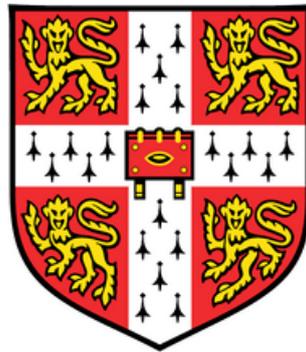


Novel roles of KAT2A chromatin
complexes ATAC and SAGA in normal
and malignant haematopoiesis



Liliana Arede

Department of Genetics & Department of Haematology
University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Lucy Cavendish College

April 2021

*“ This job is a great scientific adventure.
But it's also a great human adventure.”*

-F. Gianotti

Declaration of Originality

I hereby declare that my dissertation:

Is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

Is original and has not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University.

Does not exceed the prescribed word limit of 60,000 by the Degree Committee for the Faculty of Biology.

April 2021

Signature:

Liliana Sofia De Jesus Arede

Abstract

Haematopoiesis is a dynamic process by which the full array of blood cell types originates from multipotent haematopoietic stem cells (HSCs). The blood system has long been an attractive model to explore how cell fate decisions are conducted at the molecular level, to generate the vast diversity of blood cell types. Blood cell fate choice is for the most part regulated by transcription factors and epigenetic modifiers. The enzymatic activities of epigenetic regulators are of particular importance in malignant haematopoiesis such as leukaemia, given their reversible nature and potential for therapeutic targeting. Epigenetic histone modifiers are often found within large multi-subunit complexes, whereby their catalytic activity is exercised. However, very little is known about how epigenetic enzymes take advantage of their participating chromatin complexes in order to accomplish their functions. In the blood system, no studies have formally examined how different epigenetic complexes with the same enzymatic component coordinate cell fate choices. Here, I focus on KAT2A, a lysine acetyltransferase responsible for Histone 3 Lysine 9 acetylation (H3K9ac), and a vulnerability in Acute Myeloid Leukaemia (AML), as well as its 2 containing complexes: Ada-Two-A-Containing (ATAC) and Spt-Ada-Gcn5-Acetyltransferase (SAGA). I combine cell functional and molecular assays of human cord blood (CB) haematopoietic cells and AML cells to dissect KAT2A roles depending on its participation in ATAC or SAGA complexes. Firstly, I demonstrate that KAT2A regulates CB erythroid progenitor specification and survival. I uncover unique contributions of the ATAC complex to erythroid lineage specification, whereas SAGA acts late in differentiation. This suggests that KAT2A plays stage-specific roles, which are unique to the complexes it integrates. While KAT2A regulates specification and survival of erythroid progenitors *via* ATAC through control of biosynthetic activity, it fine-tunes progression of erythroid differentiation through participation in SAGA. Secondly, I explored the roles of ATAC and SAGA in leukaemic haematopoiesis. I found that ATAC regulates proliferation through control of biosynthetic activity. Instead, SAGA participates in maintenance of the characteristic block in AML cell differentiation, compatible with a generic control of cell identity through preservation of cell-type specific transcriptional programmes. My data is consistent with a model in which the ATAC complex has pervasive roles in maintenance of self-renewing cells, while SAGA acts to stabilise cell-type specific programmes, preserving cell identity.

Understanding ATAC and SAGA unique chromatin regulatory mechanisms has important implications for the tailored design of new drugs for clinical application.

Preface

Work done by / and assistance received by others:

Selinde Winde and Elena Foerner

Cloning of some of the shRNAs used, some of the CB experiments

Elena Foerner

ChIP-seq experiment

Dr Rashmi Kulkarni

Bioinformatic analysis of RNA-sequencing and ChIP-sequencing data

Genomics Core, CRUK Cambridge Institute

RNA-sequencing and ChIP-sequencing

Cell Phenotyping Hub NIHR BRC

Assistance in CB cell sorting

Dr Joana Cerveira at the Flow Sorting Facility of the Department of Pathology,
University of Cambridge

Assistance in sorting of cell lines

Published work related to this thesis:

Ana Filipa Domingues, Rashmi Kulkarni, George Giotopoulos, Shikha Gupta, Laura Vinnenberg, Liliana Arede, Elena Foerner, Mitra Khalili, Rita Romano Adao, Shengjiang Tan, Keti Zeka, Brian J. Huntly, Sudhakaran Prabakaran, Cristina Pina. Loss of KAT2A enhances transcriptional noise and depletes acute myeloid leukaemia stem-like cells. *eLife*, January 2020.

Liliana Arede, Cristina Pina. Buffering noise: KAT2A modular contributions to stabilization of transcription and cell identity in cancer and development. *ExpHem*, October 2020.

Liliana Arede, Elena Foerner, Selinde Wind, Rashmi Kulkarni, Ana Filipa Domingues, George Giotopoulos, Svenja Kleinwaechter, Ryan Asby, Shikha Gupta, Elisabeth Scheer, Brian J.P. Huntly, Laszlo Tora, Cristina Pina. KAT2A complexes ATAC and SAGA play unique roles in cell maintenance and identity in hematopoiesis and leukemia (*manuscript under review*).

Note: some of the results and figures in the following Chapters have previously been presented in these publications.

Acknowledgements

I couldn't have completed this project without the support and encouragement of many people. First, I would like to thank my supervisor, Dr Cristina Pina for her scientific guidance these last few years. Key to this *journey* were also members of the lab, past and present, who in a way or another all imparted some combination of their expertise, insight and inspiration. A special thank you to Filipa, for getting me started at the bench and sharing her technical knowledge, to Wade and Shikha, my PhD pals, who shared many happy hours in and outside the lab, and to Keti for her always direct approach in all things.

My research would not have been possible without the financial support of the Rosetrees Trust and Lucy Cavendish College. Likewise, I am grateful for the collaborations that made this work possible, in particular, with the labs of Dr Laszo Tora, Dr Antonio Curti and Professor Brian Huntly.

During my time at Haematology, I was lucky to meet Chiara, Rita, Joana and Pepe, their *friendship* over the years and collective comic genius has been the best! As the lab transitioned to the Department of Genetics, I want to thank particularly the first-floor teams and Dr Felipe Karam Teixeira for all the help settling in and throughout. It's been a great opportunity to get to know and to *learn* from them.

Away from the lab, I am thankful to my rowing team for kick-starting so many early mornings with *energy* and *focus* while pushing through pain (a bit like when things don't work in the lab!), as well as amusing me with *beautiful Cambridge* sunrises. I'm also thankful to my SIU friends. I'm indebted to all those who made England a *home* for me over the past couple of years, including Jay, Aeh, Marta and Danny, Laetitia, Peter, Fabio, and Stephanie, as well as all the encounters and experiences brought up by deciding to commit to this project.

I can't help but reflect on the COVID-19 pandemic which made 2020 quite surreal. Despite it all, it brought a lot of perspective and I learned just how vast my support network is.

Finally, as ever, I'm immensely *grateful* to my parents and their unwavering support as I ventured into "Terras de Sua Majestade" on what would have been a short Erasmus placement with a return ticket. Despite not really understanding what I do you always believe in me and in what I decide to follow. Mãe e Pai: OBRIGADA. Thank you also to my grandparents, who brought me close to the beauty of the natural world since I was little. Now, may I continue to "*pay attention, be astonished, and tell about it*".

Contents

Abstract	i
Preface	iii
Acknowledgments	v
List of Figures	xi
List of Tables	xvii
Abbreviations	xviii
Chapter 1 Introduction	1
1.1 Haematopoiesis	2
1.1.1 Haematopoietic stem cells (HSCs)	2
1.1.1.1 Origin, identification and isolation	2
1.1.1.2 HSCs and haematopoiesis throughout life	5
1.1.2 Regulation of haematopoietic fate decisions	6
1.1.2.1 The role of cytokines	7
1.1.2.2 Transcriptional regulation	7
1.1.3 Overview of myelopoiesis and erythropoiesis	11
1.1.3.1 Myelopoiesis and its regulation	11
1.1.3.2 Erythropoiesis and its regulation	12
1.2 Acute Myeloid Leukaemia (AML)	18
1.2.1 AML as a clinical entity	18
1.2.2 Biology of leukaemia stem-like cells	19
1.2.3 AML classification	20
1.2.4 Genomic and epigenomic landscape of AML	21
1.2.5 AML treatment	24
1.3 Epigenetics	27
1.3.1 Epigenetic mechanisms	27

1.3.2	Epigenetic regulation of haematopoiesis	32
1.3.2.1	Histone acetyltransferases / deacetylases	32
1.3.2.2	Histone lysine (K) methyltransferases / demethylases	33
1.3.2.3	DNA Methylation: the roles of DNMTs and TET proteins	36
1.3.2.4	ATP-dependent nucleosome remodelling complexes	37
1.4	KAT2A	40
1.4.1	KAT2A roles in development	40
1.4.2	KAT2A contributions to cancer	41
1.5	SAGA and ATAC – KAT2A chromatin complexes	44
1.5.1	SAGA complex	45
1.5.2	ATAC complex	46
1.5.3	SAGA and ATAC roles in cancer	48
1.6	Thesis overview	52
Chapter 2	Materials and Methods	54
2.1	Cell lines	55
2.2	Human Cord Blood CD34+ Cells	56
2.3	Primary Human AML Cells	56
2.4	Cell Cryopreservation, Thawing and Counting	57
2.5	Lentiviral Packaging and Transduction	58
2.6	Functional Cellular Assays	60
2.6.1	Flow Cytometry	60
2.6.1.1	Human CB CD34+ Cell Sorting	60
2.6.1.2	Cell Cycle Analysis	62
2.6.1.3	Apoptosis Analysis	63
2.6.1.4	Cell Divisional Tracking	63
2.6.2	Colony-Forming Cell (CFC) Assays	64
2.6.3	Erythroid Differentiation Cultures	65

2.6.4	Cytospins	65
2.7	Molecular Assays	66
2.7.1	Cloning of shRNA Constructs	66
2.7.2	RNA Extraction, cDNA Synthesis	72
2.7.3	Quantitative Real Time PCR (qRT-PCR)	72
2.8	RNA-sequencing (RNA-seq)	74
2.9	Chromatin Immunoprecipitation (ChIP)	74
2.9.1	Cross-link for suspension cells, cell lysis	74
2.9.2	Immunoprecipitation	76
2.9.3	DNA purification	76
2.10	Western Blot (WB)	79
2.10.1	SDS-PAGE	79
2.10.2	Sample preparation and WB analysis	79
2.11	Statistical Analysis	81
2.12	Data Deposition	81
2.13	Resources	81
Chapter 3	The role of KAT2A and its complexes in normal haematopoiesis	82
3.1	Introduction	83
3.2	Chapter overview	84
3.3	Results and discussion	85
3.3.1	KAT2A regulates human CB erythroid progenitor specification and survival	85
3.3.2	The ATAC complex is selectively required for erythroid specification from CB HSC, while SAGA facilitates erythroid differentiation post-commitment	96
3.3.3	KAT2A-containing ATAC and SAGA complexes have unique targets in haematopoietic cells	103
3.4	Conclusions and future work	110

Chapter 4 Cellular investigation of KAT2A-complex dependencies in human AML cell lines	117
4.1 Introduction	118
4.2 Chapter overview	120
4.3 Results and discussion	120
4.3.1 KAT2A complexes uniquely maintain proliferation and identity of MLL-AF9 AML cells	121
4.3.2 KAT2A complexes have distinct contributions to proliferation and cell identity of CD34+ AML cell lines	134
4.4 Conclusions and future work	146
Chapter 5 Functional characterisation of the role of KAT2A and ATAC and SAGA complexes in human AML primary samples	150
5.1 Introduction	151
5.2 Chapter overview	152
5.3 Results and discussion	153
5.3.1 Clonogenic growth of AML patient samples is hindered by treatment with MB-3, the KAT2A small molecule inhibitor	155
5.3.2 MS-5 co-culture system preserves growth, CFC-capacity and phenotype of human primary AML cells	156
5.3.3 KAT2A lentiviral-mediated knockdown mimics its chemical inhibition, reducing leukaemic expansion in the MS-5 co-culture system	159
5.3.4 KAT2A ATAC complex activity maintains propagation of human primary AML cells in MS-5 co-culture	163
5.4 Conclusions and future work	165
Chapter 6 Conclusions	168
Bibliography	177
Appendices	204

List of Figures

Chapter 1 Introduction

FIGURE 1.1	Classical model of haematopoiesis	4
FIGURE 1.2	Alternative models of HSC lineage commitment	9
FIGURE 1.3	Model of definitive erythroid cell differentiation	13
FIGURE 1.4	Functional categories of mutated genes involved in AML pathogenesis	21
FIGURE 1.5	Timeline of currently approved AML therapies	25
FIGURE 1.6	Epigenetic regulatory mechanisms	28
FIGURE 1.7	KAT2A participates in cancer biology with oncogene-like and tumour-suppressor roles	43
FIGURE 1.8	Schematic representation of the human SAGA (left) and ATAC (right) KAT2A-containing complexes	44
FIGURE 1.9	Timeline of relevant discoveries on KAT2A and the ATAC and SAGA complexes	47
FIGURE 1.10	Analysis of dependencies and lineage associations of ATAC and SAGA-unique elements in CRISPR drop-out screens of cancer cell lines	49

Chapter 2 Materials and Methods

FIGURE 2.1	AML cell lines used in this study and their known oncogenic mutations	55
FIGURE 2.2	pCMV-VSV-G packaging plasmid	59
FIGURE 2.3	pMD2.G VSV-G-expressing envelope plasmid	60
FIGURE 2.4	Flow cytometry gating strategy for sorting the human CB CD34+ stem and progenitor compartments	61

FIGURE 2.5	Flow cytometry gating strategy for cell cycle analysis of human CB CD34+ cells and cell lines	62
FIGURE 2.6	Flow cytometry gating strategy for apoptosis analysis of human CB CD34+ cells and cell lines	63
FIGURE 2.7	Chemical structure of α -methylene- γ -Butyrolactone 3 (MB-3), the small molecule inhibitor of KAT2A	64
FIGURE 2.8	pLL3.7 lentiviral transfer plasmid	66
FIGURE 2.9	Assessment of shRNA insertion into pLL3.7 vector using PCR and agarose gel electrophoresis. Representative gels	69
FIGURE 2.10	Assessment of shRNA insertion into pLL3.7 vector by <i>BamHI</i> restriction digestion using PCR and agarose gel electrophoresis Representative gel	70
FIGURE 2.11	Assessment of gene knockdown in lentiviral-producing 293T cells	71
FIGURE 2.12	Assessment of optimal chromatin sonication conditions for MOLM-13 (left) and K562 (right) cells using agarose gel electrophoresis	75
Chapter 3	The role of KAT2A and its chromatin complexes in normal haematopoiesis	
FIGURE 3.1	Schematic of the adopted CB experimental strategy	85
FIGURE 3.2	Progenitor activity of <i>KAT2A</i> knockdown CB CD34+ cells	86
FIGURE 3.3	CFC-efficiency of <i>KAT2A</i> lentiviral transduced, sorted, human stem (HSC/MPP) and progenitor (MEP, GMP, MLP) CB cell populations	87
FIGURE 3.4	Erythroid cell differentiation in human CD34+ CB cells progresses independently of <i>KAT2A</i> absence	89
FIGURE 3.5	Transcriptional analysis of <i>CTRLsh</i> vs <i>KAT2Ash</i> CB HSCs	91
FIGURE 3.6	Analysis of apoptosis and cell cycle status of human CB stem and progenitor cells upon depletion of <i>KAT2A</i>	94
FIGURE 3.7	<i>ZZZ3</i> (ATAC) and <i>SUPT20H</i> (SAGA) depletion in human CB CD34+ cells	96

FIGURE 3.8	CFC-efficiency of human stem and progenitor CB cells transduced with ATAC (<i>ZZZ3</i>) and SAGA (<i>SUPT20H</i>) specific subunits	97
FIGURE 3.9	Gene expression (selected region) pattern of <i>Kat2a</i> and <i>Zzz3</i> in erythroid development	99
FIGURE 3.10	Expression pattern of SAGA-specific elements during <i>in vitro</i> maturation of committed erythroid progenitors from human CB	100
FIGURE 3.11	Depletion of SAGA-DUB module <i>USP22</i> subunit in human CB CD34+ cells	101
FIGURE 3.12	ChIP-seq analysis of KAT2A-containing ATAC and SAGA complexes in K562 cells	104
FIGURE 3.13	KAT2A-containing ATAC and SAGA complexes have unique functional associations in K562 cells	105
FIGURE 3.14	KAT2A-containing SAGA complex facilitates erythroid cell differentiation in hematopoietic K562 cells	109
FIGURE 3.15	GFP+ percentage in CTRLsh, KAT2Ash and <i>ZZZ3</i> sh-K562 cells in a 2-day erythroid differentiation culture experiment	113

Chapter 4 Cellular investigation of KAT2A-complex dependencies in human AML cell lines

FIGURE 4.1	KAT2A is a genetic vulnerability in AML	118
FIGURE 4.2	Schematic of the adopted experimental strategy.	120
FIGURE 4.3	Knockdown of ATAC and SAGA-specific components hinders expansion of MOLM-13 cells	121
FIGURE 4.4	Analysis of MOLM-13 metabolism and self-renewal gene signature upon depletion of KAT2A, SAGA-specific <i>SUPT20H</i> and <i>TADA2B</i> sh ATAC-specific <i>ZZZ3</i> and <i>TADA2A</i> subunits	123
FIGURE 4.5	H3K9ac ChIP-qPCR analysis of ribosomal protein genes and self-renewal genes in MOLM-13 cells upon knockdown of SAGA (<i>SUPT20H</i>) and ATAC (<i>ZZZ3</i>) elements	124
FIGURE 4.6	Cell cycle analysis of MOLM-13 cells upon knockdown of SAGA (<i>SUPT20H</i> , <i>TADA2B</i>) and ATAC (<i>ZZZ3</i> , <i>TADA2A</i>) elements	125
FIGURE 4.7	Cell divisional tracking of MOLM-13 cells upon knockdown of SAGA (<i>SUPT20H</i>) and ATAC (<i>ZZZ3</i>) elements	126

FIGURE 4.8	Differentiation analysis of MOLM-13 cells upon depletion of KAT2A, SAGA specific SUPT20H and ATAC-specific ZZZ3 subunits	127
FIGURE 4.9	Apoptosis analysis of MOLM-13 cells upon depletion of KAT2A, SAGA specific SUPT20H and ATAC-specific ZZZ3 subunits	128
FIGURE 4.10	Quantification of flow cytometry analysis of cell cycle and apoptosis in MOLM-13 cells transduced with shRNA construct against ATAC YEATS2 component	129
FIGURE 4.11	Cell cycle analysis of THP-1 cells upon knockdown of KAT2A and SAGA (SUPT20H, TADA2B) and ATAC (ZZZ3, YEATS2) elements	131
FIGURE 4.12	Representative photographs of THP-1 cytopins upon depletion of SAGA specific SUPT20H and TADA2B, and ATAC-specific YEATS2 subunits	132
FIGURE 4.13	Apoptosis analysis of THP-1 cells upon depletion of KAT2A, SAGA-specific SUPT20H and TADA2B ATAC-specific ZZZ3 and YEATS2 subunits	133
FIGURE 4.14	Apoptosis and cell cycle analysis of MV4-11 cell line upon KAT2A depletion	135
FIGURE 4.15	Cell cycle assay in MV4-11 cell line upon knockdown of SAGA-specific SUPT20H and TADA2B and ATAC-specific TADA2A and YEATS2 subunits	136
FIGURE 4.16	Differentiation analysis of MV4-11 cells upon depletion of SAGA-specific	137
FIGURE 4.17	Apoptosis analysis of MV4-11 cells upon knockdown of SAGA-specific SUPT20H and TADA2B and ATAC-specific TADA2A and YEATS2 subunits	138
FIGURE 4.18	Quantification of flow cytometry analysis of cell cycle in KG1a AML cells transduced cells with shRNA constructs against KAT2A, SAGA (SUPT20H, TADA2B) and ATAC (ZZZ3, TADA2A) elements	139
FIGURE 4.19	Quantification of flow cytometry analysis of apoptosis by Annexin V staining in KG1a cells transduced cells with shRNA constructs against KAT2A, SAGA (SUPT20H, TADA2B) and ATAC (ZZZ3, TADA2A) elements	140
FIGURE 4.20	Quantification of flow cytometry analysis of cell cycle and apoptosis in KG1a cells transduced with shRNA construct against ATAC YEATS2 component	141
FIGURE 4.21	Analysis of differentiation status of KG1a cells upon knockdown of ATAC specific ZZZ3 and SAGA specific SUPT20H elements	142

FIGURE 4.22	Cell cycle analysis of Kasumi-1 cells upon knockdown of KAT2A, SAGA specific SUPT20H and TADA2B and ATAC-specific ZZZ3 and TADA2A	143
FIGURE 4.23	Analysis of differentiation status of Kasumi-1 cells upon knockdown of SAGA specific SUPT20H and TADA2B and ATAC specific ZZZ3 and TADA2A elements	144
FIGURE 4.24	Apoptosis analysis of Kasumi-1 cells transduced with KAT2Ash, SAGA specific SUPT20Hsh and TADA2Bsh and ATAC-specific ZZZ3sh and TADA2Ash	145

Chapter 5 Functional characterisation of the role of KAT2A and ATAC and SAGA complexes in human AML primary samples

FIGURE 5.1	Experimental strategy adopted for investigating the role of KAT2A and its participating complexes in human primary AML cells	153
FIGURE 5.2	KAT2A small molecule inhibitor, MB-3, reduces CFC-capacity of human primary AML cells	155
FIGURE 5.3	MS-5 co-culture system preserves growth, expansion and moderate cell surface phenotype of primary AML cells for up to 3 weeks	156
FIGURE 5.4	KAT2A chemical inhibition with MB-3 suppresses growth and expansion of primary AML cells for up to 3 weeks in MS-5 co-culture	157
FIGURE 5.5	Flow cytometry gating strategy used for analysis of primitive AML CD34+ and GMP-like (L-GMPs) cell populations	159
FIGURE 5.6	Flow cytometry profiling of KAT2A transduced (GFP+) cells over a 2-week co-culture period	160
FIGURE 5.7	KAT2A knockdown reduces expansion of human primary AML cells in the MS-5 co-culture system	161
FIGURE 5.8	Flow cytometry profiling of transduced (GFP+) cells over a 3-week co-culture period for AML7	163
FIGURE 5.9	ATAC-ZZZ3 knockdown reduces expansion of human primary AML cells in the MS-5 co-culture system	164

Chapter 6 Conclusions

FIGURE 6.1	Working model of cell dependence on ATAC and SAGA complexes in normal and leukemic hematopoiesis	174
------------	--	-----

Appendices **Appendix E**

FIGURE E-1	Single-cell transcriptomics identifies a KAT2A-dependent signature in Kasumi-1 AML cells	239
FIGURE E-2	pHAGE vector containing a dCas9-KAT2A-core fusion	240
FIGURE E-3	Western blot analysis of dCas9- KAT2A-Core fusion protein	241
FIGURE E-4	Region encompassing the human <i>CHEK1 locus</i> on chromosome 11 (125,586,692-125,701,556; GRCh38/hg38 assembly)	241
FIGURE E-5	CRISPR gRNA expression vector	242
FIGURE E-6	Experimental strategy	243
FIGURE E-7	Cell proliferation curves of Kasumi-1 cells transduced with dCas9-KAT2A-Core fusion and with dCas9 only, together with CHEK1 gRNAs or empty vector	244
FIGURE E-8	Cell viability of Kasumi-1 cells transduced with dCas9-KAT2A-Core fusion and with dCas9 only, together with CHEK1 gRNAs or empty vector	244
FIGURE E-9	ChIP-qPCR analysis of H3K9ac at the CHEK1 and E2F3 promoters	245
FIGURE E-10	ChIP-qPCR analysis of H3K27ac at the CHEK1 and E2F3 promoters	246

List of Tables

Chapter 1 Introduction

TABLE 1.1	Main haematopoietic transcription regulators and their roles	10
TABLE 1.2	Major erythroid transcriptional regulators	17
TABLE 1.3	WHO Classification of Acute Myeloid Leukaemia and its subtypes (revised version, 2016)	20
TABLE 1.4	Risk stratification of AML according based on its genetics	22
TABLE 1.5	Combination strategies involving epigenetic therapies in clinical trials	26
TABLE 1.6	Phenotypes of epigenetic disruption in haematopoiesis	38

Chapter 2 Materials and Methods

TABLE 2.1	Antibodies used in flow cytometry analysis and cell sorting	62
TABLE 2.2	Sequences of shRNA constructs used in this work	67
TABLE 2.3	Taqman probes	73
TABLE 2.4	Sequences of primers used for qRT-PCR analysis	74
TABLE 2.5	Buffers used in chromatin immunoprecipitation (ChIP)	77
TABLE 2.6	Antibodies used in chromatin immunoprecipitation (ChIP)	78
TABLE 2.7	Sequences of primers used for ChIP-qPCR analysis	78
TABLE 2.8	Antibodies used in Western Blot (WB)	80

Chapter 4 Cellular investigation of KAT2A-complex dependencies in human AML cell lines

TABLE 4.1	Summary of KAT2A-complex cellular effects in the AML cell lines studied	146
-----------	---	-----

Chapter 5 Functional characterisation of the role of KAT2A and ATAC and SAGA complexes in human AML primary samples

TABLE 5.1	Characteristics of AML patient samples	154
-----------	--	-----

Abbreviations

AML	Acute myeloid leukaemia
ATAC	Ada-Two-A-Containing
BM	Bone marrow
CB	Cord blood
CFC	Colony-forming cell
ChIP	Chromatin immunoprecipitation
CMP	Common myeloid progenitor
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DUB	Deubiquitination
EDTA	Ethylene diamine tetra acetic acid
ESC	Embryonic stem cells
FBS	Fetal Bovine Serum
GFP	Green fluorescent protein
GMP	Granulocyte monocyte progenitors
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HEPES	4-(2-hydroxyethyl)-1 piperazineethanesulfonic acid
HSCs	Haematopoietic stem cells
KAT	Lysine acetyltransferases
KAT2A	Lysine acetyltransferase 2A
KO	Knockout
MEP	Megakaryocyte erythrocyte progenitors
MLL	Mixed lineage leukaemia

MLP	Myelo-lymphoid progenitors
MNC	Mononuclear cells
MPP	Multipotent progenitors
PBS	Phosphate Buffered Saline
PCAF	p300/CREB-binding protein-associated factor
RNA	Ribonucleic Acid
SAGA	Spt-Ada-Gcn5-Acetyltransferase
SDS	Sodium dodecyl sulphate
shRNAs	short-harpin RNAs
TFs	Transcription factors
TSS	Transcriptional start site
USP22	Ubiquitin Specific Peptidase 22
WT	Wild type
YEATS2	YEATS Domain Containing 2
ZZZ3	Zinc Finger ZZ-Type Containing 3

Chapter 1

Introduction

1.1 Haematopoiesis

The process of blood cell development is termed haematopoiesis, deriving from the Greek: *haima*, meaning “blood”, and *poiéō*, meaning “to make”. The blood is made up of an aqueous fraction, the plasma, and of a broad repertoire of cells that carry out vital functions. Red blood cells supply oxygen/export carbon dioxide to/from tissues, megakaryocytes and their platelet progeny are involved in thrombus formation, and immune cells provide protection against pathogens. Specifically, phagocytic (macrophages and neutrophils) and lymphoid cells (natural killer T cells) employ innate immunity, whereas B and T cells are the players of the acquired immune system. In a healthy individual, approximately 300 billion blood cells are produced in bone marrow (BM) every day. Abnormalities in the production of blood cells can lead to benign complications such as anaemias, as well as severe tumours like leukaemia. Thus, the detailed study of haematopoiesis has immediate implications with regard to understanding the origin of blood-related diseases, but also to decipher the biological mechanisms by which it unfolds, allowing the different blood cell types to attain their unique functions.

1.1.1 Haematopoietic stem cells (HSCs)

1.1.1.1 Origin, identification and isolation

Haematopoiesis begins with rare multipotent stem cells - haematopoietic stem cells (HSCs), the progeny of which differentiate and proliferate into the fully functioning mature cells found in the blood [1]. Our knowledge of blood cells began with studies that sought to understand the effects of radiation following the bombings in Hiroshima and Nagasaki in 1945. Jacobson and colleagues discovered that mice could be protected from lethal irradiation by shielding the animal’s spleen with lead [2]. Lorenz *et al.* then observed the same protective effect by transplanting BM post-irradiation into mice [3]. These, and subsequent studies demonstrated that the transfused BM cells were responsible for protecting the mice from otherwise lethal irradiation, pioneering research of the cells of the marrow. The first functional description of HSCs, however, came a few years later when Tim and McCulloch showed that if BM was taken from one mouse and injected into another, rare and specialised cells would migrate to the spleen, divide and form colonies – colony forming unit spleen (CFU-S) [4], which contained a mixture of different blood cell

types. Incredibly, when the colonies were removed and re-transplanted into another animal, new colonies appeared in the spleen of the secondary recipient. Although it is now understood that CFU-S were formed not by HSCs, but by myeloid progenitor cells [5], this and successive experiments pointed out two key properties of HSCs: (1) self-renewal i.e., the ability to perpetuate itself during extended cell divisions; and (2) differentiation along a certain cell lineage (s), giving rise to specialised cells. From then onwards, HSCs became functionally defined as cells capable of reconstituting multi-lineage haematopoiesis in lethally irradiated recipient mice upon transplantation. Conversely, progenitor cells are defined by the absence of extended self-renewal and restricted lineage differentiation capacity, being lost within 2-3 weeks post-transplantation [1]

Following HSC identification, subsequent research concentrated on their isolation, first from mice and later from humans. Initially, HSCs were isolated based on physical properties in CFC assays, and *via* dye exclusion with reagents such Hoechst 33342, or by 5-fluorouracil [6–7]. Eventually, with the advent of fluorescence activated cell sorting (FACS) and monoclonal antibodies, HSCs and progenitors could be purified on the basis of specific combinations of cell surface markers [8]. Mouse HSCs were first purified and characterised by Spangrude and colleagues as a lineage-negative Sca-1⁺ c-Kit⁺ (LSK) population [9–10]. Later on, research on candidate human HSCs led to the discovery that cells expressing CD34 marker (CD34⁺) contained primitive progenitor activity *in vitro* [11] and *in vivo* [12]. Hence, CD34 was the first marker shown to enrich human HSCs and progenitors, and it is been widely used since to isolate blood stem cells for clinical purposes such as human haematopoietic stem cell transplantation [13].

The knowledge that transplantable HSCs originate progenitors that can be traced in *in vitro* assays, and *in vitro* progenitors lack multilineage repopulation capacity in an *in vivo* setting, combined with FACS, led to the eventual isolation and characterisation of the downstream HSC progeny [14–15]. As a result, a hierarchical arrangement of haematopoiesis was proposed [16–17] (**Fig. 1**). In this classical representation of haematopoietic development, HSCs, at the apex, undergo a series stepwise of bifurcations associated with progressive commitment to increasingly differentiated progenitors and ultimately to terminally differentiated cell types. HSCs initially restrict their self-renewal capacity to generate multipotent progenitors (MPPs); MPPs separate into common myeloid progenitors (CMP), or common

lymphoid progenitors (CLP). CMPs give rise to either megakaryocyte/ erythrocyte progenitors (MEP) or granulocyte/ macrophage progenitors (GMP), whereas CLPs originate Natural Killer cells, T and B cells.

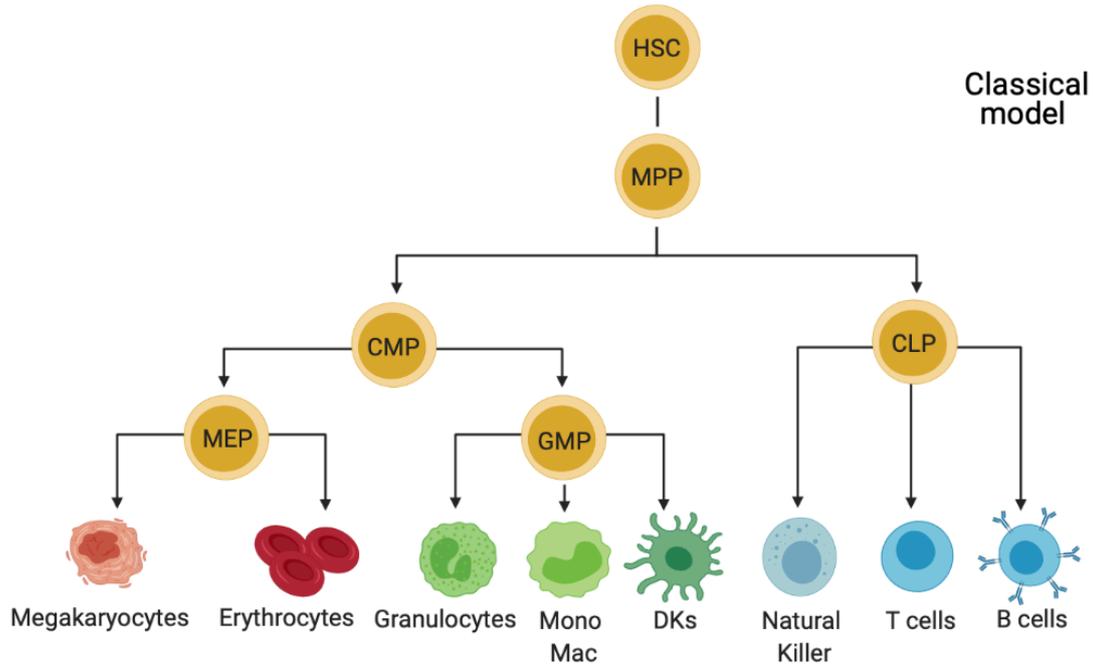


FIGURE 1.1 Classical model of haematopoiesis. This model is based on the identification of common myeloid (CMP) and common lymphoid (CLP) progenitors. Accordingly, the first lineage commitment step of HSCs results in a complete separation myelopoiesis and lymphopoiesis. Each step is associated with a binary lineage branchpoint, resulting in a tree-like model. This representation defines a set of progenitors and mature lineages and links them with a discrete set of differentiation stages. The mature cell types can be traced back to the HSCs through the hierarchical organisation. HSCs, haematopoietic stem cells; MPP, multipotent progenitor; CLP, common lymphoid progenitor; CMP, common myeloid progenitor; MEP, megakaryocyte/erythrocyte progenitor; GMP, granulocyte/macrophage progenitor.

Of note, the classical model is based on the idea that the starting HSC pool is a relatively homogeneous population and contributes equally to each blood lineage.

1.1.1.2 HSCs and haematopoiesis throughout life

Haematopoiesis is generally preserved across vertebrate evolution, although not entirely identical [18], beginning in the early embryo and continuing throughout life. It occurs in distinct temporal waves – primitive and definitive, and at different anatomical sites. In humans, the most primitive blood cells arise in the yolk sac during embryogenesis and consist mainly of nucleated erythrocytes [19]. In turn, definitive haematopoiesis from which HSCs emerge, originates in the aorta-gonad-mesonephros (AGM) from a specialised endothelium called the haemogenic endothelium – a process termed endothelial to haematopoietic transition [19]. After their emergence in the AGM region and as gestation progresses, HSCs colonise the foetal liver (FL), which remains the location for foetal haematopoiesis, before finally lodging in the BM where they reside for the adult life of the organism [20–22]. Occasionally, HSCs can also be found in the spleen [22] and lungs [23]. Moreover, even though the majority of HSCs resides in the BM, a small percentage is released in the peripheral blood with circadian-clock controlled rhythms [24].

During a lifetime, the properties of HSCs differ slightly at each anatomical organ. For instance, studies in mice indicate that FL and BM HSCs can efficiently engraft lethally irradiated animals, whereas HSCs isolated from the AGM do so less efficiently. Moreover, HSCs obtained from the foetus are capable of more robust haematopoietic reconstitution of irradiated recipients and have greater proliferative and differentiation capacity than those isolated from adult BM [25]. HSCs in the BM are mostly dormant, or quiescent. Indeed, quiescence is an important trait of HSCs and regulation of the balance between quiescence and cell cycle entry is key for healthy haematopoiesis. Other characteristics of HSCs include an autophagy-dependent, glycolytic metabolism, low mitochondrial activity and tightly controlled levels of protein synthesis [26–27].

The distinct properties of HSCs observed across different organs highlight the contribution of the microenvironment commonly referred to as ‘niche’ in the support of blood stem and progenitor cell development (reviewed in [28]). The niche is a complex system where a variety of non-haematopoietic cells are present, and thereby can interact with the cells of the blood, to provide cytokines and growth factors, oxygen tension and nutrients [29].

Two distinct HSC niches have been identified: the endosteal HSC niche, which contains osteoblasts as the main supportive cells, and the vascular niche, where HSCs associate with the sinusoidal endothelium in the BM and spleen [30]. Interestingly, a recent study showed that even within the same niche, HSCs and progenitors do not seem to occupy the same locations, indicating spatial segregation of blood lineages to distinct BM regions [31].

In addition to maintaining steady-state haematopoiesis throughout life, HSCs are able to respond to environmental stressors such as infection, and readily produce the cell type(s) in need [32].

The diversity in healthy haematopoietic cells is also shaped by factors including age and genetic mutation [33]. Both proliferative and differentiation potential of HSCs are known to decline with age [7]. Moreover, with age, a tendency for accumulation of HSCs that produce preferentially myeloid output has been reported [34]. Besides, as we age, healthy blood cells can acquire mutations, a phenomenon termed clonal haematopoiesis, and which is present in virtually all adults over 65 years old [35]. Importantly, the majority of these mutations arise in HSCs and is associated with an increased risk to develop haematological malignancies [36–37].

To perpetuate themselves and to maintain an appropriate blood production throughout life, HSCs must then coordinate alternative fates of self-renewal, multilineage differentiation, quiescence or apoptosis in response to not only extrinsic stimuli (mostly provided by the niche), but also to cell-intrinsic cues.

1.1.2 Regulation of haematopoietic fate decisions

First, it is important to clarify the notion of “cell fate decision”, and how it differs from “cell lineage potential”. Lineage potential refers to the *capacity* of HSCs to self-renew or to differentiate along a certain developmental lineage; it encompasses a range of possible fates. Cell fate decision is the *actual behaviour or identity* the cell assumes; the fate of a cell is more restricted than its potential. Understanding how a stem cell decides to remain a stem cell or to differentiate towards a certain lineage relies on the ability to interrogate not only extrinsic cues (e.g.: cytokine availability), but also cell-intrinsic properties, such as transcriptional and epigenetic regulation.

1.1.2.1 The role of cytokines

Hematopoietic cytokines were first identified and purified on the basis of their abilities to support *in vitro* formation of hematopoietic colonies in semisolid media [38]. Cytokines are a large family of extracellular ligands that bind to their respective receptors to stimulate a specific biological response, acting in concert with transcription factors (TFs) and epigenetic regulators. Some affect preferentially HSC self-renewal, whereas others stimulate the growth and proliferation of multiple haematopoietic lineages. HSC-acting cytokines include Stem cell factor (SCF), Thrombopoietin (TPO), Interleukins (ILs) 6 and 11, IL-6 and IL-11 respectively.

1.1.2.2 Transcriptional regulation

TFs bind specific DNA motifs within promoter or enhancer regions and recruit epigenetic complexes (detailed in section 1.3) to modify the structure of chromatin and activate or repress gene expression. Different classes of TFs have different consensus binding motifs. Critical TFs in HSC function are presented in **Table 1.1**. The set of interactions between TFs and their targets is often represented in gene regulatory networks that associate with a certain cell fate/identity. Thus, one mechanism underpinning cell fate choice is that lineage-determining TFs expressed in HSCs and progenitors can instruct fate specification [39]. In other words, a specific cell fate is implemented by transcriptional networks that prompt patterns of gene expression for a particular lineage – instructive model.

On the other hand, lineage-specific TFs are expressed at low levels in multipotent blood progenitors that have not yet committed to a specific fate, so-called multilineage priming [40]. This associates with the existence of multiple transcriptional possibilities that, at the point of fate decision, can be employed in a random way – stochastic model [41–42]. Indeed, stochastic activation of transcriptional regulators was shown to drive lineage commitment prior to establishment of a lineage-determining programme [43]. According to the stochastic hypothesis, commitment occurs independently of extrinsic signals. Instead, it predicts that the function of cytokines is to provide survival and proliferation signals [44]. Since for the most part, cytokine receptors are required for survival /expansion of single lineages, but not lineage commitment events, their instructive roles do not

completely diverge with stochasticity in blood differentiation. Both the instructive and stochastic models have been proposed as mechanisms that result in the upregulation of one lineage program over another. In either case, when cells make a decision, they upregulate the expression of the set of genes of their chosen fate [43].

Recent gene expression analysis of single cells (e.g., by single-cell RNA-seq, *in vivo* lineage tracing) provided further insights into the mechanisms governing lineage fate trajectories from HSCs, challenging the long-standing classical view of haematopoiesis (**Fig. 1.1**). This is to do with 3 main observations:

- (1) *Confirmation of HSC lineage biases*
- (2) *Proof of transcriptional / multilineage priming*
- (3) *Early lineage segregation from HSCs*

(1) *Confirmation of HSC lineage biases*

Back in the 1980s, experiments tracking HSC progeny using retroviral inserts proposed differences in HSC self-renewal capacity and production of mature lineages. This functional heterogeneity prompted the eventual description of lineage-biased HSCs [45]. More recently, single cell transplantation assays confirmed the variability in HSC lineage output and a “bias” towards the generation of a certain lineage. Indeed, some HSCs preferentially produce myeloid or lymphoid output [46–47]. Platelet biased HSCs have also been identified [48]. These observations defy the classical haematopoietic model in which HSCs are thought to contribute equally towards the production of all mature cell lineages.

(2) *Proof of transcriptional / multi-lineage priming*

Single cell profiling of mouse and human HSCs demonstrated transcriptional / multi-lineage priming [49–50]. These analyses suggest that transcriptional lineage programs are already existent within the HSC compartment, which is in line with functional lineage biases in (1).

(3) *Early lineage segregation from HSCs*

Unexpected commitment “shortcuts” from HSCs into mature lineages have been described [48–51–53]. This is conflicting with the classical, stepwise model of haematopoiesis. Whether direct lineage commitment is a rare occurrence, or an inherent characteristic of HSC differentiation remains controversial.

Altogether, these findings led to the realisation – and now widely accepted view in the field – that HSCs represent a heterogeneous population, resulting in revised models of HSC fate specification (**Fig. 1.2**).

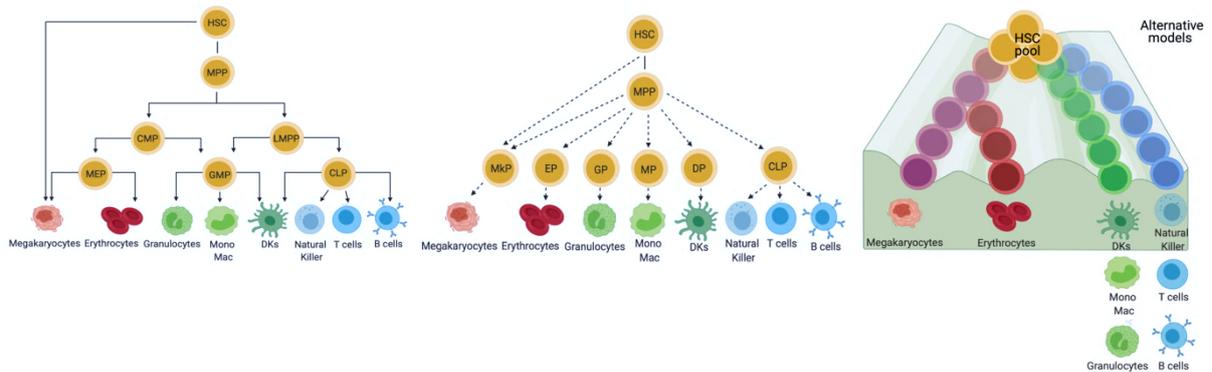


FIGURE 1.2 Alternatives models of HSC lineage commitment. (A) Combined view of the first proposed alternatives to the classical model. Here, the myeloid and lymphoid branches remain associated further down in the hierarchy *via* the lymphoid-primed multipotent progenitor (LMPP) population. Upon loss of GM potential, LMPPs generate MLP. Megakaryocyte specification can happen directly from HSCs. **(B)** Early-separation model. In this model of haematopoietic hierarchy, lineage commitment is specified very early in the HSC/MPP, as opposed oligopotent progenitors. **(C)** Continuum model. Snapshots from single-cell transcriptomics propose that cells differentiate through continuous/ smooth trajectories from a low primed HSC pool. Often portrayed in Conrad Waddington’s landscape (compare Fig. 1.1).

Table 1.1 Main haematopoietic transcription regulators and their roles.

Gene/ TF factor	Motif/ family	Regulatory function in HSCs	Ref.
<i>Erg</i>	ETS	Essential in haematopoietic development	[54]
<i>Fli1</i>	ETS	Essential in haematopoietic development; megakaryopoiesis	[55]
<i>Foxo3a</i>	-	Critical for HSC self-renewal	[56]
<i>Gata1</i>	GATA	Overexpression restricts HSC self-renewal and induces erythroid differentiation	[57–58]
<i>Gata2</i>	GATA	Essential for HSC formation in the AGM and adult HSC survival; erythropoiesis	[59]
<i>Gata3</i>	GATA	Required for HSC cell cycle entry and proliferation; roles in T-cell development	[60]
<i>Hoxa5</i>	HOX	Overexpression decreases HSC pool, inducing erythroid differentiation	[61]
<i>Hoxa9</i>	HOX	Overexpression results in expansion of HSC and early progenitors	[62]
<i>Lyl1</i>	bHLH	Essential for maintenance of adult HSCs in the absence of SCL	[63]
<i>Myb</i>	-	Required for definitive haematopoiesis – HSC self-renewal and multilineage differentiation with roles in myelopoiesis and erythropoiesis	[64]
<i>Runx1</i>	-	Essential for primitive (foetal) haematopoiesis; colocalizes with TFs specific for the HSC/MPP compartment	[65]
<i>PU.1</i>	ETS	Critical in supporting HSC self-renewal, as well as commitment and maturation of myeloid and lymphoid lineages	[66]
<i>SCL/Tal1</i>	bHLH	Essential for the formation, but not maintenance, of adult HSCs	[63]

1.1.3 Overview of myelopoiesis and erythropoiesis

1.1.3.1 Myelopoiesis and its regulation

Myelopoiesis refers to the origin of myeloid cell types, which are the dominant cellular population in the BM. As described earlier, myeloid cells are derived from HSCs, which differentiate towards MPPs, followed by CMPs and GMPs that in turn, become granulocytes and monocytes. In recent years, additional progenies of GMPs have been proposed, namely macrophage and dendritic cell (DC) precursors, and common monocyte progenitors [67]. Thus, GMPs are now thought to constitute a pool of cell types with a spectrum of differential possible fates, ranging from exclusive neutrophil, monocyte or DC progeny. Moreover, the majority of DCs was found to derive from an interleukin-7 (IL-7) receptor-positive progenitor, highlighting their predominant lymphoid origin [68].

Several molecular mechanisms are involved in myeloid cell lineage specialisation. Regulators of central importance to myelopoiesis are growth factors necessary for the proliferation and differentiation of progenitors such as granulocyte-colony stimulating factor (G-CSF), monocyte-CSF (M-CSF), and GM-CSF that foster differentiation towards their name-given lineages. Furthermore, myeloid gene expression is regulated by key TFs, including the CCAAT/enhancer-binding proteins (C/EBPs), PU.1 and its interacting proteins such as c-Jun and GATA-1), Core binding factors (CBFs), c-Myb and Egr-1 [69]. CEBPA in particular, is a major regulator of commitment into the myeloid lineage, and its deletion results in a complete block of monocyte and neutrophil (but not DCs) specification downstream of the GMP [70]. Myeloid cells also express characteristic cell surface markers, that are used for their identification and isolation. Early myeloid differentiation antigens include CD13 and CD33 [71].

Atypical myelopoiesis is prevalent in hematologic malignancies including multiple myeloma, lymphomas and leukaemias [72]. In fact, many of the same TFs that regulate normal myeloid differentiation are now known to participate in the development of leukaemia. In this respect, acute and chronic myeloid leukaemias are considered the most extreme cases of deregulated myelopoiesis, as I will be addressing later for AML.

Relevant experimental strategies to investigate the myeloid compartment under normal or malignant settings include immunologic phenotyping, functional studies e.g. gene disruption assays including inducible Cre/loxP recombination systems to control gene activity in a site and time-specific manner and cell transcriptomic approaches.

1.1.3.2 Erythropoiesis and its regulation

Red blood cells are the most abundant cell type in the blood and healthy adults produce an average of 2 million new red blood cells per second, in a process called erythropoiesis [73]. This is adjusted on demand, depending on oxygen levels, growth factor availability, iron metabolism or pathological conditions involving blood loss. Erythropoiesis occurs in 2 waves, primitive and definitive [74]. Primitive erythropoiesis is a single event that occurs during embryonic development, in the yolk sac as previously mentioned, and is essential for the survival of the embryo. These primitive erythroid cells synthesise embryonic globin forms and are nucleated and short-lived [75].

Definitive erythropoiesis, in turn, happens first in the FL and later, in postnatal BM throughout life, in different stages (**Fig. 1.3**). Classically, it evolves from HSCs to bipotent MEPs, to erythroid progenitors, erythroblast precursors and finally erythroblasts. The first, early stage of definitive erythropoiesis, is defined by the presence of two immature erythroid-restricted progenitors: the burst-forming unit erythroid (BFU-E) and colony-forming unit erythroid (CFU-E). As the name suggests, these progenitors have the ability to form colonies in semisolid media. Typically, human CFU-E progenitors require 7-10 days to form mature colonies [76]. The second stage is that of terminal erythroid differentiation, occurring in erythroblastic islands, and which begins with the appearance of immature proerythroblasts that undergo 3-4 rounds of cell division. First, they divide to become basophilic erythroblasts followed by polychromatic erythroblasts and orthochromatic erythroblasts, which expel their nucleus, originating reticulocytes [75].

These cellular transformations are accompanied by changes in cell surface marker expression, cell size, shape, and a progressive increase in adult haemoglobin forms such as β -globin (*HBB*) [77]. Reticulocytes then migrate from the BM to the bloodstream where they finally mature in erythrocytes. The lack of nucleus and their characteristic biconcave shape is important because (1) it permits an increased

storage of haemoglobin to be delivered to the body and (2) it allows flexible movement through small blood vessels for efficient tissue oxygenation.

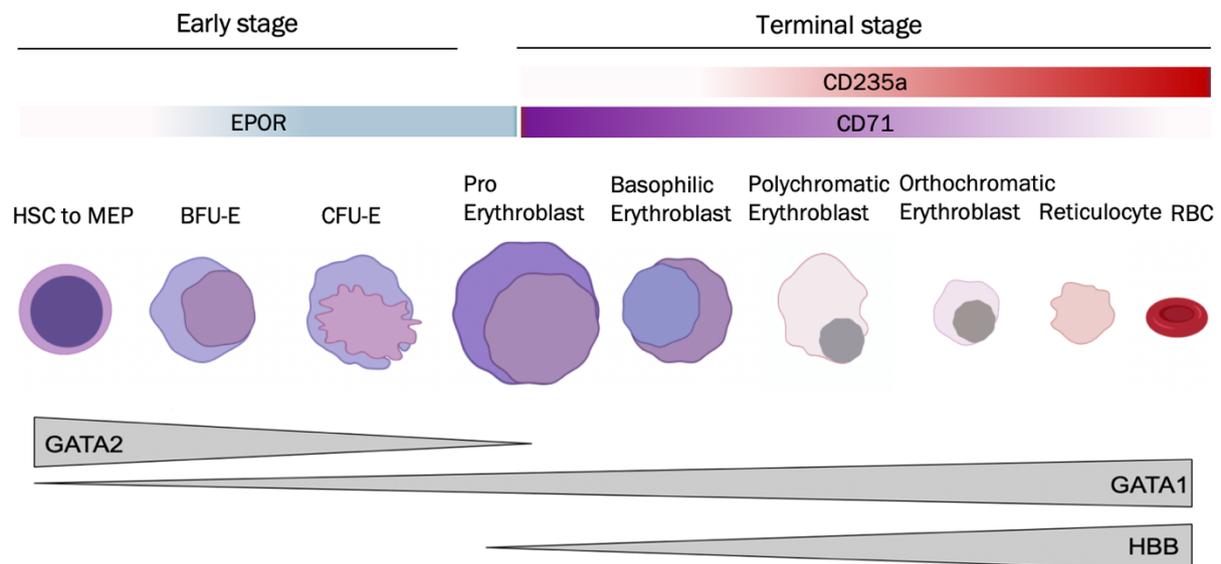


FIGURE 1.3 Model of definitive erythroid cell differentiation. The early stage of definite erythropoiesis gives rise to the burst-forming unit erythroid progenitor (BFU-E) and the colony-forming unit erythroid (CFU-E). In the terminal stage, proerythroblasts divide originating progressively smaller cells with increased chromatin condensation and increased levels of haemoglobin. Reticulocytes lose their nucleus and give rise to RBCs or erythrocytes found in blood circulation. Expression of commonly used cell surface markers to identify the various stages is indicated by the bars. Depicted are EPOR, CD71/transferrin receptor and CD235a/ glycoprotein A. EPOR is expressed in low levels in HSCs and BFU-Es and in high levels in CFU-Es and proerythroblasts. CD71 is responsible for iron uptake and expressed mainly in cells synthesising haemoglobin, with its expression decreasing during terminal differentiation. CD235a is expressed in later stages of erythroid differentiation. Opposing patterns of GATA1 and GATA2 expression during erythropoiesis are illustrated. *HBB* expression increases in terminal erythropoiesis.

