genes highly expressed in both mature monocytes and mature neutrophils (i.e., myelomonocytic genes) (Figure 5B; Tables S6 and S7).

GSEA also revealed significant overlap with expression patterns observed in murine models of leukemia associated with high-level expression of *Hoxa9* and *Meis1*. Mixed-lineage leukemia (MLL) fusion oncogenes require *Hoxa9* and *Meis1* to properly transform HSPCs and to establish a leukemia cell hierarchy. We observed that in *Hoxa9/IRX3* versus *Hoxa9/MTV* AML cells, there was significant enrichment of expression of the transcriptional sub-program that contributes to blocked differentiation and enhanced self-renewal of leukemia stem cells (LSCs) in murine MLL leukemias (LSC maintenance program; Somervaille et al., 2009) (Figure 5C; Tables S6 and S7). Likewise, the set of genes anti-correlated with the LSC maintenance program (i.e., associated with differentiated leukemia cells, downstream of the LSC) was strongly repressed in *Hoxa9/IRX3* versus *Hoxa9/MTV* AML cells (Figure 5C; Tables S6 and S7). Similar positive and negative enrichments were observed for gene sets derived from transformed progenitor cells with sustained (HOXA9_MEIS1 POSITIVELY REGULATED) or withdrawn (HOXA9_MEIS1 NEGATIVELY REGULATED) *Hoxa9/Meis1* dual expression (Hess et al., 2006) (Figure S5A). All together, these analyses demonstrate that co-expression of *IRX3* with *Hoxa9* enhances expression of genes previously associated with LSC

percentage cell type in BM at death (n = 5 per cohort). Representative images (left) are shown. (I) Representative flow cytometry plots from (H). (J) Bar chart shows the mean + SEM percentage of donor-derived cells positive for the indicated cell surface markers in BM of leukemic mice, as determined by flow cytometry. *$p < 0.05$ by unpaired t test.
See also Figures S1–S3.

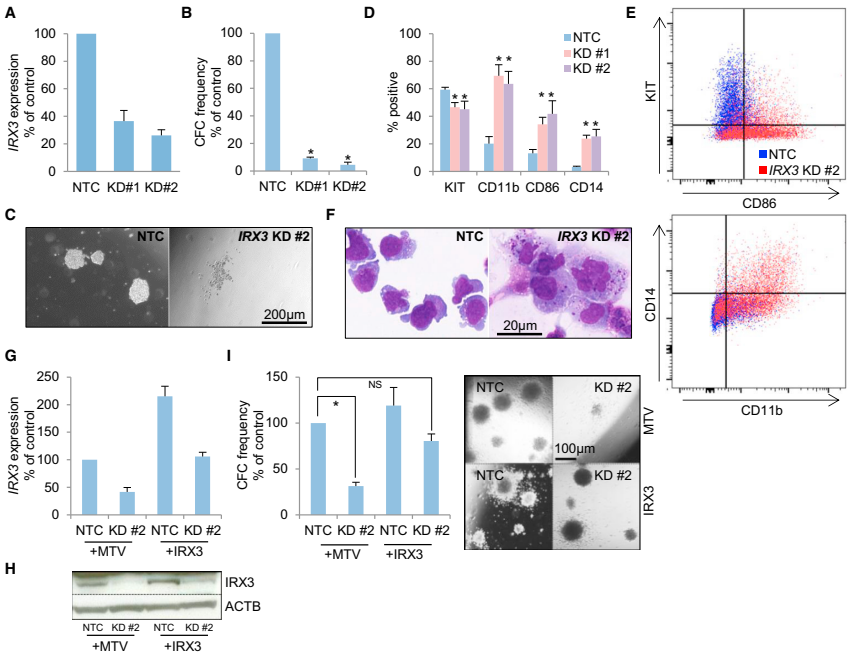


Figure 4. IRX3 Sustains the Differentiation Block and Clonogenic Potential of AML Cells

Human THP1 AML cells were infected with lentiviral vectors targeting *IRX3* for KD or a non-targeting control vector (NTC).

(A) Bar chart shows mean + SEM relative transcript expression in *IRX3* KD versus control cells (n = 3) after 48 hr.

(B) Bar chart shows the mean + SEM colony-forming cell (CFC) frequencies of KD cells relative to control cells enumerated after 10 days in semi-solid culture (n = 4).

(C) Representative images of colonies of cells from (B).

(D) Bar chart shows mean + SEM percentage of cells positive for the indicated cell surface markers, as determined by flow cytometry analysis 6 days following initiation of KD (n = 4). *p ≤ 0.05 using one-way ANOVA with Fisher's least significant difference post hoc analysis for KD conditions versus NTC.

(E) Representative flow cytometry plots from (D).

(F) Representative images of cytospins of cells from (D).

(G) Bar chart shows mean + SEM relative transcript expression in THP1 AML cells expressing either *IRX3* or a control retroviral vector (MTV) in *IRX3* KD#2 cells relative to control cells (n = 3).

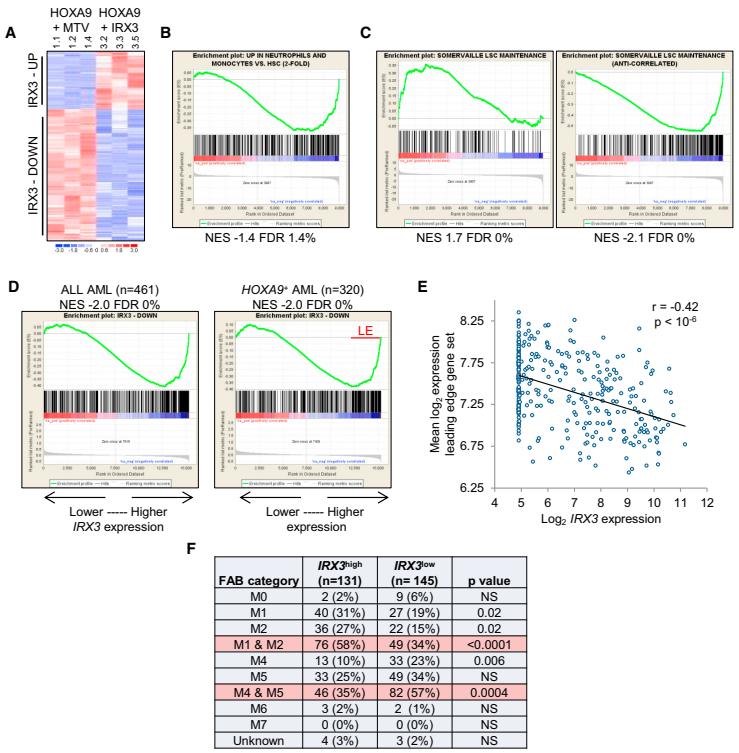
(H) Western blot shows expression of the indicated proteins in the indicated conditions.

(I) Bar chart (left) shows mean + SEM CFC frequencies of THP1 AML cells expressing either *IRX3* or a control retroviral vector (MTV) in *IRX3* KD#2 cells relative to control cells. Colonies were enumerated after 10 days in semi-solid culture (n = 4). Image (right) shows representative colonies. *p ≤ 0.05 using one-way ANOVA with Fisher's least significant difference post hoc analysis for the indicated comparison. NS, not significant.

See also Figure S4.

self-renewal and represses expression of a terminal myeloid lineage differentiation program.

We next evaluated whether a signature of IRX3 transcriptional activity could be detected in human AML. We ranked protein-coding genes in *IRX3*^{high} versus IRX3^{low} human AMLs (Wouters et al., 2009) using a signal-to-noise ranking metric (Table S6) and performed GSEA using the set of genes repressed by IRX3 in murine *Hoxa9*/*IRX3* AML cells (Table S5). There was highly significant negative enrichment of IRX3-repressed genes in human *IRX3*^{high} versus *IRX3*^{low} human AMLs whether all AMLs were considered or just those expressing high levels of *HOXA9* (Figure 5D). Indeed, in leading-edge analyses, there was a highly



F

FAB category	*IRX3*^{high} (n=131)	*IRX3*^{low} (n= 145)	p value
M0	2 (2%)	9 (6%)	NS
M1	40 (31%)	27 (19%)	0.02
M2	36 (27%)	22 (15%)	0.02
M1 & M2	76 (58%)	49 (34%)	<0.0001
M4	13 (10%)	33 (23%)	0.006
M5	33 (25%)	49 (34%)	NS
M4 & M5	46 (35%)	82 (57%)	0.0004
M6	3 (2%)	2 (1%)	NS
M7	0 (0%)	0 (0%)	NS
Unknown	4 (3%)	3 (2%)	NS

Figure 5. IRX3 Represses a Myelomonocytic Differentiation Program in Murine and Human AML

(A) Heatmap shows differentially expressed protein-coding genes ($p < 0.05$, unpaired t test, >2-fold).

(B–D) GSEA plots show enriched expression of (B) mature myelomonocytic genes, (C) LSC maintenance signature genes, and (D) IRX3-repressed genes in (B and C) *Hoxa9*/MTV versus *Hoxa9*/*IRX3* AML cells or (D) primary human AMLs ranked by *IRX3* expression, respectively.

(E) Scatterplot shows expression of *IRX3* in primary human *HOXA9*$^+$ AML samples versus mean $\log_2$ expression level for the leading-edge gene set shown in (D).

(F) Morphologic classification of *IRX3*^{high} versus *IRX3*^{low} primary *HOXA9*$^+$ AML samples (Wouters et al., 2009). p values were calculated using Fisher's exact test.

See also Figure S5 and Tables S5, S6, and S7.

significant association of higher *IRX3* expression in human *HOXA9*$^+$ AML with greater repression of IRX3-repressed genes identified in murine leukemias (Figure 5E; Table S7). Remarkably, when the morphologic classification of *HOXA9*$^+$ AMLs was considered, among cases with high *IRX3* expression, there were significantly fewer AMLs exhibiting myelomonocytic differentiation (i.e., French-American-British [FAB] M4 subtype) and significantly more AMLs exhibiting minimal differentiation (i.e., FAB M1 subtype) or maturation toward the granulocytic lineage (i.e., FAB M2 subtype) (Figure 5F). Thus, in primary human AML, as in murine AML, misexpression of *IRX3* contributes functionally to blockade of myelomonocytic lineage differentiation.