Erythropoiesis is regulated at multiple levels [78], requiring growth factors as well as the concerted action of a plethora of TFs (**Table 1.2**). Further players include hormones (thyroid, androgens, activin/ inhibin), iron metabolism, vitamins (folic acid and vitamin B12), as well as epigenetic mechanisms [79]. Here, I will briefly describe the role of erythropoietin (EPO), the master erythroid cytokine, and the main erythroid transcriptional regulators.

Erythropoietic cytokines – the role of erythropoietin (EPO)

Erythropoietin (EPO), a glycoprotein hormone, is considered the master cytokine for erythropoiesis [80]. Indeed, EPO governs proliferation, differentiation, and activation of anti-apoptotic mechanisms during erythroid development [81–82]. Low oxygen levels prompt the kidneys to release EPO, which then binds to its receptor (EPOR). EPO binding induces receptor conformational changes, triggering downstream signalling cascades through JAK2, STAT5, PI3K/AKT, and MAPK pathways [82] for activation of transcription. Both EPO and EPOR-null mice die during embryogenesis due to severe anaemia, with a complete lack of definitive erythrocytes [80]. More specifically, the defect was observed at the CFU-E stage, whereas prior stages developed normally, indicating that erythroid lineage commitment is not affected by lack of either EPO or EPOR [80]. More recently, however, EPO has been suggested to play important functions during earlier stages than previously thought [83]. Specifically, it was shown to potentiate erythroid lineage commitment through suppression of non-erythroid fate choices at the HSC/MPP compartment [83]

Major erythroid TFs

The GATA TFs

The GATA family of TFs binds the DNA consensus sequence (A/T)GATA(A/G) [84], present in a number of erythroid-affiliated genes such as globins, *KLF1* and *EPOR*. Of the 6 members of the GATA family, GATA1, GATA2 and GATA3 are involved in haematopoiesis [85]. GATA1 is considered the “master” erythropoietic TF as it regulates all aspects of erythroid maturation and function [84]. *GATA2* is most highly expressed during early blood cell development, and its expression in erythropoiesis declines as cells differentiate, whereas *GATA1* expression increases (**Fig. 1.3**). Moreover, *GATA1* inhibits *GATA2* expression during terminal erythroid differentiation [84–85]. *GATA1* KO mice exhibit embryonic lethality at around day E10.5-E11.5 due to severe anaemia [86], and *GATA1*-null cells do not progress

beyond the proerythroblast stage, undergoing apoptosis [87]. *GATA2*-null animals, in turn, exhibit severe anaemia [88]. Despite having similar binding motifs, other GATA-binding proteins cannot compensate for the absence of *GATA1*, highlighting its critical function erythroid development [86]. Typically, *GATA1* acts in multiprotein complexes with specific transcriptional repressor or activator functions, including with TAL1/SCL, LMO2, and FOG1 [84–89–91]. Besides, erythroid fate specification might centre on the antagonism between *GATA1* and *PU.1* [75–92]. Higher *GATA1* expression activates MEP molecular programs while acting to repress myeloid specification by direct interaction with PU.1, blocking its transcriptional targets [92].

TAL1/SCL

TAL1/SCL encodes a basic helix-loop-helix (bHLH) type TF pivotal for both embryonic and adult haematopoiesis [90]. *TAL1*-deficient mice die of severe anaemia between embryonic days E9.5-10.5 [90]. Specifically, in erythropoiesis, TAL1 acts at the MEP level, supporting MEP survival, downstream of *c-KIT*, and establishing a positive feedback loop in MPP and MEP [89].

KLF1

Krüppel-like factor 1, KLF1, is essential for the development and survival of erythrocytes, with *KLF1*-deficient mice dying of anaemia at E15.5 [93]. More specifically, KLF1 participates at the MEP stage, where it acts in a cross regulatory circuit with *FLI1*, to regulate megakaryocytic *vs* erythroid commitment decision [93]. Additionally, it plays critical functions at the terminal maturation stages, regulating the expression of several erythroid-specific genes including globins, heme biosynthetic enzymes, cytoskeletal proteins and blood group antigens [94]. *KLF1* expression is restricted to erythroid cells and their precursors.

FOG

The Friend of GATA (FOG) family of proteins contain a conserved zinc finger (ZF) domain that bind GATA TFs, modulating their activity. Mice lacking FOG die of anaemia between E10.5 and 12.5, although, FOG role in primitive erythropoiesis is thought occur by GATA-independent mechanisms [95]. In terminal, erythroid maturation, mutated GATA-1, unable to interact with FOG, fails to support erythroid development due to deregulated expression of GATA-1 target genes [91].

LMO2

LMO2 is a LIM domain-containing protein unable to bind DNA itself. It does so by interacting with SCL/TAL1 and GATA1, forming complexes that drive erythroid gene expression programmes. LMO2 is highly expressed in proerythroblasts, and its expression decreases with differentiation. Also, LMO2 interacts with the DNA replication complex that drives entry into S phase in erythroid precursors, preventing commitment to terminal erythroid maturation [96].

Table 1.2 Major erythroid transcriptional regulators.

Gene/ TF factor	Motif/ family	Regulatory function in erythropoiesis	Ref.
<i>Fli1</i>	ETS	Overexpression inhibits erythroid differentiation	[97–98]
<i>Fog</i>	ZF	Ablation blocks erythroid maturation at the proerythroblast stage	[99]
<i>Gata1</i>	GATA	Overexpression restricts HSC self-renewal and induces to erythroid differentiation	[84]
<i>Gata2</i>	GATA	Expressed during early stages of erythropoiesis	[85]
<i>Klf1</i>	bHLH	Severe anaemia; beta globin deficiency	[58–94]
<i>Lmo2</i>	LIM	Bloodless mice-absence of yolk sac haematopoiesis	[100]
<i>PU.1</i>	ETS	Interacts with GATA1 during erythroid differentiation, repressing GATA1 mediated transcription	[92]
<i>SCL/Tal1</i>	bHLH	Supports survival of erythroid committed progenitors at the MEP stage	[101]

1.2 Acute Myeloid Leukaemia (AML)

1.2.1 AML as a clinical entity

Early descriptions of leukaemia as “globules of pus and mucous” date back 200 years, and amount to the work of a number of physicians including Alfred Donné, John Hughes Bennett and Rudolf Virchow [102].

Acute myeloid leukaemia (AML) is one, very aggressive, type of leukaemia characterised by the uncontrolled expansion of immature myeloid cells. These cells arrested in differentiation, so called “blasts”, infiltrate BM, peripheral blood and occasionally other tissues [103]. Critically, their accumulation in the BM blocks normal haematopoiesis, ultimately leading to BM failure. As a consequence, patients with AML usually present with bruising and/or bleeding caused by anaemia or thrombocytopenia, infections due to low white blood cells, persistent tiredness, and weight loss, amongst other symptoms. AML diagnosis is usually established by the presence of >20% blasts in the BM or blood. The median age at diagnosis is 67 years old. Further diagnosis relies on testing for myeloperoxidase activity (to confirm myeloid origin), karyotyping, immunophenotyping, and more recently, on sequencing techniques [103–104].

AML is the most common type of leukaemia in adults, with an incidence of 4 per 100,000 females and 6 per 100,000 males in the UK per year. Furthermore, its incidence increases with age to 15 cases per 100,000 at the age of 75 [104]. With respect to its aetiology, AML can arise in patients with an underlying, pre-leukaemic blood malignancy, or as a consequence of prior therapy. Indeed, non-cancerous myeloproliferative neoplasms can transform into AML [105]. Likewise, treatments for unrelated malignant diseases involving radiation or cytotoxic drugs can also increase the chance of developing AML, referred to as “therapy-related AML”. Other factors associated with the likelihood of developing the disease include Down Syndrome and being a direct relative of an AML patient. In the majority of cases, however, AML appears *de novo* in previously healthy individuals.

1.2.2 Biology of leukaemia stem-like cells

At the cellular level, AML is a clonal proliferation of myeloid cells in differentiation arrest that acquired a survival advantage over their normal counterparts. In the 1990s, transplantation of AML cells (lineage-negative CD34⁺ CD38⁻) into non-obese diabetic / severe combined immunodeficiency disease mice (NOD/SCID mice) was shown to mimic the leukaemic characteristics observed in human patients [106]. Bonnet and Dick then established a hierarchical model of AML where leukaemia-initiating cells (LICs) present in the CD34⁺ CD38⁻ fraction, were able to maintain the disease long-term in SCID mice [107]. LICs comprised approximately 1 in 1 x 10⁶ leukemic blasts and possessed two key properties akin to physiological stem cells: (1) self-renewal i.e., they gave rise to leukemic engraftment that could be propagated for multiple serial transplants and (2) differentiation, i.e. they produced leukaemia progeny equivalent to the cellularity in patients, including those cells unable to engraft. Thereby, AML leukaemia-initiating cells resembled normal HSCs. In a more rigorous definition, cells that meet both criteria (1) and (2) are termed leukaemia-stem cells (LSCs) [108]. Later studies performed in better engrafting mouse strains including NOD SCID Gamma (NSG) mice, expanded the characterisation and understanding of self-renewing leukaemia stem cells. Indeed, Taussig and colleagues demonstrated that progenitor cells (CD34⁻ CD38⁺) derived from AML patients were able to recapitulate the leukaemic phenotype [109]. This and other observations, proved that both HSC-like and progenitor-like immunophenotypes could generate AML when transplanted in recipient mice [110–111]. Accordingly, Sarry *et al.* showed that AML leukaemic stem-like cells reside in CD38⁺ and CD45RA⁺ populations, which upon transplantation recapitulated the primary AML [112]. Similarly, the Vyas group demonstrated that, rather than resembling HSCs, the majority of leukaemic initiating cells from CD34⁺ AMLs, have immunophenotype and gene expression signatures characteristic of GMP or LMPP-like progenitors [113]. Together, this established the notion that most candidate LICs in CD34⁺ AML are progenitor-type cells with underlying dysregulated self-renewal programmes. Importantly, the properties and frequency of LICs vary greatly across different AML subtypes, emphasising the heterogeneity of the disease. Moreover, they can also change during the course of the disease. Indeed, at relapse the frequency of leukaemia-stem cells, able to resist chemotherapy, is often higher than it was prior treatment [114].

1.2.3 AML classification

As alluded to before, AML is heterogeneous, and it is often said that is not one disease, but “many diseases”. In fact, there are a number of distinct subgroups, which are categorised on the basis of cell morphology, cytogenetics and molecular profile. A morphological classification of AML centred on the cytological characteristics of leukaemia blasts was in use until the end of the last century. This classification by the French American British (FAB) system, consist of 8 subtypes: FAB M0-M7 [115]. Presently, the World Health Organisation (WHO) classification system, 2016, defines AML as 6 major disease entities: (1) AML with recurrent genetic abnormalities; (2) AML with myelodysplasia-related changes; (3) Therapy- related neoplasms; (4) AML not otherwise specified; (5) Myeloid sarcoma and (6) Myeloid proliferations related to Down syndrome (**Table 1.3**) [116]. Category (4) incorporates elements of the FAB classification.

Table 1.3 WHO Classification of Acute Myeloid Leukaemia and its subtypes (revised version, 2016). Adapted from Arber *et al.* 2016 [116].

AML with recurrent genetic abnormalities <ul style="list-style-type: none">○ AML with t(8;21)(q22;q22.1); RUNX1-RUNX1T1○ AML with Inv(16)(p13.1q22) or t(16;16)(p13.1;q22); CBFβ-MYH11○ AML with PML-RARA○ AML with t(9;11)(p21.3;q23.3); MLLT3-KMT2A○ AML with t(6;9)(p23;q34.1); DEK-NUP214○ AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); GATA2, MECOM○ AML (megakaryoblastic) with t(1;22)(p13.3;q13.3); RBM15-MKL1○ Provisional entity: AML with BCR-ABL1○ AML with mutated NPM1○ AML with biallelic mutations of CEBPA○ Provisional entity: AML with mutated RUNX1
AML with myelodysplasia-related changes
Therapy-related neoplasms
AML, NOS <ul style="list-style-type: none">○ AML with minimal differentiation○ AML without maturation○ AML with maturation○ Acute myelomonocytic leukaemia○ Acute monoblastic/monocytic leukaemia○ Pure erythroid leukaemia○ Acute basophilic leukaemia○ Acute panmyelosis with myelofibrosis
Myeloid sarcoma
Myeloid proliferations related to Down Syndrome <ul style="list-style-type: none">○ Transient abnormal myelopoiesis (TAM)○ Myeloid leukaemia associated with Down Syndrome

1.2.4 Genomic and epigenomic landscape of AML

In recent years, genomic sequencing has significantly advanced knowledge of the molecular landscape of AML. Critically, those studies led to the identification of new AML mutations, as well as patterns of co-occurrence or exclusivity amongst them. A landmark study by The Cancer Genome Atlas (TCGA) Research Network analysed the genomes of 200 newly diagnosed AML patients using whole genome and whole exome sequencing, RNA sequencing and DNA methylation profiling [117]. The authors categorised AML mutated genes into distinct functional categories, being: TF gene fusions; *NPM1*; tumour suppressor genes; myeloid TF genes, members of the cohesin complex, spliceosome-genes and two groups of epigenetic modifiers (**Fig. 1.4**).

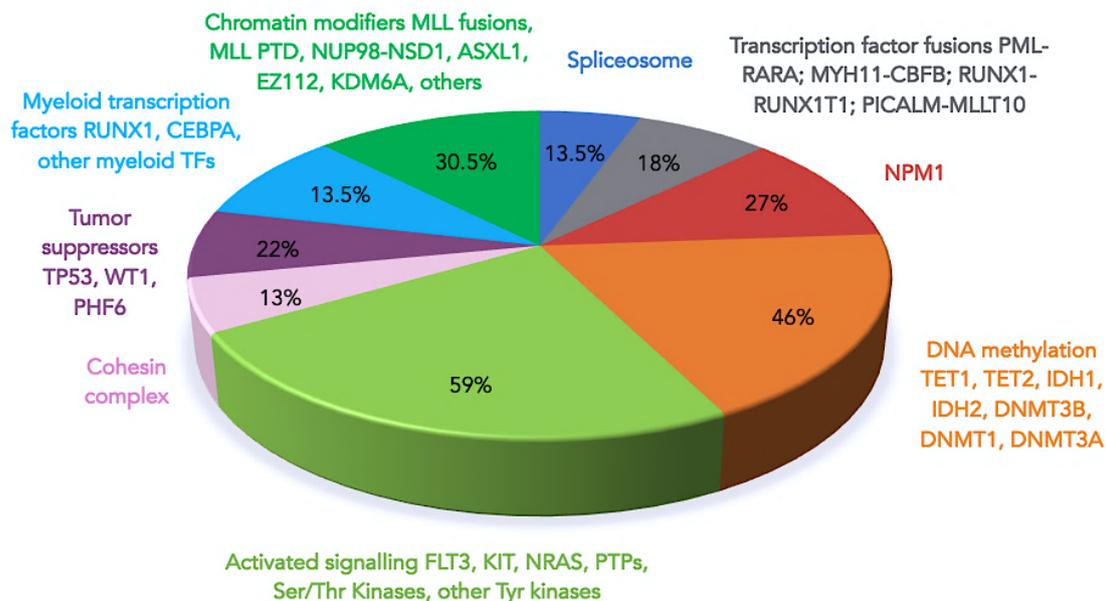


FIGURE 1.4 Functional categories of mutated genes involved in AML pathogenesis. Nine functional categories were identified. Each AML case may have several gene mutations in different functional categories. PTPs, protein tyrosine phosphatases; *MLL PTD*, *MLL* partial tandem duplication. Data from The Cancer Genome Atlas (TCGA) Research Network. Adapted from Chen *et al.* 2013 [117].

A striking observation from this study was that, despite being characterised by lower levels of genetic mutations compared to other cancers, a large percentage of AML mutations was found in epigenetic regulators (>70% AML cases). Epigenetic regulation will be discussed in detail in section 1.3. Moreover, variant-allele frequency analysis revealed that >50% of AML cases had a founding clone and

subclones, indicating possible clonal evolution during disease progression [117]. Mutations with cooperative nature were found to be those in *FLT3*, *DNMT3A* and *NPM1* genes. On the contrary, mutations in TF fusion genes, *NPM1*, *RUNX1*, *TP53* and *CEBPA* had higher mutually exclusivity patterns [117]. Interestingly, roughly 45% of the patients harbour a normal karyotype at diagnosis. Acquired genetic mutations can be identified in approximately 97.3% cases of AML.

More recently, Papaemmanuil and colleagues have profiled the genomes of more than 1000 AML patients who have undergone treatment. Importantly, they correlated newly identified genetic mutations with clinical outcome [118]. This new information on AML genetics was used to risk stratify patients, in order to inform the likely prognosis by the European Leukemia Net (ELN) (**Table 1.4**). In particular, AML with mutated *TP53*, *RUNX1* and *ASXL1* were included in the adverse risk group. About 60% of AML patients are considered of intermediate risk cytogenetics, and in these cases genetic mutations add important prognostic value. As examples, *NPM1* alone and *CEBPA* with double mutation are of favourable prognosis, whereas *FLT3-ITD* carry dismal prognosis [118].

Table 1.4 Risk stratification of AML according based on its genetics. Adapted from the European Leukaemia Net.

RISK CATEGORY	GENETIC ABNORMALITY
adverse	<ul style="list-style-type: none"> ○ t(6;9)(p23;q34.1); DEK-NUP214 ○ t(v;11q23.3); KMT2A rearranged ○ t(9;22)(q34.1;q11.2); BCR-ABL1 ○ Inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); GATA2, EVI1 ○ -5 or del(5q); -7; -17/abn(17p) ○ Complex karyotype, monosomal karyotype ○ Wild type NPM1 and FLT3-ITD^{high} ○ Mutated RUNX1 ○ Mutated ASXL1 ○ Mutated TP53
intermediate	<ul style="list-style-type: none"> ○ Mutated NPM1 and FLT3-ITD^{high} ○ Wild type NPM1 without FLT3-ITD or with FLT3-ITD^{low} ○ t(9;11)(p21.3;q23.3); MLLT3-KMT2A
favorable	<ul style="list-style-type: none"> ○ t(8;21)(q22;q22.1); RUNX1-RUNX1T1 ○ Inv(16)(p13.1q22) or t(16;16)(p13.1;q22); CBFB-MYH11 ○ Mutated NPM1 without FLT3-ITD or with FLT3-ITD^{low} ○ Biallelic mutated CEBPA

Additionally, altered chromatin status in AML associates with poor prognosis [119]. Notably, non-mutated chromatin regulators also perform critical functions in sustaining AML e.g.: *DOT1L*, *BRD4* and *LSD1* [120–121], which I will consider in more detail further ahead.

Sequencing studies also revealed that many AML patients harboured a population of mutated pre-leukaemic HSCs, evident in patients both at diagnosis and remission [122]. Further analysis shown that the clonal evolution in AML proceeds through the stepwise acquisition of mutations in self-renewing pre-leukaemic HSCs. This led to the suggestion that relapse could either happen from leukaemic clonal or sub clonal outgrowth but also from further evolution of pre-leukaemia mutated clones. Moreover, mutations in pre-leukaemic HSCs were enriched in epigenetic modifiers such as *TET2* and *DNMT3A*, underscoring a role for epigenetic dysregulation in establishing a pre-leukaemic state [123].

1.2.5 AML treatment

As previously mentioned, AML is molecularly heterogeneous, which poses significant challenges in terms of its therapy. Patients <60 years of age present curable outcomes of ~40% with standard cytotoxic chemotherapy regimens established in the 1970s, but elderly patients, unable to tolerate the intensive therapy will eventually die of the disease [103]. Remarkably, mainstay treatments for the majority of AML subtypes have remained relatively unchanged over the past 30 years. Only recently, with the advent of sequencing techniques and identification of new targets, novel therapeutics have been developed (**Fig. 1.5**).

Given the pervasive epigenetic dysregulation in AML (discussed in detail in section 1.3), therapeutic strategies aimed at modulating epigenetic mechanisms have gained a lot of interest in recent years [121–124–125]. The first group of epigenetic compounds approved for the treatment of AML was that of DNA hypomethylating agents, which include Azacytidine and Decitabine (**Fig. 1.5**) and act by depleting DNA methyltransferase enzymes (explained in section 1.3). Another way to exploit epigenetic regulators for therapeutic benefit is by changing the levels of their required metabolites. As examples, 2 drugs: Ivosidenib and Enasidenib which inhibit mutant isocitrate dehydrogenase (IDH) family of proteins - IDH1 and IDH2, respectively, have been approved for the treatment of AML (**Fig. 1.5**). Both drugs reduce levels of 2-hydroxygluturate (2-HG), abolishing inappropriate inhibition of Ten-Eleven Translocation (TET) enzymes, and ultimately resulting in lower levels of DNA methylation with release of differentiation block induced by these mutations.

Targeting epigenetic modifiers that repress differentiation programmes are another therapeutic strategy. For instance, pre-clinical studies have shown that inhibition of histone demethylase 1 (LSD1) results in LSC differentiation [126], and inhibition of histone lysine methyltransferase DOT1-like (DOT1L), reduces leukaemogenic potential particularly in MLL-rearranged leukaemias [127]. Bromo and extra terminal domain (BET) family of acetyl-lysine recognising proteins are also exciting therapeutic targets in AML, specifically for aggressive MLL-fusion leukaemias [121–124]. For example, BRD4 contains 2 acetyl-reader bromodomains (BD1 and BD2), playing a key role in gene transcription elongation. By preventing the BRD4 interaction with chromatin, BET inhibitors were shown to downregulate genes such as BCL2, c-MYC and CDK6, leading to apoptosis and cell cycle arrest of MLL–AF9

(murine) and MLL–AF4 (human) leukaemias [124]. Of note, all these targets are not mutated in AML; instead, they constitute dependencies that sustain malignant transcriptional programmes. Emerging evidence indicates that the immune system plays important roles in the efficacy of epigenetic therapies. As such, many of these epigenetic agents are being tested in combination with immunotherapy [121]. A summary of the main clinical trials assessing combination strategies which include epigenetic therapies can be found in **Table 1.5**.

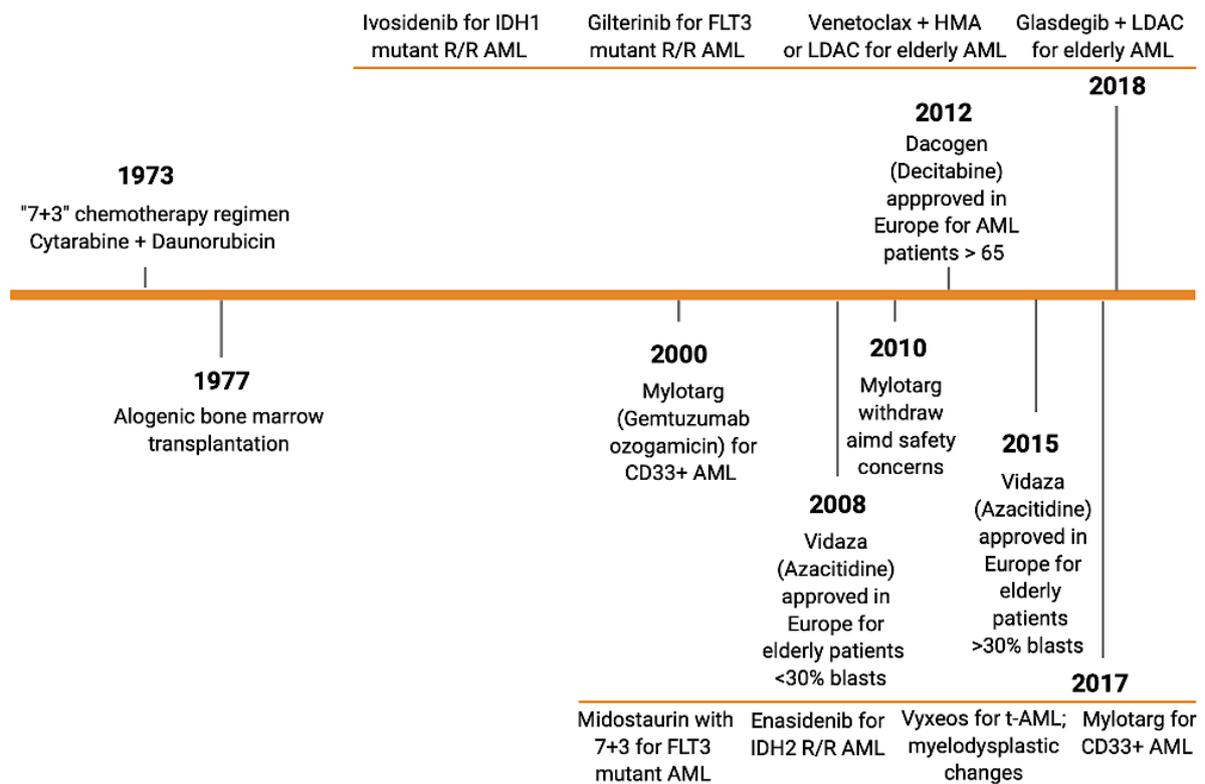


FIGURE 1.5 Timeline of currently approved AML therapies. Adapted from Fennell *et al.* 2019 [121]

Table 1.5 Combination strategies involving epigenetic therapies in clinical trials.
Source: Fennell et al., 2019 [121].

Drug	Targets	Phase	Date	Clinical trial
Combinations of epigenetic therapies				
Azacitidine + entinostat (MS275)	DNMT + HDAC	2	2011	NCT01305499
Azacitidine + FT-2102	DNMT1 + IDH1	1/2	2016	NCT02719574
Azacitidine + LSD1 inhibitor (NCB059872)	DNMT1 + LSD1	1/2	2016	NCT02712905
Azacitidine + ivosidenib (AG-120)	DNMT + IDH1	3	2017	NCT03173248
Azacitidine + pracinostat (SB939)	DNMT + HDAC	3	2017	NCT03151408
Decitabine + vorinostat before or during FLAG	DNMT + HDAC	1	2017	NCT03263936
Azacitidine + PRMT5 inhibitor (GSK3326595)	DNMT + PRMT5	1	2018	NCT03614728
Low-dose azacitidine + vorinostat after alloHSCT	DNMT + HDAC	1	2019	NCT03843528
Epigenetic therapies in combination with immunotherapies				
Azacitidine + nivolumab (MDX-1106) +/- ipilimumab (MDX-010)	DNMT + PD-1 + CTLA-4	2	2015	NCT02397720
Azacitidine + pembrolizumab (MK-3475)	DNMT + PD-1	2	2016	NCT02845297
Decitabine + ipilimumab (MDX-010)	DNMT + CTLA-4	1	2016	NCT02890329
Guadecitabine + atezolizumab (MPDL 3280A)	DNMT + PD-L1	1/2	2016	NCT02935361
Decitabine + PDR001 +/- MBG453	DNMT + PD-1 + TIM-3	1	2017	NCT03066648
Azacitidine + Hu5F9-G4	DNMT + CD47	1	2017	NCT03248479
Decitabine + CDX-1401 + poly ICLC + nivolumab (MDX-1106)	DNMT + DEC-205 + TLR-3 + PD-1	1	2017	NCT03358719
Decitabine + avelumab	DNMT + PD-L1	1	2018	NCT03395873
Azacitidine + nivolumab (ADVL1412)	DNMT + PD-1	1/2	2019	NCT03825367
Epigenetic therapies in combination with targeted therapies				
Azacitidine + milademetan	DNMT + MDM2	1	2014	NCT02319369
Decitabine + rapamycin or ribavirin	DNMT + mTOR	1/2	2014	NCT02109744
Decitabine + BI836858	DNMT + CD33	2	2015	NCT02632721
Decitabine + BP1001	DNMT + Grb-2	2	2016	NCT02781883
Azacitidine + gilteritinib (ASP2215)	DNMT + FLT3	2/3	2016	NCT02752035
Decitabine + talazoparib	DNMT + PARP	1/2	2016	NCT02878785
Decitabine + onvansertib	DNMT + PLK1	1b/2	2017	NCT03303339
Azacitidine + AZD2811 nanoparticles	DNMT + AURKB	2	2017	NCT03217838
Azacitidine + SL-401	DNMT + IL3	1	2017	NCT03113643
Decitabine + AMG-232	DNMT + MDM2	1	2017	NCT03041688
Decitabine + pevonedistat (MLN4924)	DNMT + NEDD8	1	2017	NCT03009240
Azacitidine + pevonedistat	DNMT + NEDD8	3	2017	NCT03268954
Decitabine + venetoclax (ABT-199)	DNMT + BCL-2	2	2018	NCT03404193
Azacitidine + pevonedistat	DNMT + NEDD8	2	2018	NCT03709576
Azacitidine + HMPL-523	DNMT + SYK	1	2018	NCT03483948
Decitabine + quizartanib (AC-220)	DNMT + FLT3	1/2	2018	NCT03661307
Azacitidine + enasidenib mesylate (AG-221 mesylate)	DNMT + IDH2	2	2018	NCT03683433
Azacitidine + nintedanib (BIBF-1120)	DNMT + VEGF + FGFR + PDGFR	1	2018	NCT03513484
Belinostat (PCD-101) + pevonedistat (MLN4924)	HDAC + NEDD8	1	2018	NCT03772925
Azacitidine + glasdegib (PF-04449913)	DNMT + SHH	3	2018	NCT03416179
Pracinostat + gemtuzumab ozogamicin	HDAC + CD33	1	2019	NCT03848754
Azacitidine or decitabine + venetoclax (ABT-199)	DNMT + BCL-2	3	2019	NCT03941964
Azacitidine + APR-246	DNMT + PS3	2	2019	NCT03931291

A list of currently active clinical trials that involve epigenetic therapies in combination with targeted therapies, immunotherapies, and/or other epigenetic therapies for the treatment of AML, alloHSCT, allogeneic hematopoietic stem cell transplantation; AURKB, Aurora kinase B; BCL-2, B-cell lymphoma 2; DNMT, DNA methyltransferase; FGFR, fibroblast growth factor receptor; Grb2, growth factor receptor bound protein 2; HDAC, histone deacetylase; IDH, isocitrate dehydrogenase; IL3, interleukin 3; LSD1, lysine-specific histone demethylase 1; MDM2, mouse double minute 2; mTOR, mammalian target of rapamycin; PARP, poly ADP ribose polymerase; PD-1, programmed cell death protein 1; PDGFR, platelet-derived growth factor receptor; PD-L1, programmed cell death ligand 1; PLK1, polo-like kinase 1; PRMT5, protein arginine N-methyltransferase 5; Shh, Sonic hedgehog; TIM-3, T-cell immunoglobulin and mucin-domain containing-3; TLR-3, Toll-like receptor 3; VEGF, vascular endothelial growth factor.

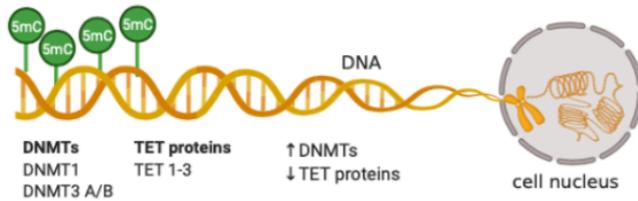
1.3 Epigenetics

As discussed in the previous section, epigenetic regulation plays a major role in the pathogenesis of AML. But what is epigenetics? In the early 1940s, developmental biologist Conrad H Waddington introduced the term “epigenetics” in an attempt to explain how gene regulation determines development [128]. He defined epigenetics as “the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being” [129]. This definition has evolved since then, and nowadays the term epigenetics is broadly used as a synonym for chromatin regulation. Indeed, epigenetic regulators are factors capable of changing chromatin structure, without altering the underlying DNA sequence, and which produce heritable changes in gene expression [130]. It has becoming increasingly clear that the action of specific epigenetic regulators, and the crosstalk between them, is essential for correct spatial and temporal gene expression.

1.3.1 Epigenetic mechanisms

In eukaryotic cells, genes are embedded in chromatin, which is composed of DNA and proteins. The fundamental repetitive unit of chromatin is the nucleosome, consisting of 147 base pairs of DNA wrapped around an octamer of core histones: 2 molecules of H2A, H2B, H3 and H4; the linker histone H1 organises nucleosomes into higher-order structures, increasing chromatin compaction [131]. Both histone proteins and DNA can be chemically modified by “epigenetic writers” that execute the modifications, and readily removed by other types of enzymes, “epigenetic erasers”. Furthermore, a collection of different proteins, termed “epigenetic readers” can recognise the different varieties of modifications [132]. As referred earlier (**Fig 1.5**) the enzymatic activities of epigenetic modifiers are reversible, rendering them as potential therapeutic targets. Epigenetic mechanisms (**Fig. 1.6**) include covalent histone modifications, DNA methylation and nucleosome remodelling by ATP hydrolysis. In recent years, microRNAs and long noncoding RNAs, miRNAs and lncRNAs, respectively, have also been shown to play a role by targeting certain mRNAs for degradation or translation repression [132].

A. DNA methylation



B. Histone modifications

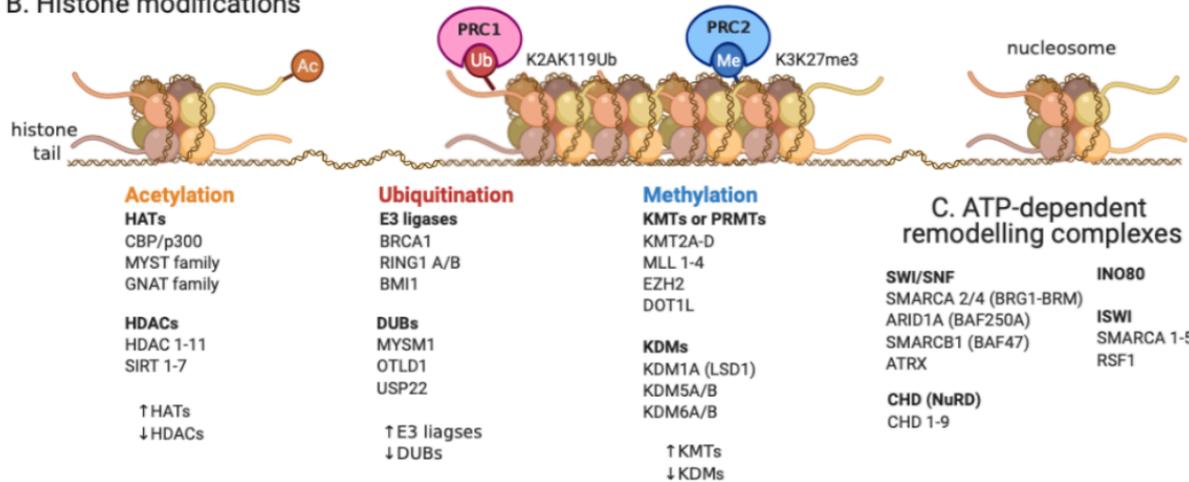


FIGURE 1.6 Main epigenetic regulatory mechanisms. DNA, in the nucleus, is wrapped around a histone octamer forming the primary functional unit of chromatin, the nucleosome. **A.** DNA methylation is achieved by DNMTs, which are responsible for the 5mC mark, associated with transcriptional silencing. TET enzymes oxidise 5mC to produce 5hmC, 5fC and 5aC. **B.** Histone modifications, such as acetylation (Ac, in orange), ubiquitination (Ub, in red) and methylation (Me, in blue) are instructive marks for both gene transcriptional activation and repression. Polycomb complexes PRC1 and PRC2 deposit H2AK119Ub and H3K27me3, respectively, both leading to repression of transcription. **C.** ATP-dependent chromatin remodelling complexes can re-position or remove nucleosomes by ATP hydrolysis. The different families of epigenetic modifiers are highlighted. HATs – histone acetyltransferases; HDACs – histone demethylases; DUBs – deubiquitinating enzymes; KMTs – lysine histone methyltransferases; PRMTs – protein arginine N-methyltransferases; KDMs – lysine methyl demethylases. Arrows pointing up next to the group of epigenetic modifiers indicate that they promote the given modification, whereas arrows pointing down indicate removal.

In general terms, these events all modulate the overall structure of chromatin, making it permissive or repressive to the binding of transcriptional machinery. Therefore, epigenetic mechanisms are key to the expression of gene programmes,

including those relevant in specifying a particular cell type or identity [130]. In this section I will describe the different flavours of epigenetic mechanisms. How they contribute to haematopoiesis and leukaemia is discussed further below.

Covalent histone modifications

Post-translational modifications of histones (PTMs) include acetylation, methylation, phosphorylation, ubiquitylation and sumoylation [133]. In 1964, Allfrey and colleagues demonstrated that histones can be modified by the addition of acetyl and methyl groups [134]. Histone acetylation and methylation are the most extensively characterised modifications. Acetylation neutralises the charge between histones (positive charge) and the DNA backbone (negative charge), promoting open chromatin structure and activation of gene expression. Levels of histone acetylation are tailored by the activity of 2 classes of enzymes: histone acetyltransferases (HATs), and histone deacetylases (HDACs), that respectively transfer and remove acetyl groups to/from lysine residues within histones [131]. HATs are also commonly referred to as KATs (lysine acetyltransferases). Both HATs and HDACs are generally conserved through evolution and are found with multisubunit complexes, with the functions of the catalytic subunit being dependent, for the most part, on the interaction with other elements present in those complexes [132–133]. For example, certain partner subunits contain bromodomains or chromodomains that recognise acetylated and methylated histones, respectively. Hence, HATs take advantage of their association with other subunits for their recruitment to distinct regions of the genome. HATs can be divided into 5 groups based on their structure and catalytic domains. These are: GNAT, MYST, p300/CBP, the general TF HATs and the nuclear hormone-related HATs. HDACs are also categorised (Class I – Class IV), a classification based on sequence and domain similarity [133–135]. Of note, HATs are also capable of modifying non-histone substrates [135].

Histone methylation, in turn, occurs by the addition of methyl groups to both lysines and arginines, and is catalysed by histone methylases (HMTs) and removed by histone demethylases (HDMs). In a similar way to that of HATs and KATs, HMTs and HDMs are also commonly named KMTs and KDMs, by convention. While lysines can take up to 3 methyl groups – me, me₂, me₃, arginines can only be mono or demethylated [132]. Furthermore, arginine methylation correlates with an open chromatin state, whereas lysine histone methylation can be associated with either open or close chromatin structure, depending on the specific residues. Methylation

of histone H3K4, H3K36, and H3K79 is associated with transcriptional activation, while demethylation and trimethylation of H3K9 and H3K27 are associated with transcriptional repression. HMTs also appear in multiprotein complexes and are classified in distinct families. Specifically, these enzymes are divided into 3 main groups, including 1 class of protein arginine methyltransferases (PRMTs) and 2 classes of lysine-specific methyltransferases (KMTs), which I consider here in more detail. The first class of KMTs contains an evolutionarily conserved SET methyltransferase domain (named after *Drosophila* Su(var)3-9, Enhancer of zeste [E(z)] and trithorax (trx)). There are multiple families of SET-domain proteins, as examples: the SUV39 family (composed of SUV39H1, SUV339H2 and G9a), the SET-1 family (including MLL1, MLL2, MLL3, MLL4, and SET1), the SET-2 family (including NSD1, NSD2, NSD3, HYPB and ASH1), the RIZ (containing PRDM1 and PRDM2), SMYD (SMYD1, SMYD3) and the SUV4-20 (SUV4-20H1 and SUV4-20H2) families [136]. The second class of KMTs, is represented by a single protein, DOT1L, the only known histone KMT without a SET domain [136]. DOT1L is the only enzyme mediating the mono-, di-, or trimethylation of histone H3 at lysine residue 79 (H3K79). In turn, several methyltransferases show specificity for H3K9, including SUV39-H1/2, G9A, PRDMs, SETDB1 and SETDB2. Amongst these, SETDB1 is the only one catalysing all three forms of methylation (mono-, di-, and tri-). SETD2 and NSD2 deposit H3K36me [136]. In turn, MLL1 and MLL2 catalyse H3K4me. In addition to its transcriptional activating functions through H3K4me activity, and by the recruitment of HATs including MOF and CBP [137], the MLLs can also repress target genes *via* recruitment of HDACs proteins and Polycomb as discussed below.

The PcG genes encode chromatin-associated proteins that assemble in 2 complexes - the Polycomb Repressive Complex 1 and 2, PRC1 and PRC2, respectively [138]. Both PRC1 and PRC2 activities involve repressive histone marks. PRC2 is responsible for H3K27 di- and trimethylation (H3K27me₂ and H3K27me₃). Its catalytic activity is mediated by two mutually exclusive KMTs, EZH1 and EZH2 [138–139]. Moreover, the complex contains two other elements, EED and SUZ12. In turn, PRC1 acts to stabilise H3K27me₃ and, monoubiquitinates histone H2A on lysine 119 (H2AK119ub1), resulting in chromatin compaction [139]. PRC1 core components include BMI1, RING1A/B, MEL18, RAE28 and at least one of the five Polycomb chromobox proteins (CBX2, 4, 6, 7, and 8), which are readers of H3K27me₃ through their chromodomain [139].

Demethylases also present specificity for different residues. Specific H3K4 demethylases are represented by LSD1 (KDM1A) and LSD2 (KDM2A) or by the JARID enzymes (JARID1A, JARID1B, JARID1C and JARID1D). H3K9me erasers include PHF8 and JHDM2 and JHDM3 families, the later also removing H3K36me. Lastly, H3K27me demethylases are UTX, UTY, JMJD3 and PHF8 [136].

Preferentially, histone modifications occur at the N-termini, so-called histone tails, but they can also be deposited in the globular domain, at the DNA binding surface of the nucleosomes [130]. The association between histone PTMs and transcriptional activity resulted in the “histone code” hypothesis, or in other words, the combination of post-translational marks on histones [130].

DNA methylation

DNA methylation consists in the stable addition of a methyl group to the 5C position of cytosines - 5-methylcytosine (5-mC), within the context of CpG dinucleotides. CpG dinucleotide enrichment occurs mainly in gene promoters and these CpG-rich locations are normally referred to as “CpG islands”. The presence of methyl groups in DNA results in closed chromatin structure and transcriptional silencing [140]. Accordingly, DNA methylation is essential during normal development, and regulates key gene silencing processes such as X-chromosome inactivation and genomic imprinting [141]. Mammalian cells balance levels of methylation *via* the positive action of DNA methyltransferases (DNMTs), which include DNMT1 and DNMT3A/B, and of negative regulators - the TET family of dehydrogenases, TET-1, -2 and -3 [141].

DNMT1 targets hemimethylated DNA at the replication *foci* during S phase of the cell cycle, thus maintaining existing patterns of methylation. On the other hand, DNMT3A and DNMT3B establish *de novo* methylation on unmethylated DNA regions during development and differentiation [142–143]. TET proteins, in turn, are responsible for the conversion of 5-mC into 5-hydroxymethylcytosine (5-hmC), preventing the binding of DNMTs to DNA, thereby keeping chromatin in an open, permissive way. Contrasting with 5-mC which is kept during cell division by DNMT1, 5-hmC is lost during DNA replication [142–144]. As mentioned in 1.2, IDH1 and IDH2 also influence DNA methylation, although in an indirect way. These enzymes produce alpha ketoglutarate (α-KG), a key intermediary of the citric acid cycle, essential for cellular metabolism. TET function requires α-KG, catalysed by IDHs.

Mutations in IDH1/2 alter the enzymatic activity of these proteins, converting α -KG to 2-HG, which cannot be used by TET proteins leading to widespread DNA hypermethylation [144].

ATP-dependent nucleosome remodelling complexes

A number of ATP-dependent enzymes act to remodel nucleosome structure - their assembly, spacing or removal - along the DNA to free or block access to TFs and polymerases. These remodelling complexes include the imitation switch-containing (ISWI), which mediates nucleosome slinging, the switch/sucrose non-fermentable (SWI/SNF), containing BRG and BRM ATPases capable of removing histones from DNA, and the nucleosome remodelling and deacetylase (NuRD). Each has a unique ATPase(s) and reader domain (sant-, bromo- and chromo-domain, respectively) able to interact with specific chromatin substrates [143].

1.3.2 Epigenetic regulation of haematopoiesis

The epigenetic mechanisms described above shape blood biology at various levels: HSC quiescence *vs* self-renewal, lineage fate specification, aging and transformation into malignancy (reviewed in [142–145–146]). Studies in mouse models have provided valuable knowledge on the physiological role of epigenetic regulators, ranging from gene KOs to the generation of conditional KO alleles due to the early lethality resultant from ablation of many of these proteins. In this section I will review what is currently known about the role of individual epigenetic regulators of normal haematopoiesis and their associations with AML, which I summarised in **Table 1.6**.

1.3.2.1 Histone acetyltransferases / deacetylases

During haematopoiesis, the expression of genes involved in HSC self-renewal and lineage-differentiation is regulated by the recruitment of HAT and HDAC complexes to their promoters. For instance, in erythropoiesis GATA1 directly recruits CBP to the β -globin locus, resulting in acetylation of H3 and H4 and globin gene expression [147]. Of note, GATA1 itself is also acetylated by CBP. The CBP/p300 family of HATs is also implicated in HSC self-renewal and differentiation by regulating c-Myb-

dependent gene expression [148]. However, whereas removal of Cbp from HSCs results in differentiation defects, p300-null HSCs are only mildly impaired in their differentiation [149]. Furthermore, Cbp^{+/-} mice show a haematopoietic defect, whereas p300^{+/-} are normal, suggesting different modes of action [149]. MOZ (MYST family) has been demonstrated to regulate HSC maintenance, proliferation and lineage commitment [150]. Specifically, *Moz* deficient mice are embryonic lethal and exhibit defects in erythroid and myeloid lineage maturation [151]. In line with its critical role in normal HSC function, MOZ-fusion proteins are a common feature in leukaemia [151] e.g.: MOZ-CBP, MOZ-p300 are generated by chromosomal translocations in AML. The interaction of HAT p300 with c-Myb is required for induction of AML, specifically, for the ability of MLL-AF9 fusion to confer self-renewal properties on myeloid progenitors [152]. Without this interaction, fusion proteins are unable to impose a differentiation block, causing, instead, terminal differentiation [153]. p300 can also enhance the self-renewal ability of LSCs through acetylating AML-1ETO, activating its target genes [149]. Recently, MOF or KAT8, was shown to be involved erythroid cell fate commitment through its H4K16ac activity [154]. Moreover, MOF interacts with MLL1 to maintain H4K16ac, activating gene transcription [154].

Unsurprisingly, HDACs are also implicated in normal and malignant HSC function. In agreement, combined deletion of HDAC1 and HDAC2 results in severe HSC defects, causing anaemia and cytopenia [155]. Finally, SIRT1 (HDAC belonging to Class III) null HSCs show defects in self-renewal and exhibit impaired lymphopoiesis [153].

1.3.2.2 Histone lysine (K) methyltransferases / demethylases

MLL has critical roles in the proliferation and lineage specification of haematopoietic progenitors during embryogenesis by maintaining (but not initiating) expression of *HOX* genes [156]. Homozygous deletion of *Mll1* results in embryonic lethality at E10.5-E12.5. In turn, *Mll1* heterozygous embryos exhibit defects in foetal haematopoiesis with reduced numbers of HSCs [157]. Chromosomal translocations involving the MLL gene and a number of different fusion partners, including AF4, AF9, ENL, AF10, ELL and AF6, are a recurrent theme in AML pathogenesis [145–157]. These oncogenic fusions aberrantly regulate MLL- target genes, such as *HOXA*

and *MEIS1*, modifying genetic programmes of proliferation and differentiation [157]. Importantly, partners of MLL fusions were reported to be found in distinct chromatin complexes. The most common fusion partners, AF9 and AF4, are present in the super elongation complex (SEC), and in the DOT1L complex (DOTCOM), in the case of AF9. Interactions with different chromatin complexes, can occur *via* specific domains of MLL itself (containing a bromodomain, PHD domain and a SET domain), or of its fusion partners. For example, the AF9 YEATS domain has been reported to be critically involved in the recruitment of DOT1L to chromatin by MLL-AF9, and to H3K79me-mediated transcriptional control [158]. MLL-driven AML types have a poor prognosis and are thus classified as high-risk AML. Moreover, about 3-5% of *de novo* AML patients display partial tandem duplication of the *MLL* gene - *MLL*^{PTD}, also acting as an oncogene by upregulating the expression of *HOX* genes, preventing cell differentiation and apoptosis. *MLL*^{PTD} AML often display mutations in other epigenetic regulators, such as TET2 (16%), EZH2 (10%), IDH1/2 (31%), and ASXL1 (6%). Moreover, a typical feature of *MLL*^{PTD} is the absence of *NPM1* mutations, and frequent co-occurrence with *RUNX1* (23%) and *STAG2* mutations (16%) [145].

Dot1l KO mice are embryonic lethal between E10.5-13.5 due to a selective defect in erythroid differentiation [159]. In agreement, Dot1L is required for *GATA2* activation (and repression of PU.1) during erythropoiesis [159]. Conditional deletion, however, results in reduction in both red and white blood cells, without complete abrogation of multilineage haematopoiesis. This suggests that DOT1L is not an absolute requirement for adult haematopoiesis [127]. Though not mutated, DOT1L has also been shown to play critical roles in leukaemia, in particular driven by MLL-fusions as mentioned above [160]. Indeed, DOT1L can be aberrantly recruited, directly or indirectly, to target genes of MLL fusion proteins such as *HOXA* and *MEIS1*, increasing expression of leukaemic programmes [161]. As referred to in section 1.2, inhibitors of DOT1L are in preclinical investigations, and were shown to reduce H3K79 methylation, together with expression of MLL-fusion target genes [121].

Polycomb complexes are also involved in HSC functionality and leukaemogenesis. KO mice of PRC1 components *Bmi1* and *Rae28* are lethal in the neonatal period highlighting a requirement in adult haematopoiesis. Furthermore, mutations in both *Bmi1* and *Rae28* decrease HSC self-renewal. Specifically, deletion of *Bmi1* results in cell cycle arrest and apoptosis, as well as de-repression of B-cell differentiation factors *EBF1* and *PAX5* [162]. KO of *Mel18*, in turn, increases HSC quiescence [162].

Interestingly, CBX proteins play contrasting roles in HSC function - CBX8 is upregulated to promote differentiation, whereas CBX7 is highly expressed in HSCs, inducing HSC renewal and repressing lineage-differentiation genes [163]. Mice lacking PRC2 components EZH2, SUZ12 or EED are all embryonic lethal. Instead, heterozygous depletion of any of these elements enhances haematopoietic stem and progenitor cell activity [162]. In addition, PRC2- EZH2 can interact with all 3 DNMTs, influencing sites for DNA methylation [139]. PRC2 complex is also involved in erythropoiesis, with mice lacking EED presenting impaired erythroid differentiation [162], as well as lymphopoiesis, in which SUZ12 is essential for T and B cell maturation [163].

In AML, CBX7 binds SETDB1 and its inhibition induces differentiation of leukaemic cells [163]. In the context of MLL fusions, CBX8 recruits histone acetyl transferase Tip60, thereby promoting fusion target gene expression, instead of mediating transcriptional repression [163]. EZH2 loss-of-function mutations have been described in myeloid malignancies, including AML, where it has been shown as a requirement for the maintenance of multiple genotypes [164]. Moreover, in the same study the authors demonstrated stage-specific roles of *Ezh2*, which functions as a tumour suppressor in AML induction, and as an oncogene once the disease is established [164]. In MLL-AF9 AML, EED is necessary for leukaemia initiation and progression [161].

Histone demethylase LSD1 (KDM1A) is a key regulator of leukaemic stem-cell potential, and it acts at genomic *loci* bound by MLL-AF9, favouring its leukaemogenic programmes, thereby preventing cell differentiation and apoptosis [165]. As mentioned in section 1.2, LSD1 has emerged as a potential therapeutic target in AML, with small-molecule inhibitors showing efficacy in AML cells bearing different MLL rearrangements but also *AML-1ETO* and *PML-RARA* [120]. LSD1 main function is demethylation of H3K4me1/2 and H3K9me1/2, which result in gene repression and activation, respectively. Reversing LSD1 activities through demethylation of H3K4 or H3K9, at the required anti-leukaemic doses, however, was shown to cause anaemia and defects in erythropoiesis [166].

1.3.2.3 DNA Methylation: the roles of DNMTs and TET proteins

Levels of DNA methylation are highly locus dependent during haematopoiesis [167]. For instance, during commitment to a specific lineage, genes that function in keeping an undifferentiated status are hypermethylated, and thus repressed. Thus, when HSCs differentiate into a lineage commitment state, the methylation of stemness genes increases, while the methylation of lineage specific genes decreases [168]. PU.1, for example, is highly expressed in HSCs and mature B-cells but is hypermethylated in CD4⁺ and CD8⁺ T-cells [168]. The process of erythrocyte differentiation is also well characterised by DNA methylation changes. Indeed, studies of the globin gene locus revealed that these genes are expressed in a sequential manner, such that during erythrocyte maturation embryonic or foetal genes are silenced, and adult globin genes are turned on [74]. A recent study by Izzo et al. proposed that lineage-determining TFs might be particularly susceptible to global DNA methylation changes. Moreover, they show that erythroid-related TFs have higher CpG content in their binding motifs than myelomonocytic-associated TFs [166].