IRX3 and FOXC1 Differentially Repress Expression of Myelomonocytic Transcription Factors

We previously reported that the mesenchymal transcription factor gene *FOXC1* is also frequently misexpressed in human AML to confer, in particular, a monocytic lineage differentiation block (Somerville et al., 2015). With regard to the morphologic classification of human AML, *FOXC1*-expressing leukemias

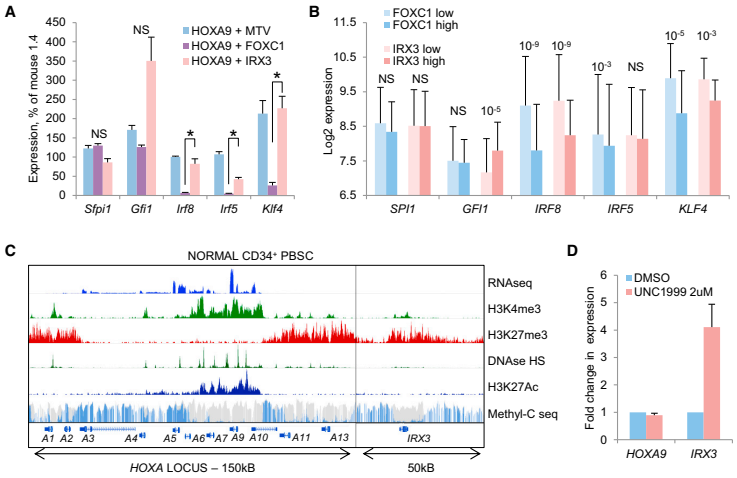


Figure 6. IRX3 Represses Myelomonocytic Genes in Human AML

(A) Bar chart shows mean + SEM relative expression of the indicated genes in flow-sorted murine KIT$^+$Gr1$^+$ BM AML cells (MTV, control, empty vector) (n = 5 or 6 per cohort). *p $\leq$ 0.05 using one-way ANOVA with Fisher's least significant difference post hoc analysis for the indicated comparison. NS, not significant.

(B) Bar chart shows mean + SD $\log_2$ array expression values for the indicated genes in *FOXC1*^{high} (n = 95) versus *FOXC1*^{low} (n = 175) and *IRX3*^{high} (n = 133) versus *IRX3*^{low} (n = 158) human *HOXA9*$^+$ AML cases (Wouters et al., 2009). p values (unpaired t test) are shown where significant. NS, not significant.

(C) Image shows high-throughput sequencing tracks from the ENCODE consortium for the *HOXA* locus and *IRX3*.

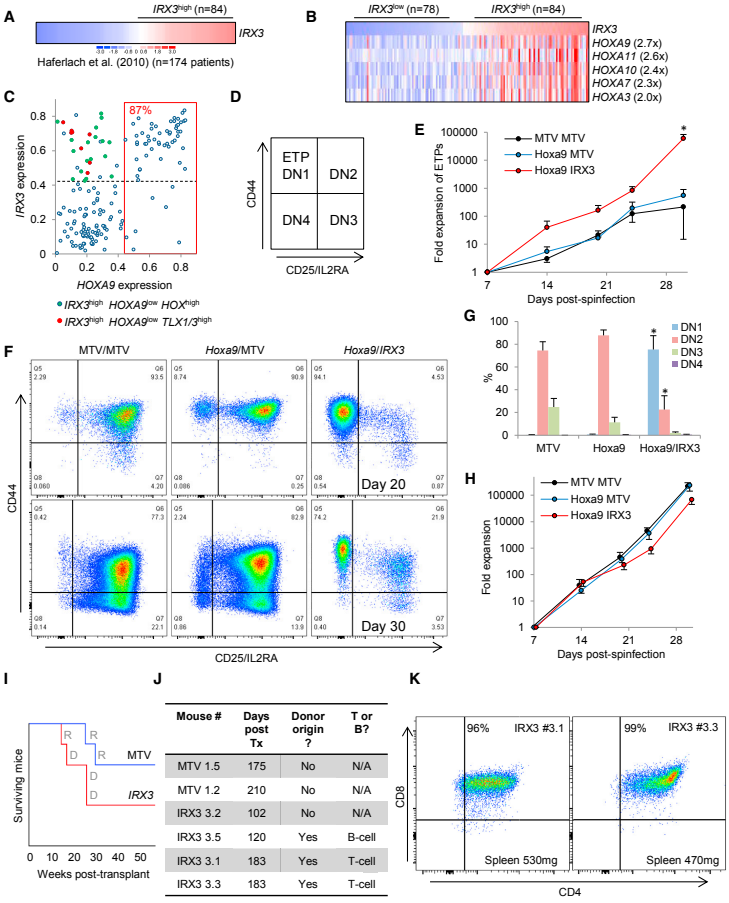
(D) Bar chart shows mean + SEM relative expression of *IRX3* and *HOXA9* in normal human CD34$^+$ cells from separate donors following 5 days of treatment with UNC1999 or DMSO in serum-free liquid culture (n = 4).

See also Figure S5.

were significantly less likely to exhibit monocytic lineage differentiation (i.e., FAB-M5) and significantly more likely to exhibit granulocytic lineage differentiation (i.e., FAB M2) (Somerville et al., 2015). The presence of high *FOXC1* expression in AML was also associated with inferior outcome, in contrast to the presence of high *IRX3* expression (Figure S5B). Murine *Hoxa9/FOXC1* leukemias also exhibited shortened latency versus *Hoxa9/MTV* leukemias (Somerville et al., 2015), in contrast to the latencies observed for *Hoxa9/IRX3* leukemias (Figure 3E).

To evaluate the consequences of misexpressed *IRX3* and *FOXC1* on expression levels of transcription factors required for normal myelomonocytic lineage differentiation, and to determine why the phenotypic consequences of *IRX3* and *FOXC1* misexpression differed one from another, we performed qPCR using murine leukemia samples and analyzed expression levels in published human AML datasets. In flow-sorted murine BM KIT$^+$Gr1$^+$ AML cells, expression levels of transcription factor genes such as *Irf8*, *Irf5*, and *Klf4* (which promote monocytic lineage differentiation) were significantly lower in *Hoxa9/FOXC1* AMLs in comparison with *Hoxa9/IRX3* and *Hoxa9*/MTV AMLs (Figure 6A). There was no significant difference in expression levels of the myeloid lineage master regulator *Sfpi1*. In the case of *Gfi1* (which promotes granulocytic lineage differentiation), there was a variable increase in expression in *Hoxa9/IRX3* AMLs compared with the other subtypes, although this did not achieve statistical significance.

In primary human AML samples, a very similar pattern was observed. In comparison with *FOXC1*^{low} or *IRX3*^{low} cases, *FOXC1*^{high} or *IRX3*^{high} cases respectively exhibited significantly lower level expression of *IRF8* and *KLF4*, and for both genes, the proportionate reduction in expression was greater for *FOXC1*^{high} cases than for *IRX3*^{high} cases (Figure 6B). *IRF5* expression was significantly lower in *FOXC1*^{high} versus *FOXC1*^{low} cases and no different between the *IRX3* groups, whereas *GFI1* expression was significantly higher in *IRX3*^{high} versus *IRX3*^{low} cases and no different between the *FOXC1* groups. Expression levels of *SPI1* did not differ. Similar analyses using a separate, smaller dataset from TCGA also gave similar results (Figure S5C), in particular with regard to expression levels of *GFI1* and *IRF8*. Together these data demonstrate that although *IRX3* and *FOXC1* misexpression contribute to the differentiation block observed in human AML, they repress myelomonocytic lineage transcription factor genes in a distinctive



Mouse #	Days post Tx	Donor origin ?	T or B?
MTV 1.5	175	No	N/A
MTV 1.2	210	No	N/A
IRX3 3.2	102	No	N/A
IRX3 3.5	120	Yes	B-cell
IRX3 3.1	183	Yes	T-cell
IRX3 3.3	183	Yes	T-cell

Figure 7. IRX3 Misexpression in Lymphoblastic Leukemia

(A) *IRX3* expression in bulk human T-ALL samples.

(B) Heatmap shows *IRX3*^{high} versus *IRX3*^{low} T-ALL cases and the most differentially expressed transcription factor genes with mean fold change.

(C) *IRX3* versus *HOXA9* expression in primary human T-ALL (Haferlach et al., 2010). Percentage in red text indicates the proportion of *HOXA9*^{high} samples (red box) with high *IRX3* expression (above the dotted gray line).

(D) Immunophenotypic definition of early T cell progenitor double-negative (DN) populations.

(E) Mean + SEM fold expansion in OP9 DL1 stromal culture of ETP/DN1 cells over input number on day 7 (n = 4–6). MTV, empty vector.

(F) Representative flow cytometry plots.

(G) Immunophenotypic developmental stage of Lin^{neg} BM HSPCs expressing the indicated gene combinations cultured for 30 days on OP9 DL1 stroma (mean + SEM percentage, n = 6).

(legend continued on next page)

manner: FOXC1 represses monocytic lineage genes more profoundly than does IRX3, whereas the opposite is the case for GFI1.

Polycomb Activity Sustains IRX3 Repression in Normal Human CD34$^+$ Stem and Progenitor Cells

IRX3 is only minimally expressed in normal CD34$^+$ cells, and its genetic locus is marked by high levels of H3K27 trimethylation (Zhou et al., 2011) (Figure 6C), suggesting that its relative transcriptional silence is maintained by Polycomb-mediated repression. To address this question, normal human CD34$^+$ HSPCs from multiple donors were treated *in vitro* for 5 days with a dual EZH1 and EZH2 inhibitor (UNC1999; Konze et al., 2013). We observed a significant increase in *IRX3* expression but no change in *HOXA9* expression (Figure 6D); the *HOXA9* locus is not marked by H3K27 trimethylation in normal CD34$^+$ HSPCs (Figure 6C). Thus, the Polycomb complex contributes to continued repression of *IRX3* in normal HSPCs.