All 3 DNMTs contribute to control of haematopoietic function, although with different effects. Specifically, Broke and colleagues have shown that constitutive DNA methylation is essential for HSC self-renewal activity [169]. Indeed, in a DNMT1 conditional knockout (cKO), DNMT1-deficient HSCs fail to self-renew and show an aberrant increase in myeloid cell output [169]. On the contrary, loss of DNMT3A was shown to promote self-renewal, resulting in expansion of the HSC population [170]. Moreover, the increase in HSC function is most apparent in serial transplantation assays, in which DNMT3A-null HSCs are able to sustain at least 12 rounds of transplantation, whereas WT HSCs undergo just 4 rounds [171]. DNMT3B cKO resembles the DNMT3A phenotype, but with milder consequences. Finally, combined DNMT3A and DNMT3B perturbation also increases HSC number and results in a more severe differentiation block [170].

DNA methylation is also critical in pathogenesis of AML, as mentioned in the previous section. Specifically, mutations in DNMT3A associate with hypomethylation, whereas mutations in TET2 and IDH1/2 result in DNA hypermethylation. Indeed, in AML, DNMT3A mutations are prevalent (17-34%), with the most common alteration being a missense mutation at arginine R882. This impairs the function of DNMT3A by

altering the flanking sequence preferences for DNA binding [172]. DNMT3A mutations are an early event in AML, and mutations in additional genes such as NPM1 or FLT3 may lead to malignant transformation of HSCs in leukaemic cells [172]. A recent study by Gebhard et al. proposes a division of AML into 2 subgroups based on methylation profile. The first group associates with PcG-mediated repression and is prevalent across AML samples. Instead, the second group is characterised by a more heterogenous behaviour and comprises both hypo and hypermethylated samples [173].

TET2 is highly expressed in HSCs [174]. Functionally, loss of TET2 in HSCs results in increased self-renewal capacity with upregulation of MEIS1 and EVI1 stemness genes as well as increased number of GMP progenitor cells [174]. TET1 was first identified as a fusion partner of MLL in the context of AML [174]. Interestingly, despite having opposing functions individually, when DNMT3A and TET2 mutations are present simultaneously, they accelerate leukaemic transformation [172–175].

1.3.2.4 ATP-dependent nucleosome remodelling complexes

The human SWI/SNF complex was initially identified as an activator of β -globin transcription by interacting with erythroid-affiliated TF KLF1 [98]. Roles for the SWI/SNF complex in myeloid and lymphoid lineages have also been demonstrated, with BAF155 subunit being a requirement in T cells [176]. Moreover, deletion of ARID1A, a core component of the same complex affects both myeloid and lymphoid lineage development. SMARCA4 is required for AML cell growth and survival by maintaining chromatin accessibility at essential enhancers [176]. NuRD ATPase Mi2b is required for HSC self-renewal [145].

Table 1.6 Phenotypes of epigenetic disruption in haematopoiesis.

Chromatin complex / epigenetic modification	Gene	Mouse models / <i>in vitro</i> inhibition	Effect on adult HSC self-renewal	Effect on HSC differentiation / progenitor function	Association with leukaemogenesis (AML)	Ref.
DNA methylases/ TET proteins						
5-mC	DNMT1	cKO	Impaired	Impaired lymphoid differentiation	Rarely mutated in AML, but through the interaction with TFs, non-coding RNAs, fusion oncogenes can affect leukemic cells biology	Wong et al., 2019
	DNMT3A	cKO	Enhanced	Progressively impaired in serial transplantations; Increased frequencies of erythroid progenitors and decreased in myelomonocytic progenitors; Associates with overexpression of DOT1L	Mutated AML; Clonal haematopoiesis	Sperling et al., 2017; Challen et al., 2011; Rau et al., 2016
	DNMT3B	KO	Unaffected	Slight increase in B cells in third serial transplantation	Rarely mutated in AML, but its overexpression associates with worse prognosis in AML patients	Wong et al., 2019
	DNMT3A and DNMT3B	2KO	Enhanced	Block in differentiation	2KO AML cells show accelerated leukaemia, suggesting a synergistic role of DNMT3A and DNMT3B in suppressing leukaemia progression	Zheng et al., 2016
5-hmC	TET2	KO catalytic mutant models	Enhanced	Impaired erythroid and lymphoid differentiation; Increased frequencies of myelomonocytic progenitors	Mutated AML; Clonal haematopoiesis; Tet2 KO mice develop both myeloid and lymphoid disorders; Tet2 catalytic mutant mice predominantly develop myeloid malignancies	Sperling et al., 2017; Ito et al., 2019
Polycomb repressive complexes						
PRC1 H2AK119Ub1	BMI1	KO	Impaired	Accelerated differentiation into B-cell lineage	Upregulated in AML; Its depletion reduces proliferation and results in apoptosis of leukaemic cell lines	Vidal and Starowicz, 2017
	RAE28	KO	Impaired	Severe delay in B-cell development	B-cell ALL; unknown roles in AML	Tokimasa et al., 2001; Ohia et al., 2002
	MEL-18	KO	Enhanced	Impaired B-cell differentiation	N/A	Vidal and Starowicz, 2017
	CBX2	KO	Impaired	Impaired B-cell differentiation	Overexpression in AML cell lines increases apoptosis and myeloid differentiation; miR-29 restricts CBX2 expression to maintain AML	Vidal and Starowicz, 2017; Ha et al., 2019
	CBX7 overexpression	OE	Enhanced	Increased myelopoiesis	Upregulated in AML; Pharmacological inhibition reduces proliferation AML cells and induces differentiation; Binds STEDB1 and its inhibition induces differentiation of AML	Jung et al., 2019
	CBX8	cKO (Cre-ERT)	Unaffected	Unaffected	Important function in KMT2A-AF9-induced HOX transcriptional activation and leukemic transformation	Vidal and Starowicz, 2017
	RING1A and RING1B	2KO (Cre-ERT)	Impaired	Hyperproliferative myeloid progenitors	Ring1A/B regulate and maintain AML stem cells in part by repressing Glis2 expression, which promotes their differentiation	Vidal and Starowicz, 2017; Shima et al., 2018
	RING1B	Mx1-Cre cKO	Impaired	Hyperproliferative myeloid progenitors	Ring1A/B regulate and maintain AML stem cells in part by repressing Glis2 expression, which promotes their differentiation	Vidal and Starowicz, 2017; Shima et al., 2018
PRC2 H3K27me3	EDD	cKO	Impaired	Pancytopenia	Overexpression in AML, associated with poor prognosis	Xie et al., 2014
	EZH1	cKO	Impaired	Impaired lymphoid differentiation	Promotes AML1-ETO transcriptional repression in leukemia	Dou et al., 2019
	EZH2	Mx1-Cre cKO	Unaffected	Impaired B-cell differentiation	Mutated in AML; Association with the development of myeloproliferative disease; Its deletion/ inhibition abrogates myeloid malignancy induced by Bap1 loss	Kuhn et al., 1995; Rizo et al., 2008; Xie et al., 2014
	EZH2	overexpression	Enhanced	Unaffected	Association with the development of myeloproliferative disease; Its deletion/ inhibition abrogates myeloid malignancy induced by Bap1 loss	Kamminga et al., 2006; Jung et al., 2019

Histone acetyltransferases / deacetylases							
H3K9ac; H3K23ac	MOZ/KAT6A	Mx1-Cre cKO	Impaired	Reduced number of CFU, but no impairment on their ability to differentiate	Involved in chromosome translocation t(8;16) (p11;p13), which is associated with the FAB M4/M5 AML subtype with monocytic arrest	Sheikh et al., 2016; Borrow et al. 1996; Katsumoto et al., 2006	
H4K16ac	MOF	Mof ^{+/-} -heterozygous mice cKO	Impaired	Anaemia; Impaired erythroid differentiation; Increase in myeloid progenitors	Decreased levels of H4K16ac and mutations in KANSL1 (MOF-containing complex)	Fraga et al., 2005; Jaskula et al., 2017	
H3K9ac	GCN5/KAT2A	cKO	Enhanced	Impaired granulocytic differentiation	Required for proper granulocytic differentiation through acetylation of CEBPA; maintains <i>MLL-AF9</i> leukemia through HAT activity	Bararis et al., 2016; Domingues et al., 2020	
H3K18ac; H3K56ac	CBP/p300	Mx1-Cre	Impaired	Impaired differentiation of HSC	Cbp is required for induction and maintenance of AML	Giotopoulos et al., 2016	
H3K27ac	HDAC1 and HDAC2	2KO	Impaired	Anaemia and cytopenia; impaired mature T cell development	Recruited by AML oncogenic fusions promoting AML pathogenesis	Wang and Liu, 2020	
H4K16ac	SIRT1	KO	Unaffected	Production of mature blood cells, and frequencies of the most primitive HSC population unaffected	Overexpressed in FLT3-ITD AML; activates a c-MYC oncogenic network to maintain drug resistance FLT3-ITD AML	Leko et al., 2012; Li et al., 2014	
Histone methyltransferases / demethylases							
MLL family (Set1/Trithorax-type) H3K4me	MLL1 / KTM2A	Mx1-Cre cKO	Impaired	All progenitors lost - severe cytopenia	MLL1 is disrupted by chromosomal translocations; SET not functionally required for MLL-AF9 transformation; Dispensable for MLL-AF9 initiation	Ernst et al., 2004	
	MLL2	N/A (germline KO only)	N/A	N/A	Can be involved in initiation and survival of MLL-AF9 AML instead of MLL1	Antunes and Ottersbach, 2020	
	MLL3	N/A	N/A	N/A	Deleted in the 7q-AML model	Yang and Ernst, 2017	
	MLL4	Mx1-Cre cKO	Enhanced, but reduced in competitive transplantation	Increased myeloid CFU output; Required for germinal B-cell formation	Drives AML through MLL-AF9 fusion and its loss limits leukaemogenesis; Acts as tumor suppressor gene in lymphomas	Yang and Ernst, 2017	
	SETD1A	Rosa26-CreER/SCL-CreERT cKO	Reduced in secondary transplants	Impaired B-cell differentiation; Erythroid-specific KO partially blocks erythropoiesis	SETD1A enhances leukemic cell growth and survival independent of its methyltransferase activity	Yang and Ernst, 2017; Tusi et al., 2015; Arndt et al., 2018	
	SETD1B	Rosa26 Cre-ER cKO	Impaired	Impaired myeloid and lymphoid lineage differentiation in CFU	Mutated in MPN polycythemia vera; transformed c-kit ⁺ Vav-Cre KO cells	MLL-ENL Schmidt et al., 2018	
H3K79me	DOT1L	Rosa26 Cre-ER cKO	Enhanced	Impaired embryonic erythropoiesis: block in proliferation and survival of erythroid progenitors; reduction in the formation of BFU-E but not CFU-E	DOT1L inhibition with EPZ5676 exerts anti-leukaemic activity in DNMT3A-mutant AML	Nguyen et al., 2011; Peng et al., 2010; Steger et al., 2008; Jones et al., 2010	

1.4 KAT2A

General control non-depressible 5 (Gcn5) or Kat2a was the first HAT identified. Brownell and Allis discovered a catalytically active HAT subunit, p55, in *Tetrahymena thermophila*, later shown to be homologous to Gcn5 in yeast [177]. This finding linked, for the first time, histone acetylation to transcription. Belonging to the GNAT family, Kat2a acetylates preferentially H3K9 and to a lesser extent H3K14 [178]. Kat2a acetylation activity was shown to occur post RNA Polymerase II recruitment to target promoters, indicating that it is not an absolute requirement for initiation of gene expression. Instead, it is involved in stabilising gene expression [179]. In addition to catalysing acetylation through its HAT domain, KAT2A contains a bromodomain, which makes it able to also recognise and be recruited to – acetylated chromatin sites [178]. More recently, Kat2a was shown to catalyse another histone lysine modification – succinylation, which also associates with active transcription [180].

1.4.1 KAT2A roles in development

Kat2a has a highly related paralogue in vertebrates: Kat2b/Pcaf [181]. However, Kat2a and Kat2b, have a largely mutually exclusive pattern of expression. Whereas Kat2a is mainly expressed in haematopoiesis and neural tissue, Kat2b dominates in skeletal muscle [181–182]. Kat2a plays essential roles in development along the evolutionary scale. It is essential for metamorphosis and oogenesis in *D. melanogaster* [183]. In the developing mouse embryo, Kat2a is ubiquitously expressed between E7.5 and E9.0, with the exception of heart and allantois, and its expression decreases after E16.5 [182], which could suggest reduced contribution in terminal differentiation. In turn, Kat2b is minimally expressed at early stages of development, and upregulated in adult tissues, particularly in the heart and skeletal muscle [182]. In line with this, Kat2a, but not Kat2b, is essential for mammalian embryonic development. *Kat2a* null mice die at E10.5, with extensive mesodermal apoptosis; embryos also display defects in notochord, somites, and in neural tube formation [182–184]. *Kat2b* null mice develop normally, but double *Kat2a/Kat2b* mutants have a more severe phenotype than single *Kat2a* null animals, and die earlier at E7.5, suggesting some functional redundancy [181]. Redundancy between the 2 paralogs has also been observed in zebrafish, where combined perturbation of

Kat2a and *Kat2b* leads to more severe heart and fin developmental defects than single gene loss [185].

In the context of stem cells, *Kat2a* is not an absolute requirement for maintenance of pluripotent mouse embryonic stem cells (ESCs) [186]. However, it stabilises pluripotency gene regulatory networks [187] and allows progression of reprogramming in induced pluripotent stem (iPS) cells [188]. Additionally, *Kat2a* is required for cell survival and correct lineage specification and differentiation in embryoid bodies, through regulation of FGF signalling [189]. *Kat2a* acts as a co-factor to Myc family proteins, namely c- and N-Myc, both of which play key roles in embryogenesis as well as in maintenance of pluripotency [190–191]. Furthermore, *Kat2a* acetylates Myc, promoting its stability [192], and is essential for activation of Myc target genes through histone acetylation [193]. Proliferation of neural stem and progenitor cells requires *Kat2a*, in which it phenocopies the role of N-Myc [194]. KAT2A has also been shown to regulate osteogenic differentiation of periodontal stem cells through inhibition of the Wnt/ β -catenin pathway by DKK1 [195]. In more detail, KAT2A mediates H3K9/K14ac of *DKK1* promoter, activating its expression [195].

In normal haematopoietic development, *Kat2a* has been implicated in the development of lymphoid blood lineages, but, interestingly, not in HSCs [196–197]. Specifically, *Kat2a* participates in maturation of B [198] and T cells [199], as well as in differentiation of invariant natural killer T (iNKT) cells [200]. *Kat2a* functions in B and T cell development involve promoter histone acetylation. On the other hand, it modulates iNKT cell differentiation through activation of the transcription factor *Egr2* via non-histone protein acetylation. Finally, contrary to the effect on *Egr2*, *Kat2a*-mediated acetylation of *Cebpa*, represses its transcriptional activity and impedes progression of terminal granulocytic differentiation [196].

1.4.2 KAT2A contributions to cancer

High KAT2A expression associates with bad prognosis in Breast Cancer [201–202], Non-Small Cell Lung Carcinoma [203–204] and Colon Cancer [205] namely through histone acetylation-mediated co-activation of E2F and MYC transcriptional targets to maintain cell proliferation and survival (**Fig. 1.7A**). In Melanoma, increased KAT2A

stability by *WDHD1* leads to enhanced H3K9 and H3K56ac levels [206]. Increased expression of *KAT2A* also associates with poor survival in Renal Cell Carcinoma (**Fig. 1.7B**), but the mechanism has not been elucidated. In contrast, high levels of *KAT2A* confer a survival advantage in Pancreatic Adenocarcinoma (**Fig. 1.7C**), and, to a lesser extent, in low grade Glioma (Cox coefficient = -0.167, FDR=0.125), suggesting that *KAT2A* may act in an oncogene-like, as well as in a tumour suppressor-like manner. Interestingly, *KAT2A* succinylation activity has recently been linked to tumour maintenance in Pancreatic Adenocarcinoma cell lines [207] (**Fig. 1.7A**). This raises the possibility that *KAT2A* may have cancer stage-specific roles, as shown for other epigenetic regulators [164], although mechanistic data in support of this hypothesis is currently lacking. Significantly, no recurrent *KAT2A* mutations have been described in cancer, suggesting that it is co-opted by tumour cells at an epigenetic level for establishment and/or maintenance of tumorigenic programs. The lack of mutations could indeed reflect conflicting roles of *KAT2A* loss or over-expression at different stages of cancer progression, with no sustained selective advantage. This is contrast with disease-specific effects, which would more likely associate with recurrent mutations in individual tumours.

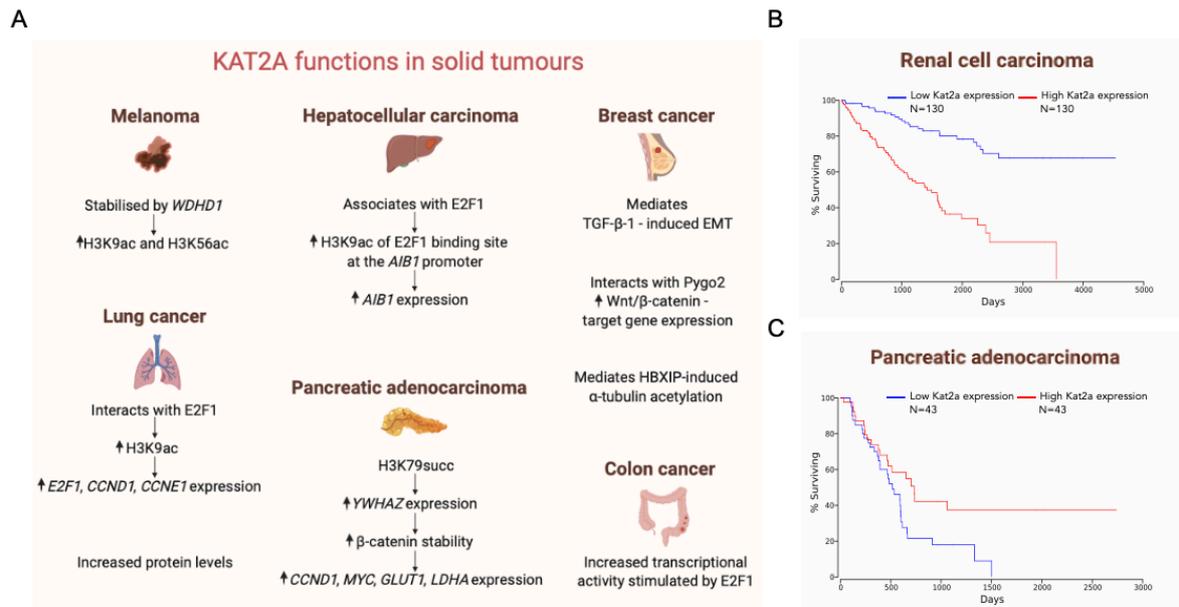


FIGURE 1.7 KAT2A participates in cancer biology with oncogene-like and tumour-suppressor roles. **A.** Described mechanisms of participation of *KAT2A* in different malignancies. **B.** *KAT2A* as an oncogene: Kaplan-Meier survival curves for Renal Clear Cell Carcinoma patients with high (25%) and low (25%) levels of *KAT2A* expression; log-rank analysis, p -value = $8.45e-11$. **C.** *KAT2A* as a candidate tumour suppressor: survival curves for Pancreatic Adenocarcinoma patients with high (25%) and low (25%) levels of *KAT2A* expression; log-rank analysis, p -value = 0.025. Data in B and C were retrieved from OncoLnc (oncolnc.org; as of August 2020), a tool that links The Cancer Genomics Atlas (TCGA) survival data to mRNA, miRNA, or lncRNAs expression. Despite *KAT2A* requirements in individual cancers, including AML, *KAT2A* levels do not systematically affect patient survival raising the possibility of stage or disease-specific effects.

1.5 SAGA and ATAC – KAT2A chromatin complexes

Both KAT2A-containing complexes, respectively Spt-Ada-Gcn5-Acetyltransferase (SAGA) and Ada-Two-A-Containing (ATAC) (**Fig. 1.8**) are key to its efficient lysine acetylation activity [208]. SAGA is evolutionary conserved from yeast to human [209–210], whereas ATAC is a smaller complex characteristic of multicellular organisms, first associated with chromatin functions in *Drosophila* [211]. Structural and functional organization of SAGA and ATAC complexes in model organisms has been reviewed in [209–210–212–213]. Herein, I will consider them briefly and include information from recent structural studies of the yeast SAGA core [214–215]. I will also present a timeline of relevant discoveries related to KAT2A and its complexes so far (**Fig. 1.8**).

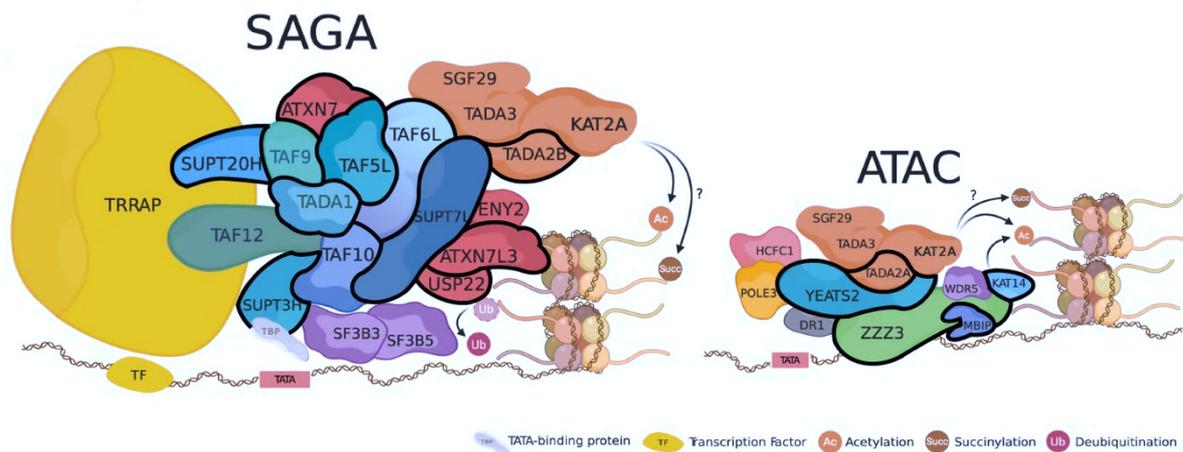


FIGURE 1.8 Schematic representation of the human SAGA (left) and ATAC (right) KAT2A-containing complexes. SAGA (left) is organized into distinct structural and functional modules. Subunits belonging to each module are coloured similarly. The KAT2A-containing histone acetyltransferase (HAT) module is depicted in orange and shared with ATAC (right), with exception of TADA2B, which is replaced by TADA2A in ATAC. The histone de-ubiquitinase (DUB) module is shown in red; the Core module in turquoise. TF-binding module, TRRAP, is shown in yellow. Splicing module is depicted in purple. TATA-binding protein (TBP), in grey, is not part of the complex architecture but it associates with SUPT3H to recruit SAGA to TATA box and facilitate transcription. Apart from the HAT domain, ATAC subunits do not have a modular organization.

1.5.1 SAGA complex

Following the discovery of SAGA multimodular complex [213] (**Fig. 1.9**), great effort has been put in unveiling its molecular architecture. It is now well established that mammalian SAGA contains 20 subunits distributed into 5 functionally distinct modules, a structure highly conserved from yeast to human [178–209]. The HAT consists of: KAT2A (or the mutually exclusive KAT2B), SGF29, TADA3 and a complex-specific TADA2 variant: TADA2B in SAGA, and TADA2A in ATAC. SGF29 is a tandem TUDOR domain protein that recruits KAT2A activity to active promoters marked by H3K4 tri or di-methylation [216]. TADA3, is a transcriptional activator adaptor required for transcriptional activity [212]. TADA2 variants are ZF proteins that can bind double-stranded DNA [212–217]. The central Core of SAGA is formed of TADA1, suppressor of Ty (SPT) elements, SUPT3H, SUPT7L and SUPT20H, and of TATA-binding protein associated factors (TAFs) - TAF5L, TAF6L, TAF9, TAF10 and TAF12 [209–212–218]. TAF5L and TAF6L are SAGA-specific, whilst the other TAFs are shared with TFIID complex, the RNA Polymerase II (PolII) General Transcription Factor.

As previously mentioned, the structure of the yeast SAGA Core has been recently deciphered [214] (**Fig. 1.9**), revealing an octamer-like fold organization consisting of Taf6l-Taf9, Taf12-Ada1 (TADA1 orthologue), and Taf10-Spt7 (SUPT7L orthologue) pairs, each contributing with a Histone-Fold (HF) domain [214]. Spt3 (SUPT3H orthologue) contributes with 2HF domains and assumes a free conformation key to binding of SAGA to TATA-binding protein (TBP) [215]. The histone octamer connects to the remaining Core *via* Taf5l [214–215]. SAGA includes a large transcription factor interaction module, TRRAP, originally identified as cofactor of c-Myc and E2F proteins [219]. TRRAP is also found in NuA4/TIP60 HAT complexes [220] and in the ATP-dependent chromatin remodelling p400 complex [221], with possible complex-unique as well as complex-shared or redundant functions. SAGA also contains a Splicing module composed of SF3B3 and SF3B5, which is only observed in multicellular organisms [222–223]. SF3B3 is a component of the splicing factor SF3B essential for spliceosome assembly, and links SAGA to the general splicing machinery [223–224]. Finally, SAGA comprises a lysine de-ubiquitination (DUB) module, catalysed by USP22, which targets H2B and H2A as well as non-histone proteins [225–226]. All 4 members of the DUB module – USP22, ATXN7L3, ATXN7

and ENY2 – are needed for full deubiquitinating activity [227–228]. ENY2 participates in another chromatin complex, TREX-2, involved in mRNA nuclear export [229]. In yeast, SAGA enzymatic modules HAT and DUB, associate with the central Core in a flexible way [214], which may be suggestive of functional independence. Indeed, previous studies showed that loss of DUB components does not impair SAGA complex integrity or HAT activity [230]. Additionally, HAT and DUB can exist as separate entities [231–233], supporting the notion of alternative HAT and DUB SAGA-independent roles.

1.5.2 ATAC complex

KAT2A was found to integrate a second epigenetic complex, ATAC, originally purified in human HeLa and HEK293T cell lines [234] (**Fig. 1.9**). ATAC shares three subunits with the SAGA, all components of the HAT module: KAT2A, TADA3 and SGF29 [212]. Furthermore, ATAC structure includes: YEATS2, a H3K27ac reader [235], which allows for integration of additional transcription activation signals; DR1/NC2 β , also present in the NC2 complex, which heterodimerizes with YEATS2 to interact with TBP; ZZZ3, a ZF protein that specifically binds H3 tails [236], and WDR5, shared with the MLL methyltransferase complexes [237], thus potentially important in establishing a crosstalk between H3K4me and H3K9ac chromatin marks, which complements or extends the roles of SGF29 double TUDOR domain. Moreover, Wdr5 and Kat2a are bound by the WNT-interacting protein Pygo2, which brings together Mll2-containing and Kat2a-containing complexes to activate expression of WNT targets, including Myc in mammary epithelial cells [238]. Within the context of MLL complexes, WDR5 was identified as a candidate therapeutic target in CEBPA N-terminal leukaemia [239] but its functions within ATAC, particularly in the context of cancer, including leukaemia, remain poorly understood. In an interesting additional link, two long non-coding RNAs were shown to promote gastric cancer through scaffolding of KAT2A and WDR5 at a subset of promoters and activation of gene expression [240–241]. The remaining ATAC components are MBIP, a MAP3K regulator exclusive to mammalian cell complexes; and a second HAT activity – KAT14 [242]. YEATS2, MBIP and KAT14 are required for complex integrity [242]. In mammalian cells, ATAC acetylates both H3 and H4, preferentially through KAT2A and KAT14 respective activities [242], whereas in *Drosophila* the complex preferentially targets H4 [243]. ATAC binds Host Cell Factor 1 (HCFC1), a scaffold to

multiple chromatin-modifying complexes [244], and POLE3, the DNA polymerase epsilon subunit 3 [242].

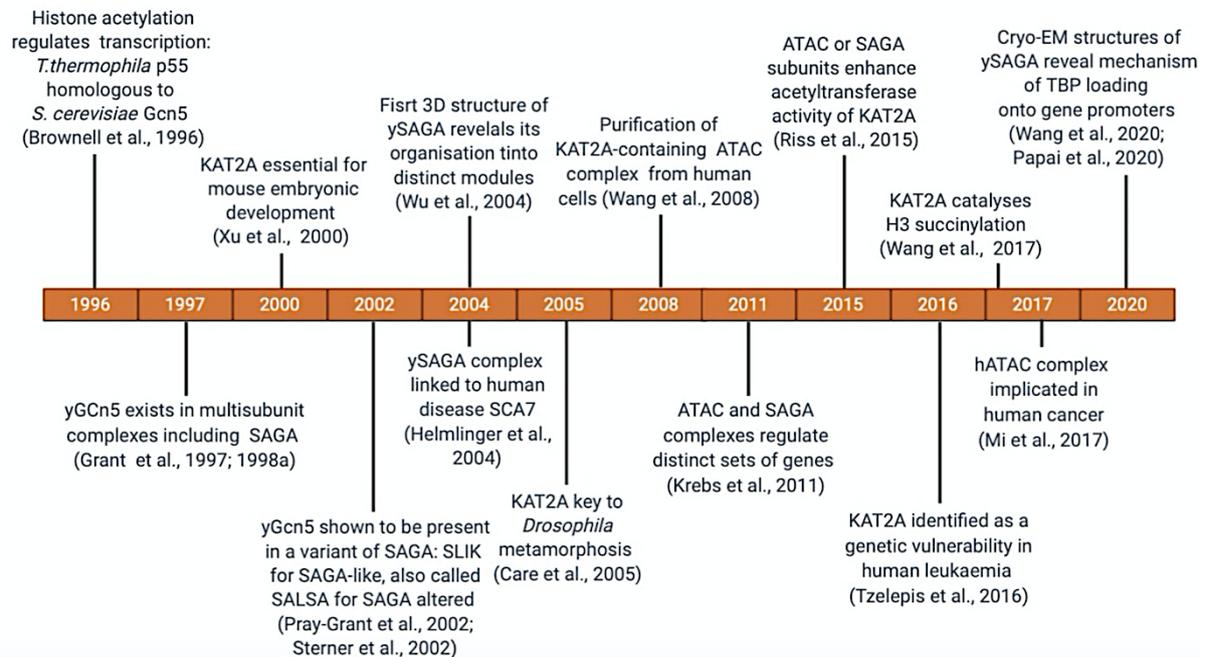


FIGURE 1.9 Timeline of relevant discoveries on KAT2A and the ATAC and SAGA complexes.

1.5.3 SAGA and ATAC roles in cancer

SAGA and ATAC have distinct sets of chromatin targets and, like KAT2A, participate in regulatory mechanisms that are ubiquitous amongst cells [245]. Thus, it is possible that both complexes contribute to transcription and cell identity – that is, the maintenance of the cell role in the long term. However, their contributions to identity of cancer cells remain poorly understood.

Cancer cells depend on high protein production [246–247] and it is perhaps unsurprising that regulators of ribosome biogenesis constitute dependencies across multiple malignancies. Global analysis of CRISPR drop-out screens through interrogation of the Cancer Dependency Map (DepMap) Project at the Broad Institute, shows that 4 ATAC components – TADA2A, YEATS2, WDR5 and DR1 –, the first two of which are exclusive to ATAC, are called “essential”, as they constitute genetic vulnerabilities in >75% of the cell lines studied. In contrast, SAGA-specific elements were not called “essential”, suggesting more specific and potentially more targetable roles in cancer [248–249].

However, ATAC and SAGA-specific subunits had distinct cancer lineage associations, as determined by statistical strength of dependencies across cell lines representing the same tissue. ATAC-specific elements *ZZZ3* and *YEATS2* have a strong statistical association with blood cells (**Fig. 1.10A**), specifically lymphocytic malignancies, likely reflecting the association between KAT2A and lymphoid cell differentiation.

SAGA-specific elements also associate with blood malignancies of lymphocytic lineages – Non-Hodgkin Lymphoma and Multiple Myeloma but have stronger associations with tumours of the Central Nervous System and Renal Cell Carcinoma (**Fig. 1.10B**). Indeed, they mimic KAT2A dependencies, suggesting that ATAC may exert KAT2A-independent effects or compensate for its loss with other HAT activities. KAT2B does not constitute a dependency on any cell line analysed, while KAT14 is required in a small number of cell lines with B-lymphocyte bias, making it a more likely candidate for KAT2A redundancy.

Effects on control of transcriptional variability may nevertheless be exclusive to KAT2A, allowing for more subtle effects on cancer cell maintenance that do not translate into an absolute requirement.

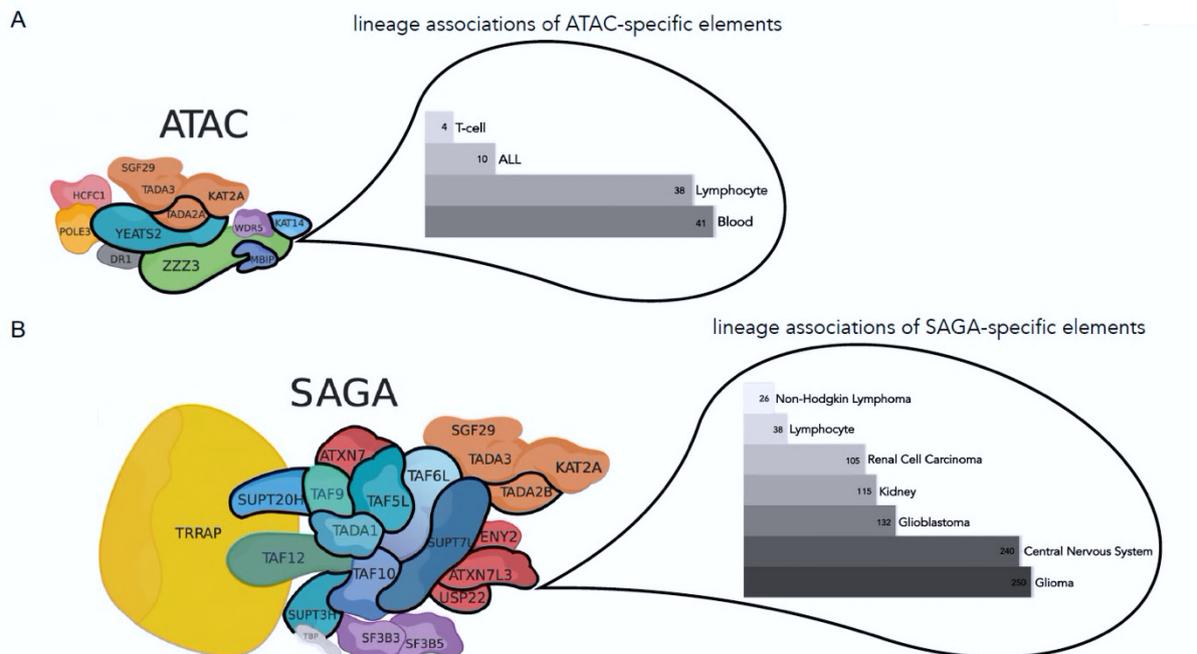


FIGURE 1.10 Analysis of dependencies and lineage associations of ATAC and SAGA-unique elements in CRISPR drop-out screens of cancer cell lines. Data was extracted from the Cancer Dependency Map - DepMap (<https://depmap.org/portal/>) as of August 2020. Unique elements in each complex are highlighted with a black edge. **A.** ATAC elements (TADA2A, YEATS2, ZZZ3, MBIP, KAT14) are required in most cell lines and are called as 'common essential genes'. Upon grouping of cell lines into specific tissues, or cell lineages, lineages for which requirements of individual genes were determined at $p\text{-value} < 0.0005$ are considered as a lineage association bias. Bars reflect the total number of cell lines in each lineage associated with the various ATAC-unique elements. **B.** SAGA unique elements (TADA2B, SUPT20H, SUPT3H, SUPT7L, TAF5L, TAF6L, TADA1, ATXN7, ATXN7L3, USP22) are highlighted with a black edge. None of the elements was called 'common essential'. Their lineage associations are depicted in the bar graph, as per the criteria in A.

When compared to KAT2A, SAGA-cancer dependencies are more similar with subunits of the DUB module, specifically with USP22 and ATXN7. Both associate with Renal Cell Carcinoma, and ATXN7 with Lymphoid malignancies. Indeed, ATXN7 has been demonstrated to associate with prognosis of Renal Cell Carcinoma [250], and it confers susceptibility to Breast Cancer [251], where ATXN7 gene variants have been linked to post-operative prognosis of HBV-related Hepatocellular Carcinoma [252], all of these tumours to which KAT2A makes oncogene-like contributions [253]. USP22 acts as an oncogene in multiple malignancies [254–255], including in NSCLC [256] and in Gastric Cancer [257]. However, it also displays tumour-suppressor roles, specifically in Colorectal Cancer through decreased mTOR activity [258], and surprisingly, in AML [259]. In AML, USP22 deubiquitinates and stabilizes PU.1. In the absence of USP22, degradation of PU.1 leads to a block in myeloid cell differentiation, which promotes K-Ras mutant leukaemia progression [226].

Interestingly, this is in contrast with Kat2a oncogenic functions in AML demonstrated by our lab [197–260]. A possible explanation is that this has to do with the relative independence between the DUB and HAT modules in SAGA. In alternative, the contrasting effects of Kat2a and Usp22 loss may be explained by the fact that different forms of AML were analysed in both studies. These may rely on distinct leukaemia-initiating cells and be driven by different mutational signatures, which may determine specific dependencies on activation or repression of individual genes. However, analysis of mouse ESCs also supports a degree of independence between Usp22 and Kat2a: Usp22 is uniquely required for appropriate differentiation into the 3 germ layers through repression of pluripotency master regulator Sox2 [226], a role that exceeds the Kat2a requirement in pluripotent cells. Likewise, although Usp22 and Kat2a KO mouse models exhibit embryonic lethality at E10.5 [261], Usp22 KO embryos present generalized apoptosis, which is not restricted to mesodermal structures as in Kat2a KO [182–184–262]. ATAC Kat14 KO mouse model [242], is also embryonic lethal at E10.5, with cell cycle defects and localised apoptosis that does not completely overlap with Kat2a, particularly in mesodermal structures.

Together, these observations are in agreement with independent functions of the complexes, and within complexes, of the catalytic units. Additionally, there may be compensation by Kat2b, which like Kat2a can function in SAGA and ATAC complexes.

On the other hand, SAGA Core *Supt20* hypomorphs, are more similar to *Kat2a* hypomorphic animals [263], with neural tube and axial skeleton developmental defects. Similarly, in mouse ESCs, Core components *Taf5l* and *Taf6l* maintain low levels of differentiation through activation of a MYC regulatory network [264], which is largely shared with *Kat2a* [194]. Therefore, loss of SAGA Core elements aligns better with loss of with *Kat2a* itself, resembling the pattern observed in lineage affiliation of cancer dependencies.

However, in fly, KAT2A essential roles are closely aligned with ATAC. Accordingly, KAT2A is required in metamorphosis [183], and so is ATAC-*Ada2a*, whereas SAGA-*Ada2b* or the SAGA DUB module, are dispensable at this stage [265].

Taken together, the examples above suggest that ATAC is a generic requirement in cancer, whereas SAGA may have lineage specific functions. ATAC regulates biosynthetic functions, which are conveyed through *Kat2a* with some degree of participation of *Kat14*, which compensates for *Kat2a* loss, at least in some lineages. On the other hand, SAGA plays lineage-specific functions, with minimal compensation of *Kat2a* activity within the complex, and with some independence of function between HAT and DUB modules.

It is likely that ATAC-mediated *Kat2a* acetylation activity maintains self-renewal and/or survival of proliferative and metabolically active cells, while *Kat2a* functions in the context of SAGA to maintain cell identity and prevent deviation from existing transcriptional programs.

1.6 Thesis overview

Hypothesis

KAT2A contributes to normal and leukaemic haematopoiesis with differential roles of the multi-modular HAT complexes it integrates.

Objectives

I investigate this hypothesis through development of 3 specific goals:

- I. Define the role of KAT2A complexes and the molecular programmes they regulate in normal haematopoiesis
- II. Define the role of KAT2A complexes and their targets in maintaining AML
- III. Explore KAT2A and complex-specific dependencies of patient AML cells with translational intent

Overall, this project aims to investigate the cellular and molecular mechanisms by which both ATAC and SAGA- KAT2A complexes attain their uniqueness in the context of normal and leukaemic haematopoiesis.

Thesis structure

Following the literature review presented in this Chapter, I move on to presenting the experimental materials and methods in Chapter 2. This Chapter describes in detail the experimental design, techniques and tools used throughout this work. The experimental results are divided in Chapters 3 to 5. Each Chapter includes a brief introduction and short summary of the main findings. Chapter 3 explores the functions of KAT2A in normal haematopoiesis. I describe the effects of inhibiting KAT2A *via* shRNA in normal human cord blood (CB) specification. Furthermore, using the same genetic knockdown approach I characterise the roles of ATAC and SAGA in stem and progenitor blood development. To do so, I use complex-unique subunits as surrogates for the function of one or the other complex. *ZZZ3* is representative of ATAC, whereas *SUPT20H* symbolises SAGA activities. In Chapter 4, I move on to investigate the roles of KAT2A and its-containing ATAC and SAGA complexes in blood malignancy, specifically in acute myeloid leukaemia (AML). For this I make use of a panel of cell lines representative of different AML genetic

backgrounds. In Chapter 5 I explore the functions of KAT2A in human patient leukaemic cells by inhibiting its catalytic activity with a small drug, MB-3, or again *via* shRNA knockdown. To do so, I use of an *in vitro* co-culture protocol to sustain the leukaemic blasts for up to 3 weeks. Moreover, I attempted to dissect the functions of ATAC and SAGA complexes in the primary AMLs. Finally, all these findings are brought together in Chapter 6, where I discuss the main conclusions, potential improvements and possible future implications of this work.

Chapter 2

Materials and Methods

This Chapter details the materials, protocols and resources used for the experimental work described in this thesis.

2.1 Cell Lines

AML cell lines used throughout this work were **THP-1** (carrying a *MLL-AF9* fusion), **Kasumi-1** (with a t(8;21) chromosome translocation, giving rise to the *AML1-ETO* or *RUNX1-RUNX1T1* fusion gene, and a c-kit mutation), **MOLM-13** (carrying fusions *FLT3- internal tandem duplication (FLT3-ITD)* and *MLL-AF9*), **MV4-11** (*FLT3-ITD* and *MLL-AF4* fusions), and **KG-1 α** (*FGF1OP2-FGFR1* fusion) (**Figure 2.1**).

	THP-1	Kasumi-1	MOLM-13	MV4-11	KG-1 α	
			■	■		FLT3-ITD
				■		MLL-AF4
	■		■			MLL-AF9
		■				AML1-ETO t(8;21)
		■				C-KIT
						del(11)(q23.1q23.2)
					■	FGFR1OP2-FGFR1 (OP2-FGFR1)

FIGURE 2.1 AML cell lines used in this study and their known oncogenic mutations.

All AML cell lines were maintained in Roswell Park Memorial Institute-1640 media (RPMI, Sigma) supplemented with 20% FBS (Foetal Bovine Serum, Invitrogen), 1% Penicillin /Streptomycin /Amphotericin (P/S/A, Invitrogen) and 2mM L-Gln – R20. Cells were subcultured at the appropriate density and maintained in a 37°C, 5% CO₂ incubator. K562 cell line is a chronic myelogenous leukaemia cell line, which was maintained in RPMI supplemented with 20% FBS (Foetal Bovine Serum, Invitrogen), 1% P/S/A, and 2mM L-Gln as above. K562, THP-1, Kasumi-1, MOLM-13 and MV4-

11 were a kind gift from Professor Brian Huntly, Cambridge, UK. KG1a were a kind gift from Dr Joanna Baxter, Cambridge Blood and Stem Cell Biobank. Human embryonic kidney (HEK-293T) cells were grown in Dulbecco's Modified Eagle's Medium (DMEM, Sigma) supplemented with 10% FBS, and P/S/A and 2mM L-Gln - *D10*. The stromal cell line MS5 obtained from the Bonnet Lab (Francis Crick Institute, London, UK), was maintained in Iscove's Modified Dulbecco's medium (IMDM, Sigma) supplemented with 10% FBS, 1% P/S/A and P/S/A and L-Gln - *I10* and used between passage 3 to 5.

2.2 Human Cord Blood CD34+ cells

CB samples were obtained with informed consent under local ethical approval (REC 07- MRE05-44). Samples were processed within 24h after collection and kept at 4°C prior processing. CB samples were first diluted 1:1 in Phosphate Buffer Saline (PBS, Sigma-Aldrich) supplemented with 2mM EDTA (Ethylene Diamine Tetra Acetic Acid, Sigma-Aldrich) to prevent mononuclear cell (MNC) aggregation. 30-70mL of blood was layered onto 15mL of Lymphoprep (Stem Cell Technologies) and MNC isolated by density gradient centrifugation. CB MNC were then enriched for CD34⁺ cells using the EasySep Human CB CD34+ Selection Cocktail (Stem Cell Technologies), 100 µL/mL sample, and Dextran RapidSpheres (50 µL/mL sample). Cell-bead mix was subjected to magnetic incubation (MojoSort, BioLegend) for 3 minutes and washed with PBS-EDTA-0.5% Bovine Serum Albumin (BSA) for a total of 4 times until high purity CD34+ cells were obtained.

2.3 Primary Human AML Cells

AML cells were obtained at Addenbrooke's Hospital (Cambridge, United Kingdom) in collaboration with Professor Brian Huntly, and at Sant'Orsola Malpighi Hospital (Bologna, Italy), in collaboration with Dr Antonio Curti, University of Bologna, Italy. Details of patient samples, obtained under ethical approval (REC 07-MRE05-44), are listed in Table 5.1 (Chapter 5).

Cells were maintained in a MS5 co-culture system for medium-term culture, which reflects AML engraftment in immunodeficient animals, and is thus of clinical significance [266]. MS5 monolayers were seeded onto 48-well plates pre-coated with Collagen (Thermo) at a density of $1.5 \times 10^4 - 3 \times 10^4$ cells/mL until reaching 50-60% confluency, prior to culturing AML patient cells.

AML cells were thawed in PBS 2% FBS, 1% P/S/A and pre-stimulated for approximately 4h with H5100 (Stem Cell Technologies) supplemented with recombinant human IL-3, G-CSF, and TPO (20 ng/mL each, Peprotech), 1% HEPES (Gibco, 1M), 1%P/S/A – 3GT, herein called pre-stimulation media (PSM). Cells were then transduced with the desired lentiviral constructs in transduction media: H5100 supplemented with SCF (100 ng/mL), Flt-3 (100 ng/mL), IL-3 (60 ng/mL), TPO (10 ng/mL), 1% HEPES and 1% P/S/A. Cells were washed twice in PBS 2% FBS, 1% P/S/A and once in H5100 1% P/S/A and seeded onto MS5 stroma. Cultures were kept at 37°C, 5% CO₂ for up to three weeks, with twice-weekly demi-population and fresh addition of H5100-3GT. For KAT2A inhibition experiments, AML cells were seeded onto MS5 stroma in H5100 in the presence of 100 µM MB-3 (Abcam) or DMSO (0.1%), and demi-populated as above with fresh addition of MB-3. Cells were stained for flow cytometry analysis once a week and surface phenotyping analysis performed on a Gallios instrument (Beckman Coulter).

2.4 Cell Cryopreservation, Thawing and Counting

All cells were cryopreserved in 80% FBS with 20% (v/v) Dimethyl Sulfoxide (DMSO, Thermo) and stored at -80°C in a Mr Frosty container (Thermo) prior to transferring to long term storage in liquid N₂. Cells were rapidly thawed by gentle agitation in a 37°C water bath and transferred to 10mL of pre-warmed complete media. Cells were then centrifuged at 400g for 5 minutes, re-suspended in complete media and seeded at the appropriate cell density. Cell counting and viability check were performed manually by dilution with Trypan Blue (Sigma) within 3-5 minutes of preparation.

2.5 Lentiviral Packaging and Transduction

A 2nd generation lentiviral production system was used for stable cell transductions. Lentiviral particles were produced by mixing 1.28µg DNA of the packaging plasmid pCMV-VSV-G (Addgene 8454) (**Fig. 2.2**) and of the envelop plasmid pMD2.G (Addgene 12259) (**Fig. 2.3**), with 1.92µg DNA of the pLL3.7 shRNA-expressing vector (Addgene 11795) (**Fig. 2.7**). DNA mixtures in serum-free DMEM were then added to the transfection reagent Turbofect (Thermo) or TransIT (Mirus). Complexes were left to form at for 30 minutes at RT prior to adding dropwise to adhered HEK-293T cells (about 80% confluency) in D10. Media was changed at 24h and the lentivirus-containing media harvested at 48h and 96h by pooling the media from the desired amount of T75 flasks followed by a 5-minute centrifugation 400g and transfer to a new collection tube until concentration overnight at 3,220g, 4°C. Supernatant was discarded, and viral pellets re-suspended in PBS, in the minimum volume required to achieve complete pellet dispersion. Lentivirus were used straight away or stored at -80°C until use. Cell lines were transduced overnight with 1-2 T75 packaging flask-equivalents (FE)/10⁶ cells and washed the following day. For transduction of CD34+ CB, cells were pre-stimulated in serum-free medium (HSC expansion medium XF, Miltenyi Biotec) with SCF, TPO and Flt3L (respectively 200, 20 and 20 ng/mL) for up to 24h and transduced overnight using 2FE/ 2-3 x 10⁵ cells. Cells were washed the following day and cultured in half the cytokine concentration for an additional 2-3 days prior to sorting. The shRNA constructs express green fluorescent protein (GFP+), which was quantified on a Gallios flow analyser (Beckman Coulter) to determine the transduction efficiency prior to sorting.

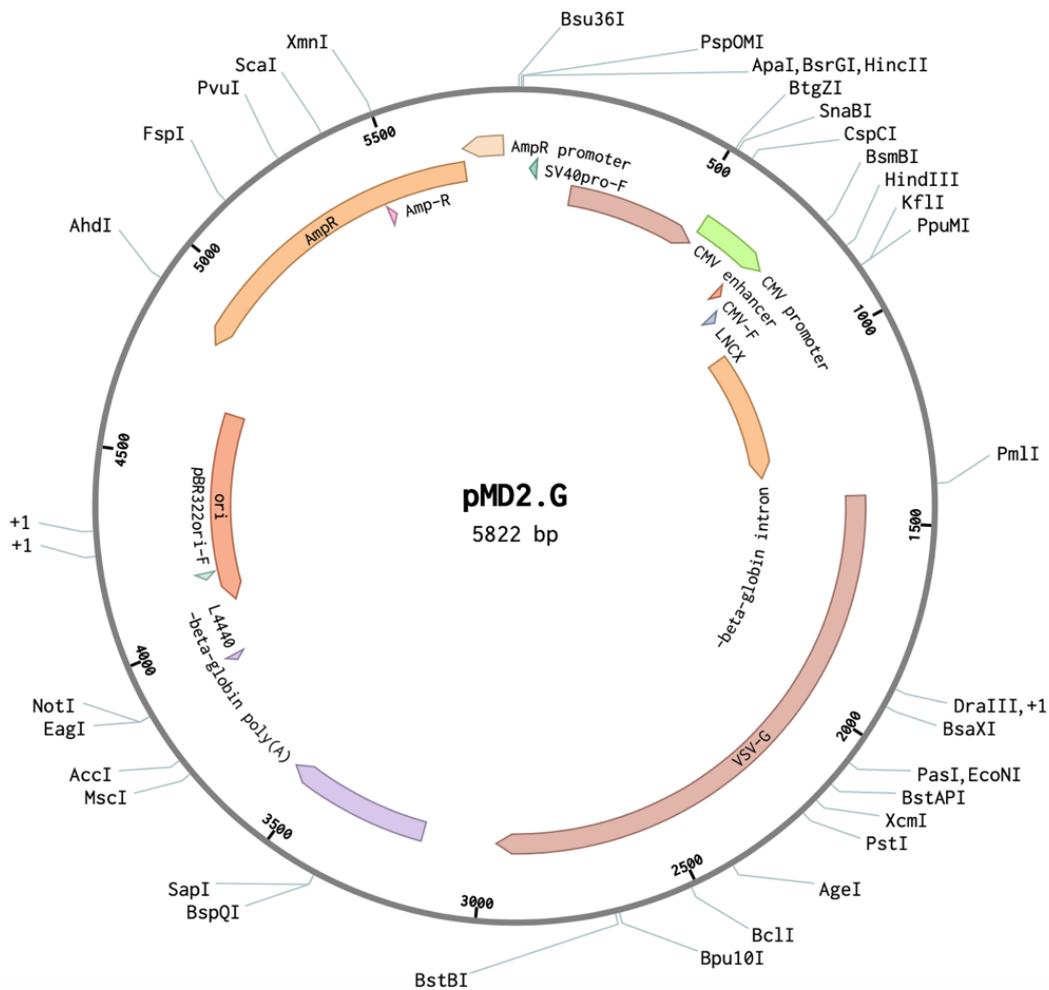


FIGURE 2.3 pMD2.G VSV-G-expressing envelope plasmid.

2.6 Functional Cellular Assays

2.6.1 Flow Cytometry

2.6.1.1 Human CB CD34+ Cell Sorting

CD34+ stem and progenitor cells were stained on ice for 20 minutes using the cell surface antibodies listed in **Table 2.1**. FcR blocking reagent (Miltenyi) was included. Cells were washed in PBS, re-suspended in 500 μ L PBS-EDTA-0.5% BSA and sorted on a BD FACSAria™ Fusion (BD Biosciences). Applied gating strategy used to sort

the different stem and progenitor compartments is shown in **Fig. 2.4**. Compensation was determined using single stains, prepared from pooled samples.

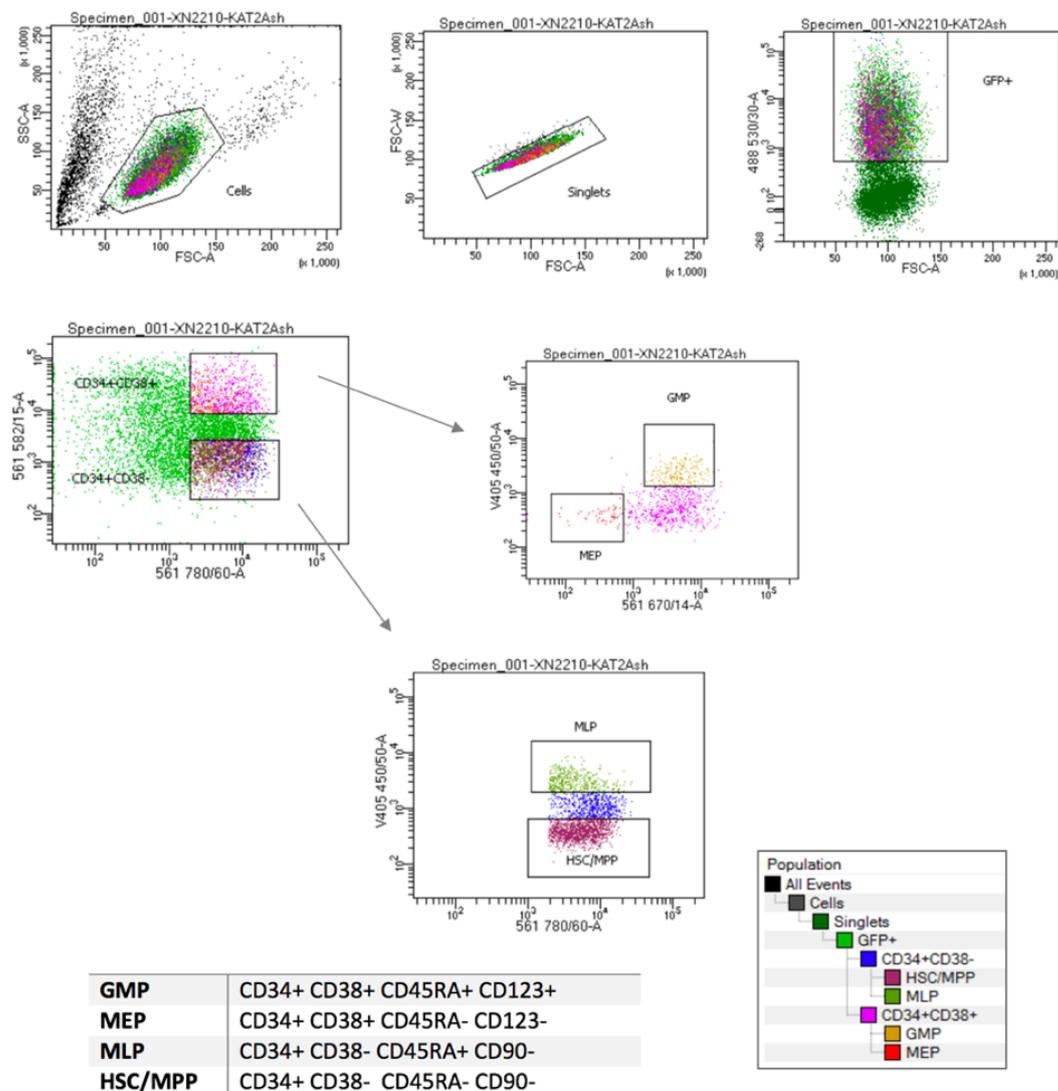


FIGURE 2.4 Flow cytometry gating strategy for sorting the human CB CD34+ stem and progenitor compartments. Cells were first gated for live cells (SSC-A vs FSC-A), from which a second gate was drawn to separate singlets from doublets (FSC-W vs FSC-A). Singlets were gated for GFP positivity. The GFP+ gate was further analysed on CD34 and CD38 positivity. Of the CD34+38+ cells, surface expression was determined for CD45RA and CD123 to further isolate GMPs and MEPs. CD34+38- cells were subdivided in CD45RA positivity and CD90 positivity. CD45RA+90- cells are MLP and CD45RA-90- are HSC/MPP.

Table 2.1 Antibodies used in flow cytometry analysis and cell sorting.

Antibody	Fluorochrome	Catalogue #	Clone	Dilution	Supplier
CD34	PE-Cy7	343515	581	1:200	BioLegend
CD38	PE	356603	HB-7	1:100	BioLegend
CD45RA	APC	304118	HI100	1:100	BioLegend
CD123	PE-Cy5	306008	HI264	1:100	BioLegend
CD71	PE	334105	GY1G4	1:100	BioLegend
CD235a	PB	349103	H1264	1:100	BioLegend
Annexin V	APC	640941	-	1:100	BioLegend
Hoechst 33342	-	H3570	-	1:10000	LifeTechnologies
Tag-it Violet	APC	425101	-	5nM	BioLegend

2.6.1.2 Cell Cycle Analysis

Hoechst 33342 (Life Technologies) was used to determine the fraction of cells in different stages of the cell cycle. CD34⁺ stem and progenitor cells were stained with 4 μ g/mL of Hoechst 33342, a fluorescent nucleic acid stain, and incubated for 2h at 37 °C in a water bath, protected from the light, and vortexing every 15 minutes. Cells were then washed and re-suspended in 300 μ L of 0.2 μ g/mL Hoechst 33342 maintenance solution in *R20* (cell lines) or *I10* (human CB), prior to examination by FACS (Gallios, Beckman Coulter). Gating strategy is shown in **Fig. 2.5**.

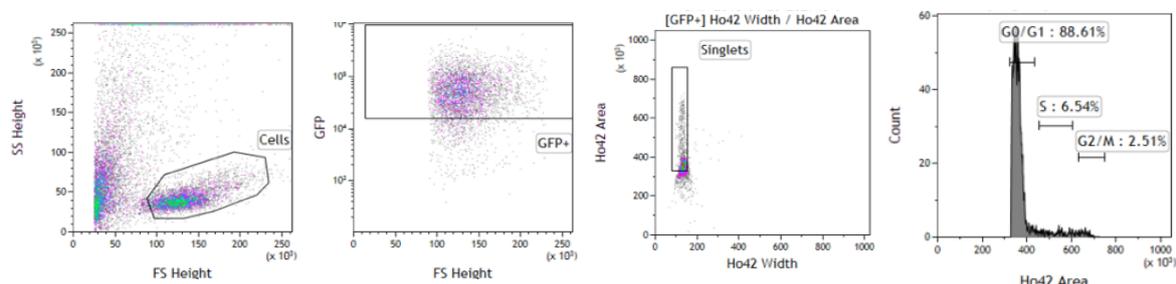


FIGURE 2.5 Flow cytometry gating strategy for cell cycle analysis of human CB CD34⁺ cells and cell lines.

2.6.1.3 Apoptosis Analysis

Annexin-V was used to detect apoptotic cells based on its ability to bind to phosphatidylserine, a marker of apoptosis when present on the outer leaflet of the plasma membrane. Cells were treated with 3 μ L Annexin-V-APC (BD Biosciences) / 100 μ L sample. Staining was performed for 15 minutes at room temperature (RT), protected from the light, along with counterstaining with a non-permeable fluorescent nucleic acid stain Hoechst 33258 (Life Technologies) 1:10,000 or DAPI (Thermo) 1:10,000. The combination of the two stainings allows for a distinction between viable cells (double negative), cells in early apoptosis (Annexin-V positive), cells in late apoptosis (double positive) and dead cells (Hoechst positive). The total number of Annexin-V positive cells was determined by FACS (Gallios, Beckman Coulter) and expressed as a percentage of dead cells. Gating strategy is shown in **Fig. 2.6**.

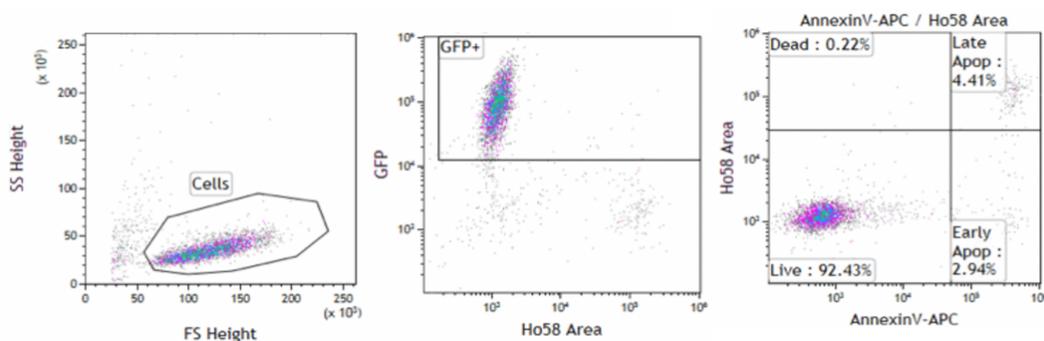


FIGURE 2.6 Flow cytometry gating strategy for apoptosis analysis of human CB CD34+ cells and cell lines.

2.6.1.4 Cell divisional tracking

For cell divisional tracking and proliferation, 1E6 cells/mL of cell suspension were stained as per manufacturer's instructions using 1 μ L of 5 μ M Tag-it Violet (BioLegend) in PBS. Cells were incubated for 20 min at RT, protected from light. Staining was quenched by adding 5 mL of cell culture medium R20. Pelleted cells were re-suspended in pre-warmed R20 incubated for 10 min at 37°C and analysed daily by flow cytometry on Day0 (loading) to Day3.

2.6.2 Colony-Forming Cell (CFC) Assays

The colony-forming cell (CFC) assay was used to study the proliferation and differentiation pattern of CB stem and progenitor cells, as well as the proliferation ability of leukaemic patient blasts. CFC assays were performed in methyl-cellulose-based semi-solid medium StemMACS-HSC-CFU complete with EPO (Miltenyi Biotech), which supports the growth of granulocyte (G) and monocyte (M) precursor cells, CFU-GM (Colony-Forming Unit Granulocyte Monocyte), erythroid progenitor cells or BFU-E (Burst-Forming Unit Erythroid), and of Mix or CFU-GEMM (Colony-Forming Unit Granulocyte, Erythrocyte, Monocyte, Megakaryocyte) type colonies. For CB CD34⁺ cells, CFC-assays were set up immediately after sorting the different progenitor compartments. Preferably, 200 to 300 cells were re-suspended in *I10* and embedded in falcon tubes containing 3mL of StemMACS-HSC-CFU with EPO. Tubes were vortexed and left at RT until bubbles disappear. Cells were then plated in duplicate onto 35 mm-dishes and incubated at 37 °C 5% CO₂. Proliferation capacity of leukaemic patient cells post treatment with KAT2A inhibitor (**Fig. 2.7**), alpha-methylene-gama-Butyrolactone 3 (MB-3) [267], or vehicle dimethyl sulfoxide (DMSO) (Sigma), was determined using 5x10⁵ blasts per CFC-assay as above. The number and morphology of the colonies were scored under the light microscope 10-12 days after plating.

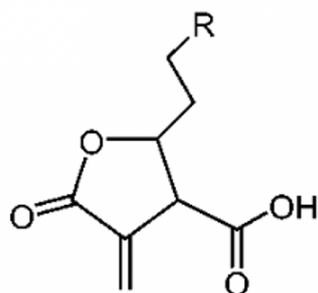


FIGURE 2.7 Chemical structure of α -methylene- γ -Butyrolactone 3 (MB-3), the small molecule inhibitor of KAT2A. MB-3 has an IC₅₀ of 100 μ M. The length of the aliphatic side chain of MB-3 is critical for its KAT inhibitory activity. The drug was dissolved in an appropriate volume of DMSO to provide a 100mM solution and stored in aliquots at -80°C until use. R- alkyl chain. Adapted from [267].

2.6.3 Erythroid differentiation cultures

Transduced HSC were tested in erythroid differentiation conditions in serum-free liquid culture (HSC expansion medium XF, Miltenyi Biotec) in the presence of hydrocortisone (10^{-5}M), SCF (100 ng/mL), TPO (10 ng/mL) and EPO (3U/mL). Cultures were followed up for 9 days with daily live and dead cell counts and regular flow cytometry analysis for CD34, CD71 and CD235a markers (Supplemental Table 4). K562 cells were differentiated into the erythroid lineage in the presence of 1.5% DMSO as described²⁹. Cells were monitored for differentiation markers (CD71 and CD235a) and/or activation of erythroid-affiliated molecular programmes during a 6-day culture period.

2.6.4 Cytospins

Cells were centrifuged onto slides for 5 minutes at 700rpm, stained with rapid Romanowsky stain pack as per manufacturer's instructions and fixed with Depex (Sigma) mounting medium. Maturation of myeloid cells was assessed by examining the Romanowsky-stained cytospin smears for morphologic changes. In more detail, myeloid differentiated cells were classified if presenting an enlarged, irregular shape and when visible if the nucleus of the cell was multilobed and/or if it presented a 'band' or 'kidney bean' shape. In turn, the myeloblasts or undifferentiated /blast-like cells, were classified if presented a round or oval shape with a large nucleus occupying pretty much the whole cell. Slides were classified blindly, both by me and by another person from the lab, as an independent observer. Alternatively, myeloid-specific surface protein expression was analysed by flow cytometry, particularly expression of CD13 by antibody staining.

CTCGAG to **GGGATCC** to permit a diagnostic restriction digestion with *Bam*HI (NEB) and one **T** base was added to the poly T-tail, to match the Lab's shRNA library. shRNA constructs for *KAT2A*, *USP22* and *ZZZ3* had previously been cloned in the Lab. All sequences of shRNA constructs used throughout this work can be found in **Table 2.1**.

Table 2.2 Sequences of shRNA constructs used in this work.

Oligonucleotide	Sequence (5' to 3')
<i>KAT2A</i>	TGCTGAACTTTGTGCAGTACAAGGGATCCGTTGTACTGA CAAAGTTCAGCTTTTT T C
<i>SUPT20H</i> sh1	TCCATCAAGTATTCCTCGGAAAGGGATCC TTTCCGAGGAATA CTTGATGGTTTT T
<i>SUPT20H</i> sh2	TGCGGATGTGTCATAGCAGAAAGGGATCC TTTCTGCTATGAC ACATCCGCTTTTT T
<i>TADA2B</i> sh1	TCGTGACTGTGAAGACTATTATGGGATCC ATAATAGTCTTCA CAGTCACGTTTT T G
<i>TADA2B</i> sh2	TATGATTACGAGATCGAGTATGGGGATCC CATACTCGATCTC GTAATCATTTTT T G
<i>TADA2A</i> sh1	TTTGAAGATGACTCGGACATTTGGGATCC AAATGTCCGAGTC ATCTTCAATTTTT T G
<i>TADA2A</i> sh2	TGCACTATATGAAGCATTTTATGGGATCC ATGAAATGCTTCA TATAGTGCTTTTT T
<i>YEATS2</i> sh1	TCGTCAGAGTTCAAGTTCATTTGGGATCC AAATGAACTTGAA CTCTGACGTTTT T G
<i>YEATS2</i> sh2	TTCATTGACCAGCGACTGATTGGGGATCC CAATCAGTCGCT GGTCAATGATTTTT T G
<i>USP22</i>	TAGCTACCAGGAGTCCACAAAGGGGATCC CTTTGTGGACTC CTGGTAGCTTTTT T C
<i>ZZZ3</i>	TGCATCAGATGACGAAAGTATTGGGATCC AATACTTTCGTCA TCTGATGCTTTTT T C

pLL3.7 digestion

The vector was digested at 37°C for 3h with the restriction enzymes *Hpa*I and *Xho*I (NEB). *Hpa*I cuts the 3' cloning site bluntly, where *Xho*I leaves a sticky end when cutting the 5' cloning site, allowing for the binding of the designed oligos to complement the restriction ends. Digested DNA was run on a 0.8% agarose gel in 1x TAE (Tris-Acetate-EDTA) with SYBR Safe gel stain (Invitrogen) for visualisation. DNA was then extracted from the gel and purified using the QIAquick gel extraction kit (Qiagen) as per manufacturer's instructions.

Annealing, ligation and transformation

Sense and antisense shRNA oligos were reconstituted at 100 μ M in nuclease-free water. A total reaction volume of 10 μ L, containing 0.12 μ M of each oligo, 1x concentrated *EcoRI* buffer (NEB) and nuclease-free water, were annealed using a thermal cycler (G-Storm GS0004M). The program was set to start at 95°C and drop ~0.1°C/s down to RT. Ligations into the digested vector were performed overnight at RT, using 0.02pmol of the linearized backbone, 0.06pmol of the annealed oligos, 1x buffer (NEB), of T4 DNA Ligase (NEB, 20,000 U) and water to a total volume of 10 μ L per reaction. Two controls were included: (1) insert replaced by water, which controls for single-cut and re-ligated, as well as undigested vector; (2) ligase replaced by water, which controls for undigested vector. Transformation of NEB-Stable chemically competent *E. Coli* was performed on ice for 30 minutes using 2 μ L of the ligation product, followed by a heat shock at 42°C for 30 seconds and another incubation on ice for 5 minutes. 200 μ L of pre-warmed Luria-Bertani (LB) broth media was added to the bacteria, followed by a 1h incubation at 30°C, 250rpm (Thermoshaker, Thermo). Cells were plated onto LB-agar plates with 100 μ g/mL ampicillin (Amp, Sigma) and cultured for 48h at RT. If transformation was deemed successful, 3-12 bacterial colonies were inoculated into 5mL LB with 5 μ L Amp and incubated overnight at 37°C, 250rpm Thermoshaker (Thermo). Bacterial colonies were minipreped to isolate plasmid DNA using QIAprep Spin Miniprep Kit 50 (Qiagen), as per manufacturer's instructions. DNA concentration was measured by NanoDrop Lite Spectrophotometer (Thermo).

Screening for positive clones

Screening for positive clones was performed by 2 methods: (A) PCR and (B) *BamHI* restriction digestion, prior to validation by sequencing.

A. PCR screen

Successfully inserted clones were screened by PCR in a thermal cycler (G-Storm GS0004M). Primers used were as follows: FW: CTGTTCTTTTAACTAGC and RV: TATTAGGTCCCTCGACCTGCTGC. PCR reactions were set up following the HotStarTaq Plus Master Mix kit (Qiagen) protocol in a 20 μ L reaction volume. PCR controls included a non-template control (NTC) to test for contamination; a positive control using the pLL3.7-vector expressing the *KAT2A* shRNA to recognise positive clones and a negative control consisting of a pLL3.7 empty vector to distinguish negative clones. Thermal cycler profile was as follows: 95°C for 15 minutes, 35 \times [94°C

for 30 seconds, 48°C for 30 seconds, 72°C for 45 seconds], 72°C for 5 minutes and lastly 25°C for 10 minutes. Amplified PCR products were resolved on a 2% agarose gel in 1x TAE with 10µL SYBR Safe gel stain (Invitrogen) for visualization with a DNA ladder (Thermo).

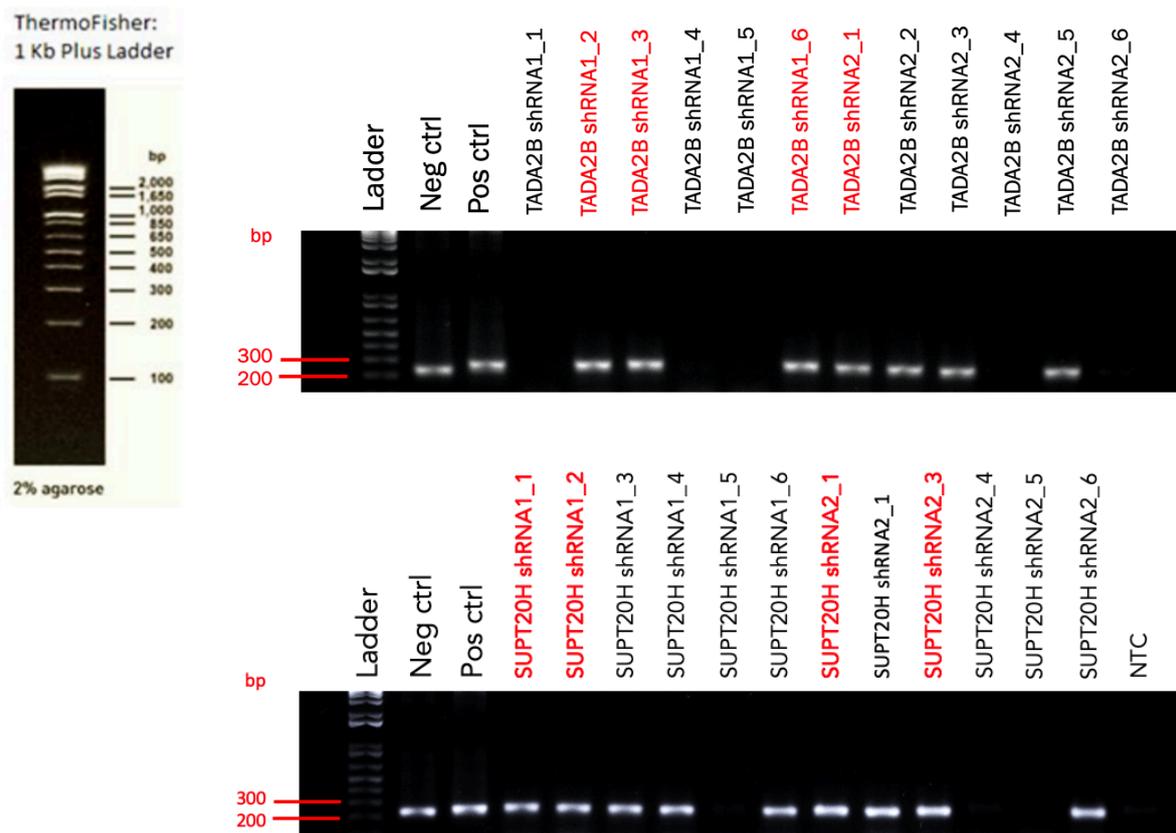


FIGURE 2.9 Assessment of shRNA insertion into pLL3.7 vector using PCR and agarose gel electrophoresis. Representative gels. Positive ligated clones (red) were sent for sequencing validation.

B. *Bam*HI restriction digestion

A diagnostic screen was performed for 600 ng purified bacterial DNA with *Bam*HI (NEB) in a total reaction volume of 20 µL for 3h, at 37°C. Digested material was then resolved on a 0.8% agarose gel in 1x TAE with 10 µL SYBR Safe gel stain (Invitrogen) for visualization with a DNA ladder (Thermo).

Positive clones identified by a combination of both methods were sent for Sanger Sequencing (Source Bioscience), using the FW and RV primers at concentration of 10pM. 5ng of DNA from successfully validated clones were transformed into 20µL of competent cells. Following a two-day incubation at RT, single colonies were inoculated in 3mL of LB with 3µL Amp (100µg/mL) and incubated 6-8h at 37°C, 225 rpm. Bacterial colonies were then transferred to a 50 mL LB with 50µL Amp (100 µg/mL) and cultured overnight at 37°C, 225rpm. Isolation of plasmid DNA was performed with the PureLink HiPure Plasmid Filter Midiprep kit (Invitrogen) as per manufacturer's instructions, and DNA concentration measured by NanoDrop Lite Spectrophotometer (Thermo).

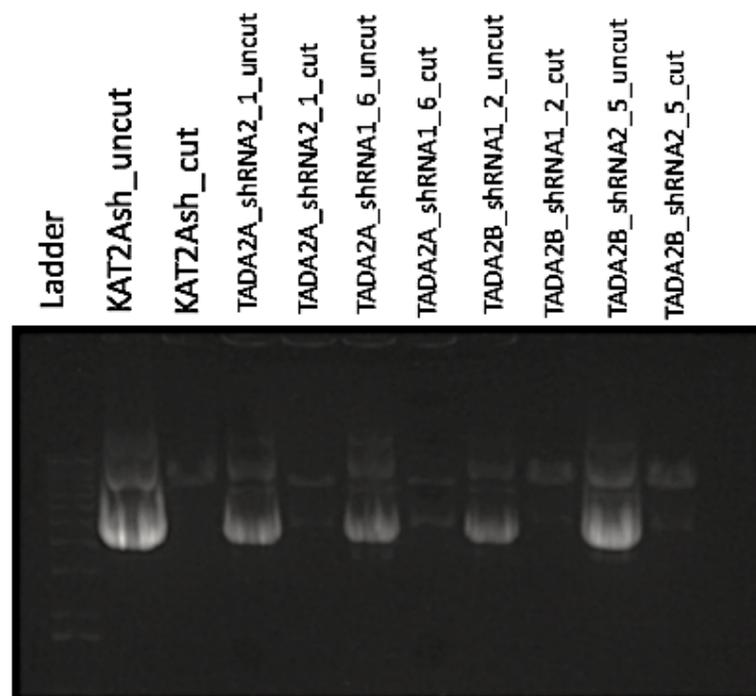


FIGURE 2.10 Assessment of shRNA insertion into pLL3.7 vector by *Bam*H1 restriction digestion using PCR and agarose gel electrophoresis. Representative gel. *Bam*H1 cuts the shRNA loop sequence, which was adjusted from CTCGAG to GGGATCC to permit the diagnostic restriction digestion. Circular (uncut) plasmid DNA produces 3 bands on the gel, representing nicked (highest band), linear (middle band), and supercoiled (lowest band), whereas linear (cut) DNA appears as a single band, only if shRNAs are successfully inserted.

Assessment of gene knockdown in lentiviral producing HEK-293T cells

When first producing lentivirus for the cloned shRNAs, prior to transducing into the experimental cells, mRNA was isolated from lentiviral-pancaking 293T cells to tested knockdown efficiency (**Fig, 2.11**).

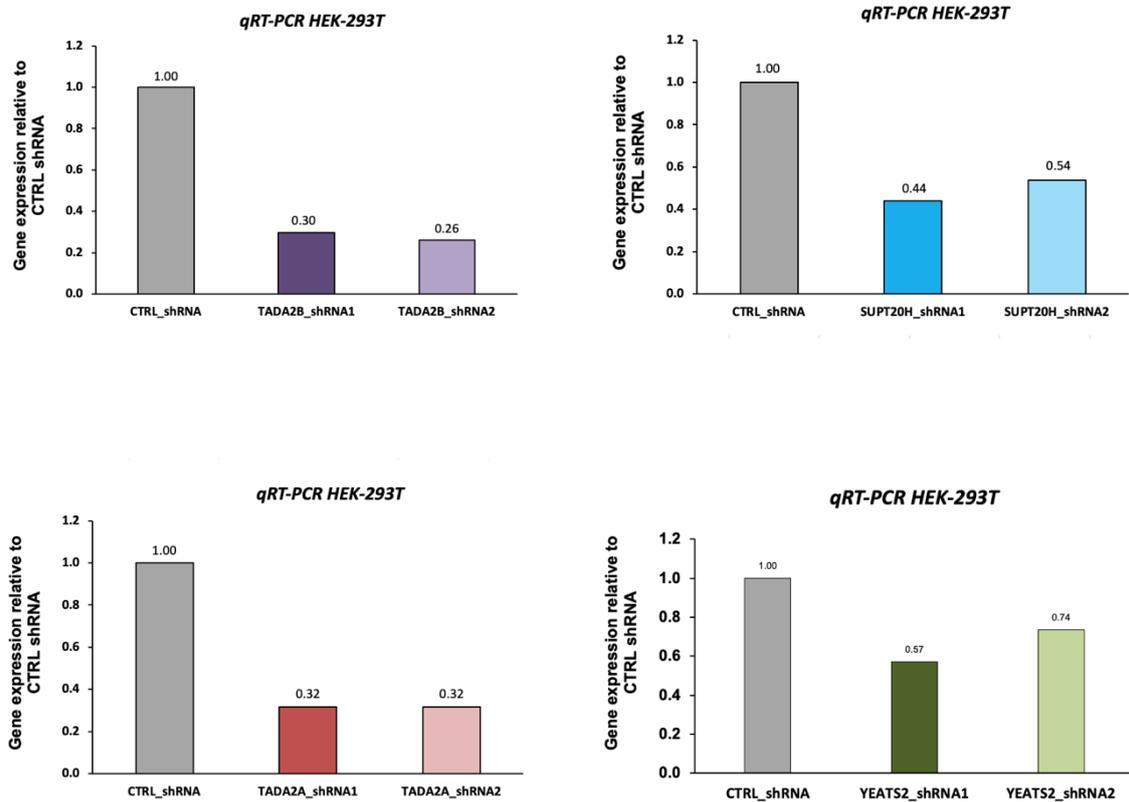


FIGURE 2.11 Assessment of gene knockdown in lentiviral producing 293T cells.

2.7.2 RNA Extraction, cDNA Synthesis

RNA extraction from Trizol (Invitrogen) homogenates was performed as follows: 200 μ L of Chloroform (Sigma) was added to cells in 1mL of Trizol, mixed by hand for 15 seconds and phase separation performed by centrifugation at 16, 000g for 15 minutes at 4 °C. The upper aqueous phase was transferred into a new tube using 0.5-1 μ L GenElute LPA (25 mg/mL, Sigma) or GlycoBlue (15 mg/mL, Invitrogen) as a carrier, to which 5/6 of the volume of the upper phase of Isopropanol (Sigma) was added and mixed well by hand. Samples were incubated for 30 minutes at RT and RNA precipitated at 16,000g for 10 minutes at 4 °C. Excess salt was removed with 80% Ethanol (Sigma) by centrifugation at 16, 000g for 5 minutes at RT; pellet was released by vortexing. Residual Ethanol was removed by centrifugation at 16, 000g for 30 seconds at RT. RNA was finally resuspended in 3-5 μ L of RNase free water and incubated in a Thermoshaker (Thermo) at 50 °C, for 10 minutes at 300rpm. Complementary (c)DNA synthesis was performed from total RNA from equal numbers of cells using the SuperScript III Reverse Transcriptase Kit (Invitrogen) as per manufacturer's instructions.

2.7.3 Quantitative Real Time PCR (qRT-PCR)

cDNA was analysed in triplicate by qPCR using Taqman Gene Expression Assays and Taqman Gene Expression Mastermix (Applied Biosystems). Additionally, qRT-PCR was performed using SYBR Green Mastermix (Takyon Rox SYBR Core Kit dTTP Blue, Eurogentec or Brilliant II SYBR Green Low Rox Master Mix, Agilent Technologies). Taqman probes and primers are listed in **Table 2.3** and **Table 2.4**, respectively. Relative gene expression was calculated by the $2^{-\Delta\Delta C_t}$ method using *HPRT1* as reference. NTC was used as control. Assays were carried out in 96-well, semi-skirted, white, PCR plates for Roche LightCycler (Starlab) on a LightCycler 480 System (Roche), or on a MxPro - Mx3000P Multiplex qPCR System (Agilent Technologies). Thermal cycler profile for SYBR qRT-PCR was as follows: 50°C for 2 minutes, 95°C for 10 minutes, 40 \times [95°C for 15 seconds, 60°C for 1 minute].

Table 2.3 Taqman probes.

Gene	Catalogue#	Supplier
<i>CEBPA</i>	<i>Hs00269972_s1</i>	Applied Biosystems
<i>EPOR</i>	<i>Hs00959427_m1</i>	Applied Biosystems
<i>GATA1</i>	<i>Hs01085823_m1</i>	Applied Biosystems
<i>GATA2</i>	<i>Hs00231119_m1</i>	Applied Biosystems
<i>HPRT1</i>	<i>Hs02800695_m1</i>	Applied Biosystems
<i>KAT2A</i>	<i>Hs00221499_m1</i>	Applied Biosystems
<i>KLF1</i>	<i>Hs00610592_m1</i>	Applied Biosystems
<i>MPO</i>	<i>Hs00924296_m1</i>	Applied Biosystems
<i>CCDC101</i>	<i>Hs00540812_m1</i>	Applied Biosystems
<i>ID2</i>	<i>Hs04187239_m1</i>	Applied Biosystems
<i>MED13</i>	<i>Hs00192834_m1</i>	Applied Biosystems
<i>RPL15</i>	<i>Hs04334752_g1</i>	Applied Biosystems
<i>RPL3</i>	<i>Hs01581771_g1</i>	Applied Biosystems
<i>TAL1</i>	<i>Hs01097987_m1</i>	Applied Biosystems

Table 2.4 Sequences of primers used for qRT-PCR analysis.

Gene	Forward	Reverse
<i>EPOR</i>	<i>GAGCATGCCAGGATACCTA</i>	<i>TACTCAAAGCTGGCAGCAGA</i>
<i>GATA2</i>	<i>TGTAGTTCCTGCCCTCTCT</i>	<i>CCGACTCCCAGACCGTTCC</i>
<i>HBB</i>	<i>AGGAGAAGTCTGCCGTTACTG</i>	<i>CCGAGCACTTTCTTGCCATGA</i>
<i>HOXA10</i>	<i>GAGAGCAGCAAAGCCTCGC</i>	<i>CCAGTGTCTGGTGCTTCGTG</i>
<i>HOXA9</i>	<i>GGTGACTGTCCCACGCTTGAC</i>	<i>GAGTGGAGCGCGCATGAAG</i>
<i>HPRT1</i>	<i>CCTGGCGTCGTGATTAGTGAT</i>	<i>TCGAGCAAGACGTTTCAGTCC</i>
<i>JARID2</i>	<i>AGAGGAGGAAGATGCGCTCG</i>	<i>GGAATCCCATCACTGTCATCCTGG</i>
<i>KAT2A</i>	<i>CCCGCTACGAAACCACTCAT</i>	<i>GCATGGACAGGAATTTGGGGA</i>
<i>MYC</i>	<i>CACTCTCCCTGGGACTCTTG</i>	<i>TCTCCCTTTCTCTGCTGCTC</i>
<i>RPL13</i>	<i>CGCAGGAGCCGCAGG</i>	<i>CTGCCAGTCCTTGTGGAAGT</i>

<i>RPS7</i>	<i>CCCAGGAGCCGTA</i>	<i>GCCATCTAGTTTGACGCGGA</i>
<i>SUPT20H</i>	<i>CCCTTAAATCTACTCCAGCTTCCAG</i>	<i>TTGACTGGTTGAACCTTGCTC</i>
<i>TADA2A</i>	<i>CGCCTTAAACGCACTATGCTC</i>	<i>GGCCGGAATCAATGTCAGCTT</i>
<i>TADA2B</i>	<i>GCTACCACGGCTACCAGC</i>	<i>AGCCATATCTTCCCAGTTTCCG</i>
<i>USP22</i>	<i>GAGGCCATGGACGCCG</i>	<i>AGATACAGGACTTGGCCTTGC</i>
<i>YEATS2</i>	<i>AACCCAGGGTCTGAATTTATTG</i>	<i>CGGACGCATACACGTTCT</i>
<i>ZZZ3</i>	<i>GGACAGCAAAACAGGTTGCC</i>	<i>GTGCTGTCGTCTGCTTGTTG</i>

2.8 RNA-sequencing (RNA-seq)

RNA was extracted from *CTRLsh* or *KAT2Ash* HSCs obtained from 2 individual CB donors as described in 2.7.2. RNA-seq libraries were prepared at the Cambridge Stem Cell Institute Genomics Core Facility using the Ovation RNA-seq kit (NuGen) with incorporated DNase treatment, as per manufacturer's instructions. Libraries were sequenced on an Illumina HiSeq4000 instrument at the CRUK Cambridge Research Institute Genomics Core Facility using 50bp single-end reads. The raw fastq files were processed as per the RSEM v1.2.31 workflow and aligned to the reference human genome assembly GRCh37. Differentially expressed genes were obtained at 10% FDR using the R package edgeR. Gene signatures of down-regulated genes in *KAT2Ash* HSC for gene set enrichment analysis were obtained from MSigDB [268]. Erythroid differentiation (NES= -1.79, q-val=0.018) and platelet biology (NES= -2.05, q-val=0.004).

2.9 Chromatin Immunoprecipitation (ChIP)

2.9.1 Cross-link for Suspension Cells, Cell Lysis

Cross-linking of 1×10^7 suspension cells was performed with 278 μ L 37% v/v Formaldehyde (Sigma) per 10mL of PBS. 571 μ L of 2.5M Glycine were added and samples incubated for 10 minutes, 50rpm, at RT prior to centrifugation at 2,000g for 5 minutes at 4°C. Supernatant was discarded and 1mL ice-cold PBS added per 1×10^7 , with gentle shaking to re-suspend pellets. Samples were further aliquoted into 1.5mL tubes (Eppendorf) and centrifuged at 2,000g for 5 minutes at 4°C.

Supernatant were removed and cross-linked pellets stored at -80°C or used straight after.

Lysis buffer was added to either frozen or freshly cross-linked pellets such that the final concentration of the cell suspension was 15×10^6 cells/mL. Chromatin was sheared using 1.5mL TPX tubes (Diagenode) to a medium fragment size of 300bp on a Bioruptor XL (Diagenode). Sonication was performed at 4°C and conditions were optimised for each particular cell type: 9 sonication cycles, 30 seconds ON/30 seconds OFF for K562 cells and MOLM-13 cell lines. Optimisation of sonication conditions is shown in **Fig. 2.12**.

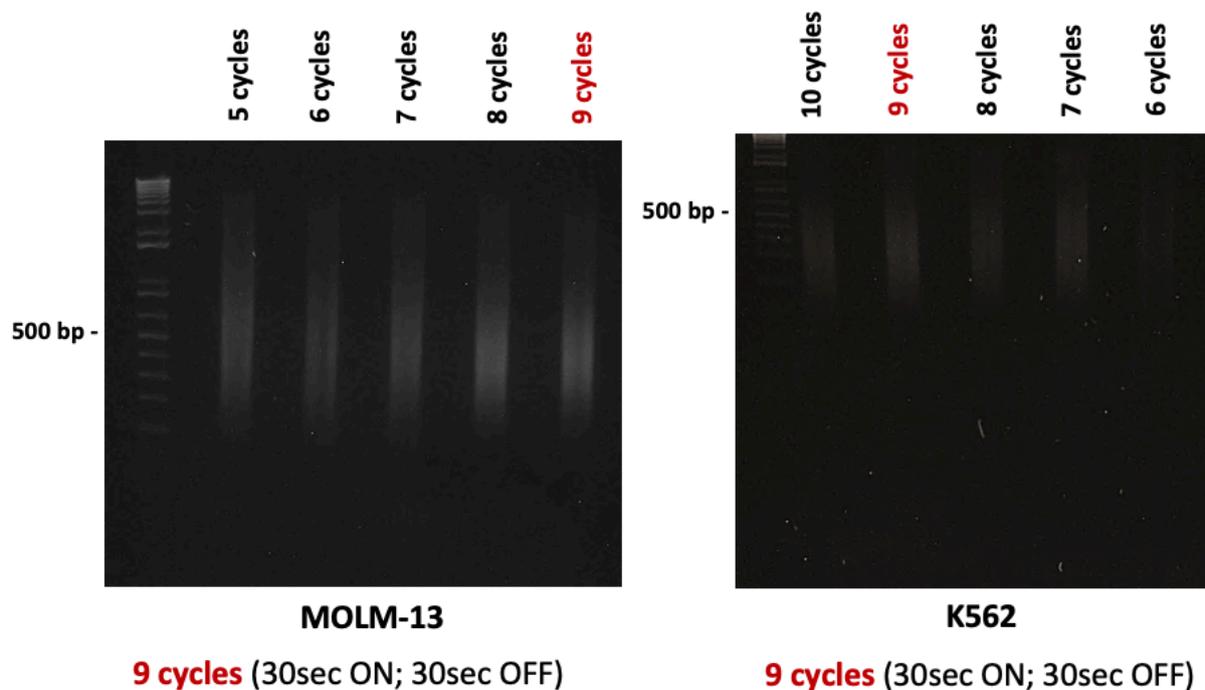


FIGURE 2.12 Assessment of optimal chromatin sonication conditions for MOLM-13 (left) and K562 (right) cells using agarose gel electrophoresis. The number of cycles producing fragment sizes of about 300bp was selected.

Following sonication, chromatin was transferred to regular 1.5mL tubes and samples span down for 5 minutes at 13,000rpm at RT. 2% of supernatant was saved as input and stored at -80°C.

2.9.2 Immunoprecipitation

For ChIP-qPCR, chromatin was immunoprecipitated in dilution buffer, protease inhibitor cocktail (PIC, Sigma) and with the antibodies listed in **Table 2.6**, in a total reaction volume of 3mL for H3K9ac, H3K27ac, rabbit-IgG, and 10mL for Myc, mouse-IgG. Chromatin was incubated 4 h-overnight with rotation at 4°C. The following day, ChIP Grade Protein A/G beads (Thermo) were prepared by washing with 500µL of dilution buffer (+ 0.15% sodium dodecyl sulfate (SDS), + 0.1% BSA). Beads were span down, placed in a magnetic rack (Diagenode) for 1 minute and supernatant removed. Beads were then dissolved in 15-20µL dilution buffer (+0.15% SDS, +0.1% BSA) per ChIP reaction and added to the chromatin/antibody mix, followed by incubation for 1h, with rotation at 4°C. Washes (consisting of 1x wash buffer 1; 2x wash buffer 2; 2x wash buffer3) were performed with 2mL of each buffer for 5 minutes on ice, with short spin down steps in between washes. Samples were eluted in 200µL of freshly prepared elution buffer with 5M NaCl and Proteinase K (Biolone) and left rotating for 20 minutes at RT together with inputs retrieved from -80°C storage. Samples were briefly span down, placed in the magnetic rack for 1 minute and supernatants collected. Cross-links were reversed overnight at 65 °C, 1,000rpm.

2.9.3 DNA purification

DNA was purified with a DNA Clean & Concentrator kit (Zymo Research) as per manufacturer's instructions. Eluted DNA was diluted and quantified by SYBR green qPCR (see 2.6.3) using 2µl DNA per triplicate reaction (primers in **Table 2.7**). Peak enrichments relative to rabbit IgG or mouse IgG were determined using the $-2^{\Delta\Delta Ct}$ method with a reference gene or reference intergenic region. Detailed ChIP buffers/solutions are listed in **Table 2.5**.

For ChIP-sequencing (ChIP-seq), chromatin was prepared in duplicate from K562 cells (9 sonication cycles, 30 seconds ON/30 seconds OFF) and immunoprecipitated using anti-SPT20 and anti-ZZZ3 sera prepared in the Tora Lab [245]. Raw ChIP-seq reads were analysed on the Cancer Genomics Cloud (CGC) platform [269].

Table 2.5 Buffers used in chromatin immunoprecipitation (ChIP).

Lysis Buffer Stock	Final concentration	Volume stock solution (10mL)
1M	20mM HEPES	0.25
10%	1% SDS	1
PIC	1:100	0.1
H2O		Fill up to 10ML
Dilution Buffer Stock	Final concentration	Volume stock solution (100mL)
10%	0.15% SDS	1,5
20%	1% Triton	5
0.5M	1.2mM EDTA	0.24
1M	16.7mM Tris (pH 8.0)	1.76
5M	167mM NaCl	3.34
H2O		Fill up to 100ML
Wash Buffer 1 Stock	Final concentration	Volume stock solution (400mL)
0.5M	2mM EDTA	1.6
1M	20mM Tris (pH 8.0)	8
20%	1% Triton	20
10%	0.1% SDS	4
5M	150mM NaCl	12
H2O		Fill up to 400ML
Wash Buffer 2 Stock	Final concentration	Volume stock solution (400mL)
0.5M	2mM EDTA	1.6
1M	20mM Tris (pH 8.0)	8
20%	1% Triton	20
10%	0.1% SDS	4
5M	500mM NaCl	40
H2O		Fill up to 400ML
Wash Buffer 3 Stock	Final concentration	Volume stock solution (1000mL)
0.5M	1mM EDTA	2

1M	10mM Tris (pH 8.0)	10
H2O		Fill up to 1000ML
Elution Buffer Stock	Final concentration	Volume stock solution (5mL)
10%	1% SDS	5
1M	0.1M NaHCO ₃	5
H2O		Fill up to 50ML

Table 2.6 Antibodies used in chromatin immunoprecipitation (ChIP).

Antibody	Catalogue#	Supplier
H3K27ac	Ab4729	Abcam
H3K9ac	07-352	Millipore
Myc	sc-764 x	Santa Cruz Biotechnology
rabbit IgG	12-370	Millipore
mouse IgG	Sc-2025	Santa Cruz Biotechnology

Table 2.7 Sequences of primers used for ChIP-qPCR analysis.

	Forward	Reverse
<i>Eif4e</i>	GCAGACCACATCAACGACTCT	TCTTTTCGCCTCCCACCATT
<i>EPOR</i>	TGGCACATAGCGAACATTCCA	GGCTGGGAAGAGAATGCTGATT
<i>Gadd45g</i>	GGCATCGACTCTGACCTTGT	CGCTATGTCGCCCTCATCTT
<i>HBB</i>	GCCATCCATTTTCTTAATTCTGAG	TGAGGGCACCATTAGCCAG
<i>HOXA10</i>	GTTTATAGCGGCGCATTCCA	CGGGTTTGATTTCTGAGCCC
<i>HOXA9</i>	CGCTCTCATTCTCAGCATTG	TTAAACCTGAACCGCTGTCTG
<i>Intergenic region</i>	TGGTTTGGAGTGGGTGCT	TCCTGCTCTCCGTCACCT
<i>JARID2</i>	CATGCATTGAGCTAGGGAGCC	AACCAAGACACCCGAGCGAC
<i>Kdelr2</i>	CCTTGAGTGTGGCCGTCTAA	TCAATGGTGACGTGGAGCAA
<i>KRT5</i>	AGGTTGTAGAGGCTCCGGCT	CAGCTTCACCTCCGTGTCCC
<i>mChr1 (intergenic)</i>	CATAGATGAAGCTGCCACATAGGT	GTGGGCAAGGACAAAGCATT

<i>MEIS1</i>	<i>CCAGAAGAAGACAGAGCGGA</i>	<i>CCCTCAGACCCAACTACCAA</i>
<i>MYC</i>	<i>TACAACACCCGAGCAAGGAC</i>	<i>AGAGCCTTTCAGAGAAGCGG</i>
<i>Pcbp1</i>	<i>AGAGCGCCTTGTGCTTTCTT</i>	<i>CTGGTCCTTTCGGCCAAGTA</i>
<i>Ralbp1</i>	<i>GTGTTGACTTGCGGGAAACT</i>	<i>GCGGCTTAACTCGGGTATG</i>
<i>RPS7</i>	<i>CCTGCTCTCCGACAGAACTT</i>	<i>CGGGTAATCGGCTGTATCCC</i>

2.10 Western Blot (WB)

2.10.1 Sodium Dodecyl Sulphate – PolyAcrylamide Gel Electrophoresis (SDS-PAGE) Preparation

For preparing the running gel, casting frames (Bio-Rad) were set on the casting stands (Bio-Rad) and two glass plates (Bio-Rad) clamped into the casting frames. Running gel ((30% (w/v) Acryl/Bis Solution, 1M Tris-HCl pH 8.8, (w/v) 10% SDS, ddH₂O, 10% (w/v) ammonium persulfate (APS), Tetramethylethylenediamine (TEMED)), was poured into the gap between the glass plates. Isopropanol was added into the gap until overflow, in order to make the top of the running gel even. After letting the running gel solidify for about 40 minutes, the stacking gel (30% (w/v) Acryl/Bis Solution, 0.5M Tris-HCl pH 6.8, (w/v) 10% SDS, ddH₂O, 10% (w/v) APS, TEMED) was prepared and poured above the running gel until overflow. A 15mm 15-well gel comb (Bio-Rad) was inserted and following complete solidification of the stacking gel (about 30 minutes), the comb and glass plates were removed. Running buffer 1x (Bio-Rad) was poured into the electrophoresis tank (Bio-Rad) until reaching the required level for the number of gels being run.

2.10.2 Sample Preparation and WB Analysis

For western-blot analysis, collected cells were lysed in RIPA buffer (50mM Tris-HCl pH 7.4, 150mM sodium chloride, 2mM EDTA, 1% (v/v) IGEPAL CA-630, 0.5% (w/v) sodium deoxycholate, 0.1% (w/v) SDS and protease inhibitor (Roche) for 20 minutes at 4°C with rotation. Protein extracts from equal cell numbers (1x10⁶ cells/extract) were mixed with Trident 6x Laemmli SDS sample buffer (GeneTex) and reducing

agent Tris (2-carboxyethyl) phosphine - TCEP 0.5M (Sigma) and denatured at 95°C for 5 minutes or at 70°C for 10min. Proteins were then resolved by SDS-PAGE (cast to a final acrylamide percentage between 7.5-10%) using a tris/glycine/SDS buffer system (Bio-Rad) for about 1h at 120V. Precision Plus Protein Kaleidoscope (Bio-Rad) was used as standard. Subsequently, transfer membranes (Immobilon-P PVDF pore size 0.45 µm, Millipore) were activated with Methanol (Sigma) for 5 minutes. Gels were transferred onto the transfer membrane for about 90 minutes at 100V. Membranes were stained with Ponceau for 1 minute, blocked in blocking buffer (5% skimmed milk in TBS-T (Tris 10 mmol/L, NaCl 50 mmol/L, Tween 0.005%)) for 1h at RT, washed and immunoblotted with primary antibodies diluted in blocking buffer overnight, at 4°C with rotation. The following day, membranes were washed in TBS-T 3x for 10 minutes, and incubated with the appropriate fluorescently-labelled secondary antibody diluted in 1.5% skimmed milk in TBS-T for 1h at RT. Membranes were washed 2x in TBS-T and 1x in TBS for 10 minutes, subjected to 5 minute SuperSignal West Pico PLUS Chemiluminescent Subtract (Thermo) for detection of HRP, and developed on a SRX-101A Processor using the hypercassette method. After drying, membranes were taped onto a hypercassette (Amersham Biosciences) and exposed to X-ray film (exposure dependent on the antibody strength) prior developing. WB antibodies used are listed in **Table 2.8**.

Table 2.8 Antibodies used in Western Blot (WB).

Antibody	Catalogue #	Clone	Dilution	Supplier
Cas9	C15200203-100	Mouse monoclonal	1:6000	Diagenode
CHEK1	EP691Y	Rabbit monoclonal	1:10 000	Abcam
Gcn5	ABE1934	Rabbit polyclonal	1:1000	Millipore
Gcn5 (A11)	sc-365321	Mouse monoclonal	1:1000	Santa Cruz
Gcn5 (H75)	sc-20698	Rabbit polyclonal	1: 2000	Santa Cruz
ACTB	ab8227	Rabbit polyclonal	1:1000	Abcam
CCDC101	ab204367	Rabbit polyclonal	1:1000	Abcam
USP22	ab195289	Rabbit monoclonal	1:2000	Abcam

2.11 Statistical Analysis

Statistical analysis was performed in GraphPad Prism 8.1.2 (GraphPad Software). Data are reported as mean \pm SEM. Significance calculated by 2-tailed t-test at $p < 0.05$, details are specified in the figure legends.

2.12 Data Deposition

ChIP-seq and RNA-seq data have been deposited in GEO (accession numbers are GSE128902 and GSE128512).

2.13 Resources

<https://software.broadinstitute.org/software/igv/>

<https://biorender.com>

<https://www.snapgene.com>

<https://cellline.molbiol.ox.ac.uk/eryth/index.html>

https://kleintools.hms.harvard.edu/paper_websites/tusi_et_al/

Chapter 3

The role of KAT2A and its chromatin complexes in normal haematopoiesis

3.1 Introduction

Our lab has previously shown that chemical inhibition of KAT2A activity had not impacted colony-formation from total CB CD34⁺ cells, suggestive of preservation of healthy blood development [260]. Moreover, our lab and the Tenen group did not find a requirement for *Kat2a* in mouse BM haematopoiesis [196]. However, re-inspection of *Kat2a* expression in BM subpopulations of our conditional KO model found that gene expression ablation in MEP was less extensive than in myelo-monocytic cells. Mean *Kat2a* expression in *Kat2a* KO MEP cells is 0.39 ± 0.20 (SD) of WT expression. However, it is 0.13 ± 0.08 in GMP, compatible with selective preservation of unexcised *Kat2a* allele-carrying MEP cells, and a possible partial requirement for *Kat2a* expression (n=4; 2-tailed t-test p-val = 0.054). Thus, I went on to investigate in detail if KAT2A was required in normal human haematopoietic progenitor specification or differentiation. For that, I combined functional and transcriptomic analysis of human CB, a well-established model of human blood development [1]. I exploited RNA interference (RNAi) - shRNA-mediated lentiviral delivery - to silence *KAT2A* expression and assess phenotypes of functional, transcriptomic and ChIP-seq assays.

In vitro functional studies of cell lineage potential can be done by two main strategies, employed here: (1) CFC-assays and (2) liquid differentiation cultures. CFC assays consist in plating cell populations in semi-solid methylcellulose-based media, which viscosity allows cells to grow separately as clonal entities that can be scored visually. Both stem and progenitor cells are able to form colonies in a CFC assay, as it measures differentiation and proliferative potential and not self-renewal (which can be assessed by serial re-plating). In liquid differentiation culture assays, lineage output is assessed by flow cytometry, providing more mechanistic detail into proliferative capacity of bulk populations.