IRX3 Is Frequently Co-expressed with HOX Genes in Human Acute Lymphoblastic Leukemia

To provide a more comprehensive evaluation of the role of *IRX3* in human acute leukemia, we analyzed published expression datasets from patients with acute lymphoblastic leukemia (ALL). In a cohort of T-ALL patients (Microarray Innovations in Leukemia [MILE] study; Haferlach et al., 2010), *IRX3* transcripts were detected at high level (i.e., with a probeset [229638_at] value of $\geq$ 0.42, approximating to a value among the top 25% of array probeset values) in 84 of 174 cases (48%) (Figure 7A). To identify transcription factor genes concordantly expressed with *IRX3* in T-ALL, we compared *IRX3*^{high} (array signal intensity > 0.42, n = 84) with *IRX3*^{low} (array signal intensity < 0.3, n = 78) cases and observed that only *HOXA* genes were upregulated by greater than a mean 2-fold change in array signal intensity (and with $p < 10^{-5}$, by unpaired t test) (Figure 7B). Of the 84 *IRX3*^{high} cases, 52 (62%) expressed *HOXA9* at high level, 18 (21%) were *HOXA9*^{low} but expressed one or more alternate HOX genes, and 8 (14%) were HOX^{low} but instead expressed the homeodomain transcription factors *TLX1* or *TLX3* (Figure 7C). Of the *HOXA9*^{high} T-ALL cases (i.e., 60 of 174), 52 of 60 (87%) were *IRX3*^{high} (Figure 7C). High HOXA expression is a feature of human T-ALL associated with *CALM-AF10* fusions, *SET-NUP214* fusions, *MLL* gene rearrangements, or inv(7) or t(7;7) translocations (resulting in apposition of the HOXA locus to *TCR-β* regulatory elements; Soulier et al., 2005; Van Vlierberghe et al., 2008).

In B-acute lymphoblastic leukemia (B-ALL), high *IRX3* expression was detected in 116 of 563 cases (21%) in the MILE cohort (Figure S6A). There was a particular association of high *IRX3* expression with the presence of an *MLL* gene rearrangement (31 of 70 cases [44%]), or the presence of t(12;21), the cytogenetic hallmark of the *ETV6-RUNX1* fusion (44 of 58 cases [76%]) (Figure S6B). To identify transcription factor genes exhibiting concordant expression with *IRX3* in B-ALL, we compared *IRX3*^{high} (array signal intensity > 0.42, n = 116) with *IRX3*^{low} (array signal intensity < 0.3, n = 398) B-ALLs. Only *HOXA9*, *MEIS1*, and *SOX11* were upregulated by greater than a mean 2-fold change in array signal intensity (and with $p < 10^{-5}$, by unpaired t test) in *IRX3*^{high} B-ALLs (Figure S6B). Of the 57 *MLL* rearranged *HOXA9*^{high} cases, 27 (47%) were *IRX3*^{high} (Figure S6C). *IRX3*^{high} cases with a t(12;21) did not express HOX genes at high level (Figure S6B). Together these data demonstrate that *IRX3*, which is expressed at very low levels in normal hematopoiesis, is frequently highly expressed in human ALL.

IRX3 Impedes Phenotypic Differentiation of T Cell Precursors and Induces Lymphoid Leukemias

To evaluate the functional consequences of *IRX3* expression in early stage lymphoid development, Lin^{neg} murine BM HSPCs were double-transduced with pairwise combinations of *IRX3*, *Hoxa9*, or control retroviral vectors and, following drug selection, co-cultured on stromal layers. Where BM cells were co-cultured on OP9 stroma (which supports B-lineage differentiation), we observed no significant difference in upregulation of the B-lineage markers B220 and CD19 or expansion of cell numbers (Figure S6D). In contrast, where cells were co-cultured on OP9 DL1 stroma (which ectopically expresses the Notch ligand DLL1 and supports T-lineage differentiation), we observed a highly significant block in differentiation of *Hoxa9*/*IRX3* co-expressing cells at the early thymic progenitor (ETP) stage. Differentiating T cell progenitors exhibit sequential expression of CD44 and CD25, which together define developmental stages ETP/DN1 through to DN4 (Figure 7D) (Yui and Rothenberg, 2014). Over the 4 week assay, there was a mean 100-fold greater expansion of cells with an ETP/DN1 immunophenotype in the presence of *Hoxa9*/*IRX3* co-expression, whereas cells expressing *Hoxa9* alone or control cells readily progressed to DN2 and DN3 downstream differentiation stages (Figures 7E–7G). Overall expansion of cell numbers was similar (Figure 7H). Of note, in contrast to BM HSPCs cultured in myeloid conditions (Figure 2), *IRX3*/MTV expressing Lin^{neg} cells failed to expand on OP9 DL1 stroma (n = 4).