RNA-seq was used in order to detect transcripts showing differential expression in control (CTRLsh) vs *KAT2Ash* CB HSCs, thus aiming to understand transcriptional programmes regulated by KAT2A in normal blood. Briefly, RNA-seq involves isolation of RNA, preparation of RNA-seq library (reverse-transcribing the RNA to cDNA, fragmenting or amplifying cDNA molecules, ligating sequencing adaptors), and sequencing it on an NGS platform (Kukurba and Montgomery, 2015).

Through ChIP-seq, a technique for genome-wide mapping of protein-DNA interactions, unbiased examination of the DNA binding sites of the KAT2A-chromatin complexes, ATAC and SAGA in haematopoietic K562 cells was performed. Specifically, by targeting unique subunits of either complex, ZZZ3, in case of ATAC and SPT20, in case of SAGA. K562 is a human cell line derived from a chronic myelogenous leukaemia patient in the blastic phase, in 1970 [270]. These cells have the capacity to differentiate towards the erythroid lineage upon treatment with DMSO, and thereby, can be used as a surrogate for normal haematopoietic development [271–272].

3.2 Chapter overview

This chapter details the results of experiments designed to investigate the role of acetyltransferase KAT2A to cell fate decisions in normal human haematopoiesis. Also, it dissects contributions of the two multimolecular complexes in which KAT2A participates, ATAC and SAGA. First, I describe the functional and molecular characterisation of KAT2A in normal human CB blood specification. By combining functional and transcriptomic (RNA-seq) analysis, I show that KAT2A is required for specification and survival of erythroid progenitor cells. I further demonstrate that lymphoid and myeloid lineages are not affected upon KAT2A loss. Additionally, KAT2A controls a common erythroid and platelet transcriptional program in blood stem and multipotent progenitor cells. Dissection of the macromolecular complex dependencies in normal blood production revealed that ATAC is selectively required for erythroid lineage specification, while SAGA may act later in erythroid cell differentiation, and more specifically contribute to cell identity. Finally, I demonstrate that ATAC and SAGA have unique sets of target genes in haematopoietic cells: ATAC controls biosynthetic activity including ribosomal protein machinery, while SAGA regulates transcription.

3.3 Results and discussion

3.3.1 *KAT2A* regulates human CB erythroid progenitor specification and survival

I started by obtaining human CB from healthy donors, and isolating CD34+ cells, which were then transduced with a *KAT2A* shRNA construct. Transduced (GFP+) cells were flow-sorted for stem and multipotent progenitor cells (HSC), lineage-restricted myelo-lymphoid (MLP), megakaryocytic-erythroid (MEP) and granulocytic-monocytic (GMP) progenitors (**Fig. 3.1**). Experiments were performed against a control shRNA that does not target eukaryotic genes (CTRLsh).

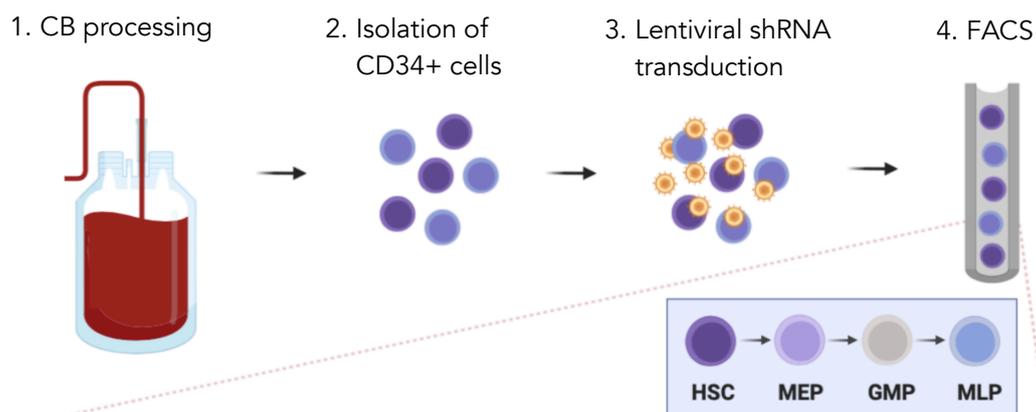


FIGURE 3.1 Schematic of the adopted CB experimental strategy.

I initially validated the *KAT2A* knockdown by WB, which showed robust reduction of protein levels (**Fig. 3.2A**). Inspection of progenitor activity of *KAT2A* knockdown CB CD34+ cells in CFC assays (**Fig. 3.2B**) detected a reduction in erythroid (E) colony formation (**Fig. 3.2C**). This was also apparent upon depletion of the Tudor-domain protein SGF29 (gene *CCDC101*) (**Fig. 3.2D**), a *KAT2A* partner common to its participating SAGA and ATAC complexes, and which KD was confirmed by WB in erythroid-affiliated K562 cells (**Fig. 3.3E**). Equally, differentiation of *KAT2A* knockdown CD34+ cells in mixed-lineage differentiation cultures revealed the relative loss of differentiated erythroid cells (**Fig. 3.3F**). This was evident by reduced percentage of CD235+ cells in *KAT2A*sh cells after seven days in culture, compatible with a defect in specification, differentiation, or survival of the red blood cell lineage.

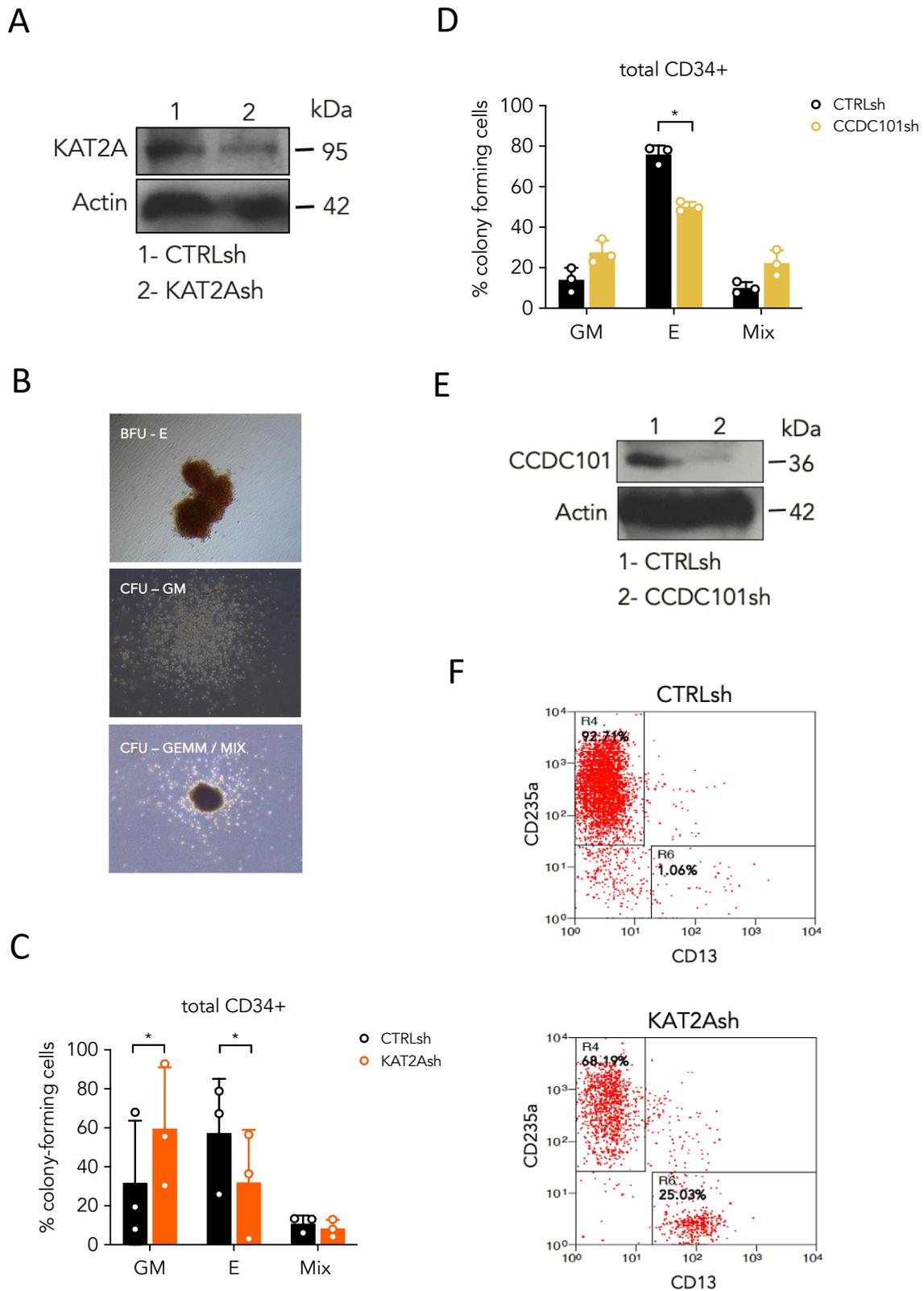
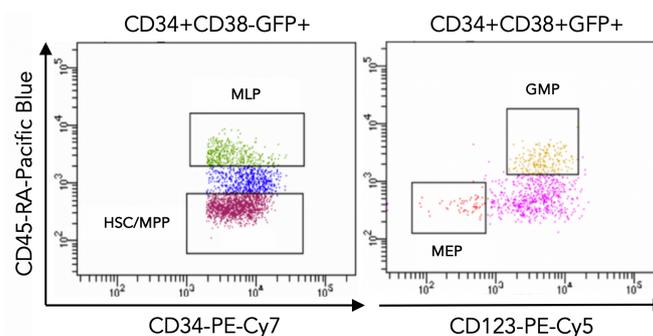


FIGURE 3.2 Progenitor activity of *KAT2A* knockdown CB CD34⁺ cells. (A) Western blot analysis showing *KAT2A* (95kDa) knockdown efficiency in erythroid-affiliated K562 cells. ACTB (42kDa) was used as loading control. **(B)** Photographs of representative colony morphologies from transduced human

CB HSC and progenitors. BFU-E: Boost-forming unit – erythrocyte colony. CFU-GM: Colony-forming unit - granulocyte monocyte colony; CFU-GEMM or mixed: Colony-forming unit – granulocyte erythrocyte monocyte megakaryocyte colony. **(C)** Relative representation of colony-forming cells present in total CB CD34+ cells transduced with *KAT2Ash*. Mean \pm SEM of 3 individual CB samples. Paired two-tailed t-test for significance * $p < 0.05$. **(D)** Relative representation of colony-forming cells present in total CB CD34+ cells transduced with *CCDC101sh*. Mean \pm SEM of 3 individual CB samples. Paired two-tailed t-test for significance * $p < 0.05$. **(E)** Western blot analysis showing CCDC101 (36KDa) knockdown efficiency in erythroid-affiliated K562 cells. ACTB (42KDa) was used as loading control. **(F)** Flow cytometry analysis of CB HSC transduced with *CTRLsh* (top) or *KAT2Ash* (bottom) after 7 days of liquid culture in erythroid differentiation conditions. Cells plotted were gated as live GFP+ (transduced) CD34-. Data are representative of 3 independent experiments.

When investigating the CFC capacity of HSC and the different progenitor compartments upon KAT2A knockdown, at the point of sorting (**Fig. 3.3A**), I observed a significant decrease in the proportion of KAT2A-depleted MEP in total CD34+ CB cells (**Fig. 3.3B**), suggesting a defect in the specification, differentiation or survival of progenitors committed to the erythroid lineage, in agreement with the re-analysis of BM from the KAT2A KO mouse model. Moreover, I verified that loss of *KAT2A* affected erythroid lineage output from both HSC and MEP (**Fig. 3.3, top**). CFC ability of progenitors affiliated to the granulocyte-monocyte and lymphoid lineages, in turn, was unaffected (**Fig. 3.3, bottom**).

A



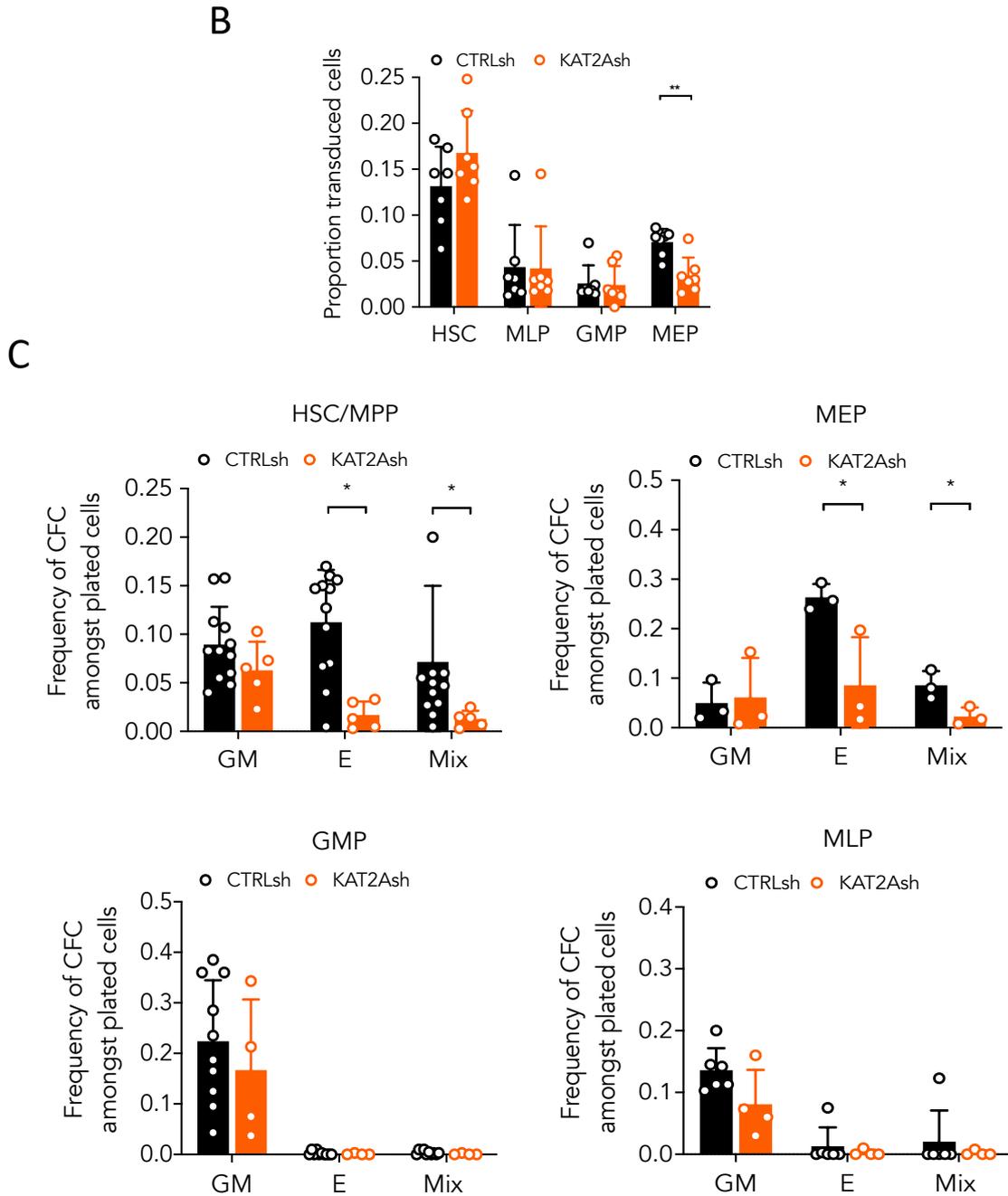
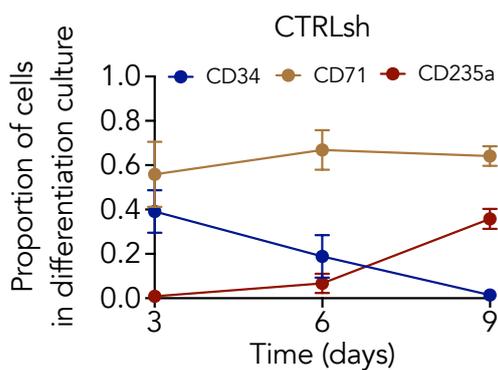


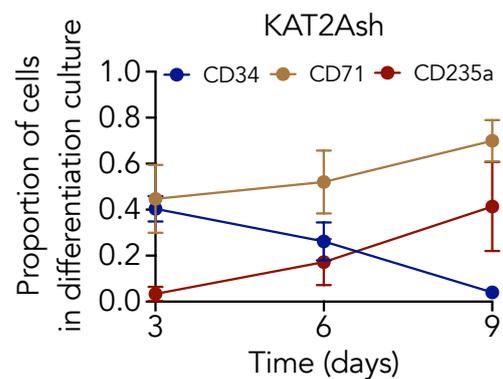
FIGURE 3.3 CFC-efficiency of KAT2A lentiviral transduced, sorted, human stem (HSC/MPP) and progenitor (MEP, GMP, MLP) CB cell populations. (A) Representative sorting plot for transduced HSC and progenitor cells from CB CD34⁺ cells. **(B)** Proportion of *KAT2Ash* transduced CB HSC and progenitors. Mean \pm SEM of >7 individual sorting experiments. Two-tailed paired t-test for significance; * $p < 0.05$, ** $p < 0.01$. **(C)** Frequency of CFC efficiency in the HSC/MPP, MEP (top), GMP and MLP (bottom) compartments transduced with *KAT2Ash*. Mean \pm SEM of 7 individual CB samples. Two-tailed paired t-test for significance; * $p < 0.05$, ** $p < 0.01$.

CFC assays, however, are end-point experiments that do not permit a detailed dissection of mechanisms of cell specification or differentiation. As an attempt to overcome this limitation, a liquid culture system was used over the course of 9 days. Differentiation of HSC/MPP transduced with *CTRLsh* in the presence of hydrocortisone, SCF, TPO and EPO resulted in progressive loss of stemness marker CD34 and accumulation of cells expressing glycophorin A (CD235a), a marker characteristic of late erythroid maturation stages. In turn, the level of CD71+ (early erythroid marker) cells remained relatively constant (**Fig. 3.4A**). This result indicated effective progression of erythroid differentiation. When analysing the results of *KAT2Ash* cultures, it was possible to observe a highly similar trend (**Fig. 3.4B**), which culminated in a low proportion of CD34+ cells and high proportion of CD235a+ cells at Day9. Furthermore, cultures initiated from *KAT2A*-depleted HSC cells were severely restricted in their expansion (**Fig. 3.4C**) compared to control, with relative preservation of GFP positivity in both conditions, at least in the experiment where GFP % was assessed (**Fig. 3.4D**). However, the achievement of similar end-point levels of CD235a-positivity in (A) and (B), suggests that erythroid differentiation can progress unrestricted in the absence of *KAT2A*.

A



B



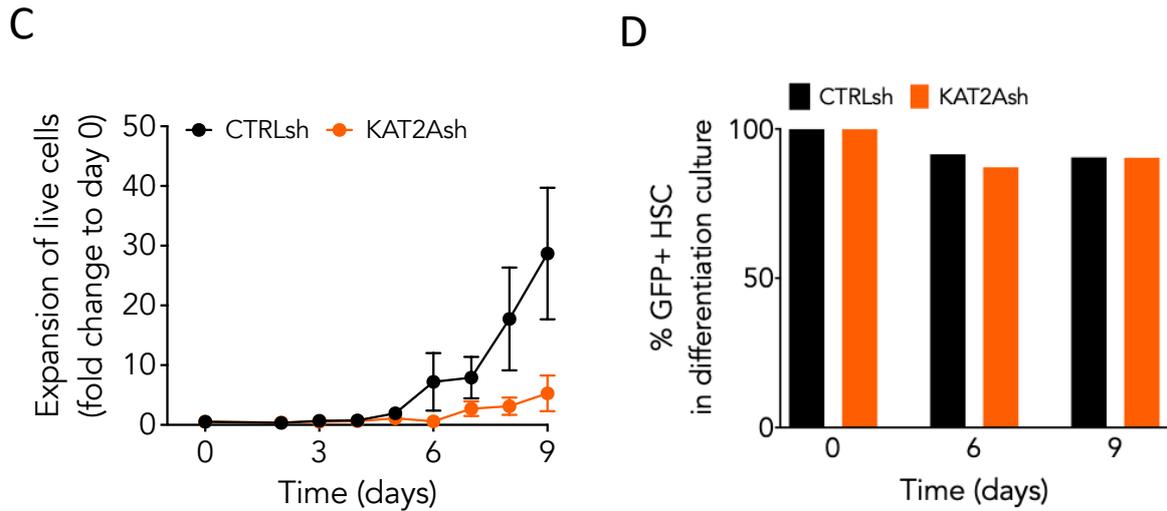


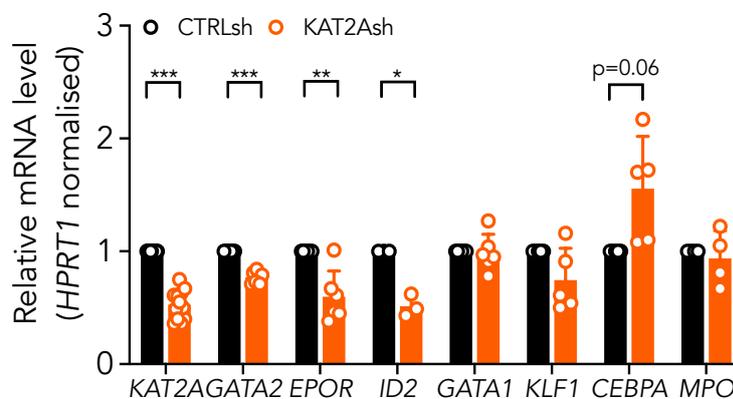
FIGURE 3.4 Erythroid cell differentiation in human CB HSCs progresses independently of KAT2A absence. **(A)** Erythroid differentiation of HSC/MPP sorted from CD34+ cells of individual CB samples transduced with *CTRLsh* in liquid cultures. **(B)** Erythroid differentiation of HSC/MPP sorted from CD34+ cells of individual CB samples transduced with *KAT2Ash* in liquid cultures. CD34 marks HSC and progenitors; CD71 and CD235a mark early and late differentiated erythroid cells, respectively. Data summarise mean \pm SEM of 4 independent differentiation experiments. **(C)** Expansion of *CTRLsh* and *KAT2Ash* HSC in erythroid differentiation cultures in (A). **(D)** GFP+ percentage in *CTRLsh* and *KAT2Ash* HSC in a representative erythroid differentiation culture experiment.

Together, the data suggests that cells that commit to the erythroid lineage can progress unrestricted in their differentiation. However, most *KAT2Ash* cells don't seem to get to that stage, in agreement of my earlier observations of impaired specification or survival of MEP.

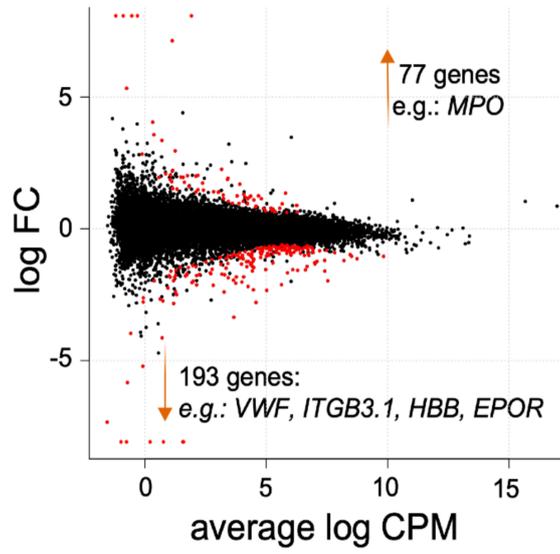
Quantitative RT-PCR analysis of *KAT2A*-knockdown HSC showed downregulation of early erythroid regulatory genes including *GATA2*, *EPOR* and *ID2* (**Fig. 3.5A**). Assessing for any unbiased differences in gene expression through RNA-sequencing in order to better characterise which transcriptional programmes were being subverted by loss of *KAT2A*, indicated a global reduction of erythroid and megakaryocytic programmes (**Fig. 3.5B**). This was in line with the outcome of CFC assays, portraying an erythroid defect. Indeed, amongst the down-regulated genes were erythroid-affiliated factors *EPOR*, *HBB*, *ID2*, *GFI1B*, *HBG1* and *HBG2*. Examples of up-regulated genes are *MPO* and *ERG*. The full list of differentially expressed genes is given in **Appendix A - RNA-seq differentially expressed genes in CTRLsh vs KAT2Ash HSCs (10% FDR)**.

Furthermore, gene ontology (GO) associations as calculated by GOrilla analysis, identified gas transport and blood coagulation as enriched biological processes (**Fig. 3.5C**). Likewise, gene Set Enrichment Analysis of down-regulated genes in *KAT2Ash* HSC revealed an enrichment for erythroid differentiation (**Fig. 3.5D, left**) and platelet biology (**Fig. 3.5D, right**), thus confirming a novel role for *KAT2A* in erythroid/megakaryocytic biology. The genes up-regulated upon *KAT2A* knockdown, in turn, did not capture any enriched GO categories. However, considering for example the case of *MPO*, it is possible that its upregulation is a consequence of *GATA2* downregulation, as previously demonstrated [273].

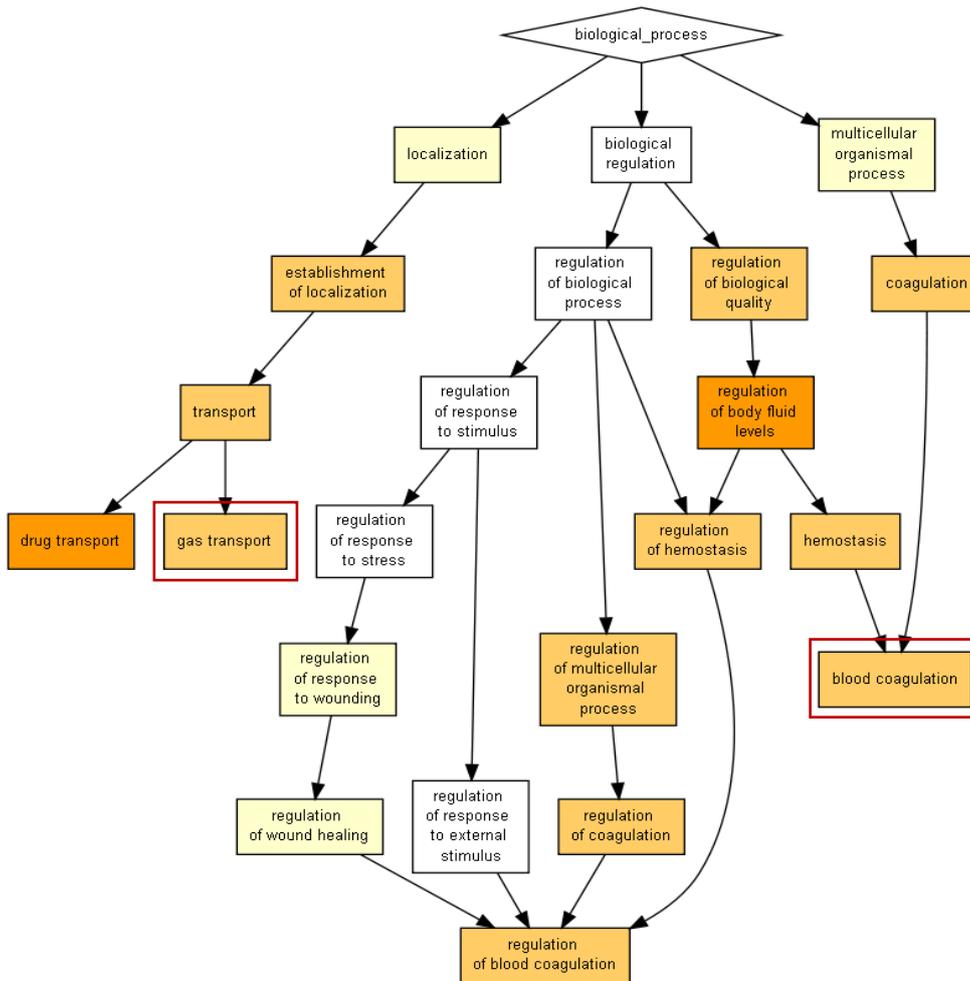
A



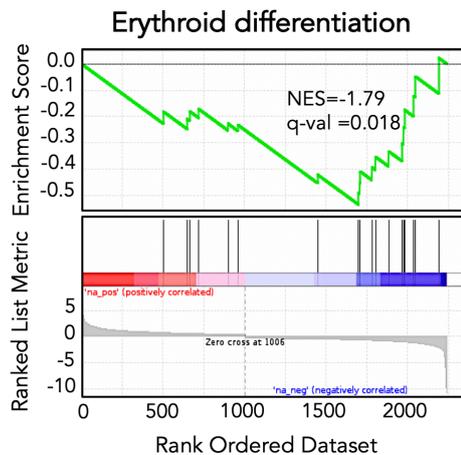
B



C



D



E

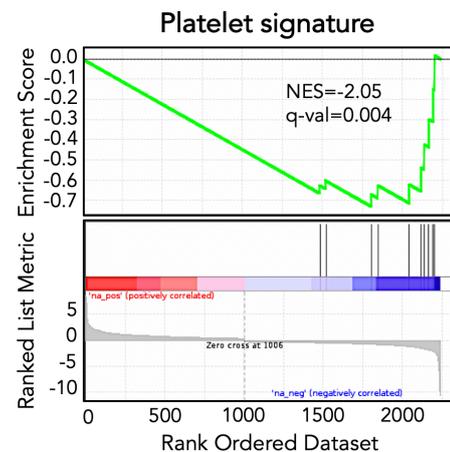
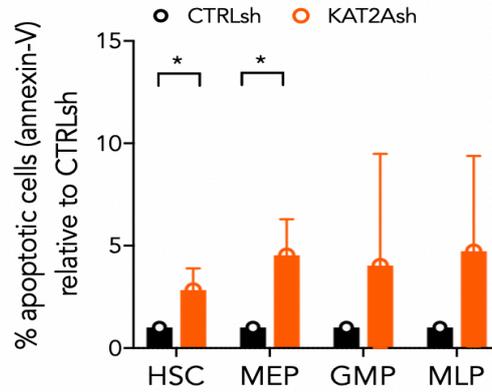


FIGURE 3.5 Transcriptional analysis of *CTRLsh* vs *KAT2Ash* CB HSCs. (A) Quantitative RT-PCR analysis of expression of erythroid-associated genes in *KAT2Ash* transduced HSC from individual CB samples. $N \geq 3$ independent experiments, mean \pm SEM of gene expression relative to *CTRLsh*, normalised to *HPRT1* housekeeping gene. Two-tailed paired t-test for significance * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **(B)** MA plot of RNA-seq gene expression analysis of differentially-expressed (DE) genes (red) in *CTRLsh* vs *KAT2Ash* transduced HSC/MPP cells at a false discovery rate of 10%. The average value of logCMP (counts per million) is plotted against the logFC (fold change). **(C)** Top Gene Ontology (GO) associations of *KAT2A* DE genes on biological processes as calculated by the GOrilla tool. **(D)** Gene Set Enrichment Analysis (GSEA) plot for erythroid differentiation (right) and platelet signature (left) in the RNA-seq data in (D). Gene signature identifiers are ADDYA Erythroid Differentiation by HEMIN and GNATENKO Platelet signature, respectively, as obtained from the UC San Diego and BROAD Institute Molecular Signatures Database (MSigDB).

To further understand the erythroid defect observed, I conducted cellular investigation by analysing the apoptosis and cell cycle status of *KAT2A* sorted, transduced (GFP+) stem and progenitor populations. Quantification of apoptotic cells revealed specific increases in apoptotic HSC and MEP, suggestive of putative effects on cell survival peri-erythroid commitment, and which did not extend to other lineages (**Fig. 3.6A**). This is in line with a reduction in erythroid and megakaryocytic-associated gene signatures in *KAT2A*-knockdown HSC (**Fig. 3.5**). There were no changes in cell cycle (**Fig. 3.6B**).

A



B

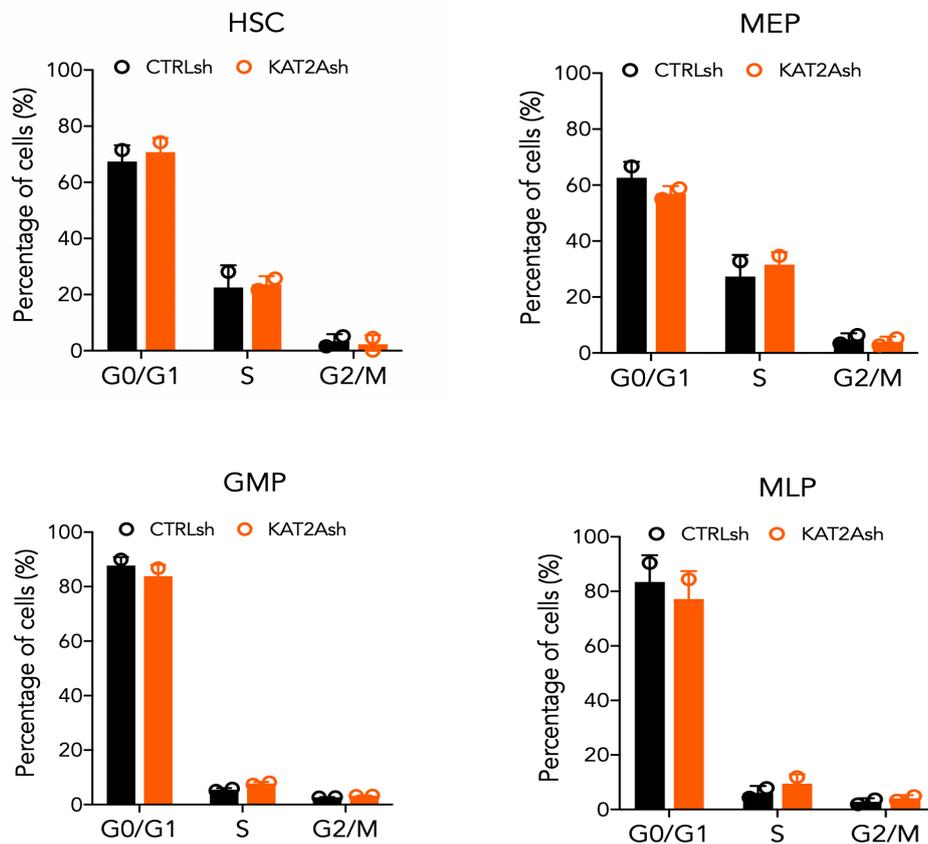


FIGURE 3.6 Analysis of apoptosis and cell cycle status of human CB stem and progenitor cells upon depletion of *KAT2A*. (A) Flow cytometry analysis of apoptosis by Annexin-V staining. N=3 individual cord blood samples; mean \pm SEM. Two-tailed paired t-test for significance; * $p < 0.05$. (B) Flow cytometry analysis of cell cycle in HSC/MPP, MEP, GMP and MLP transduced cells with KAT2Ash vs CTRLsh. N=3 individual cord blood samples; mean \pm SEM. Two-tailed paired t-test not significant.

Altogether, these observations support a function of KAT2A in human CB red blood cell specification and/or survival, with no impact on lymphoid or myeloid potential, sustaining a shared erythroid and platelet transcriptional programme.

Given that KAT2A is present in two distinct macromolecular complexes, it leaves open the possibility of ATAC or SAGA being specifically required in erythroid development. I thus decided to explore this hypothesis next.

3.3.2 The ATAC complex is selectively required for erythroid specification from CB HSC, while SAGA facilitates erythroid differentiation post-commitment

To dissect KAT2A-complex dependencies in normal haematopoiesis, I transduced CD34⁺ CB cells with ATAC and SAGA specific shRNA constructs, *ZZZ3sh* and *SUPT20Hsh*, respectively. CD34⁺ CB cells transduced (GFP⁺) with *ZZZ3* or *SUPT20H* shRNAs (**Fig. 3.7A**) were flow-sorted into stem and multipotent progenitor cells (HSC) and lineage-restricted myelo-lymphoid (MLP), megakaryocytic-erythroid (MEP) and granulocytic-monocytic (GMP) progenitors as described (**Fig. 3.3A**). Upon sorting, no differences in the proportion of stem and progenitor cells were observed in CB CD34⁺ cells transduced with either construct (**Fig. 3.7B-C**).

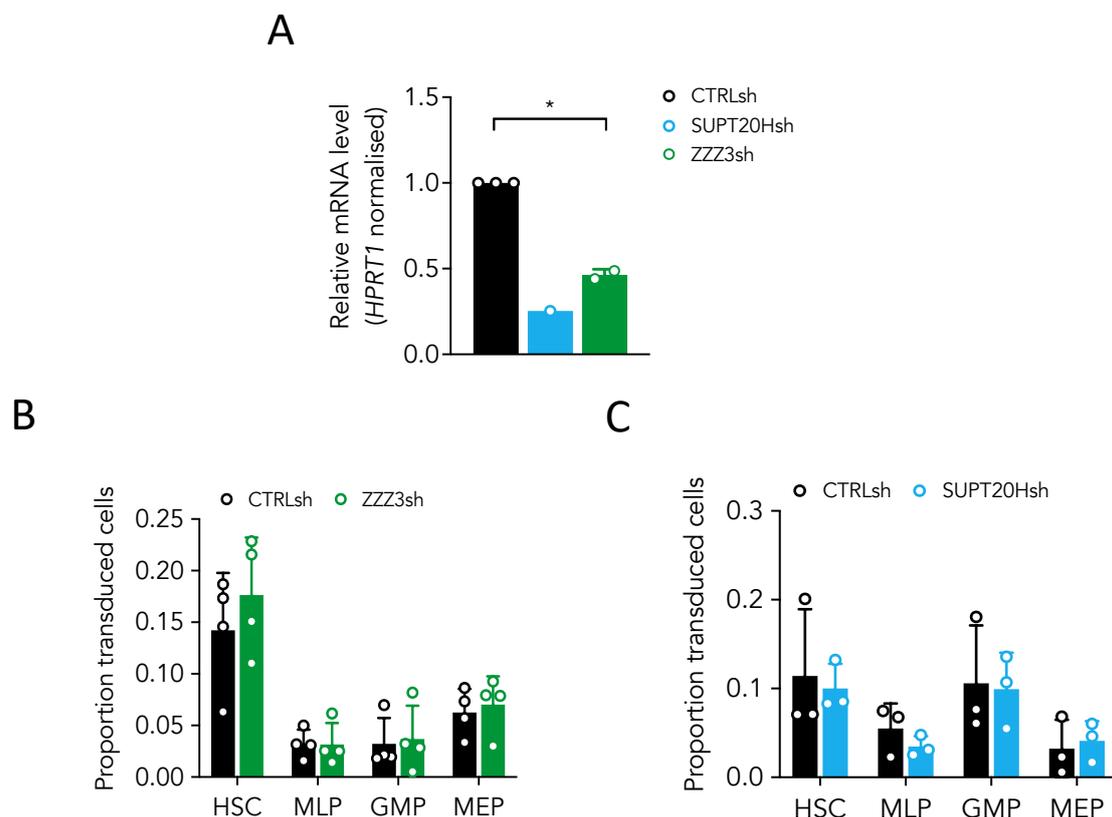
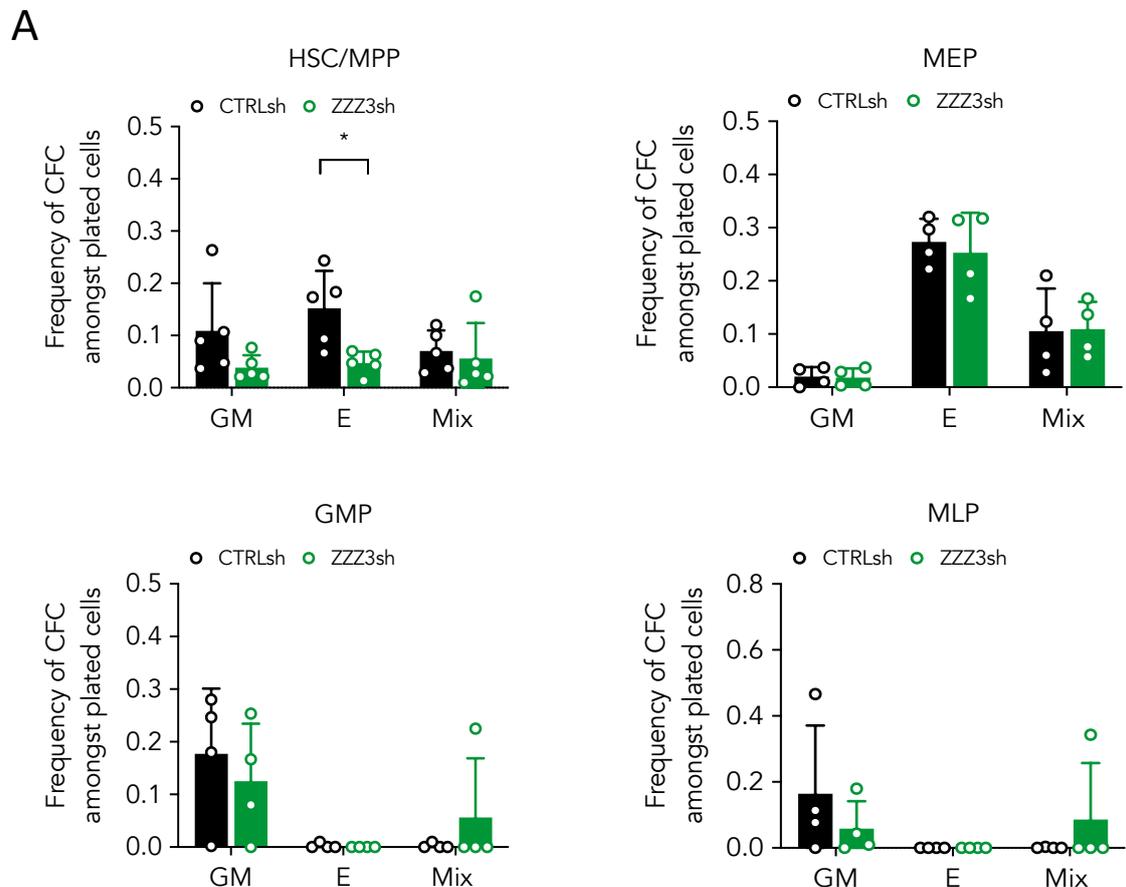


FIGURE 3.7 *ZZZ3* (ATAC) and *SUPT20H* (SAGA) depletion in human CB CD34⁺ cells.

(A) Quantitative RT-PCR validation of *SUPT20H* and *ZZZ3* knockdown in human cord blood HSC. Representative experiment for *SUPT20H*; mean \pm SEM of 2 individual experiments for *ZZZ3*; gene expression relative to CTRLsh, normalised to *HPRT1* housekeeping gene. Paired two-tailed t-test for

significance $*p < 0.05$. **(B)** Proportion of *ZZZ3sh* transduced CB HSC and progenitors. Mean \pm SEM of >3 individual sorting experiments. Two-tailed paired t-test for significance; no significant differences. **(C)** Proportion of *SUPT20Hsh* transduced CB HSC and progenitors. Mean \pm SEM of >3 individual sorting experiments. Two-tailed paired t-test for significance; no significant differences.

In contrast, CFC assay output revealed a unique defect in erythroid specification from HSC upon *ZZZ3* loss (**Fig. 3.8A**), with no changes to generation of mixed-lineage (Mix) or GM colonies. Colony formation from downstream lineage-restricted progenitors was not affected, suggesting a unique requirement of *ZZZ3* in early erythroid commitment, and a dispensable role for the ATAC component post-erythroid commitment, as well as in the myelo-monocytic lineages. *SUPT20H* expression knockdown, on the other hand, did not result in significant changes in CFC efficiency from either HSC or lineage-restricted progenitors (**Fig. 3.8B**).



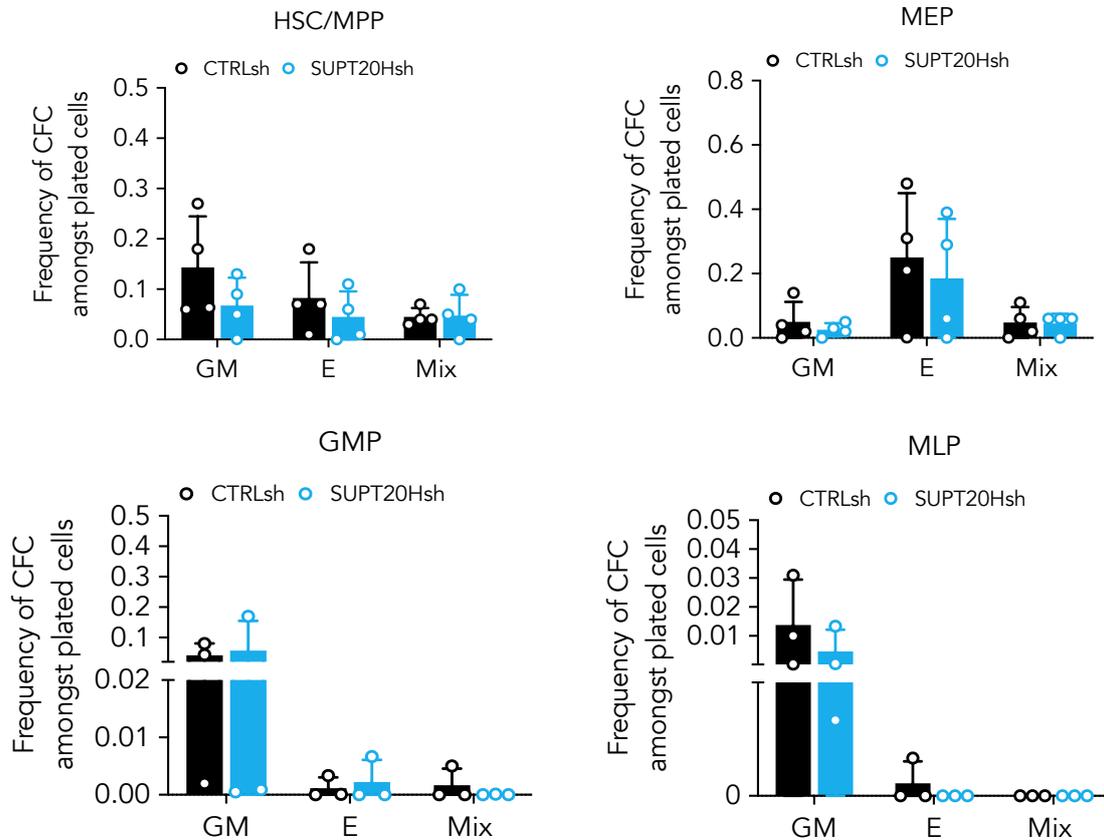
B

FIGURE 3.8 CFC-efficiency of human stem and progenitor CB cells transduced with ATAC (*ZZZ3*) and SAGA (*SUPT20H*) specific subunits. (A) Frequency of colony forming efficiency (CFC) in the HSC/MPP, MEP (top), GMP and MLP compartments (bottom) transduced with *ZZZ3sh*. Mean \pm SEM of 5 individual CB samples (4 for GMP, MLP). Two-tailed paired t-test for significance; * $p < 0.05$, ** $p < 0.01$. **(B)** Frequency of CFC efficiency in the HSC/MPP, MEP, GMP and MLP compartments transduced with *SUPT20Hsh*. Mean \pm SEM of 4 individual CB samples. Two-tailed paired t-test for significance; no significant differences.

nspection of a detailed single-cell profiling of erythroid lineage development by Tusi et al. (2018) captured *Kat2a* and *Zzz3* enrichment, but no elements of the SAGA complex, at the transition of multipotent progenitors to the erythroid and megakaryocytic lineages (**Fig. 3.9; Appendix B - Enriched genes in erythroid-basophil-megakaryocyte-biased progenitors (EBMP) as per detailed single-cell profiling of erythroid development by Tusi et al. (2018).**

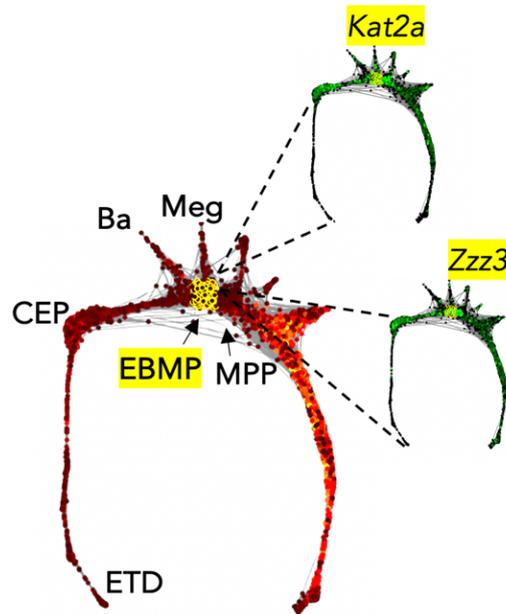


FIGURE 3.9 Gene expression (selected region) pattern of *Kat2a* and *Zzz3* in erythroid development. Visualisation obtained using SPRING tool (https://kleintools.hms.harvard.edu/paper_websites/tusi_et_al/). As per **Appendix B**, other genes in this region include regulators of E/Meg cell fate commitment including *Gata2*, *Zfpm1* and *Myb*. MPP: multipotent progenitors. EBMP: erythroid-basophil-megakaryocyte-biased progenitors. Meg: megakaryocyte. Ba: basophil. CEP: committed erythroid progenitors. ETD: erythroid terminal differentiation.

This result supports the notion that KAT2A regulation of human early erythroid specification is likely dependent on ATAC activity.

Interestingly, examination of the expression pattern of SAGA-specific elements during *in vitro* maturation of committed erythroid progenitors from human CB [274], showed that several elements associate with late differentiation, including members of the SAGA Core (TAF5L and SUPT7L), but also the SAGA-specific HAT subunit *TADA2B* and the H2B de-ubiquitinase, *USP22* (**Fig. 3.10; Appendix C – Erythroid differentiation-associated genes within cluster 17**).

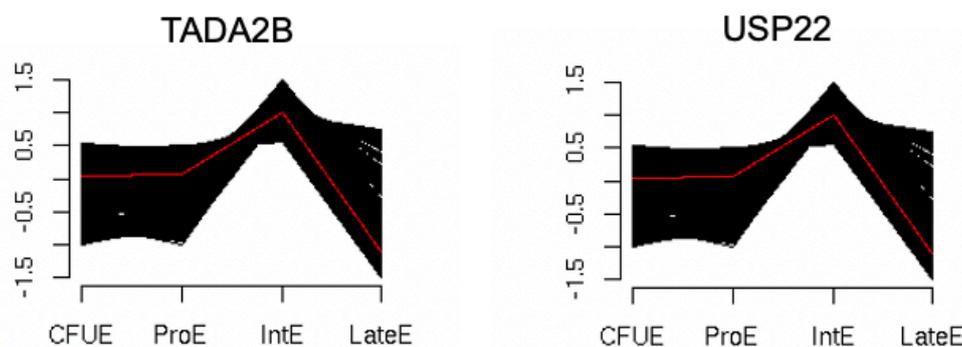


FIGURE 3.10 Expression pattern of SAGA-specific elements during *in vitro* maturation of committed erythroid progenitors from human CB. SAGA-specific elements peak at the Intermediate (IntE) phase of late erythroid differentiation. Representation of SAGA-specific *TADA2B* (left) and *USP22* (right) within Cluster 17 extracted from visualisation database: https://cellline.molbiol.ox.ac.uk_Details of individual genes can be found in **Appendix C**.

In light of these data, I investigated if loss of SAGA DUB module-*USP22* would mimic the defect observed in colony-formation at later E maturation (MEP stage) verified upon *KAT2A* knockdown in human CB (**Fig. 3.3**). As before, CB CD34+ cells were transduced with a lentiviral-delivered *USP22* targeting shRNA and HSC, MLP, GMP and MEP sorted for downstream analysis. *USP22* depletion was verified by both qRT-PCR (**Fig. 3.11A**) and WB in haematopoietic cells (**Fig. 3.11B**). There was no detectable difference in the relative proportion of stem and progenitor GFP+ cells post-transduction with *USP22sh* (**Fig. 3.11C**). However, unlike *SUPT20H* loss, knockdown of *USP22* did result in loss of erythroid colony formation from committed MEP (**Fig. 3.11D**). It did not significantly affect erythroid output from HSC and multipotent progenitors (**Fig. 3.11C**), suggesting that the MEP loss observed upon *KAT2A* knockdown may be due to synergistic effects of *KAT2A* participation in ATAC (early-stage roles) and SAGA (post-erythroid commitment). Colony output from MLP and GMP, and indeed mixed-lineage or GM colony formation from HSC, were not affected (**Fig. 3.11C**), indicating a unique contribution of *USP22* to expansion, differentiation, or survival of MEP.

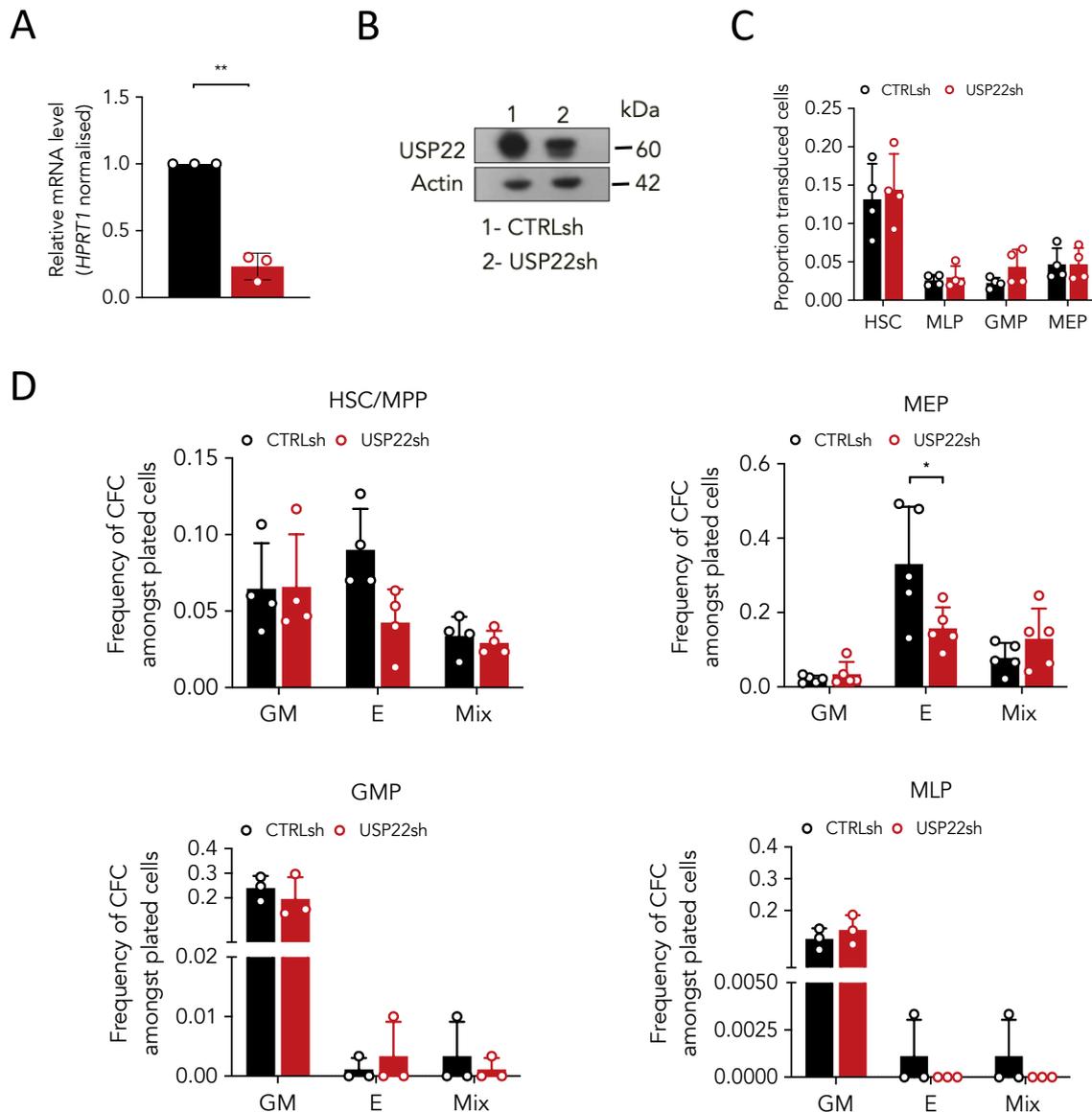


FIGURE 3.11 Depletion of SAGA-DUB module *USP22* subunit in human CB CD34+ cells.

(A) Quantitative RT-PCR validation of *USP22* knockdown in CB HSCs. Mean \pm SEM of 3 individual experiments; gene expression relative to *CTRLsh*, normalised to *HPRT1* housekeeping gene. Paired two-tailed t-test for significance $**p < 0.01$. **(B)** Western blot analysis showing *USP22* (60KDa) knockdown efficiency in K562 cells. *ACTB* (42KDa) was used as loading control. **(C)** Proportion of *USP22sh* transduced CB HSC and progenitors. Mean \pm SEM of >4 individual sorting experiments. Two-tailed paired t-test for significance; no significant differences. **(D)** Frequency of CFC efficiency in the HSC/MPP, MEP (top), GMP and MLP (bottom) compartments transduced with *USP22sh*. Mean \pm SEM of 4 individual CB samples. Two-tailed paired t-test for significance; $*p < 0.05$.

Altogether, colony-forming progenitor assays of CB CD34⁺ cells aligned loss of KAT2A-containing ATAC complexes with impairment of early erythroid and megakaryocytic specification. It also suggested that KAT2A-containing SAGA complexes play a role in erythroid differentiation post-commitment, as evidenced by *USP22* depletion.

3.3.3 KAT2A-containing ATAC and SAGA complexes have unique targets in haematopoietic cells

In order to characterise KAT2A complex-specific roles in the blood system, chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) of ATAC and SAGA-specific subunits was performed in self-renewing K562 cells. Specific sera against human ZZZ3 (ATAC-specific) and SPT20 (SAGA-specific) prepared in the Tora Lab [245], was used for ChIP-seq experiments, performed in duplicate.

As in previous reports, there was limited overlap between ATAC and SAGA targets (**Fig. 3.12A; Appendix D – ChIP-seq targets in human K562 cells**). Bound peaks were preferentially found in the vicinity of the transcriptional start site (TSS) (**Fig. 3.12B**) and were robustly enriched for known ZZZ3 and SPT20 targets, respectively (**Fig. 3.12C-D**). Interestingly, ZZZ3 peaks were more proximal than SPT20, contrary to the previously described enhancer association of ATAC complexes in lymphoblast and HeLa cells [245]. However, consistent with these results, a distinct ZZZ3 ChIP-seq experiment in human NSCLC cells revealed that ZZZ3 peaks were strongly enriched at regions ± 1 kb of TSS [235]. This suggests that ATAC-complex enhancer region occupancy may be cell type or context dependent. ENCODE experimental enrichment [275] of sequence-specific transcription factors within ZZZ3 and SPT20 peaks was clearly distinct (**Fig. 3.12C-D**), highlighting unique complex biologies. These were reflected in distinct GO categories associated with complex-specific peaks, which encompass RNA and ribosomal metabolism in the case of ATAC (**Fig. 3.12E**) and transcriptional activity for SAGA (**Fig. 3.12F**).

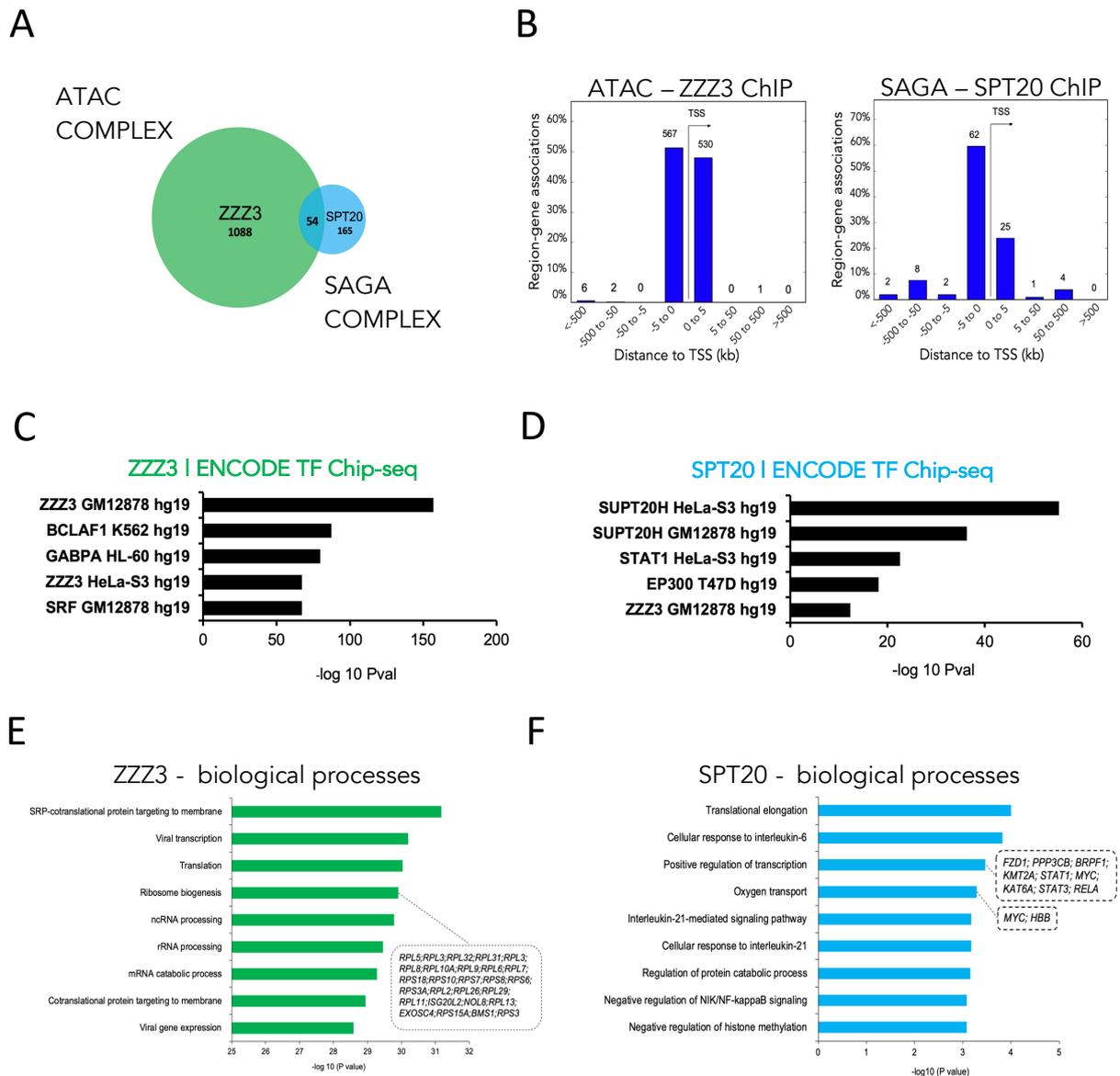
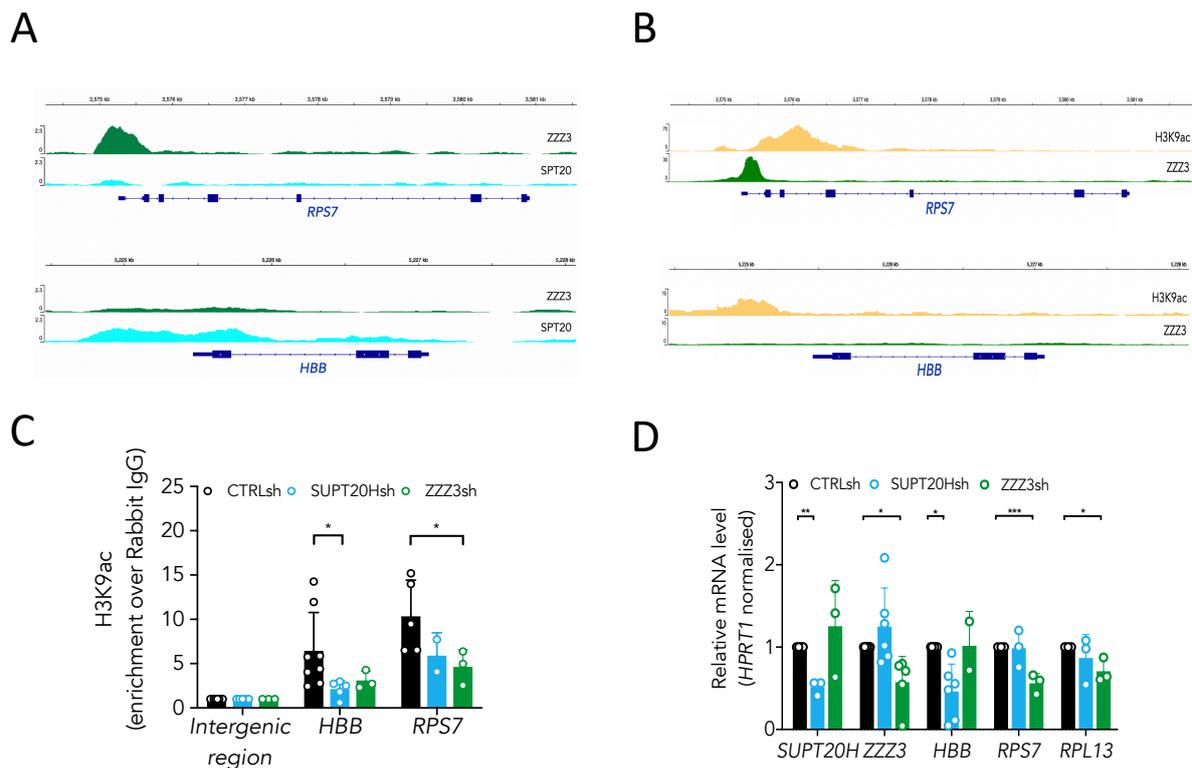


FIGURE 3.12 ChIP-seq analysis of KAT2A-containing ATAC and SAGA complexes in K562 cells. **(A)** Venn-diagram of consensus ZZZ3 and SPT20 ChIP-seq binding from 2 independent experiments. **(B)** Genomic location of ZZZ3 (left) and SPT20 (right) ChIP-seq binding in K562 cells. Summary of consensus peaks from 2 independent ChIP-seq experiments is shown. **(C)** Specificity of ATAC (ZZZ3) and SAGA (SPT20) ChIP-seq targets against ENCODE; data as retrieved by EnrichR online annotation tool [275] **(C,D)**. Specificity of ATAC (ZZZ3) and SAGA (SPT20) ChIP-seq targets against ENCODE; data as retrieved by EnrichR online annotation tool [275] **(E,F)**. Top Gene Ontology (GO) associations of ZZZ3 and SPT20 ChIP-seq targets on biological processes as calculated by EnrichR [275].

Moreover, SAGA specifically binds red blood cell-associated genes such as *HBB* (**Fig. 3.13A**), which may indicate a unique role in erythroid differentiation or identity. *ZZZ3*, on the other hand, binds ribosomal protein genes such as *RPS7* (**Fig. 3.13A**). Inspection of publicly available K562 ChIP-seq datasets in the ENCODE database confirmed that SPT20 and *ZZZ3*-bound regions coincided with H3K9ac peaks, a hallmark of *KAT2A* enzymatic activity [177] (**Fig. 3.13B**). Accordingly, loss of H3K9ac was observed at gene promoters upon SPT20 or *ZZZ3* shRNA-mediated knockdown (**Fig. 3.13C**). Locus-specific loss of H3K9ac coincided with down-regulation of gene expression, which on the selected genes, was unique to loss of SAGA or ATAC complex elements (**Fig. 3.13D**). In other words, downregulation of ATAC function had no effect on SAGA-bound genes and vice versa. Gene expression changes could be partially recapitulated by *KAT2A* gene expression knockdown (**Fig. 3.13E-F**), particularly in respect of *ZZZ3*-bound elements, which could reflect differential dependence and/or redundancy of histone acetyltransferase activity in either complex. Indeed, K562 cells were more clearly dependent on *ZZZ3* than on SPT20 for propagation in culture (**Fig. 3.13F**), which could potentially explain the greater number of target genes (**Fig. 3.12A**), and loss of *KAT2A* had an intermediate effect (**Fig. 3.13F**) which did not involve changes in cell cycle (**Fig. 3.13G**) or apoptosis (**Fig. 3.13H**). Finally, I also looked at H3K9ac changes at *EPOR* which was downregulated up *KAT2A* knockdown in CB, but did not observe significant changes, at least in SUPT20Hsh cells.



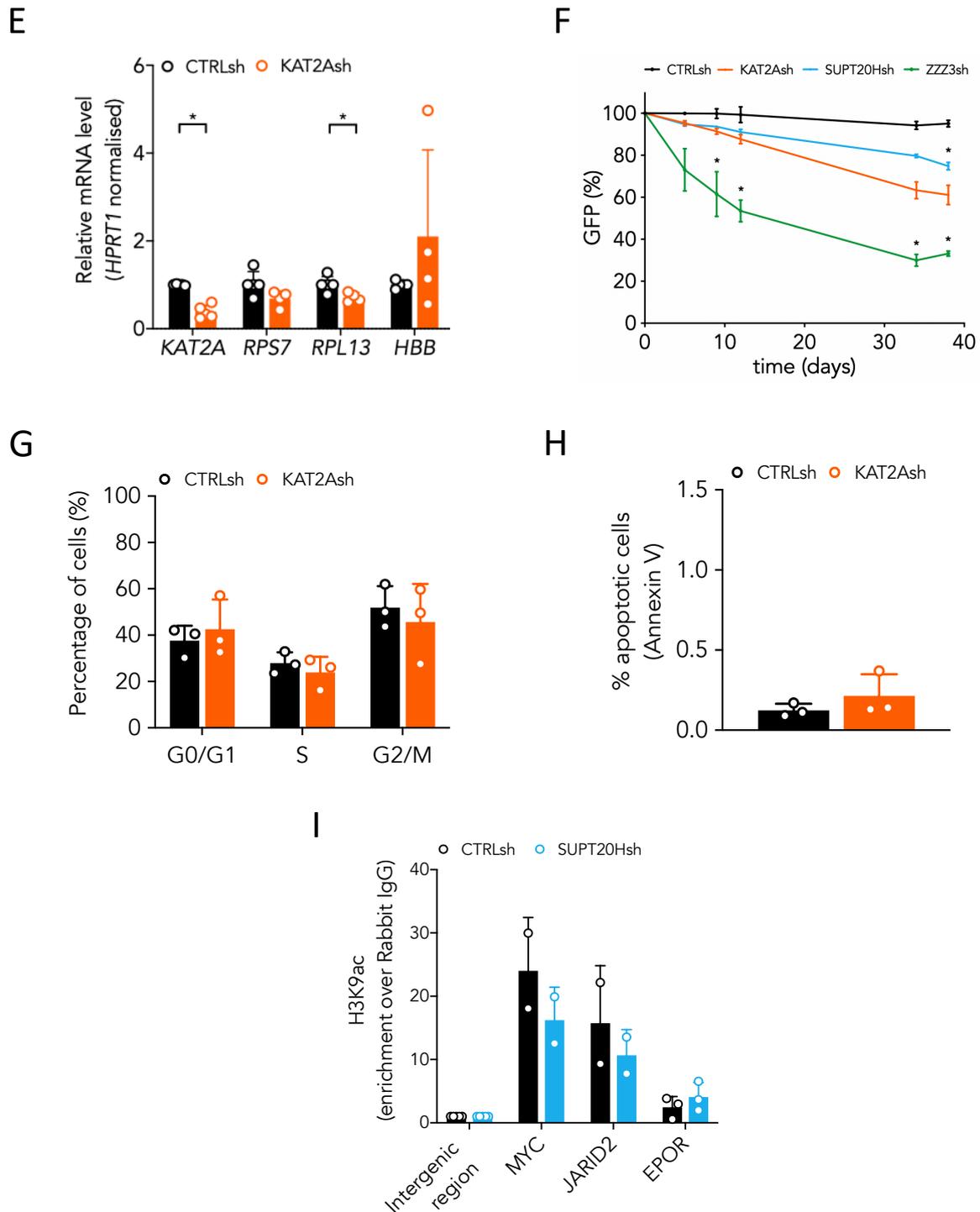


FIGURE 3.13 KAT2A-containing ATAC and SAGA complexes have unique functional associations in K562 cells. (A) Representative ChIP-seq peak for *ZZZ3* target in K562 cells (*RPS7*) and representative ChIP-seq peak for SPT20 target in K562 cells (*HBB*). **(B)** Publicly-available ChIP-seq tracks for H3K9ac (ENCFF257CLC) and *ZZZ3* (ENCFF856KCV) in K562 cells at the *RPS7* and *HBB* loci retrieved from the ENCODE project portal (www.encodeproject.org). The *RPS7* and *HBB* loci in (A) are represented confirming the presence of

H3K9ac peaks and reproducing the selective *ZZZ3* binding at *RPS7* also observed in our data. **(C)** H3K9ac ChIP-qPCR analysis of representative *SPT20* and *ZZZ3* targets upon knockdown in K562 cells. $N \geq 3$ independent experiments. Mean \pm SEM of enrichment relative to rabbit-IgG, with normalisation to control intergenic region with no significant H3K9ac enrichment. Two-tailed t-test for significance * $p < 0.05$. **(D)** Quantitative RT-PCR analysis of expression of ATAC and SAGA complex targets in K562 cells. $N \geq 3$ independent experiments, mean \pm SEM of gene expression relative to *CTRLsh*, normalised to *HPRT1* housekeeping gene. Two-tailed t-test for significance * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **(E)** Quantitative RT-PCR validation of *KAT2A* knockdown in K562 cells. Mean \pm SEM of 8 individual experiments; gene expression relative to *CTRLsh*, normalised to *HPRT1* housekeeping gene. Paired two-tailed t-test for significance *** $p < 0.001$. **(F)** Growth curve of K562 cells transduced with shRNA constructs against *ZZZ3*, *SUPT20H* and *KAT2A*. Mean \pm SEM of 3 independent experiments. ANOVA for mixed effects analysis - Tukey's multiple comparisons test significance * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **(G)** Flow cytometry analysis of cell cycle in K562 cells transduced with *CTRLsh* and *KAT2Ash*. Mean \pm SEM of $N = 3$ independent experiments. Two-tailed t-test for significance; no significant changes. **(H)** Quantification of Annexin V+ apoptotic cells in K562 cultures transduced with *CTRLsh* and *KAT2Ash*. Mean \pm SEM of $N = 3$ independent experiments. Two-tailed t-test for significance. No significant difference. **(I)** H3K9ac ChIP-qPCR analysis of representative EPOR, and *SPT20* targets *MYC* and *JARID2* upon knockdown in K562 cells. $N \geq 3$ independent experiments for EPOR; $N = 2$ independent experiments for EPOR for *MYC* and *JARID2*. Mean \pm SEM of enrichment relative to rabbit-IgG, with normalisation to control intergenic region with no significant H3K9ac enrichment. No significant differences.

Together, in line with independent activities of each complexes in hematopoietic cell maintenance and/or identity, which might be balanced by the action of *KAT2A*.

As in primary CB cultures, I inspected the progression of erythroid differentiation in K562 cells upon depletion of ATAC and SAGA members *ZZZ3* and *SPT20*. Erythroid differentiation is indicated by the gain of CD235a and loss of CD71 surface expression. Exposure of K562-*CTRLsh* (GFP+) cells to 1.5% DMSO resulted in accumulation of CD235a+ cells and progressive loss of the early erythroid-expressing CD71+ cells over a 6-day time-course (**Fig. 3.14A**). Moreover, it was accompanied by downregulation of early erythroid regulator *GATA2* from Day 0 to Day6, with up-regulation of *EPOR* and *TAL1* (**Fig. 3.14B**). Thus, these results suggest that DMSO treatment effectively promotes erythroid differentiation programs. DMSO-induced differentiation cultures initiated by *KAT2A* or *ZZZ3* knockdown cells followed a similar progression to control, as demonstrated by molecular analysis. As shown in **Fig. 3.14C**, *ZZZ3* knockdown resulted in decreased *GATA2* expression from Day 0 to Day6, whereas *TAL1* expression augmented from Day 0 to Day6. The same was verified upon *KAT2A* depletion, for which a change in *EPOR* was also detected, its expression increasing in the last day. This is compatible with unhindered erythroid differentiation post commitment. In contrast, *SUPT20H* depletion significantly perturbed erythroid differentiation of K562 cells, which abnormally retained expression of *GATA2*, and failed to consistently up-regulate *EPOR* and *TAL1*, associating with a differentiation block (**Fig. 3.14C**).

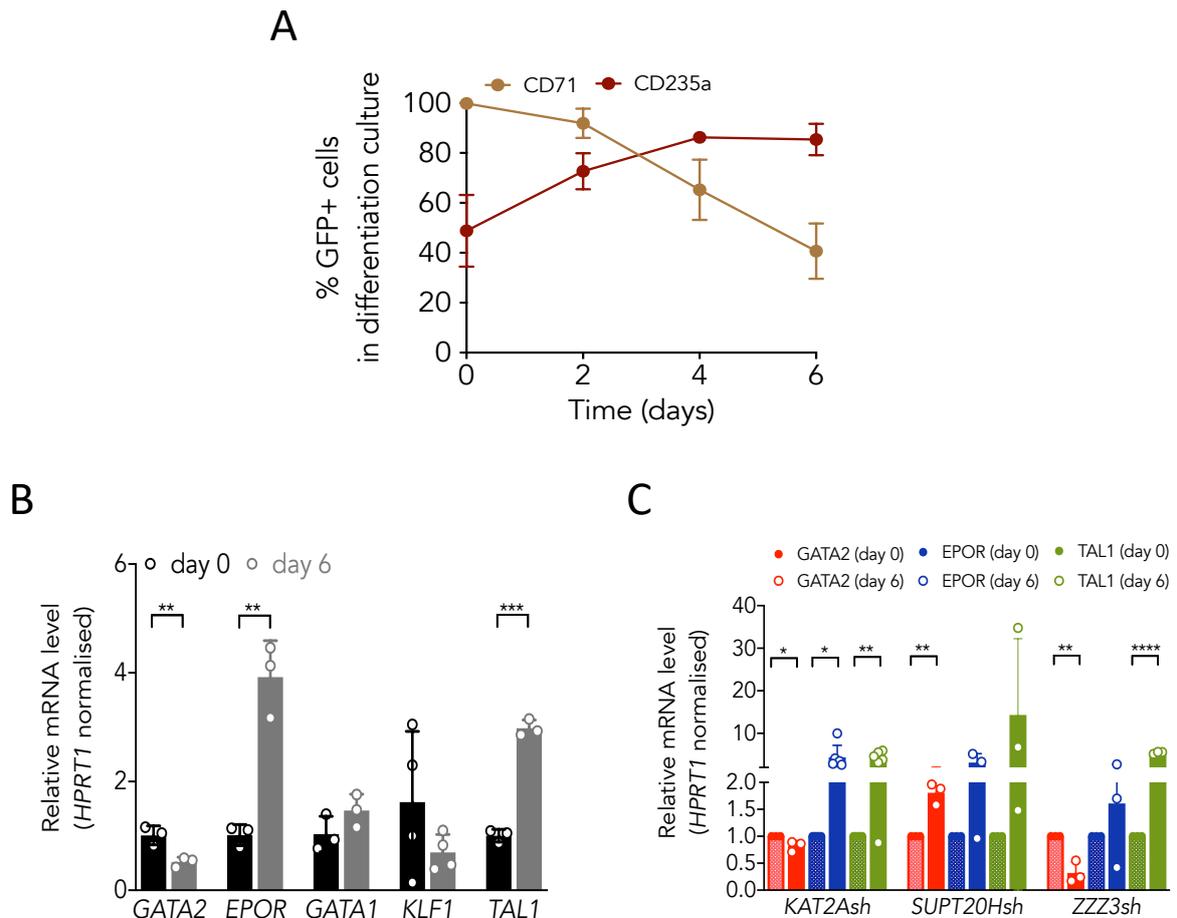


FIGURE 3.14 KAT2A-containing SAGA complex facilitates erythroid cell differentiation in hematopoietic K562 cells. (A) Time course of erythroid marker profiling by flow cytometry in K562 cells transduced with *CTRLsh* treated with 1.5% DMSO to induce erythroid differentiation. Proportion of cells expressing early (CD71) and late (CD235a) differentiation markers is shown at each time-point. Data summarise mean \pm SD of 6 independent experiments. **(B)** Quantitative RT-PCR analysis of erythroid gene expression of *CTRLsh*-transduced K562 cells at the start (day 0) and day 6 of 1.5% DMSO-induced erythroid differentiation. Mean \pm SD of N>3 independent experiments; data are represented relative to day 0 normalised to HPRT1 housekeeping gene. Two-tailed paired t-test for significance * p <0.05, ** p <0.01, *** p <0.001. **(C)** Quantitative RT-PCR analysis of erythroid gene expression in K562 cells transduced with *KAT2Ash*, *SUPT20Hsh* and *ZZZ3sh* at day 6 of 1.5% DMSO-induced erythroid differentiation. Mean \pm SEM of gene expression relative to day 0, normalised to HPRT1 housekeeping gene. Two-tailed paired t-test for significance * p <0.05, ** p <0.01.

Overall, this result is compatible with a role for SAGA in facilitating late erythroid differentiation programmes, which supports the requirement for KAT2A and USP22 in erythroid colony-formation from CB MEP. ATAC, on the other hand, is dispensable for this process, despite being required for maintenance of undifferentiated K562 cells. The stage-specific requirement of either complex is broadly compatible with my observations in CB.

3.4 Conclusions and future work

The work herein, set out to investigate KAT2A functions in normal human haematopoiesis, demonstrated that KAT2A depletion resulted in loss of erythroid cell-containing colony types from both HSC and MEP CB compartments. This indicates that KAT2A is specifically required for appropriate specification and/or survival of erythroid/platelet committed progenitors, as well as in differentiation of red cell and platelet lineages. The removal of *KAT2A* had no impact on lymphoid and myeloid lineage potential from CB. It had been previously demonstrated that loss of *Kat2a* promoted terminal granulocytic differentiation *in vitro*, in the mouse blood system [196]. Whilst I failed to see an increase in GM colony formation upon KAT2A-depletion, this could be due to differences between mouse and human haematopoietic systems. It should be said that CB-derived colonies were counted/scored visually under the microscope not-blindly. As such, it is arguable that this constitutes a limitation when compared to blind non-biased scoring, as well as more automated methods and/or the scoring by multiple operators. Any possible bias associated would have been excluded if the plates had been scored blindly. No defects in cell proliferation were noted upon *KAT2A* knockdown in any of the CB populations studied, discarding cell cycle as a prevailing mechanism by which KAT2A fosters erythroid development. In contrary, *KAT2A*-depletion resulted in increased apoptosis of HSC and MEP fractions, suggesting a role for KAT2A in survival of these cell populations. The results from CFC assays were matched by transcriptomic analysis of stem/multipotent CB cells in which *KAT2A* expression had been knocked down. Indeed, there was evident downregulation of erythroid (e.g.: *HBB*, *GFI1B*, *HBG1* or *HBG2*) and platelet (e.g.: *VWF*, *ITGB3*) associated genes upon *KAT2A*-depletion in CB HSC, indicating that KAT2A regulates a shared erythroid and megakaryocytic transcriptional programme. Moreover, GSEA identified ‘erythroid differentiation’ and ‘platelet biology’ as enriched signatures of the RNA-seq data. In

future, it would be valuable to test some of the genes downregulated upon *KAT2A* depletion e.g.: *EPOR*, *HBB*, *ID2* or *TAL1*, in phenocopy and rescue experiments to see (1) if the erythroid defect is indeed mimicked at the expected developmental stage (2) if the erythroid defect can be rescued by enforced expression of the candidate(s) downregulated gene(s). This would be a better indication that *KAT2A* directly regulates these genes, and it would be carried out by shRNA and overexpression constructs attempting to elicit or rescue phenotypes in CFC or other functional assays. Whereas the erythroid lineage is more obvious to track in CFC-assays, detection of megakaryocyte-only colonies would require adding TPO to the media. Still, megakaryocyte colonies may be difficult to detect morphologically, as they may contain very few cells. It would have been interesting, however, to follow megakaryocytic/platelet cell fate more thoroughly, for example by performing specialised CFU-Mk colony assays which relies on immunostaining for the megakaryocyte marker CD41.

Given that *KAT2A* exerts its function *via* two distinct transcriptional co-activator complexes, ATAC and SAGA, I explored their contribution to normal blood development. Whilst I found that loss of ATAC (*ZZZ3*) resulted in a specific defect in erythroid colony-formation exclusively from HSC, no effects in lineage specification were noted when depleting SAGA (*SUPT20H*) from CB. This result points to a unique role of ATAC in human erythroid specification and align *KAT2A* early requirement in HSC with its participation in ATAC-complexes. Moreover, this seems to be the case not only at the level of lineage specification, but potentially also in survival of the erythroid lineage. Indeed, by examining the single-cell profiling of erythroid development by Tusi and colleagues, I found that both *Kat2a* and *Zzz3* were enriched at the transition of multipotent progenitors to the erythroid and megakaryocytic lineages, where no elements of the SAGA complex were found [276]. Additionally, in the same study, gene set enrichment on dynamic gene clusters in early erythroid differentiation revealed cell essential biosynthetic pathways (RNA processing factors and ribosomal proteins) as early developmental events, which one could speculate is how ATAC-mediated *KAT2A* activities participate in cell survival [276]. This is speculative, and it would have been further understood by looking at apoptosis in CB compartments upon specific ATAC and SAGA- subunit depletion. Interestingly, when investigating the role of SAGA DUB module by downmodulating *USP22* in human CB, I verified loss of erythroid colonies exclusively from MEP, bringing SAGA activities back into play. This same phenotype had also been noted upon *KAT2A*-

depletion. Two hypotheses can be raised to try and explain this result. On one hand, it is possible that both SAGA enzymatic modules - HAT (KAT2A) and DUB (USP22) – cooperate to regulate erythroid development post-commitment, in which SAGA Core (SPT20) roles may be dispensable. In other words, regulation of fully committed erythroid progenitors could involve the combinatory action of both histone/protein acetylation and deubiquitination. USP22 catalyses the deubiquitination of histone H2B and H2A, thereby acting as a transcriptional coactivator [277–278]. It is recruited to specific gene promoters by activators such as MYC and p53 [277]. As a result of USP22 loss, the level of H2B/A deubiquitination is lowered, leading to decreased transcription at target gene *loci*. As mentioned in the previous Chapter, transcriptionally repressive states can involve the action of PRC1 and PRC2 complexes, which catalyse H2A ubiquitination and H3K27me3, respectively. It is possible that late erythroid-affiliated TFs would recruit SAGA to remove ubiquitination marks deposited by PRC complexes at histones *via* USP22. In turn, upon relief of repressive chromatin status, this would allow other epigenetic players, for example MLL complexes to promote SAGA-dependent acetylation, maintaining gene expression programmes relevant to erythroid development. Moreover, USP22 deubiquitinates and stabilises SIRT1, likely within the context of SAGA [261], a histone deacetylase with specificity for a number of proteins including histone H3. Once deubiquitinated, SIRT1 is not targeted to proteasome degradation, in turn, suppressing transcriptional programmes [261]. Further investigation is certainly needed to understand the potential crosstalk between these chromatin-remodelling complexes and their implications to transcription of erythroid genes. As a second possibility, KAT2A and USP22 roles are not strictly dependent on the association with the SAGA Core. Accordingly, it was previously demonstrated that SAGA HAT module can exist as a separate complex, at least in *Drosophila* [232], which may explain some level of independence from the Core module containing SPT20, and the differential role in late erythropoiesis in CB. Likewise, USP22 does not seem to require SAGA incorporation in order to be directed to target genes [278]. Also, the recent elucidation of the structure of yeast SAGA suggests that upon binding of a nucleosome, SAGA displaces the HAT and DUB modules from the Core module [214], again suggesting some degree of independence that could be conserved from yeast to human.

The fact that SAGA did not impact lineage fate commitment from HSC but was shown to regulate progression of erythroid differentiation post-commitment (via USP22),

suggests a role in controlling, rather than establishing transcriptional programmes and cell identity. In agreement, USP22 functions have been associated with lineage-specific differentiation programmes [226]. Yet SAGA requirements may be contributory, rather than absolute, given that the same was not verified with the other SAGA element investigated, SPT20. To investigate this further, it would be of interest to assess what happens to the expression of late erythroid regulators upon USP22 knockdown and see if it correlates with the loss of H2B deubiquitination. Testing additional specific elements of Core or DUB modules, for example SUPT3H, that binds TATA-binding protein, and ATXN7, respectively would also be informative.

This interpretation from the CB experiments, suggesting that ATAC, but not SAGA, is relevant for the cell fate decision towards erythropoiesis and not to terminal differentiation, would have been further supported if the GFP+ levels had been assessed at the end point of CFC-assays to confirm whether the shRNA expression was, in fact, silenced. The GFP+ levels had, however, been evaluated in the erythroid differentiation experiment in CB upon KAT2A depletion (**Fig. 3.4**), confirming gene silencing throughout the differentiation. Moreover, from a 2-day erythroid differentiation experiment performed in K562 cells in which *KAT2A* and *ZZZ3* had been knocked down (data not presented in this thesis), the percentage of GFP+ cells had also been assessed and shown to be maintained (**Fig. 3.15**).

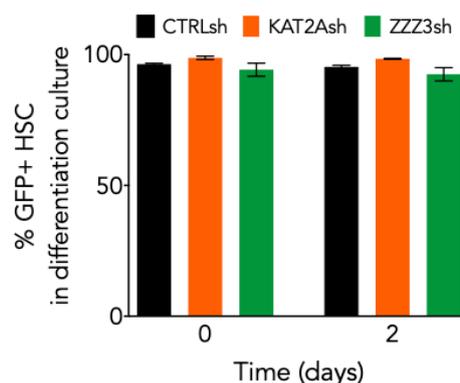


FIGURE 3.15 GFP+ percentage in *CTRLsh*, *KAT2Ash* and *ZZZ3sh* K562 cells in a 2-day erythroid differentiation culture experiment.

In addressing the genome binding pattern of the two human KAT2A-containing complexes in haematopoietic K562 cells, the results showed that ATAC and SAGA have distinct chromatin substrates. Specifically, ATAC (*ZZZ3*) binds RNA metabolism

and ribosomal protein genes, while SAGA (SPT20) associates with DNA-binding and transcription related *loci*. First, this was interesting because, in previous work, namely in yeast, SAGA was found to be recruited mainly to stress regulated genes [279–280], which was not verified in this, as well as another study in human cell lines [245]. Thus, the data suggests diversification of SAGA complexes, or of KAT2A function itself, from yeast to human. Furthermore, it was demonstrated that both ATAC and SAGA have a narrow subset of target genes, 1088 and 165 bound-peaks, respectively. This, particularly in the case of SAGA, alludes to the possibility that it functions as coactivator for specific gene expression programmes, contrary to previously reported global acetylation functions [281]. This is supported by analysis of SAGA-bound genes, which associate with DNA-binding and regulation of transcription, as examples: *MYC*, *E2F4*, *MLL1*, *STAT1*, *STAT3*, *MBD3* and *RELA*. The fact that K562 cell line was used to perform the ChIP-seq experiment to investigate normal haematopoiesis could pose some limitations, particularly if we aim to compare with the RNA-seq datasets, which derive from experiments performed in primary cells. As previously mentioned, K562 are a chronic myelogenous leukaemia cell line with multilineage potential [271]. These cells display aspects of molecular differentiation into erythroid, megakaryocytic and macrophage identities under defined cytokine conditions; in steady state, K562 cells represent immature leukaemia blasts with self-renewal properties. K562 cells have, however, been repeatedly used in past research as a model for understanding molecular mechanisms of normal erythroid, and malignant blood specification [282–285].

The highly specific binding sites likely reflect different modes of recruitment of the two complexes, which could in principle associate with a stage specific TF/chromatin-complex relationship. In human CB, both ATAC and SAGA co-activator complexes were found to be required in regulation of developmental stages in erythropoiesis - ATAC acting early, at the lineage specification level, and SAGA facilitating late differentiation, a phase devoid of ATAC function. In self-renewal K562 cells, SPT20 (but not ZZZ3) was found to bind late erythroid-*loci* including *HBB*, for which expression was exclusively down-regulated upon *SUPT20H* knockdown. Indeed, knockdown of ATAC ZZZ3 or SAGA *SUPT20H* clearly affected the specific genes predicted to be bound by the respective complexes, which associated with loss of H3K9ac in both cases, suggesting that the expression of the regulated genes is dependent on HAT activity. If the limited number or target genes was to be a technical limitation of the ChIP experiment, functional qPCR (RT and CHIP-) analysis wouldn't

have shown a correlation between knockdown of ATAC (*ZZZ3*) or SAGA (*SUPT20H*) and downregulation of the specific predicted bound genes by either complex. Interestingly though, knockdown of shared ATAC and SAGA - KAT2A in K562s, lead to reduced expression of ATAC target genes, but not of SAGA, at least the selected *HBB* locus. This may be due to the need for additional SAGA subunits in regulation of the *HBB* locus, for example its DUB module as suggested in erythroid maturation from CB. On the other hand, the fact that SPT20-bound peaks are less close to the TSS than ATAC, could suggest that SAGA is recruited to tissue stage-specific/proximal enhancers. Although, reports suggest that the action of HAT complexes is more specific at enhancers, but more redundant at promoters [286], in which case maybe SAGA relies on KAT2B, and not KAT2A in final erythroid maturation stages. Notably, *SUPT20H* specifically interferes with the establishment of late differentiation programmes, a role that apparently exceeds KAT2A itself, since in its absence K562 and CB cells reach a similar CD235a+ levels as control cells. Again, it is possible that SAGA complexes in late erythroid differentiation utilise the orthologue KAT2B thus compensating for KAT2A loss. Increased *KAT2B* expression in late erythroid differentiation (**Appendix C**) is compatible with this view, but this hypothesis needs examination. Another line of evidence to suggest redundancy by KAT2B in late erythroid development is that it was shown to interact with, and acetylate, TAL1, promoting erythroid differentiation in murine erythroleukemia (MEL) cells [287]. TAL1 associates with complexes containing master erythroid regulators, such as GATA2. The fact that progression of erythroid differentiation in K562 cells was uniquely impaired upon deletion of *SUPT20H* through failure in downregulating GATA2, could suggest redundancy by KAT2B, potentially involving prior TAL1-acetylation. Investigating combined requirements of subunits from both Core / DUB modules, Core / HAT or HAT / DUB modules would help elucidate the exact epigenetic and transcriptional mechanisms at place.

Overall, analysis of normal human haematopoietic development indicates that KAT2A plays stage-specific roles in erythropoiesis, which are unique to the complexes it integrates. On the one hand, KAT2A regulates specification and survival of erythroid/ megakaryocytic progenitors likely through participation in the ATAC complex, which associates with biosynthetic activity pathways. On the other hand, KAT2A fine-tunes progression of erythroid differentiation through participation in SAGA, which is only required in fully committed progenitors and may play a contributory rather than an absolute role. This investigation emphasises the

importance of detailed analysis to unambiguously understand the biological roles of epigenetic complexes sharing the same enzyme in healthy, as well as malignant settings, with novel implications for follow-up studies.

Chapter 4

Cellular investigation of KAT2A-complex dependencies in human AML cell lines

4.1 Introduction

Having established differential requirements for ATAC and SAGA activity in normal blood and in the multipotent K562 line, I went on to investigate their contribution to AML biology, where our lab has previously established a role for *KAT2A* through a CRISPR drop-out screen (**Fig. 4.1**) [260]. In more detail, the authors studied five commonly used AML cell lines (MOLM-13, MV4-11, HL-60, OCI-AML2 and OCI-AML3) in which they generated Cas9-expressing cells, which were then transfected with a lentiviral gRNA library to induce gene knockouts genome-wide. Following a period of continued culture, cells with diminished survival upon the loss in ‘essential genes’ were eventually outgrown by the larger number of cells whose survival was unaffected. The remaining cells were then harvested to determine stably integrated gRNA content by PCR amplification, thus allowing the identification of the ‘drop-out’ genes, when compared to the original gRNA library.

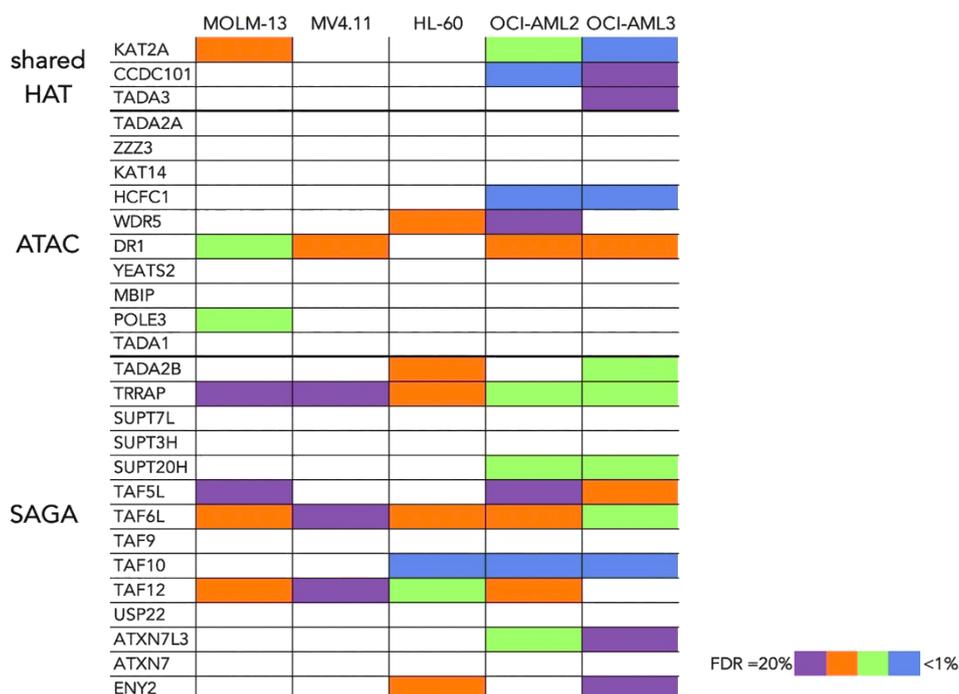


FIGURE 4.1 *KAT2A* is a genetic vulnerability in AML. Table generated based on our lab’s CRISPR drop-out screen in AML cell lines across different genetic backgrounds [260]. Genes depleted in one or more AML lines are considered genetic vulnerabilities. MOLM-13 (*FLT3-ITD*, *MLL-AF9*); MV4-11 (*FLT3-ITD*, *MLL-AF4*); HL-60 (*NRAS* p.Q61L); OCI-AML2 (hyperdiploidy and *DNMT3A* R635W); OCI-AML3 (*NRAS* p.Q61L, *NPM1* p.W288fs*12, *BAX* p.E41fs*33, *DNMT3A*).

Thus, human cell lines derived from AML bearing recurrent genetic abnormalities represent important *in vitro* models for the identification of novel genetic vulnerabilities of this disease. In order to investigate the cellular requirements of the two KAT2A complexes in human AML, I selected a panel of five AML cell lines ranging from MLL-rearranged leukaemia such as MOLM-13, MV4-11 (included in the original screen) and THP-1, as well as less commonly studied models, KG1a and Kasumi-1. MOLM-13 is a human AML cell line established from the peripheral blood of a 20-year-old male at relapse, which had evolved from myelodysplastic syndrome (MDS). It harbours *MLL-AF9* and *FLT3-ITD* fusion genes [288], and, phenotypically, corresponds to a myelomonocyte type leukaemia. THP-1 designates a monocyte-like cell line, derived from the peripheral blood of a childhood case of AML, also carrying an *MLL-AF9* translocation [289], which occurs in approximately 20% of acute leukaemias. Model cell line MV4-11 was established from blasts cells of 10-year-old male with B-myelomonocytic leukaemia, harbouring a *MLL-AF4* fusion gene and a *FLT3-ITD* mutation [290]. In turn, Kasumi-1 represents a less differentiated type of AML. It was established from peripheral blast cells from a 7-year-old Japanese male carrying a 8; 21 chromosome translocation, which gives rise to the fusion gene *AML1-ETO* or *RUNX1-RUNX1T1*, and an associated *KIT* mutation [291]. KG1a, a subline of KG1, is a promyeloblast type cell line [292]. KG1 was established from a 59-year-old patient in whom AML had developed following an earlier phase of erythroleukaemia. Its derivative, KG1a, has the same karyotype abnormalities, including *FGFR1OP2-FGFR1* fusion, but appears undifferentiated, failing to express the myeloid characteristics of the parent line KG1. Thus, the KG1a line represents a very immature type AML. Whereas MOLM-13 and MV4-11 were used in the CRISPR screen that established KAT2A as a vulnerability in AML, none of the remainder cell lines were investigated, and its dependencies on elements of ATAC and SAGA complexes is currently unknown. AML remains a lethal disease for the majority of patients. A block of cellular differentiation is a hallmark of this disease, in which myeloid blasts fail to differentiate into fully functional mature myeloid cells. Preservation of the typical undifferentiated leukaemic cell identity and/or function depends on cellular mechanisms balancing cell death and cell proliferation. As such, cell processes including apoptosis, proliferation and differentiation are key mechanisms shaping leukaemogenesis. Here, I aimed to investigate how they are regulated by KAT2A and its complexes in the AML cell lines described above.

4.2 Chapter overview

In this Chapter, I describe the results of functional *in vitro* experiments aimed at dissecting the contributions of ATAC and SAGA-KAT2A complexes to human AML. I used a panel of five human AML cell lines and a shRNA-mediated gene knockdown approach throughout. I targeted ATAC specific components *ZZZ3*, *TADA2A* and *YEATS2*, as well as *SUPT20H* and *TADA2B*, unique SAGA subunits. I demonstrate that both ATAC and SAGA functions are required to sustain AML cells in culture, but with separate effects on the leukaemia biology. Specifically, ATAC regulates biosynthetic activity necessary to maintain proliferating cells. On the other hand, SAGA mediates preservation of the characteristic undifferentiated identity of AML blasts, its loss promoting leukaemic differentiation. The unique roles of both complexes were observed across the different AML cell lines, independently of their molecular background. Together, these results explain the requirement of *KAT2A* in AML, which participates in one or the other complex to mediate distinct activities.

4.3 Results and discussion

In order to investigate *KAT2A* complex-specific requirements to AML, I knocked down the expression of unique ATAC (*ZZZ3*, *TADA2A* and *YEATS2*) and SAGA (*SUPT20H*, *TADA2B*) subunits in a panel of five AML cell lines. I explored the phenotypic consequences of depleting the ATAC and SAGA elements by conducting, cell cycle differentiation and apoptosis assays, complemented by gene expression analysis (**Fig. 4.2**).

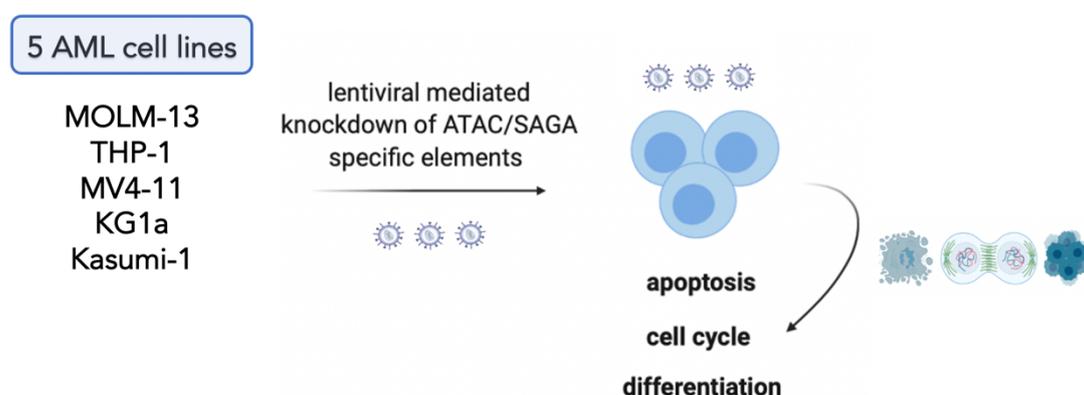


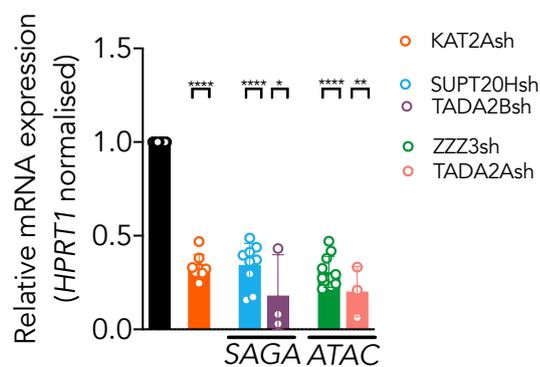
FIGURE 4.2 Schematic of the adopted experimental strategy.

4.3.1 KAT2A complexes uniquely maintain proliferation and identity of MLL-rearranged AML cells

MOLM-13

I started by investigating the model AML cell line MOLM-13, previously shown to be dependent on *KAT2A* expression and activity [260]. Knockdown of ATAC (*ZZZ3* and *TADA2A*), and of SAGA (*SUPT20H* and *TADA2B*) specific elements was verified by RT-qPCR (**Fig. 4.3A**), together with the shared HAT *KAT2A*. Knockdown of *KAT2A* and of the subunits of its containing chromatin complexes restricted expansion of MOLM-13 cultures as shown in **Fig. 4.3B-C**. Moreover, loss of *KAT2A*, at least when examined in parallel with *ZZZ3sh* and *SUPT20Hsh* transduced lines, seems to have a more profound effect than that of ATAC (*ZZZ3sh*) and SAGA (*SUPT20Hsh*) complexes alone, over a 5-day culture period.