In keeping with a role for IRX3 in promoting the development of lymphoid leukemias, we found that irradiated congenic mice transplanted with *IRX3*-expressing KIT$^+$ BM HSPCs developed lymphoid leukemias with incomplete penetrance. In the 12 weeks after transplantation, although there was significantly reduced donor:recipient chimerism in comparison with animals receiving

(H) Mean + SEM fold expansion in OP9 DL1 stromal culture of total cell number over input numbers on day 7 (n = 4–6). For (G) and (H), *p $\leq$ 0.05 using one-way ANOVA with Fisher's least significant difference post hoc analysis for *Hoxa9*/*IRX3* versus *Hoxa9* and MTV conditions.
Murine CD45.1$^+$ KIT$^+$ BM cells were infected with *IRX3*-expressing or control retroviral vectors, and 96 hr later, 10^6 drug-resistant cells were transplanted into irradiated CD45.2$^+$ congenic recipients.
(I) Survival curves. D, donor-derived leukemia; R, recipient-derived leukemia.
(J) Survival times and cause of death.
(K) Flow cytometry plots show immunophenotype of CD45.1$^+$ BM cells at death in the indicated mice.
See also Figures S6 and S7.

control cells (Figure S6E) (as observed in mice transplanted with *IRX3*/*Hoxa9*-expressing cells; Figure 3D), there was no significant proportional difference in myeloid, B-lineage, and T-lineage engraftment (Figures S6E and S6F). Three of six *IRX3*-expressing HSPC recipients developed donor-derived lymphoid leukemia (Figures 7I and 7J); in two cases, the leukemia was T-lineage, and mice exhibited splenomegaly and near total BM involvement (Figures 7K and S7A). These cases expressed *Hoxa* genes at comparable levels to KIT^+ BM HSPCs (Figure S7B). In the third case, although the mouse was found dead and detailed autopsy could not be completed, flow cytometry analysis of blood cells performed 8 days before death revealed a $CD45.1^{lo}B220^+CD19^+$ population accounting for 92% of donor-derived leucocytes (Figure S7C). This population was not present 4 weeks earlier, suggesting that this mouse died of a B-lineage leukemia. Three other mice died during the 400 day follow-up period, but all succumbed to recipient-derived hematologic malignancies likely induced by irradiation at transplant conditioning (Figures 7I, 7J, and S7A). At experiment termination, there was no evidence of incipient hematologic neoplasms in remaining mice. Together these data demonstrate that *IRX3* expression impedes normal T-progenitor differentiation *in vitro* and induces T-lineage leukemias *in vivo*.

DISCUSSION

Our studies demonstrate that tissue-inappropriate misexpression of *IRX3* is both frequent and functional in human acute leukemias of multiple lineages. The lack of a major role for *IRX3* and its paralog *IRX5* in normal hematopoiesis is emphasized by the observation of minimal or absent expression in human BM cell populations and genetic knockout experiments that demonstrate that $Irx3^{-/-}$ and $Irx5^{-/-}$ mice are viable and fertile (Smemo et al., 2014). At least in AML, *IRX3* misexpression is found in both bulk and the putative stem cell compartments, suggesting functional contribution throughout the leukemia clone. With regard to co-regulated gene expression within the *IRX3* topologically associated domain, our qPCR and published dataset analyses indicate that the set of *IRX5*-expressing AMLs is a subset of the *IRX3*-expressing cases, with lower expression levels of *IRX5* than *IRX3*.

Why is *IRX3* expressed so extensively in acute leukemia? Our *in vitro* and *in vivo* studies demonstrate that misexpression of *IRX3* contributes to the cardinal pathologic feature of acute leukemia, the differentiation block. In cultures supporting myeloid lineage differentiation, *IRX3* expression alone or in combination with *Hoxa9* enhanced clonogenic potential and impeded differentiation of normal HSPCs. *In vivo*, *IRX3* co-expression with *Hoxa9* dramatically enhanced the degree of differentiation block in murine AMLs, even though the onset of AML was delayed. Critically, the same transcriptional signature of IRX3-mediated repression of myelomonocytic differentiation was readily identified in human AML, confirming that *IRX3* misexpression is both frequent and functional in human acute leukemia. The paradigm that misexpression of *IRX3* confers a differentiation block is further supported by our observation that co-expression of *IRX3* with *Hoxa9* in T-lineage cultures impeded differentiation of ETPs into downstream developmental stages and that *IRX3*-expressing HSPCs generated lymphoid leukemias *in vivo*. The expanded ETP population may serve as a reservoir for acquisition of genetic mutations required for full-fledged leukemia. The observations that HSPCs expressing *IRX3* alone were immortalized in myeloid culture, failed to expand significantly in OP9 stromal culture, but generated *Hoxa*-expressing lymphoid leukemias *in vivo* reflect the importance of the cellular microenvironment in supporting phenotypic outcome. The outcomes also emphasize the importance of the interaction of misexpressed IRX3 with cell type-specific patterns of chromatin accessibility for transcription factor binding. Cells of different lineages, and of different differentiation states, express different repertoires of transcription factor genes; it is likely that IRX3 binds to chromatin and interferes with gene expression in distinct ways in cells of different lineages.

How is *IRX3* gene expression turned on in hematopoietic cells? Although its expression is positively associated with NPM1 and FLT3 mutations in AML, the link is not absolute; for example, more than a third of NPM1 mutant cases do not express *IRX3* (Wouters et al., 2009). The positive association with other genetic lesions that near invariably lead to high-level HOX gene expression (i.e., a t[6;9] or translocations targeting *MLL* at chromosome 11q23) raises the possibility that *IRX3* is positively regulated by HOX transcription factors. In keeping with this, when *HOXA9* or *HOXA10* is expressed in human $CD34^+$ HSPCs, *IRX3* is upregulated (Ferrell et al., 2005). Relatedly, in *Xenopus* development, HOXB4 and IRX5 have overlapping patterns of expression, and the latter is a direct target of the former (Theokli et al., 2003). However, high-level HOX gene expression alone in AML is not sufficient to result in *IRX3* expression, because only 40% of $HOXA9^+$ cases express *IRX3*. This suggests that additional factors act combinatorially to induce *IRX3*. For example, the Wnt signaling pathway, which is active in AML, induces *IRX3* in forebrain development (Braun et al., 2003). In addition to these candidate positive regulators, it seems likely that loss of repressor activity makes a significant contribution. Our observation that the *IRX3* locus is marked by H3K27me3 in normal $CD34^+$ HSPCs and that treatment of cells with a dual EZH1 and EZH2 inhibitor led to *IRX3* upregulation demonstrates that its repression in normal hematopoiesis is dependent on Polycomb.

As well as the strong association of *IRX3* with HOX gene expression in acute leukemias of multiple lineages, there was also high-level expression of *IRX3* in ~90% of cases of APML and in ~75% of cases of t(12;21) B-ALL, leukemias that do not express HOX genes. It is possible that *IRX3* is induced as a direct consequence of PML-RARA or ETV6-RUNX1 fusions, respectively, although the close association of co-expressed *SOX11* in the latter case suggests potential collaboration.

The molecular consequences of *IRX3* misexpression in the acute leukemias remain unclear. It is known that TALE family transcription factors such as MEIS1 and PBX can form triple complexes with HOXA9 that bind to PBX-HOXA9 consensus sequences to regulate gene expression (Shen et al., 1999). We speculate that misexpressed IRX3 might alter the function or stability of HOX transcription factor heterotrimeric complexes, perhaps to prevent downregulation of self-renewal genes or upregulation of transcription factors required for terminal

differentiation. Alternatively, it may redirect HOX transcription factors to new binding sites or function on its own to activate or repress key transcription regulators.

Like *IRX3*, the Forkhead transcription factor gene *FOXC1* is also frequently misexpressed in AML, although the phenotypic consequences in mouse models and primary human AMLs are quite distinct. This is likely related to distinct mechanisms of action and sites of genomic binding; for example, Forkhead and Iroquois transcription factors have different consensus binding motifs and will bind different sites in the genome to regulate overlapping but fundamentally distinct gene sets. In particular, FOXC1 seems more effective than IRX3 at suppressing expression of monocytic lineage transcription factor genes such as *IRF8*, *IRF5*, and *KLF4*. In contrast, in comparison with *IRX3*^{low} AML cases, in *IRX3*^{high} cases there is increased expression of the granulocytic lineage regulator gene *GFI1*. In some cases both *FOXC1* and *IRX3* are misexpressed, and here the resulting cellular phenotype will represent the integrated consequence of the prevailing nuclear transcription factor milieu.

In summary, we demonstrate that the Iroquois homeodomain transcription factor IRX3 is frequently misexpressed in human acute leukemia to contribute to the differentiation block that is the pathognomonic feature of the disease. Future investigations will identify approaches to target these transcription factors for pro-differentiation therapies to improve patient outcomes.

EXPERIMENTAL PROCEDURES

Human Tissue and Ethical Approval

Normal CD34$^+$ HSPCs surplus to requirements were from patients undergoing autologous transplantation for lymphoma. Their use was authorized by the Salford and Trafford Research Ethics Committee and, for samples collected since 2006, following written informed consent from 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. Primary human AML samples were from Manchester Cancer Research Centre's Tissue Biobank (approved by the South Manchester Research Ethics Committee). Their use was authorized by the Tissue Biobank's scientific sub-committee, with the informed consent of donors.

Murine Experiments

Experiments using mice (female, aged 6–12 weeks) were approved by Cancer Research UK Manchester Institute's Animal Ethics Committee and performed under a project license issued by the United Kingdom Home Office, in keeping with the Home Office Animal Scientific Procedures Act of 1986. C57BL/6 (CD45.2$^+$) mice were from Envigo. B6.SJL-*Ptprc*^{a}*Pepc*^{b}/BoyJ (CD45.1$^+$) mice were from Jackson Laboratories and bred in house. Details of transplantation procedures are in the Supplemental Information.