A



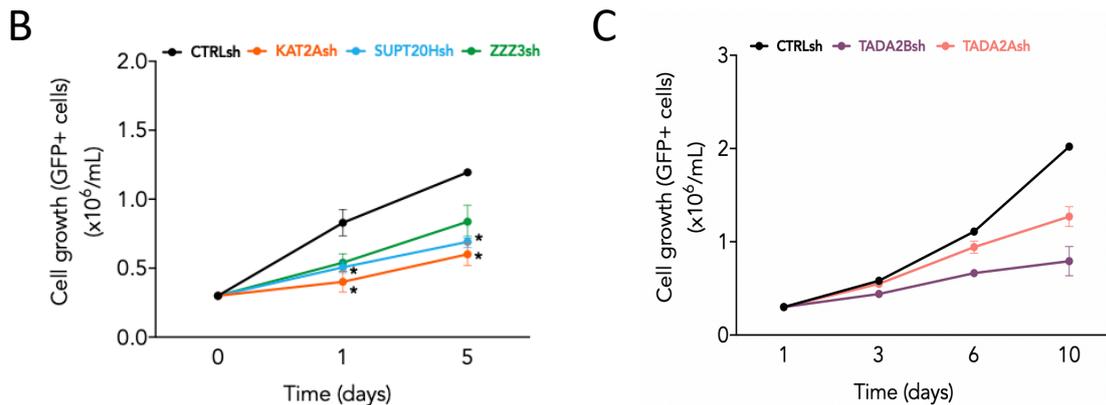


FIGURE 4.3 Knockdown of ATAC and SAGA-specific components hinders expansion of MOLM-13 cells. (A) Quantitative RT-PCR validation of *KAT2A*, *SUPT20H* and *TADA2B* (SAGA-specific subunits) and *ZZZ3* and *TADA2A* (ATAC-specific subunits) knockdown in MOLM-13 cells. $N \geq 3$ independent experiments, mean \pm SEM of gene expression relative to *CTRLsh*, normalised to *HPRT1* housekeeping gene. Two-tailed t-test for significance * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **(B)** Growth curve of MOLM-13 cells transduced with shRNA constructs against *KAT2A*, *SUPT20H* and *ZZZ3*. Mean \pm SEM of 3 independent experiments. ANOVA for mixed effects analysis - Tukey's multiple comparisons test significance * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **(C)** Growth curve of MOLM-13 cells transduced with shRNA constructs against *TADA2B* and *TADA2A*. Because two of the replicates from the *CTRLsh* group were lost due to technical failure, statistical analysis could not be performed.

At a molecular level, depletion of *KAT2A* and ATAC components *ZZZ3* and *TADA2A*, but not of SAGA elements (*SUPT20H* and *TADA2B*), led to reduced expression of ribosomal protein genes, *RLP13* and *RPS7* (**Fig. 4.4A**). This result in line with a pervasive control of protein biosynthetic activity by the ATAC complex suggested in the previous Chapter. The same function was also reported in lung cancer cell lines depleted of *ZZZ3* and *YEATS2* (ATAC complex) [235]. As shown in **Fig. 4.4B**, expression of MOLM-13 self-renewal signature genes *HOXA9* and *HOXA10*, showed a more extensive association with ATAC elements. This was evidenced by reduced expression of *HOXA9* upon depletion of *ZZZ3* and *TADA2A*, and of *HOXA10* following *ZZZ3* knockdown (**Fig. 4.4B**). *KAT2A* loss in turn, had no impact on the expression of the tested *HOXA* genes (**Fig. 4.4B**). In accordance, inspection of our lab's *Kat2a* KO leukaemia mouse model [197], carrying a similar *MLL-AF9* leukaemia fusion oncogene, showed that progressive depletion of leukaemia stem-like cells does not strictly depend on the typical *Hoxa* gene signature.

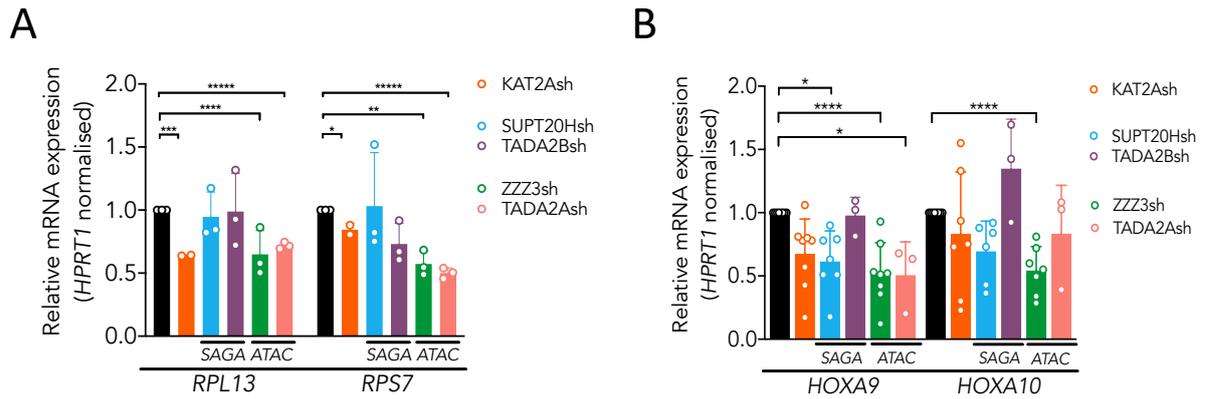


FIGURE 4.4 Analysis of MOLM-13 metabolism and self-renewal gene signature upon depletion of *KAT2A*, SAGA-specific *SUPT20H* and *TADA2Bsh* ATAC-specific *ZZZ3* and *TADA2A* subunits. (A) Quantitative RT-PCR analysis of ribosomal protein gene expression in MOLM-13 cells transduced with *KAT2Ash*, SAGA-specific *SUPT20Hsh* and *TADA2Bsh* and ATAC-specific *ZZZ3sh* and *TADA2Ash*. $N \geq 3$ independent experiments, mean \pm SEM of gene expression relative to *CTRLsh*, normalised to *HPRT1* housekeeping gene. Two-tailed t-test for significance * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **(B)** Quantitative RT-PCR analysis of self-renewal gene signature in MOLM-13 transduced with *KAT2Ash*, SAGA-specific *SUPT20Hsh* and *TADA2Bsh* and ATAC-specific *ZZZ3sh* and *TADA2Ash*. $N = 2$ biological replicates, each run as 2 or 3 technical repeats; mean \pm SEM of relative gene expression relative to *CTRLsh*, normalised to *HPRT1* housekeeping gene. Two-tailed nested t-test for significance * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Given that *KAT2A* mediates H3K9 acetylation activity through either ATAC or SAGA complexes, to facilitate transcriptional activation, the next step was to assess differences in H3K9ac changes at the promoters of the genes analysed earlier. Loss of *ZZZ3* resulted in decreased levels of H3K9ac in ribosomal protein genes and *HOXA* genes (**Fig. 4.5**) in line with the reduced gene expression (**Fig. 4.4**), which now can be attributed to *KAT2A* histone acetylation activity. Surprisingly, despite the selective impact of ATAC elements on gene expression, this analysis also showed reduced H3K9ac at ribosomal protein genes, *HOXA* genes and *MEIS1* promoters upon loss of SAGA *SUPT20H* (**Fig. 4.5**).

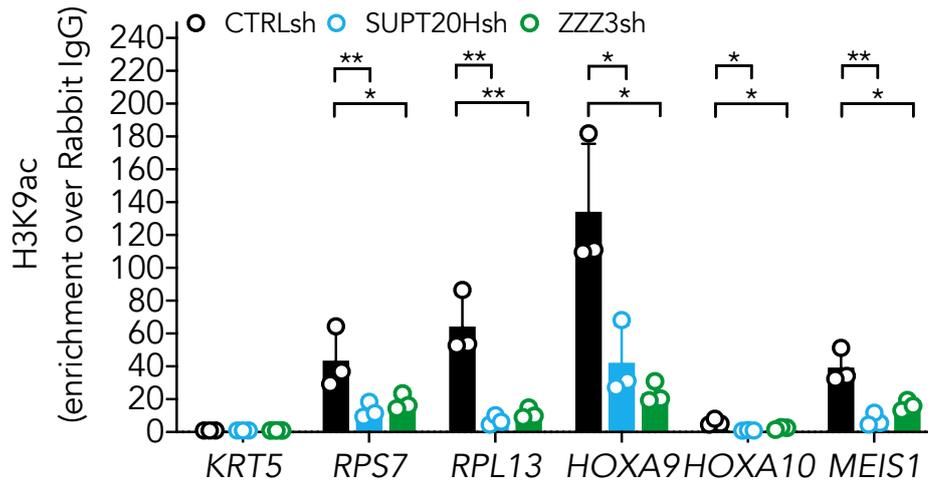
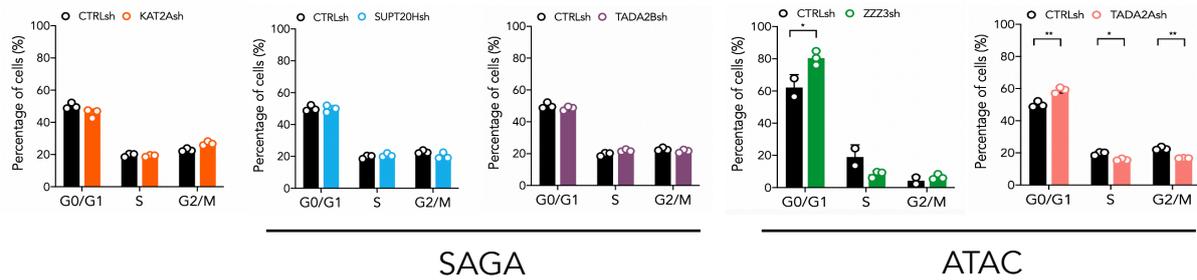


FIGURE 4.5 H3K9ac ChIP-qPCR analysis of ribosomal protein genes and self-renewal genes in MOLM-13 cells upon knockdown of SAGA (*SUPT20H*) and ATAC (*ZZZ3*) elements. $N \geq 3$ independent experiments. Mean \pm SEM of enrichment relative to rabbit-IgG, with normalisation to control region in *KRT5* locus with no significant H3K9ac enrichment. Two-tailed t-test for significance * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

HOX genes are highly expressed in stem cells, and their expression decreases with differentiation [293]. Whilst a reduced expression of *HOXA9* and *HOXA10* genes was verified predominantly upon loss of ATAC elements (**Fig. 4.4, right**), H3K9ac levels were reduced at their promoter following loss of both elements of ATAC (*ZZZ3*) and SAGA (*SUPT20H*). This suggests that although promoter occupancy may be more promiscuous than observed, for example, in K562 cells (Chapter 3), loss of H3K9ac *per se* is insufficient to decrease gene expression upon knockdown of SAGA.

Analysis of the cellular consequences of gene knockdown, however, suggest that ATAC and SAGA act on distinct aspects on the biology of MOLM-13 cells. As shown in **Fig. 4.6A-B**, loss of ATAC *ZZZ3* and *TADA2A* specifically arrested cell cycle progression in G₀/G₁, whereas depletion of *KAT2A* or SAGA-specific components allows normal progression through the cell cycle. In agreement with this finding, knockdown of ATAC *TADA2A* was previously shown to prevent normal cell cycle progression [294]

A



B

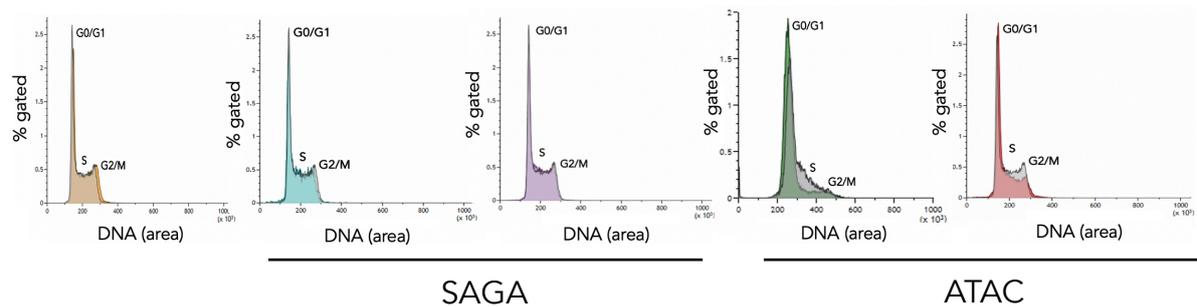


FIGURE 4.6 Cell cycle analysis of MOLM-13 cells upon knockdown of SAGA (*SUPT20H*, *TADA2B*) and ATAC (*ZZZ3*, *TADA2A*) elements. (A) Quantification of flow cytometry analysis of cell cycle in MOLM-13 cells transduced with *KAT2Ash*, SAGA-specific *SUPT20Hsh* and *TADA2Bsh* and ATAC-specific *ZZZ3sh* and *TADA2Ash*. Mean \pm SEM of 3 independent experiments. Two-tailed t-test for significance * $p < 0.05$, ** $p < 0.01$. (B) Representative flow cytometry plots of cell cycle analysis of MOLM-13 cells transduced with *CTRLsh*, *KAT2Ash*, SAGA-specific *SUPT20Hsh* and *TADA2Bsh*, and ATAC-specific *ZZZ3sh* and *TADA2Ash*.

To clarify the proliferation defect upon ATAC loss, I went on to perform cell divisional tracking following downmodulation of *ZZZ3* and *SUPT20H*. In agreement with the lack of proliferation induced upon loss ATAC elements (**Fig. 4.6**), I observed a reduced number of *ZZZ3* knockdown cells entering cell division (**Fig. 4.7A-B**) over a 3-day culture period (**Fig. 4.7C**), as captured by divisional tracking (**Fig. 4.7D**).

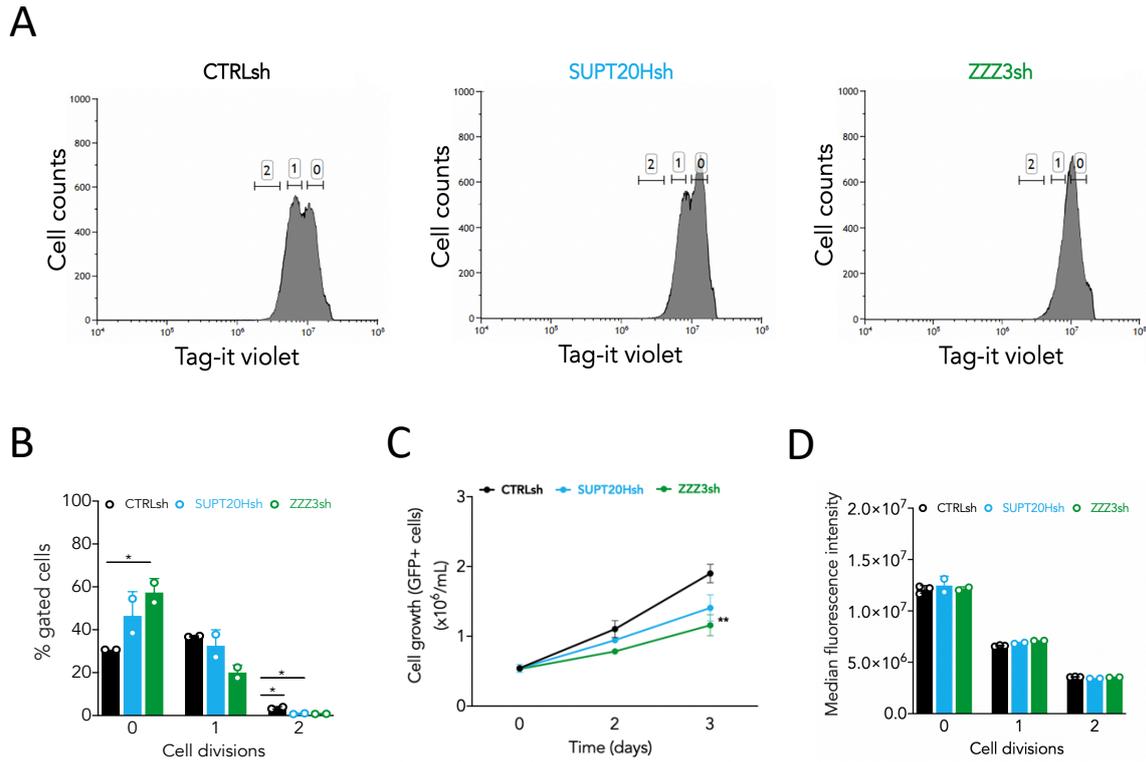


FIGURE 4.7 Cell divisional tracking of MOLM-13 cells upon knockdown of SAGA (*SUPT20H*) and ATAC (*ZZZ3*) elements. (A) Representative flow cytometry plots of divisional tracking of MOLM-13 cells transduced with *CTRLsh*, *SUPT20Hsh* and *ZZZ3sh* and loaded with the Tag-IT violet dye after 3 days of culture. Regions 0, 1 and 2 represent the number of cell divisions relative to initial loading control. **(B)** Quantification of results in (B) representing the distribution of MOLM-13 cells transduced with *CTRLsh*, *SUPT20Hsh* and *ZZZ3sh* undergone 0-2 cell divisions after 3 days in culture. Mean \pm SEM of 3 independent experiments. Two-tailed t-test for significance * $p < 0.05$. **(C)** Growth curve of MOLM-13 cells transduced with *CTRLsh* and shRNA constructs against *SUPT20H* and *ZZZ3* and stained with Tag-it Violet cell tracking dye. Mean \pm SEM of 3 individual experiments. ANOVA for mixed effects analysis - Tukey's multiple comparisons test significance * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **(D)** Flow cytometry analysis of median fluorescence intensity of MOLM-13 cells stained with Tag-it Violet proliferation and cell tracking dye. As expected, the fluorescence is halved at each division.

The data thus far positions the ATAC complex as an important regulator of self-propagation of MOLM-13 AML cells. It is in agreement with previous observations that associate ATAC with correct cell division [235–294].

In contrast, loss of SAGA *SUPT20H* resulted in extensive differentiation of MOLM-13 cells as observed morphologically (see methods section) (**Fig. 4.8A-B**). Similarly, loss of *KAT2A* also promoted cell differentiation (**Fig. 4.8A-B**), thus likely through its participation in SAGA complexes.

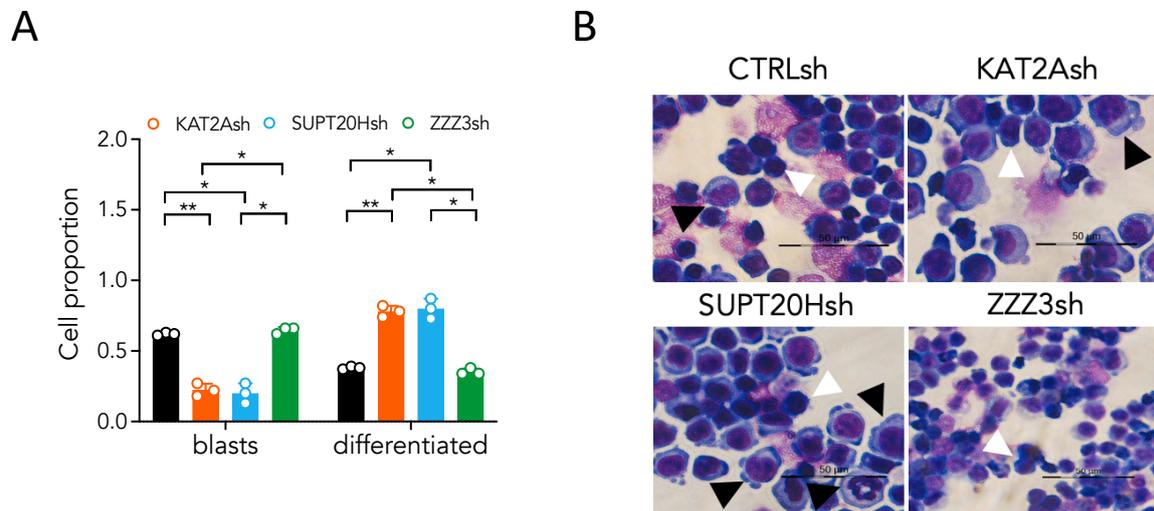


FIGURE 4.8 Differentiation analysis of MOLM-13 cells upon depletion of *KAT2A*, SAGA-specific *SUPT20H* and ATAC-specific *ZZZ3* subunits. (A) Quantification of blast-like and differentiated cells in MOLM-13 cultures transduced with *CTRLsh*, *KAT2Ash*, *SUPT20Hsh* and *ZZZ3sh*. Scoring of 3 randomly selected fields of >100 cells; Two-tailed t-test for significance; * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$. **(B)** Representative photographs of MOLM-13 cytopspins. White arrow heads denote blast-like cells; black arrow heads denote differentiated cells. Scale 50 μm .

This result points to a possible role of SAGA in preservation of cell identity as the blast cells are prompted into differentiation following loss of *SUPT20H*, and indeed of *KAT2A*. Interestingly, SAGA-mediated impact on MOLM-13 cell identity may not be exclusively exerted through the typical self-renewal *HOXA* gene signature, as suggested in **Fig 4.4**, where I observed loss of H3K9ac at *HOXA9*, *HOXA10* and *MEIS1* promoters followed *ZZZ3* (ATAC) knockdown. This could suggest possible cooperation from other regulators, such as *MYC* in this AML, thus differentiation changes observed at cellular level may not be depended on H3K9ac. On the other hand, *KAT2A* activity in MOLM13 may involve H3K79 succinylation or other yet uncharacterised acylations.

When analysing apoptosis, there was no consistent apoptotic response to knockdown of any of the SAGA (*SUPT20H* and *TADA2B*), or ATAC (*ZZZ3* and *TADA2A*) specific elements (**Fig. 4.9A-B**). This suggests that control of cell survival is not central to KAT2A complex-mediated maintenance of MOLM-13 AML cells. As for depletion of KAT2A itself, there is a trend towards enhanced apoptosis, which could be an end-point consequence of enhanced differentiation (**Fig. 4.8**).

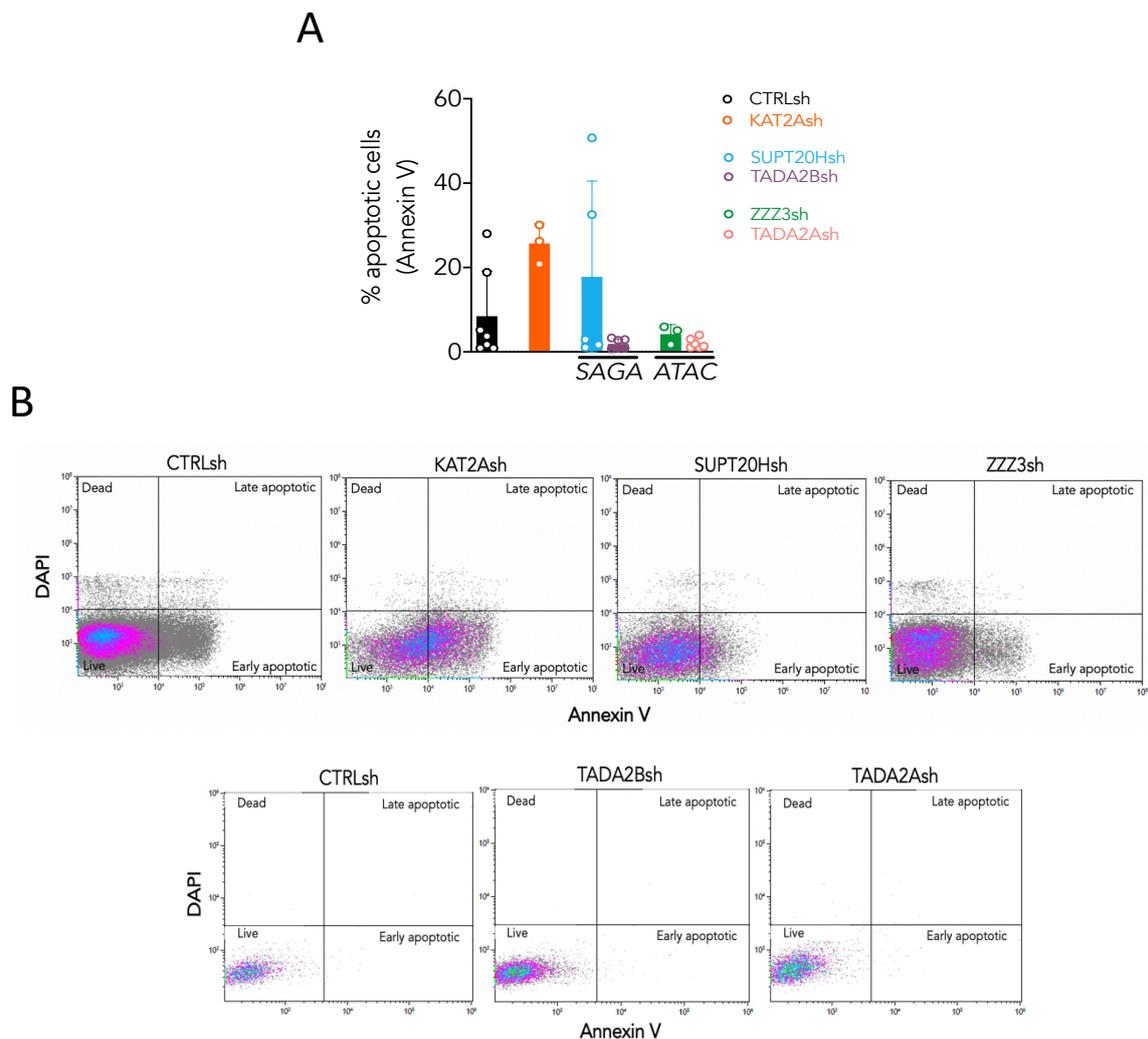
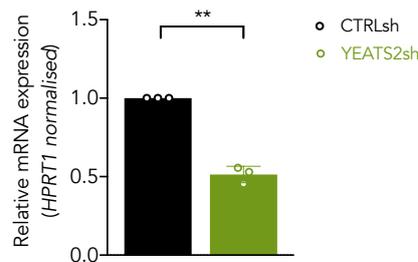


FIGURE 4.9 Apoptosis analysis of MOLM-13 cells upon depletion of KAT2A, SAGA-specific *SUPT20H* and ATAC-specific *ZZZ3* subunits. (A) Quantification of Annexin V+ apoptotic cells in MOLM-13 cultures analysed by flow cytometry. Mean \pm SEM of N>3 independent experiments; 2 biological replicates; 3 technical replicates for *KAT2Ash*, *TADA2Bsh*, and *ZZZ3sh* groups; 6 technical replicates for *SUPT20Hsh*) Two-tailed t-test for significance. **(B)** Representative flow cytometry analysis of apoptosis by Annexin V staining in MOLM-13 cells transduced with *CTRLsh*, *KAT2Ash*, *SUPT20Hsh*, *ZZZ3sh* (top) and *CTRLsh*, *TADA2Bsh* and *TADA2Ash* (bottom). The combination of Annexin V and DAPI

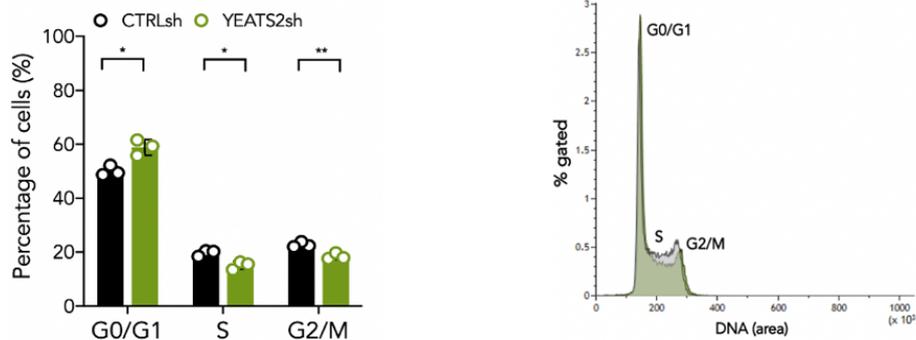
staining allows for a distinction between viable cells (double negative), cells in early apoptosis (Annexin V positive), cells in late apoptosis (double positive) and dead cells (DAPI positive). $N \geq 3$ individual samples; mean \pm SEM. Two-tailed t-test for significance; no significant differences.

Analysis of an additional ATAC-subunit, YEATS2 (**Fig. 4.10A**), also resulted in defects in cell cycle progression (**Fig. 4.10B**). Consistent with this observation, a compelling study showed dysregulation of cell cycle genes in human cancer cell lines upon shRNA-mediated depletion of *YEATS2* [235]. No changes on cellular apoptosis were noted (**Fig. 4.10C**).

A



B



C

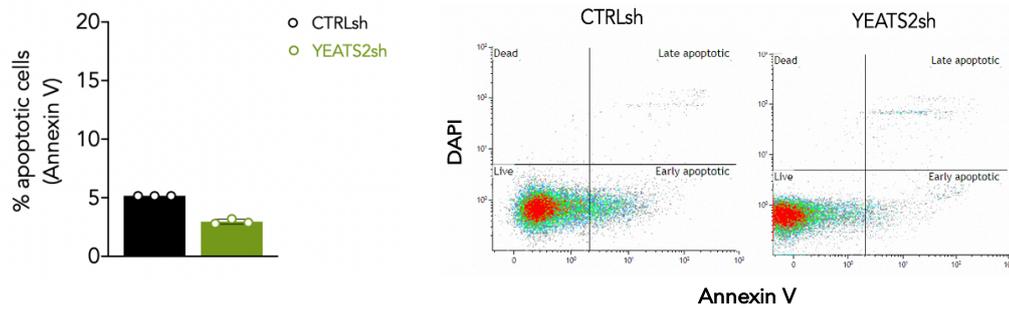


FIGURE 4.10 Quantification of flow cytometry analysis of cell cycle and apoptosis in MOLM-13 cells transduced with shRNA construct against ATAC *YEATS2* component. (A) Quantitative RT-PCR validation of *YEATS2* (ATAC-specific subunit) knockdown in MOLM-13 cells. $N \geq 3$ independent experiments, mean \pm SEM of gene expression relative to *CTRLsh*, normalised to *HPRT1* housekeeping gene. Two-tailed t-test for significance * $p < 0.05$, ** $p < 0.01$. **(B)** Flow cytometry analysis of cell cycle in KG1a cells transduced with ATAC-specific *YEATS2sh*. Mean \pm SEM of 3 independent experiments. Two-tailed t-test for significance; no significant changes. **(C)** Flow cytometry analysis of apoptosis in KG1a cells transduced with shRNA constructs against ATAC-*YEATS2*. Mean \pm SEM of % Annexin V positive cells in 3 independent experiments. Two-tailed t-test for significance * $p < 0.05$.

Overall, the data supports distinct roles for ATAC and SAGA in the model AML cell line MOLM-13. ATAC regulates proliferation/cell division likely through molecular control of biosynthetic machinery. SAGA, on the other hand, preserves the undifferentiated status of AML cells, its loss promoting differentiation. In light of the CRISPR screen, these results unveil previously uncharacterised dependencies of ATAC and SAGA elements in MOLM-13, serving to illustrate the importance of functional hit validation.

THP-1

I next investigated cellular consequences of perturbing ATAC and SAGA elements in another *MLL-AF9* cell line model, THP-1. Analysis of cell cycle revealed abnormalities upon perturbation of both ATAC and SAGA elements (**Fig. 4.11**), including shared KAT2A.

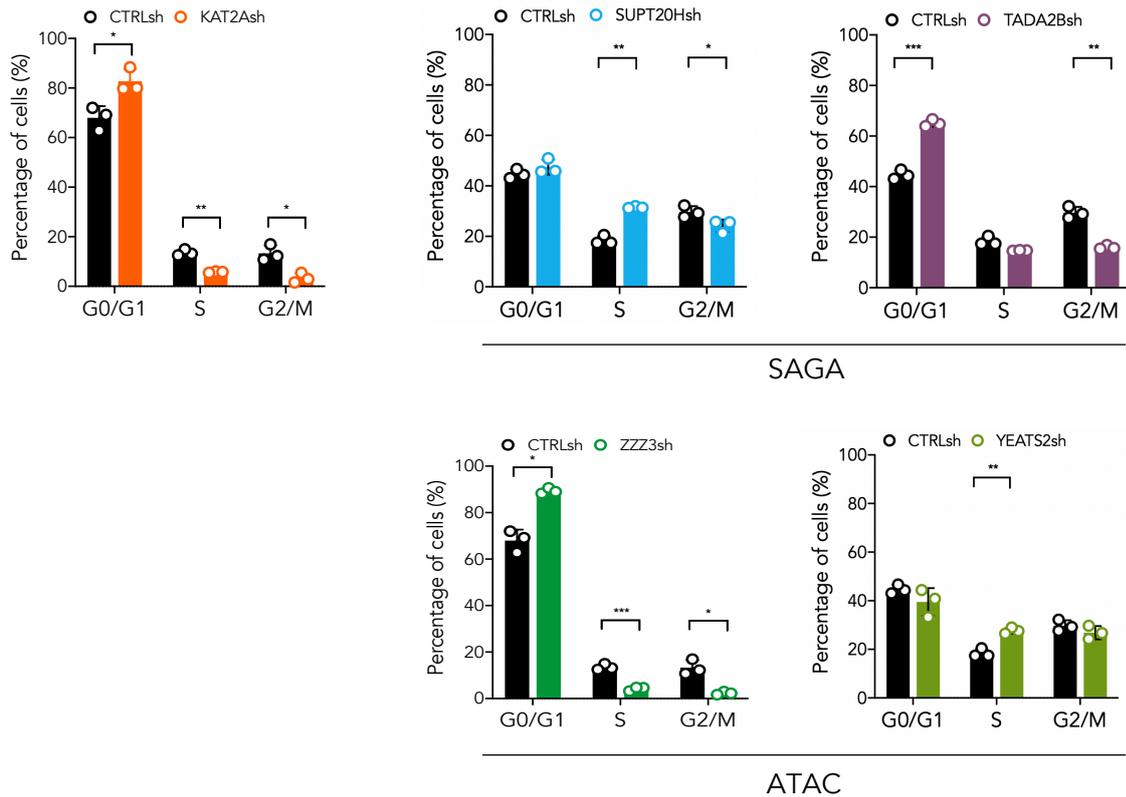


FIGURE 4.11 Cell cycle analysis of THP-1 cells upon knockdown of *KAT2A* and SAGA (*SUPT20H*, *TADA2B*) and ATAC (*ZZZ3*, *YEATS2*) elements. Quantification of flow cytometry analysis of cell cycle in THP-1 cells transduced with *KAT2Ash*, SAGA-specific *SUPT20Hsh* and *TADA2Bsh* and ATAC-specific *ZZZ3sh* and *YEATS2sh*. Mean \pm SEM of 3 independent experiments. Two-tailed t-test for significance * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Although the effects of *SUPT20H* and *ZZZ3* knockdown are similar to those observed in MOLM-13 (**Fig. 4.6**), the cell cycle changes do not seem to be captured by other members of the KAT2A complexes SAGA and ATAC in THP-1, *SUPT20H* and *YEATS2*, respectively.

On the other hand, analysis of differentiation (see methods section) is consistent with observations in MOLM-13 (**Fig. 4.8**), whereby depletion of SAGA complex subunits, *SUPT20H* and *TADA2B*, promotes cell differentiation (**Fig. 4.12**), which is not visible upon loss of ATAC-*YEATS2*.

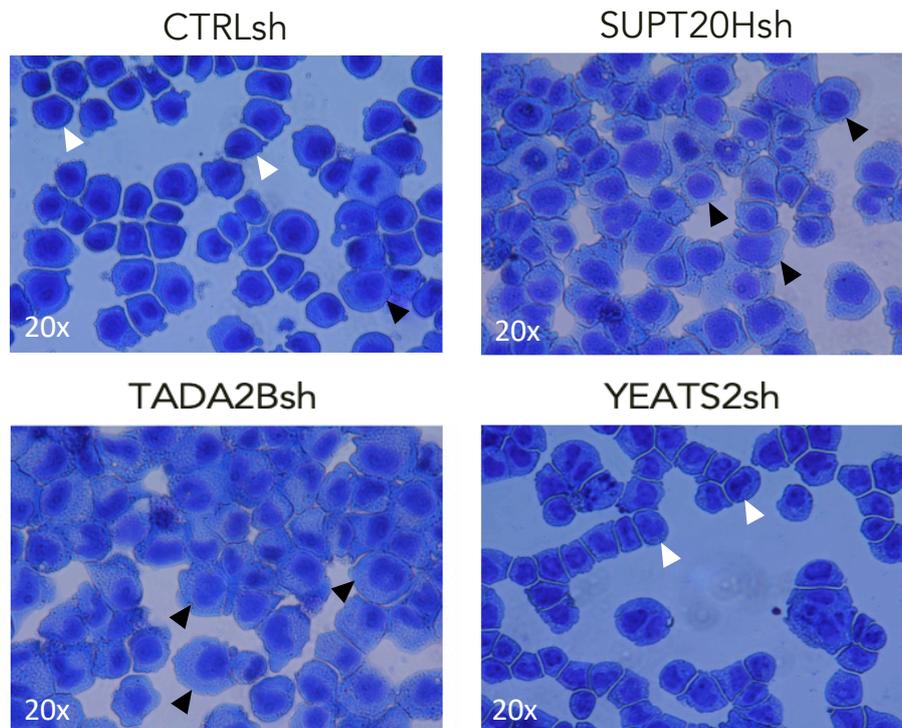


FIGURE 4.12 Representative photographs of THP-1 cytopins upon depletion of SAGA-specific *SUPT20H* and *TADA2B*, and ATAC-specific *YEATS2* subunits. White arrow heads denote blast-like cells; black arrow heads denote differentiated cells.

Figure **Fig. 4.13A-B** shows analysis apoptosis in THP-1 cells, for which no significant differences were found upon knockdown of unique SAGA or ATAC members.

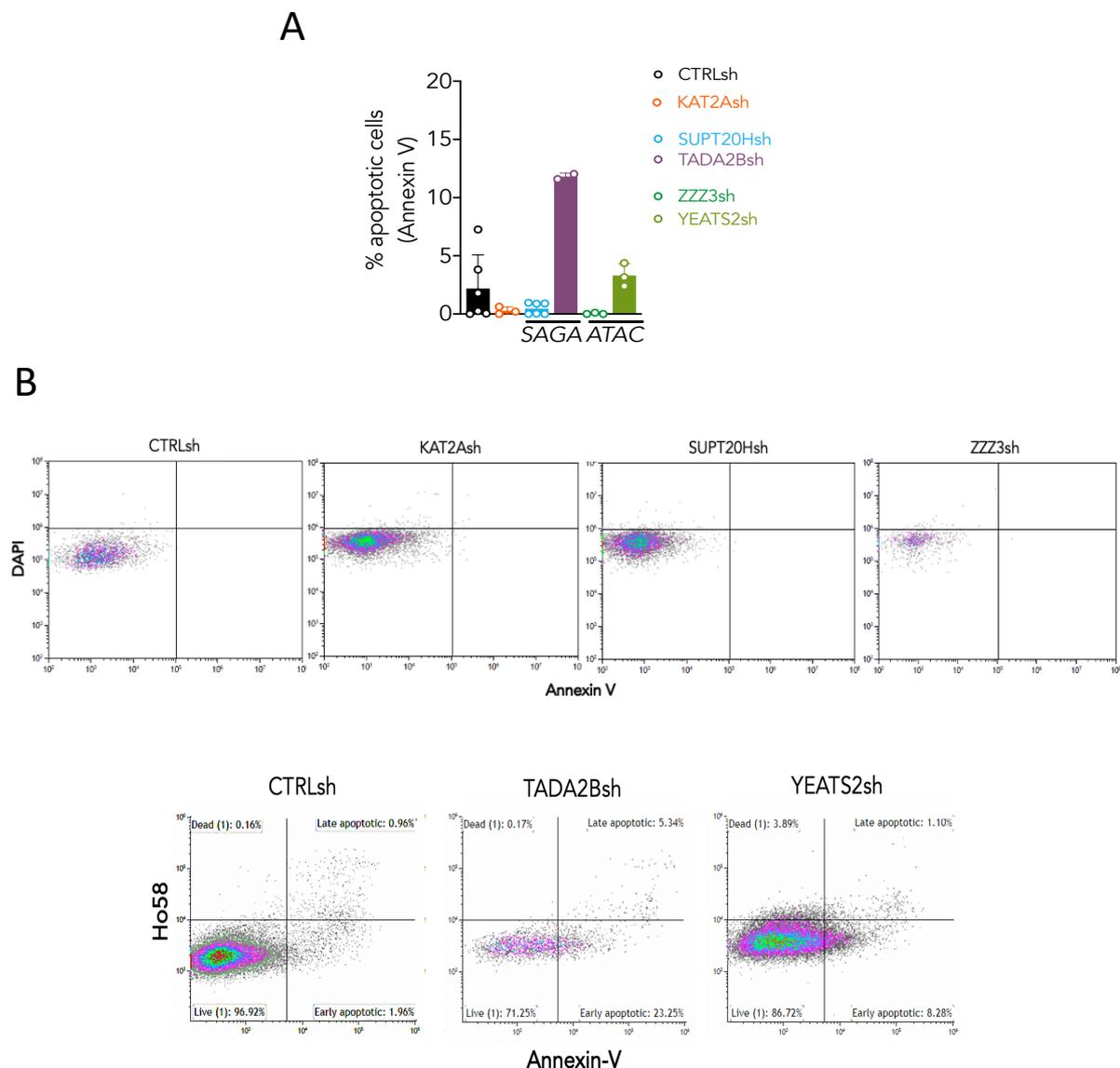


FIGURE 4.13 Apoptosis analysis of THP-1 cells upon depletion of *KAT2A*, SAGA-specific *SUPT20H* and *TADA2B* ATAC-specific *ZZZ3* and *YEATS2* subunits. (A) Quantification of Annexin V+ apoptotic cells in THP-1 cultures analysed by flow cytometry. Mean \pm SEM of N>3 independent experiments; 2 biological replicates for *SUPT20Hsh*; 3 technical replicates *KAT2Ash*, *TADA2Bsh* (1 replicate was lost due to technical failure), and the *ZZZ3sh* and the *YEATS2sh* groups. Two-tailed t-test for significance. **(B)** Representative flow cytometry analysis of apoptosis by Annexin V staining in THP-1 cells transduced with *CTRLsh*, *KAT2Ash*, *SUPT20Hsh*, *ZZZ3sh* (top) and *CTRLsh*, *TADA2Bsh* and *YEATS2sh* (bottom).

In summary, these data suggests that THP-1 survival is not dependent on the function of KAT2A and its complexes, although it is arguable that as this analysis was performed shortly after transduction, the cells may not have been kept in culture for a sufficient period of time to observe differences. Moreover, whilst depletion of ATAC *ZZZ3* lead to cell cycle arrest as previously observed in MOLM-13, the same effect was not observed upon knockdown of a second ATAC element, *YEATS2*. The effects in cell differentiation, on the other hand, are consistent with a role of SAGA elements in promoting differentiation of cultured cells upon its knockdown. Given that differentiated cells eventually exit the cell cycle [295], it is possible that the increase in cell differentiation shown in THP-1 followed by loss of *TADA2B*, reflects the cell cycle profile, in which cells abnormally accumulated in G0/G1 (**Fig 4.11**).

MV4-11

MV4-11 is another mix-lineage leukaemia, carrying a *MLL-AF4* translocation. In addition, it has a *FLT3-ITD* mutation, one of the most common gain-of-function mutations in AML, occurring in nearly 30% cases. With the caveat that this experiment was only performed once, knockdown of *KAT2A* did not seem to impact cell apoptosis (**Fig. 4.14 A-B**) nor cell cycle progression (**Fig. 4.14 C-D**). Despite an absolute need for repeating, this experiment matched the result of the AML CRISPR screen, where *KAT2A* did not come up as a dependency in this AML.

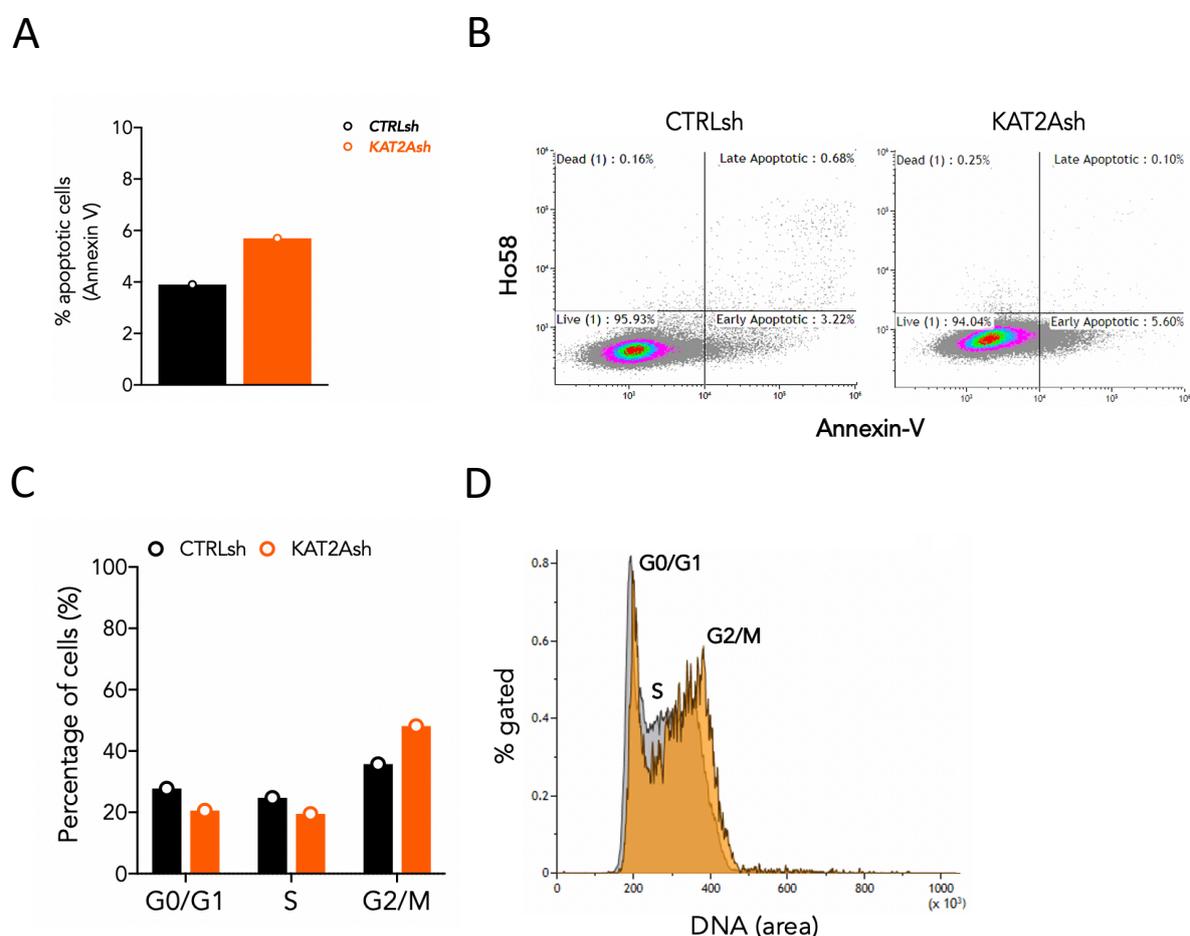
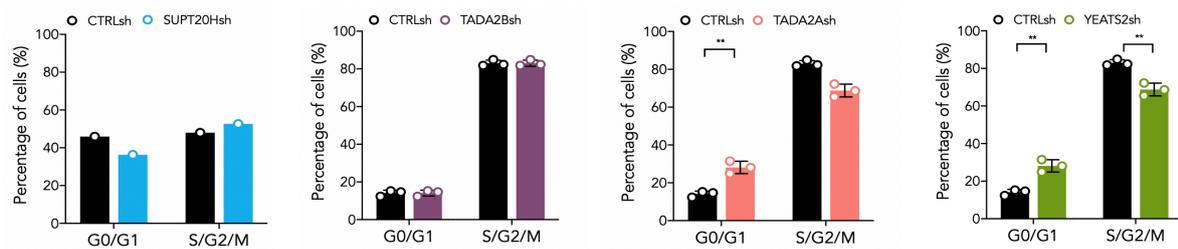


FIGURE 4.14 Apoptosis and cell cycle analysis of MV4-11 cell line upon *KAT2A* depletion. **(A)** Flow cytometry analysis of apoptosis in MV4-11 cells transduced with *KAT2Ash*; representative experiment. **(B)** Representative flow cytometry plots of apoptosis analysis in (A). **(C)** Flow cytometry analysis of cell cycle in MV4-11 cells transduced with *KAT2Ash*; representative experiment. **(D)** Representative flow cytometry overlays of cell cycle analysis in (C).

Yet, I proceeded to dissecting complex specific essentialities by looking at the cellular consequences of depletion of *TADA2B* (SAGA) and *TADA2A* and *YEATS2* (ATAC). Surprisingly, cell proliferation analysis revealed a G0/G1 defect upon loss the two ATAC elements, *TADA2A* and *YEATS2* (**Fig. 4.15 A-B**). It was somewhat unexpected to see changes associated with depletion of members of the complexes, given that loss *KAT2A* itself (although n=1), did not affected MV4-11 cell status.

A



B

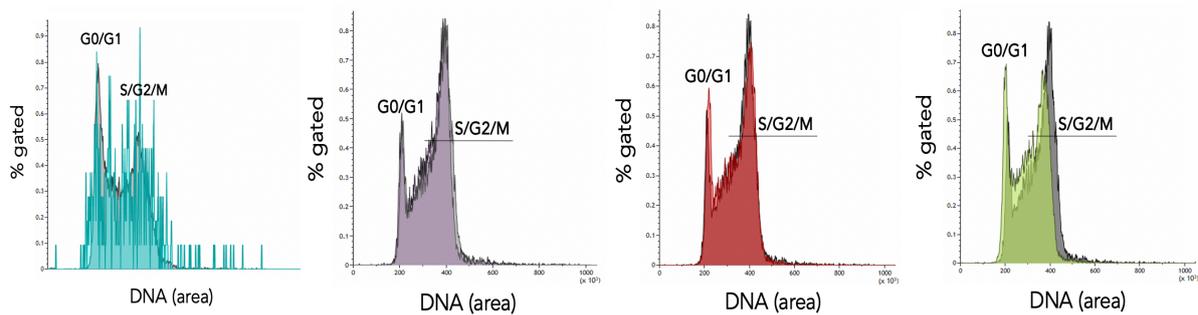


FIGURE 4.15 Cell cycle assay in MV4-11 cell line upon knockdown of SAGA-specific *SUPT20H* and *TADA2B* and ATAC-specific *TADA2A* and *YEATS2* subunits.

(A) Flow cytometry analysis of cell cycle in THP-1 cells transduced with SAGA-specific *TADA2Bsh*, and ATAC-specific *TADA2Ash* and *YEATS2sh*. Mean \pm SEM of 3 independent experiments. Two-tailed t-test for significance; no significant changes. **(B)** Representative flow cytometry overlays of cell cycle analysis in MV4-11 cells transduced with shRNA constructs against *TADA2B*, *TADA2A* and *YEATS2*.

I next assessed differentiation changes, by staining the cells with a CD13 antibody, and looking at its expression by flow cytometry in MV4-11 cells transduced with *SUPT20Hsh* (SAGA) and *YEATS2sh* (ATAC). It was possible to observe a mild increase in the percentage of cells expressing CD13 uniquely upon *SUPT20H* knockdown, which might be indicative to induced differentiation. In a similar vein, *SUPT20Hsh* resulted in increased cell differentiation in both MOLM-13 (**Fig. 4.8**) and MV4-11 (**Fig. 4.12**).

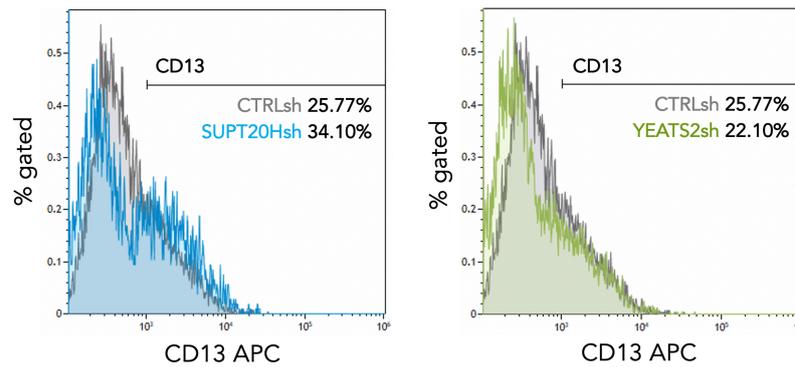


FIGURE 4.16 Flow cytometry analysis of CD13 expression in MV4-11 cells transduced with CTRLsh, SAGA-specific SUPT20Hsh and ATAC-specific YEATS2sh. Representative experiment.

No significant changes were noted in cell survival (**Fig. 4.17A-B**) followed by knockdown of any of the subunits analysed.