Reagents, Cell Culture, and Flow Cytometry

Details are in the Supplemental Information.

RNA Preparation, qPCR, RNA Sequencing, Bioinformatics, and Statistics

Details are in the Supplemental Information.

Statistical Analyses

Statistical analyses were performed using StatsDirect software version 1.9.7 (StatsDirect), Microsoft Excel, or SPSS for Mac version 22 (IBM). Survival curves were generated using Prism software version 6.0 (GraphPad Software).

DATA AND SOFTWARE AVAILABILITY

The accession number for the RNA sequencing data files reported in this paper is GEO: GSE97450.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and seven tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2017.12.063.

ACKNOWLEDGMENTS

We thank John Weightman, Jeff Barry, Toni Banyard, Abi Johnson, and staff at the Biological Resources Unit for technical assistance; Dan Wiseman for assistance with figures; and Ruud Delwel for sharing the survival data from the Dutch AML cohort. This work was supported by Cancer Research UK (grant C5759/A20971).

AUTHOR CONTRIBUTIONS

T.D.D.S., F.S., J.A.C., E.L.W., G.J.S., and T.C.P.S. performed experiments and analyzed data. K.B., E.T., and R.J.B. generated the tissue array and performed histopathological stains and analyses. C.W. performed bioinformatics analyses. T.D.D.S. and T.C.P.S. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 16, 2017
Revised: October 22, 2017
Accepted: December 20, 2017
Published: January 16, 2018

REFERENCES

Braun, M.M., Etheridge, A., Bernard, A., Robertson, C.P., and Roelink, H. (2003). Wnt signaling is required at distinct stages of development for the induction of the posterior forebrain. Development *130*, 5579–5587.

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.

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. *373*, 895–907.

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.

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.

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.

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.

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.

Haferlach, T., Kohlmann, A., Wieczorek, L., Basso, G., Kronnie, G.T., Béné, M.C., De Vos, J., Hernández, 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.

Hess, J.L., Bittner, C.B., Zeisig, D.T., Bach, C., Fuchs, U., Borkhardt, A., Frampton, J., and Slany, R.K. (2006). c-Myb is an essential downstream target for homeobox-mediated transformation of hematopoietic cells. Blood *108*, 297–304.

Houweling, A.C., Dildrop, R., Peters, T., Mummenhoff, J., Moorman, A.F., Rüther, U., and Christoffels, V.M. (2001). Gene and cluster-specific expression of the Iroquois family members during mouse development. Mech. Dev. *107*, 169–174.

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.

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.

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.

Ley, T.J., Miller, C., Ding, L., Raphael, B.J., Mungall, A.J., Robertson, A., Hoadley, K., Triche, T.J., Jr., Laird, P.W., Baty, J.D., et al.; Cancer Genome Atlas Research Network (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N. Engl. J. Med. *368*, 2059–2074.

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.

Martorell, Ò., Barriga, F.M., Merlos-Suárez, A., Stephan-Otto Attolini, C., Casanova, J., Batlle, E., Sancho, E., and Casali, A. (2014). Iro/IRX transcription factors negatively regulate Dpp/TGF-β pathway activity during intestinal tumorigenesis. EMBO Rep. *15*, 1210–1218.

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.

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.

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*, 17ra9.

Shen, W.F., Rozenfeld, S., Kwong, A., Köm ves, L.G., Lawrence, H.J., and Largman, C. (1999). HOXA9 forms triple complexes with PBX2 and MEIS1 in myeloid cells. Mol. Cell. Biol. *19*, 3051–3061.

Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, 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., 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.

Somerville, T.D., Wiseman, D.H., Spencer, G.J., Huang, X., Lynch, J.T., Leong, H.S., Williams, E.L., Cheesman, E., and Somervaille, T.C. (2015). Frequent derepression of the mesenchymal transcription factor gene FOXC1 in acute myeloid leukemia. Cancer Cell *28*, 329–342.

Soulier, J., Clappier, E., Cayuela, J.M., Regnault, A., García-Peydró, M., Dombret, H., Baruchel, A., Toribio, M.L., and Sigaux, F. (2005). HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). Blood *106*, 274–286.

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.

Van Vlierberghe, P., Pieters, R., Beverloo, H.B., and Meijerink, J.P. (2008). Molecular-genetic insights in paediatric T-cell acute lymphoblastic leukaemia. Br. J. Haematol. *143*, 153–168.

Wiseman, D.H., Greystoke, B.F., and Somervaille, T.C. (2014). The variety of leukemic stem cells in myeloid malignancy. Oncogene *33*, 3091–3098.

Wouters, B.J., Löwenberg, 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.

Yui, M.A., and Rothenberg, E.V. (2014). Developmental gene networks: a triathlon on the course to T cell identity. Nat. Rev. Immunol. *14*, 529–545.

Zhang, S.S., Kim, K.H., Rosen, A., Smyth, J.W., Sakuma, R., Delgado-Olguín, 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. USA *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.