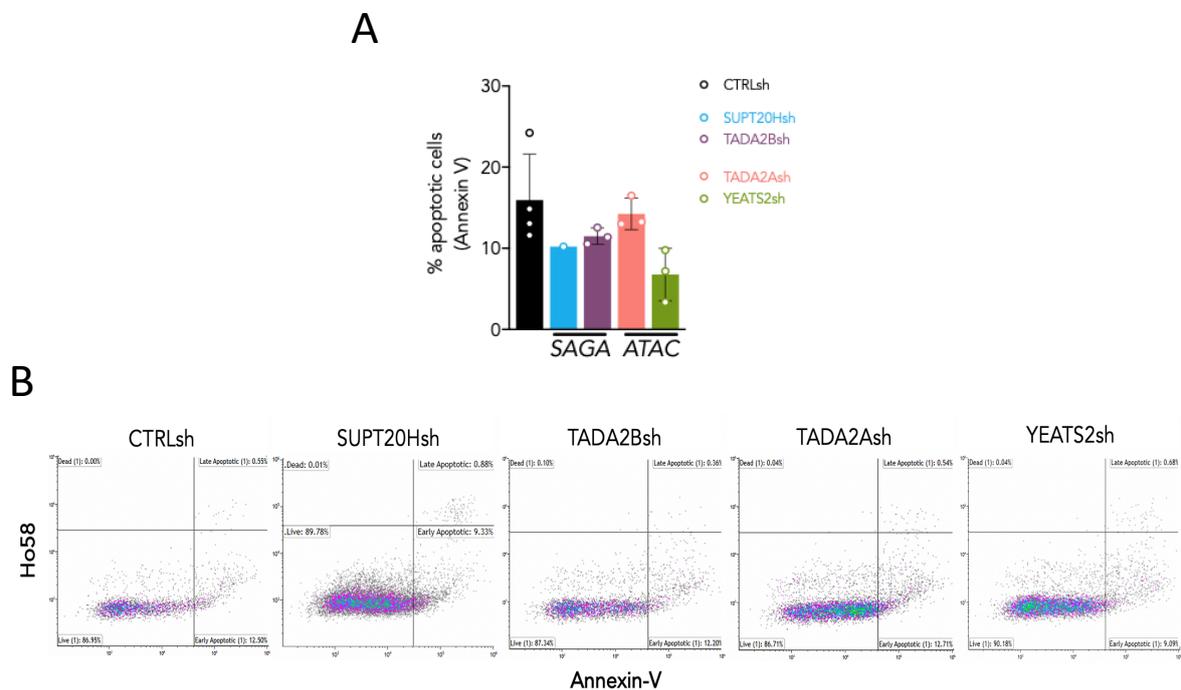


FIGURE 4.17 Apoptosis analysis of MV4-11 cells upon knockdown of SAGA-specific *SUPT20H* and *TADA2B* and ATAC-specific *TADA2A* and *YEATS2* subunits. (A) Flow cytometry analysis of apoptosis in MV4-11 cells transduced with shRNA constructs against SAGA-TADA2B, and ATAC-TADA2A and YEATS2. Mean \pm SEM of % Annexin V positive cells in 3 independent experiments. Two-tailed t-test for significance; no significant changes. **(B)** Representative flow cytometry plots of apoptosis analysis of MV4-11 cells transduced with *CTRLsh*, *TADA2Bsh*, *TADA2Ash* and *YEATS2sh*. N=2 biological replicates; 3 technical replicates for *TADA2Ash*, *TADA2Bsh* and *YEATS2sh*.

These results implicate the ATAC complex in regulation of the cell cycle, as observed in MOLM-13, and a potential role for SAGA in keeping MV4-11 phenotype. Whilst a role for KAT2A itself in MV4-11 biology cannot be excluded, and needs further assessment, the experiment performed here and the result of the CRISPR screen do not indicate a dependency on KAT2A.

I then went on to extend the cellular analysis of KAT2A and its complexes in CD34+ leukaemia, making use of KG1a and Kasumi-1 cell lines.

4.3.2 KAT2A complexes have distinct contributions to proliferation and cell identity of CD34+ AML cell lines

KG1a

KG1a carry a *FGFR1OP2-FGFR1* fusion, involving fibroblast growth factor receptor 1 (FGFR1) gene, and represent a very undifferentiated AML type. As shown in **Fig. 4.18**, depletion of *TADA2A* (ATAC), resulted in cell cycle arrest in G0/G1, which I had also observed in MOLM-13 and MV4-11 cell lines. Thus, ATAC control of cell proliferation seems to be a consist feature, despite differences in genetic background between cell lines.

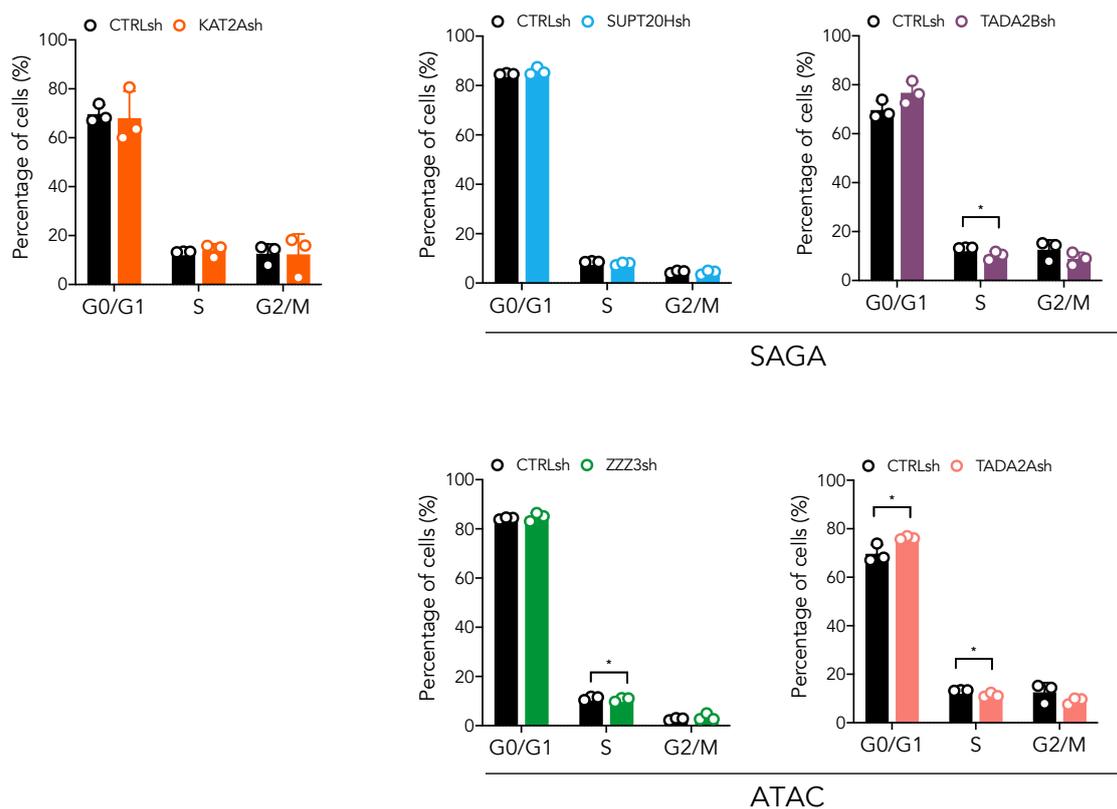


FIGURE 4.18 Quantification of flow cytometry analysis of cell cycle in KG1a AML cells transduced cells with shRNA constructs against *KAT2A*, *SAGA* (*SUPT20H*, *TADA2B*) and *ATAC* (*ZZZ3*, *TADA2A*) elements. N=3 individual experiments; mean \pm SEM. Two-tailed t-test for significance * $p < 0.05$.

Also consist with observation in the previous AML models, no changes in apoptosis were observed upon loss of either ATAC or SAGA components (**Fig. 4.19**).

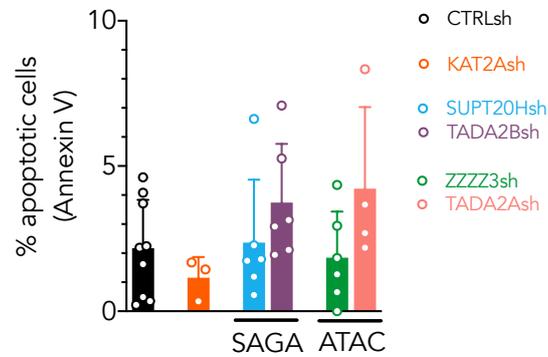


FIGURE 4.19 Quantification of flow cytometry analysis of apoptosis by Annexin V staining in KG1a cells transduced cells with shRNA constructs against *KAT2A*, SAGA (*SUPT20H*, *TADA2B*) and ATAC (*ZZZ3*, *TADA2A*) elements. $N \geq 3$ individual samples mean \pm SEM; 2 biological replicates for *SUPT20Hsh*, *TADA2Bsh* and *ZZZ3sh* groups (6 technical replicates each); 3 technical replicates for *KAT2Ash* and 4 technical replicates for *TADA2Ash*. Two-tailed t-test for significance; no significant differences.

Furthermore, when investigating shRNA-mediated knockdown of *YEATS2* (ATAC complex), I observed a trend towards G0/G1 arrest ($p=0.06$), suggesting reduced cell proliferation (**Fig. 4.20A**), with no impact on cellular apoptosis (**Fig. 4.20B**).

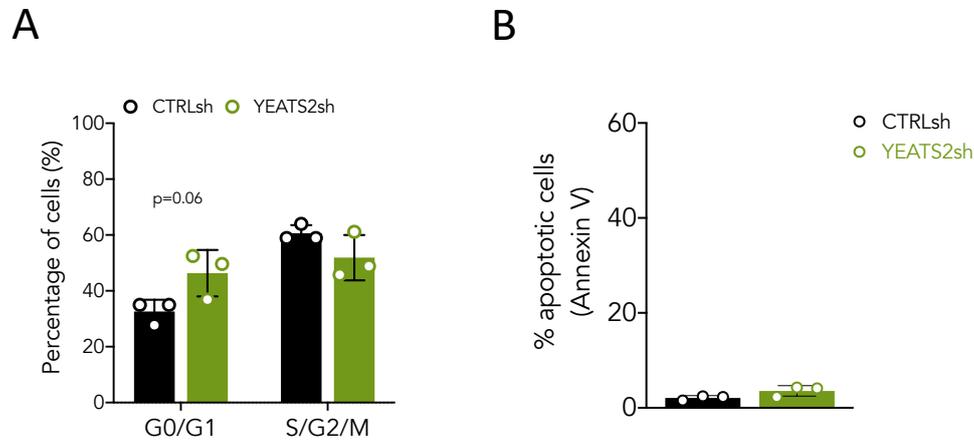


FIGURE 4.20 Quantification of flow cytometry analysis of cell cycle and apoptosis in KG1a cells transduced with shRNA construct against ATAC *YEATS2* component. (A) Flow cytometry analysis of cell cycle in KG1a cells transduced with ATAC-specific *YEATS2sh*. Mean \pm SEM of 3 independent experiments. Two-tailed t-test for significance; no significant changes. **(B)** Flow cytometry analysis of apoptosis in KG1a cells transduced with shRNA construct against ATAC-*YEATS2*. Mean \pm SEM of % Annexin V positive cells in 3 independent experiments. Two-tailed t-test for significance; no significant changes.

When assessing differentiation consequences of loss of ATAC or SAGA components, and I observed downregulation of the CD34 marker by flow cytometry upon loss of both *SUPT20H* (SAGA) and *ZZZ3* (ATAC) (**Fig. 4.21A-B**). However, morphological differentiation along the monocytic lineage (see methods section) (**Fig. 4.21C-D**) was only verified when depleting *SUPT20H*.

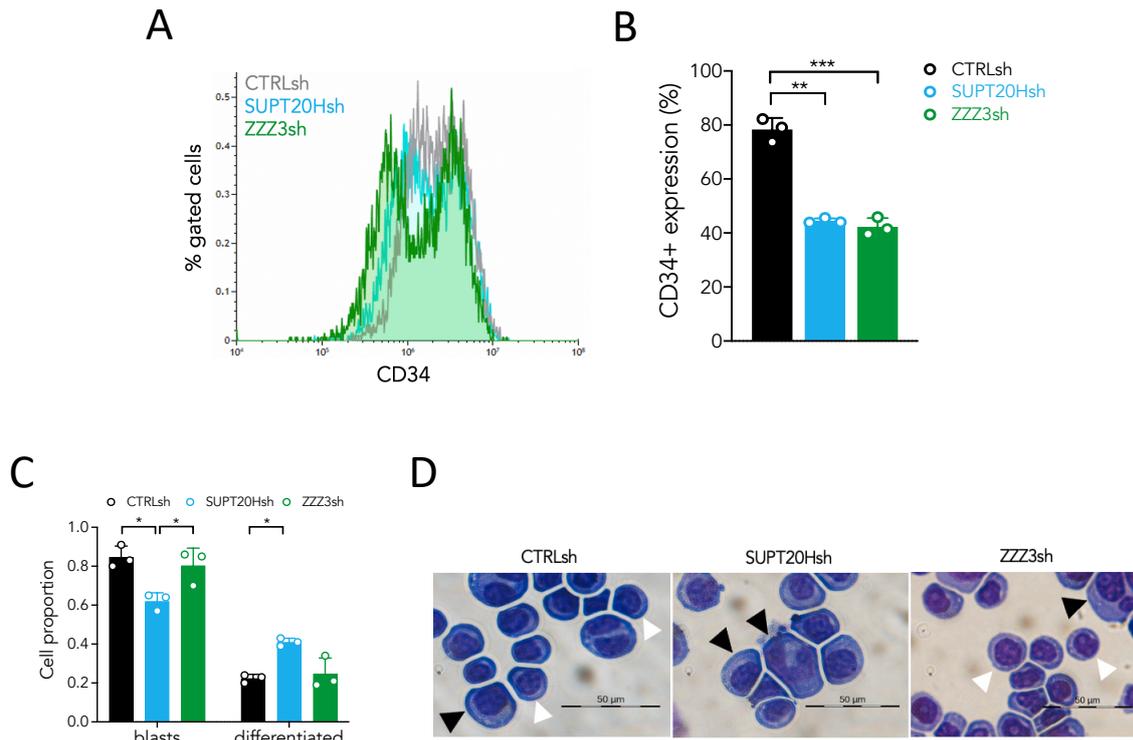


FIGURE 4.21 Analysis of differentiation status of KG1a cells upon knockdown of ATAC specific ZZZ3 and SAGA specific SUPT20H elements. (A) Representative analysis of undifferentiated marker CD34 in KG1a cells transduced with *CTRLsh*, *SUPT20Hsh* and *ZZZ3sh*. **(B)** Quantification of CD34+ cells in transduced KG1a cells in (A). N=3 independent experiments. Two-tailed t-test for significance * $p < 0.05$, ** $p < 0.01$. **(C)** Quantification of blast-like and differentiated cells in KG1a cultures transduced with *CTRLsh*, *SUPT20Hsh* and *ZZZ3sh*. Scoring of 3 randomly selected fields of >100 cells; Two-tailed t-test for significance; * $p < 0.05$. **(D)** Representative photographs of KG1a cytopins. White arrow heads denote blast-like cells; black arrow heads denote differentiated cells. Scale 50 μm .

Differentiation following depletion of *SUPT20H* was apparent also in the previously investigated cell line models, supporting a predominant role for SAGA activities in sustaining the characteristic undifferentiated state/identity of AML cells. On the other hand, the data points to a pervasive role for ATAC elements in cell cycle control of AML carrying *FGFR1OP2-FGFR1*.

Kasumi-1

Lastly, I sought out to investigate Kasumi1 which like KG1a represents a minimally differentiated AML. Kasumi-1 harbour a *RUNX1-RUNX1T1* fusion, one of the most frequent chromosomal rearrangements in AML, and an associated *KIT* mutation [296]. Similar to what had been observed in MOLM-13 and THP-1, loss of *ZZZ3*, restricted cell cycle progression with cells accumulating in G0/G1 (**Fig. 4.22**).

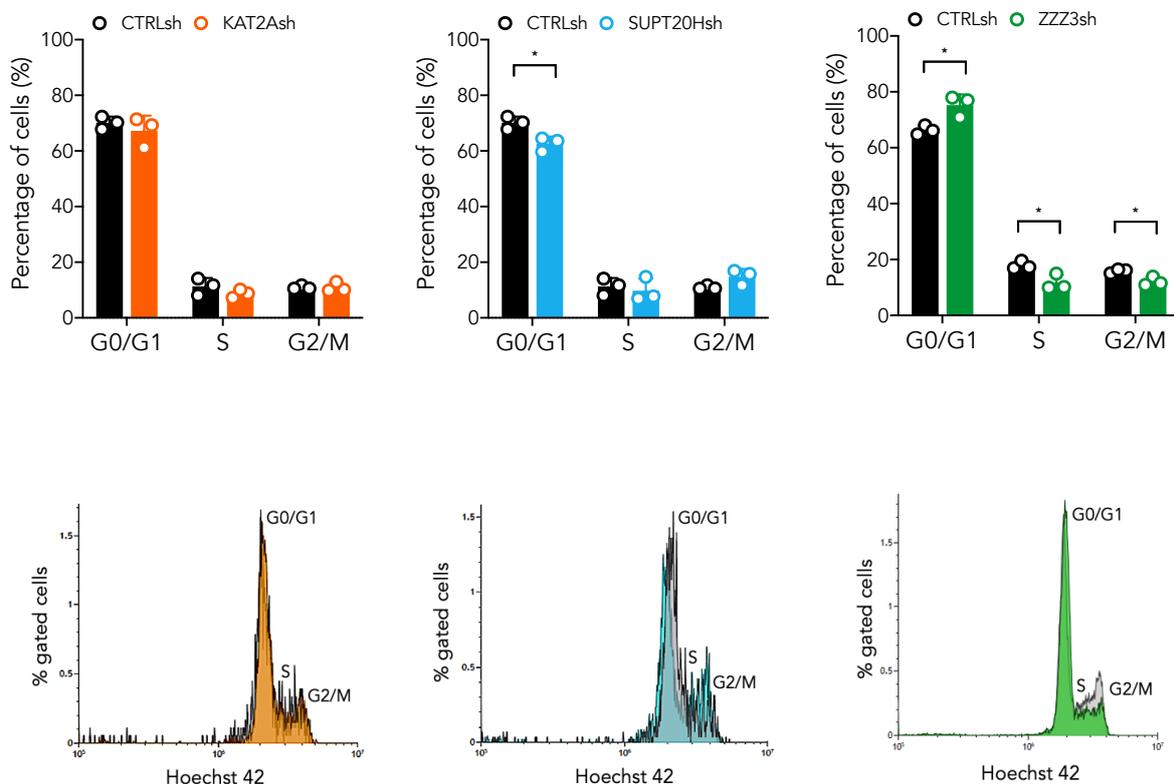


FIGURE 4.22 Cell cycle analysis of Kasumi-1 cells upon knockdown of *KAT2A*, *SAGA*-specific *SUPT20H* and *ATAC*-specific *ZZZ3*. Quantification and representative of flow cytometry plots of cell cycle in Kasumi-1 cells transduced with *CTRLsh*, *KAT2Ash*, *SAGA*-specific *SUPT20Hsh* and *ATAC*-specific *ZZZ3sh*. Mean \pm SEM of 3 independent experiments. Two-tailed t-test for significance * $p < 0.05$, ** $p < 0.01$.

When assessing the differentiation status of Kasumi-1 cells following knockdown of SAGA (*SUPT20H* and *TADA2B*) or ATAC (*ZZZ3* and *TADA2A*), no differences in CD34 surface expression were seen, when compared to CTRLsh (**Fig. 4.23A**).

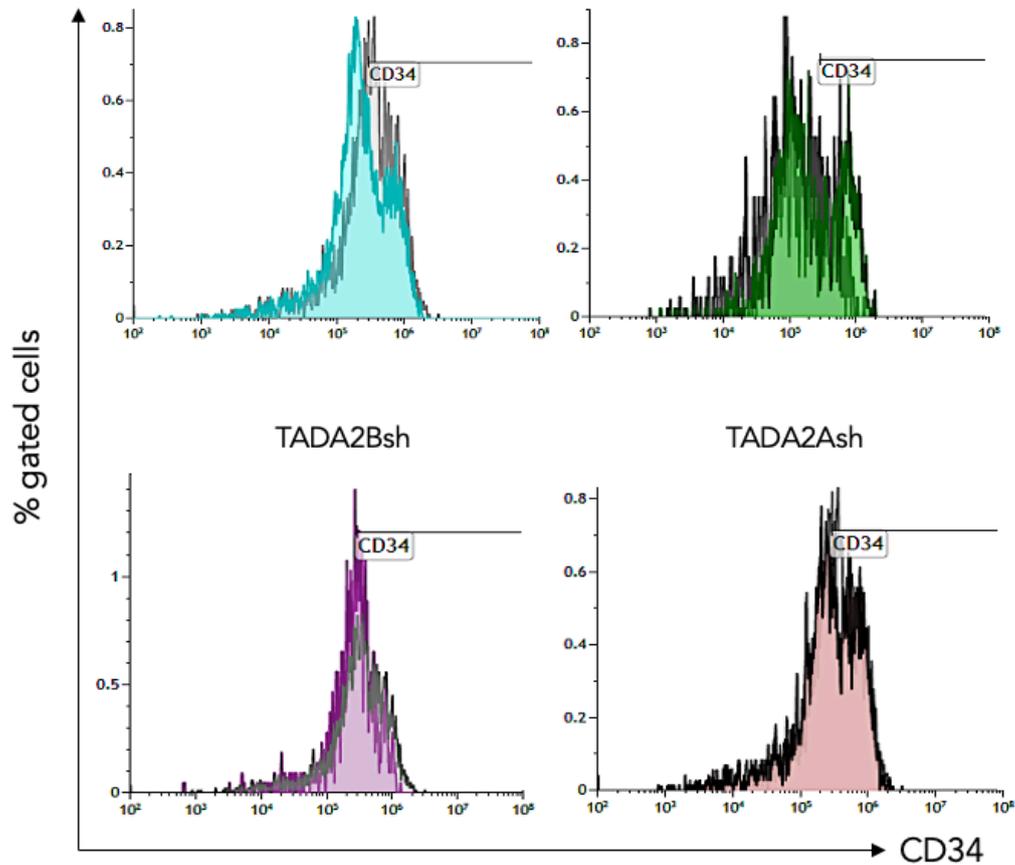


FIGURE 4.23 Analysis of differentiation status of Kasumi-1 cells upon knockdown of SAGA specific *SUPT20H* and *TADA2B* and ATAC specific *ZZZ3* and *TADA2A* elements. Representative analysis of undifferentiated marker CD34 in Kasumi-1 cells transduced with *CTRLsh*, *SUPT20Hsh*, *TADA2Bsh*, *ZZZ3sh* and *TADA2Ash*.

As in previous cell lines, Kasumi-1 did not display changes in cell apoptosis in response to loss of KAT2A or any of the complex-specific elements (**Fig. 4.24**).

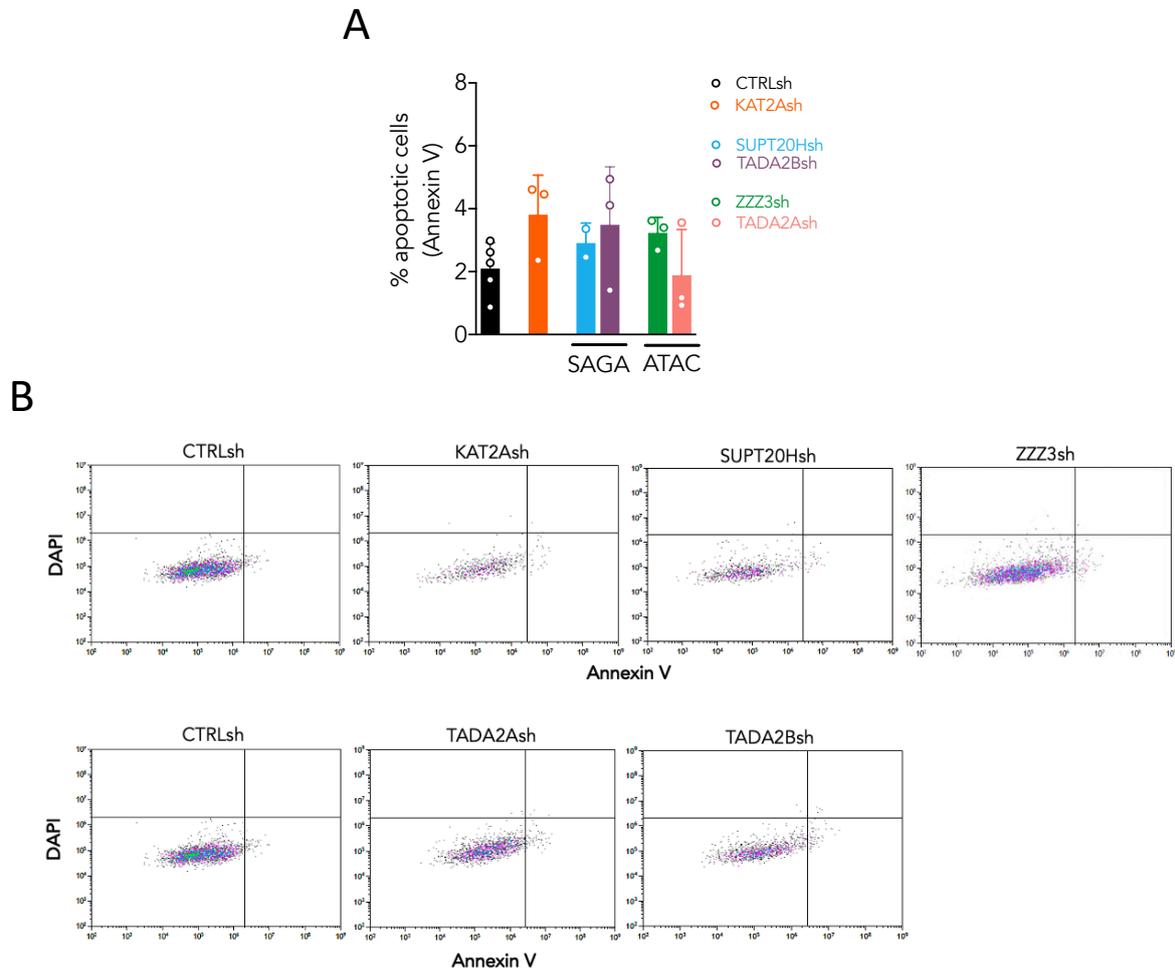


FIGURE 4.24 Apoptosis analysis of Kasumi-1 cells transduced with *KAT2Ash*, SAGA-specific *SUPT20Hsh* and *TADA2Bsh* and ATAC-specific *ZZZ3sh* and *TADA2Ash*. (A) Quantification of Annexin V⁺ apoptotic cells in Kasumi-1 cultures analysed by flow cytometry. Mean \pm SEM of N>3 independent experiments. N=2 biological replicates; 3 technical replicates for *KAT2Ash*, *SUPT20Hsh*, *TADA2Bsh*, *TADA2Ash* and *ZZZ3sh*. Two-tailed t-test for significance; no significant differences. (B) Representative flow cytometry analysis of apoptosis by Annexin V staining in Kasumi-1 cells transduced with *CTRLsh*, *KAT2Ash*, *SUPT20Hsh*, *ZZZ3sh* (top) and *CTRLsh*, *TADA2Ash* and *TADA2Bsh* (bottom). The combination of Annexin V and DAPI staining allows for a distinction between viable cells (double negative), cells in early apoptosis (Annexin V positive), cells in late apoptosis (double positive) and dead cells (DAPI positive).

Altogether, these cellular analyses show some discrepancies in regulation of Kasumi biology by the two chromatin complexes. Whereas cell cycle regulation seems dependent on ATAC (*ZZZ3*) thus likely to mediate proliferation, as previously observed, members of SAGA do not contribute to t(8;21) AML phenotypes.

4.4 Conclusions and future work

In this Chapter, I performed cellular characterisation of the roles of KAT2A complexes, ATAC and SAGA, in cultured AML cell lines, describing the phenotypic readouts of knocking down specific elements of either complex. The deliberate use of five different AML cell lines, each with markedly different mutations, and which clinically associate with a range from good to poor prognosis, allowed for a better experimental representation of the considerable human AML heterogeneity. Both complexes were found to contribute to AML biology, although mediating different cellular activities. Results are summarised in **Table 4.1**.

Table 4.1 Summary of KAT2A-complex cellular effects in the AML cell lines studied.

	MOLM-13			THP-1			MV4-11			KG1a			Kasumi-1		
	CC arrest	apoptosis	differentiation												
KAT2Ash															
TADA2Ash															
ATAC															
ZZZ3sh															
YEATS2sh															
SAGA															
TADA2Bsh															
SUPT20Hsh															

It became apparent that removing elements of the ATAC complex including *ZZZ3*, *YEATS2* and *TADA2A* limited cell proliferation capacity in different AML cell lines. In the case of MOLM-13, the most extensively characterised cell line here, it was further demonstrated by cell divisional tracking that less cells enter cell division, uniquely when depleting *ZZZ3*. Although, the exact mechanism by which ATAC causes the cell cycle phenotype, remains to be understood. For example, it was previously shown that KAT2A plays roles in cell cycle progression through direct acetylation of CDC6 [297] and cyclin A (G1/S cell cycle transition) the latter, within its participation in the ATAC complex [294]. On the other hand, as demonstrated in MOLM-13, ATAC-*ZZZ3* depleted cells had lost H3K9ac at promoters of ribosomal protein genes, *RPS7* and *RPL13*, which associated with a decrease in their gene expression. This is in

agreement with a role for ATAC-mediated H3K9ac activity in controlling ribosomal protein and translation associated genes, which are essential pathways to cell survival and proliferation. Signer and colleagues have previously shown that AML stem-like cells are highly dependent on control of translational activity [298]. Thus, it is possible that the increased fraction in G0/G1 observed across AML cell lines followed by ATAC loss, reflects downregulation of the biosynthetic/ribosomal protein machinery targeted by the ATAC complex. Indeed, ATAC histone acetylation activity has also been associated with regulation of biosynthetic pathways in other cancer types beyond leukaemia such as NSCLC [235–236]. Thus, ATAC likely contributes to AML through preservation of metabolism required to maintain leukaemia stem-like cells (LSC). Given that the cell cycle experiments performed here did not distinguish between G0 and G1 phases, this is something that could be looked at in future experiments to investigate regulation of LSC quiescence. One possible approach would be to enforce cells into G0 quiescence by serum deprivation and see if cells in which ATAC elements have been knockdown down accumulated at that stage or could progress to G1. It would also be interesting to measure protein synthesis activity in the absence of ATAC and SAGA, to understand if leukaemia cells are indeed less translationally active when removing ATAC members. This could be achieved by an OP-puro incorporation assay.

On the other hand, the cellular events observed upon knockdown of SAGA-specific elements associated with enhanced AML differentiation. This was verified both in differentiated/monocytic-type leukaemic cell lines, such as MOLM-13 and THP-1, but also in a more immature or less differentiated AML, KG1a. While this suggests that SAGA contributions to the characteristic AML differentiation block are irrespective of the exact cell differentiation state, which could be an indication that SAGA makes use of the transcriptional machinery available on specific cell types rather than having a fix set of target genes. An exception to the pro-differentiative effects of SAGA shRNA treatment was Kasumi-1, for which no changes were noted upon depletion of SAGA elements. Again, this discrepancy could be an indication that SAGA operates in a cell-type/context manner to favour the AML maturation arrest. In other words, maybe the genes targeted/regulated by SAGA are less likely to be shared between leukaemia types. Mechanistic studies are essential to understand this. *SUPT20H* loss of function, in particular, showed minimal consequences to normal haematopoiesis, which may suggest potential therapeutic targeting of SAGA Core module for differentiation-based therapies in AML. However,

based on its role in erythropoiesis (K562), the context in which its inhibition for therapeutic benefit is proposed is clearly of significant importance. This warrants further studies to rule out the possibility that dynamic interactions with subunits of other SAGA modules would have detrimental consequences in normal blood.

Considering the molecular results in MOLM-13, loss of H3K9ac had a more profound effect at promoters of both ribosomal protein and *MLL-AF9 HOXA* target genes, than their relative mRNA levels followed by loss of *ZZZ3* or *SUPT20H*. Two interpretations can be drawn for this observation: (1) regulation by *KAT2A* complexes may be cell type specific, as this was not observed in K562 cells; (2) this may suggest that specific functions on transcription regulation by SAGA or ATAC at a given promoter region, may require additional enzymatic activities of their neighbouring subunits (H2B deubiquitination by *USP22*) and (H4 acetylation by *KAT14*), respectively. In a similar way, they could also involve regulation by the non-catalytic elements. For example, in the case of disruption of ATAC complexes in the absence of *ZZZ3*, this may affect H3K27ac recognition by *YEATS2* and/or, interaction with *MLL* complexes *via* *WDR5*, resulting in a more dramatic transcriptional effect. Alternatively, *KAT2A* activity in MOLM13 may involve H3K79 succinylation [180] or other yet uncharacterised acetylation marks. Furthermore, it is possible that full activity requires protein-protein interaction and/or acetylation, which may be different between complexes. This emphasises the intricacy of multi-modular chromatin complexes to regulation of transcription and further studies investigating combined knockdowns of these subunits in AML would help understanding how they drive the leukaemic potential.

Our lab has previously shown a requirement for *Kat2a* in sustaining leukaemia stem-like cells (LSC) in a *MLL-AF9* mouse model [197]. In that context, it was demonstrated that loss of *Kat2a* led to reduction in LSCs in a manner associated with increased transcriptional variability (stochasticity in gene expression). Genes regulated by *Kat2a* through promoter H3K9ac, and which had lost the chromatin mark upon *kat2a* KO with increased transcriptional variability, specifically associated with ribosomal /translation activity, which in this thesis I show to be targets of the ATAC complex. Also interesting, is the fact that one TF prevented from binding at its target gene promoters in *Kat2a* KO cells, was *GABPA*, which is bound by ATAC in K562 (**Appendix D**). Additionally, *Gabpa* has previously been reported as a regulator of CML self-renewal [299–300], which is potentially related with ATAC functions described here. Thus, it is possible that *Kat2a* control of transcriptional variability

in AML is exerted through ATAC, and it would be interesting to test this hypothesis in the future. On the other hand, control of transcriptional variability by Kat2a is well recognised in yeast, which contains SAGA but not ATAC complexes. Indeed, in our mouse *kat2a* KO model, TF Myc, which is bound by SAGA in K562s (**Appendix D**), is prevented of binding in promoters that lost H3K9ac in *kat2a* KO cells [197]. Myc is a known partner of Kat2a [301], and has been implicated in different AML subtypes, including MLL-driven leukaemia [302]. As such, both complexes may in principle make contributions to transcription stability and variability control to regulate the self-renewal and differentiation activities in AML. Future studies investigating this hypothesis could be done by single cell RNA-seq or by single molecule RNA FISH (a method for detecting individual RNA molecules within cells by fluorescence microscopy) in response to parallel ATAC and SAGA-specific knockdowns in the same AML type and across different AMLs.

None of the cell lines studied here showed cell survival defects when knocking down KAT2A, or specific ATAC or SAGA members. Except MOLM-13, where most extensive characterisation was carried out, this comes with the caveat that cells may not have been cultured for a sufficient period of time in order for me to observe differences in apoptosis, since the experiments were performed shortly after transduction.

Together, this work offers functional assessment of the ATAC and SAGA complexes in AML biology. The data implicates SAGA in the AML differentiation block, whereas ATAC plays a role in maintaining leukaemia propagation. Functional investigation of KAT2A, SAGA and ATAC requirements in human AML patient cells is presented in the next Chapter.

Chapter 5

Functional characterisation of the role of
KAT2A and ATAC and SAGA complexes
in human AML primary samples

5.1 Introduction

Having assessed the sensitivity of human AML cell lines to downmodulation of KAT2A and specific ATAC and SAGA members, I aimed to investigate their roles in human AML primary cells, which better reflect *in vivo* circumstances.

Whereas cell lines are relatively easy to manipulate, maintaining primary AML cells in culture is more difficult, particularly without an adequate supportive niche, as AML blasts rapidly differentiate and enter apoptosis *ex vivo*. Moreover, while cell lines are reasonably homogenous (unless cultured for prolonged periods, in which they can acquire mutations), patient samples have an inherent clonal and heterogeneous nature, adding a degree of unpredictability when trying to investigate. Culturing primary AML cells *ex vivo* has been made possible mainly from adapting standard protocols for normal HSCs [266]. In 1977, Dexter *et al.* established a ‘niche-like’ methodology using BM stromal cells obtained from femurs of mice [303]. This co-culture method has been refined and currently involves the use of immortalised feeder layers, such as murine MS-5, osteoblast derived SaOs-2 and human umbilical vein endothelial (HUVEC) cells [266–304–307]. The choice of the supportive stromal cell type selects for a range of different secreted cytokines (reviewed in [308]), needed for stimulating the growth of AML cells. Interestingly, a cross-comparison between stromal layers showed MS-5 to be the most supportive for AML blasts [266]. Additionally, MS-5 are advantageous because they display contact-inhibition and do not require irradiation [308]. More recently, patient-derived BM stromal cells have also been used in co-culture assays with patient samples [309]. Others investigated co-culture systems using mesenchymal stromal cells (MSC) from healthy human donors [310]. In the same study, by comparing the supportive capacity of normal MSC with those of AML patients, Azadniv and colleagues showed that patient derived cells better supported survival of leukaemic blasts [310].

AML arises from the transformation of a single cell and it is maintained by a rare population of leukaemia stem-like cells [106]. These have been found to reside in the CD34⁺ compartment [106], and particularly in CD38⁺ and CD45RA⁺ population, resembling GMP-like progenitors [112], and more rarely, in CD34⁻ compartment [311].

Ex vivo co-culture systems work as a surrogate for animal transplantation experiments in which these populations are normally investigated due to their ability to recapitulate disease. Importantly, they more closely approximate with *in vivo* human biology, allowing the investigation of leukaemic stem like/initiating cells for a few weeks, while monitoring phenotypes.

5.2 Chapter overview

This Chapter details the results of experiments aimed at testing KAT2A requirements in primary human AML cells, as well as an attempt to validate ATAC and SAGA complex dependencies. I tested a small array of AML patient samples directly obtained from the AML clinic, with the aim to capture diverse genetic backgrounds and reflect the heterogeneous nature of the disease. I used an adaptation of the MS-5 stromal co-culture system by Schuringa and Schepers to validate the functions of KAT2A and its complexes in the AML patient samples obtained. I showed that the MS-5 AML co-culture system preserves expansion of leukaemic cells for up to 3 weeks. As the assay proved suitable, I performed KAT2A inhibition studies with a small molecule inhibitor, MB-3, in the co-culture system, and demonstrate that the drug suppresses growth of leukaemic blasts and results in a reduction of primitive leukemic CD34+ marker over time. Furthermore, genetic ablation of *KAT2A* reproduced the results obtained by chemical inhibition. Lastly, I was able to transduce one patient sample with lentiviral shRNAs against ATAC - *ZZZ3* and SAGA – *SUPT20H*, in which knockdown of *ZZZ3* uniquely reduced expansion of AML cells, a finding compatible with a role of the ATAC complex in maintaining leukaemic proliferative capacity.

5.3 Results and discussion

I started by obtaining a small cohort of seven human AML patient samples representative of different genetic backgrounds. Patient sample characteristics can be found in **Table 5.1**.

As detailed in the methods section (Chapter 2), samples were thawed, pre-stimulated with cytokines and either transduced with lentiviral-delivering shRNAs against KAT2A or complex-specific elements, or treated with KAT2A catalytic inhibitor, MB-3. Patient cells were then either tested for colony-forming activity in methylcellulose-based assays or seeded and maintained onto MS-5 layers for a few weeks, with hemipopulation twice a week (**Fig. 5.1**).

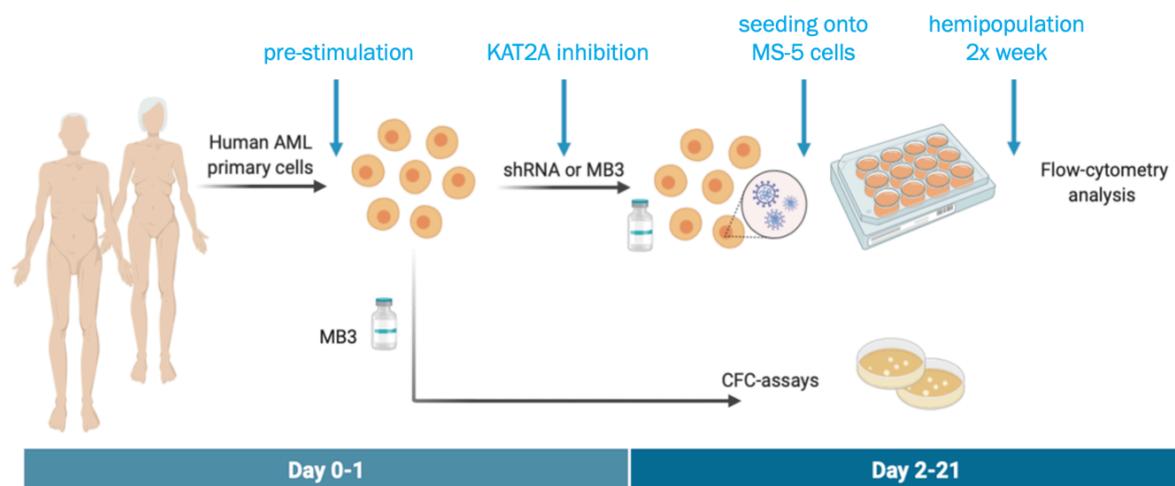


FIGURE 5.1 Experimental strategy adopted for investigating the role of KAT2A and its participating complexes in human primary AML cells.

TABLE 5.1 Characteristics of AML patient samples.

Designation	Age	WHO Classification	Karyotype	FLT3	NPM1	IDH1	IDH2	WT1	ELN	CD34+%	BM blasts	WBC
AML1	51	AML	nd	wt	nd	nd	nd	nd	nd	+	100	165,000
AML2	79	AML	nd	nd	nd	nd	nd	nd	nd	+	90-95	145,000
AML3	31	NPM+	normal	ITD	mut	wt	wt	4482.27	IR	+	90-95	46000
AML4	42	Mixed phenotype AL	t(6;14) , BCL11B rearrangement , t(11;12)	mut	wt	wt	wt	2756.17	IR	+	90-95	62400
AML5	nd	AML	+8	nd	nd	nd	nd	nd	nd	+	nd	nd
AML6	nd	AML	nd	wt	wt	nd	nd	nd	nd	-	65	nd
AML7	50	NPM+	nd	ITD	nd	nd	nd	nd	nd	+	>80	nd

nd - non defined

5.3.1 Clonogenic growth of AML patient samples is hindered by treatment with MB-3, the KAT2A small molecule inhibitor

In a first set of experiments, I tested the effects of inhibiting KAT2A by treating AML patient samples with 100mM MB-3. Three out of four samples shown reduced CFC-capacity compared to vehicle DMSO (**Fig. 5.2A-B**). In line with these results, our lab has previously demonstrated that pharmacological ablation of KAT2A with MB-3 enhances differentiation and apoptosis of AML cells *in vitro* [260].

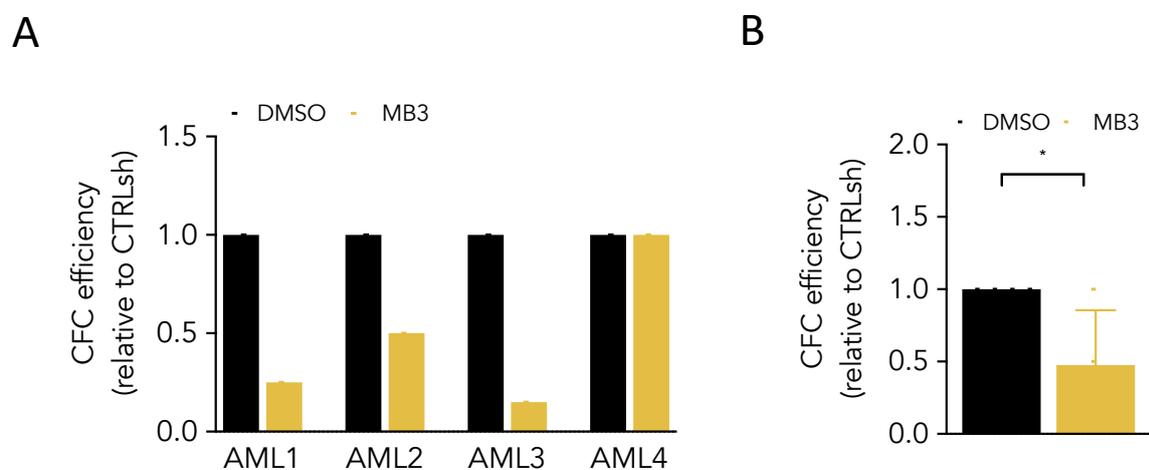


FIGURE 5.2 KAT2A small molecule inhibitor, MB-3, reduces CFC-capacity of human primary AML cells. (A) CFC-assay of 4 primary AMLs treated with 100 mM MB-3 and vehicle DMSO. AML1, AML2 and AML3 are responsive to MB-3 treatment. **(B)** Mean values of CFC-assays in (A). Two-tailed unpaired t-test for significance; * $p < 0.05$.

As per **Table 5.1**, the inhibitory effect on clonogenic potential was observed across AML samples of different genotypes, and appeared independent of sex, age and cytogenetic abnormality. Short-term methylcellulose-based assays, however, do not necessarily capture the most primitive candidate leukaemic stem cells, and medium-to-long term cultures would help identify these rare populations and their ability to sustain the leukaemia over longer periods. With this in mind, next I aimed to test a co-culture system that would allow maintenance of AML primitive cells for longer, in the presence of competent stromal MS-5 cells.

5.3.2 MS-5 co-culture system preserves growth, CFC-capacity and phenotype of human primary AML cells

When testing of the MS-5 co-culture system (Schuringa and Schepers, 2009), it was found that growth of AML patient cells (AML5) was sustained for up to 3 weeks (**Fig. 5.3A**). In parallel, I also tested the sample for colony initiation capacity, which was preserved for up to three weeks, suggesting maintenance of progenitor-like leukaemic cells (**Fig. 5.3B**). Moreover, co-culture on MS-5 layers broadly supports maintenance of the leukaemic phenotype (**Fig. 5.3C**).

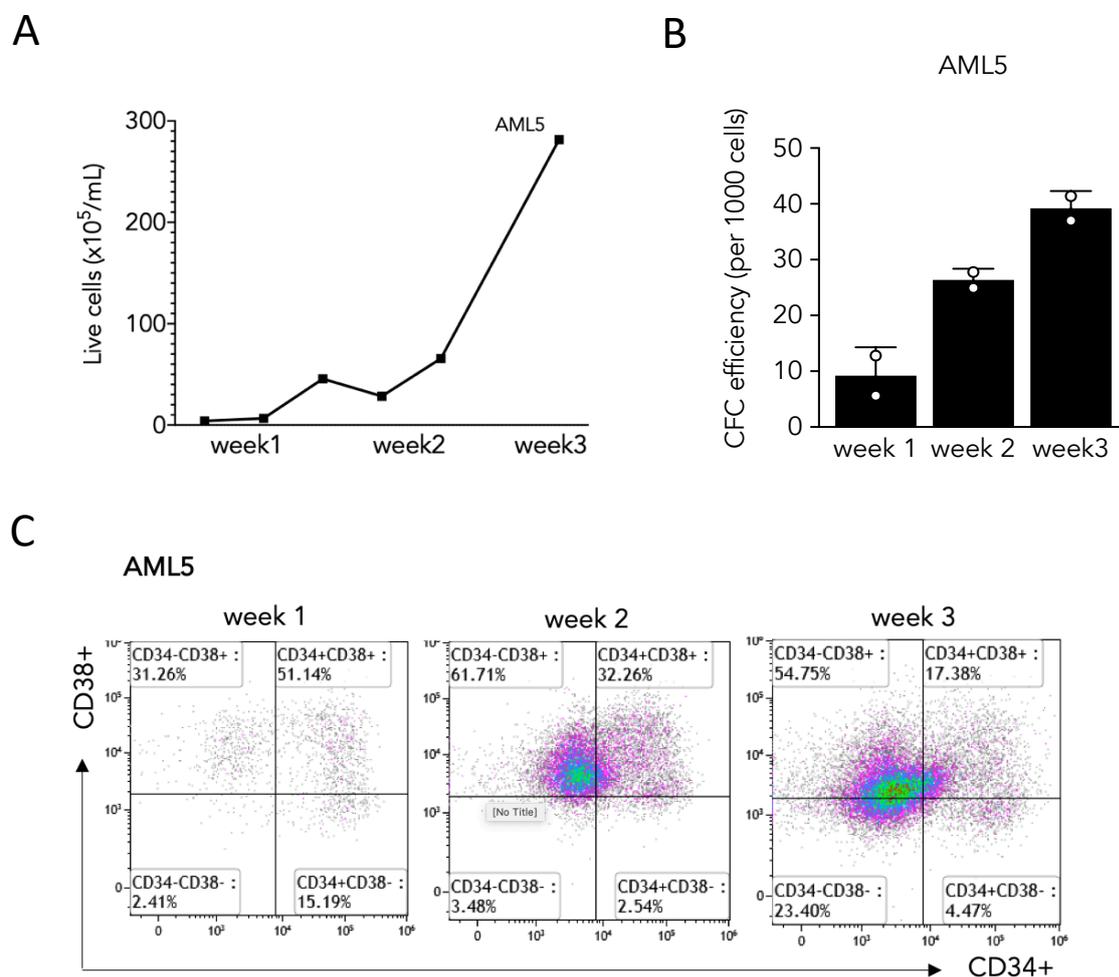


FIGURE 5.3 MS-5 co-culture system preserves growth, expansion and moderate cell surface phenotype of primary AML cells for up to 3 weeks. (A) Growth curve of human primary AML cells in the MS-5 co-culture system. **(B)** CFC-capacity of primary human AML cells is preserved in the MS-5 co-culture system for up to 3 weeks. **(C)** Maintenance of cell surface phenotype of AML patient cells for up to 3-weeks.

Having validated the suitability of the co-culture method to experiment on primary human AML cells, I then tested the effects of inhibiting KAT2A activity using MB-3 in the culture system. Drug treatment led to reduced growth of CD34+ AML cells when compared to vehicle, DMSO (**Fig. 5.4A**), impacting the proportion of candidate phenotypic leukaemia stem-like (L-)GMP cells (L-GMPs) (**Fig. 5.4B-C**).

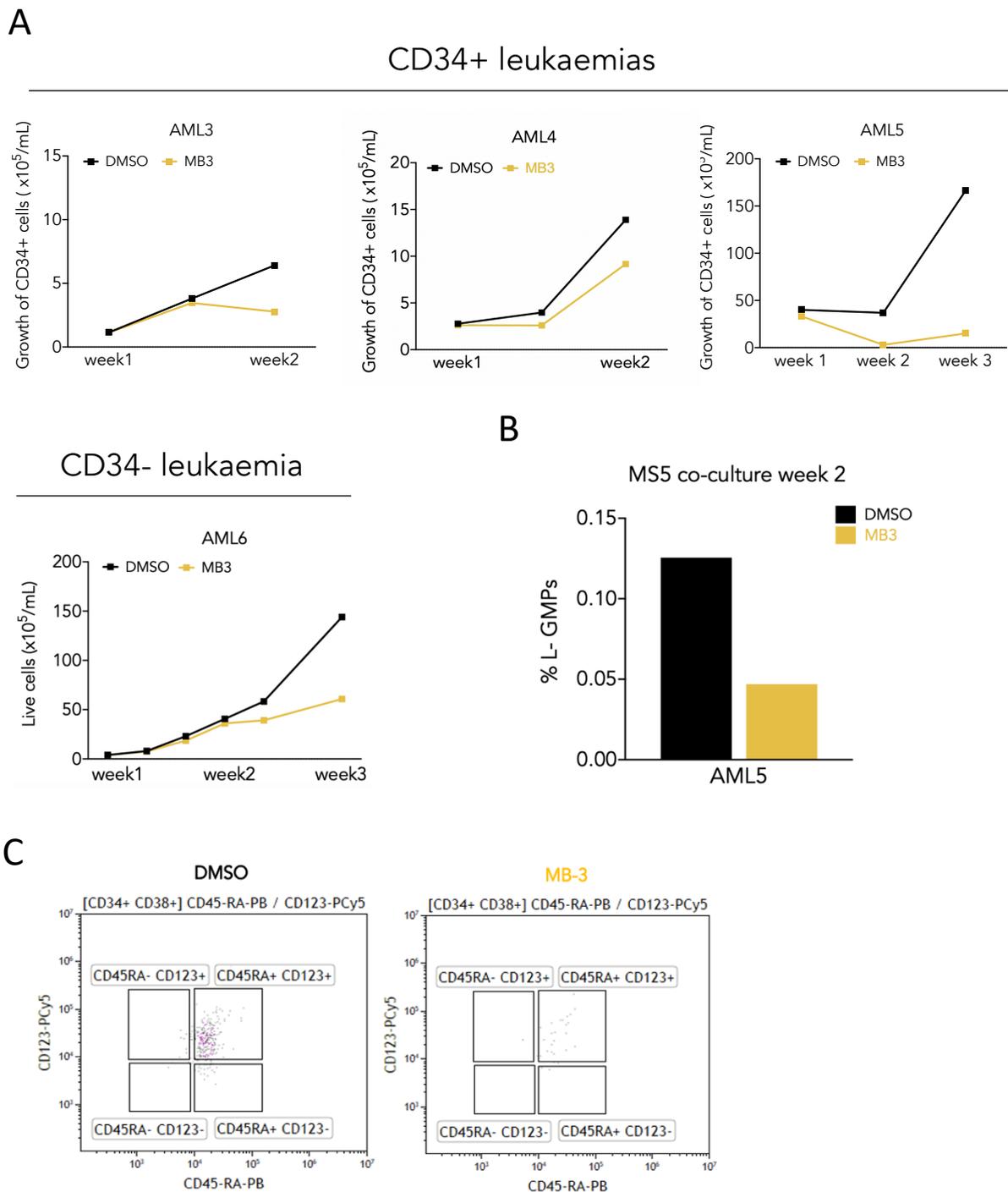


FIGURE 5.4 KAT2A chemical inhibition with MB-3 suppresses growth and expansion of primary AML cells for up to 3 weeks in MS-5 co-culture. (A) Growth of human

primary CD34+ AML cells treated with KAT2A inhibitor MB-3 vs vehicle DMSO in the MS-5 coculture system for up to 3 weeks. Individual CD34+ patient samples (AML3, AML4 and AML5) and one CD34- sample (AML6) are shown. **(B)** Percentage of L-GMPs at week 2 of MS-5 co-culture for AML2 cells treated with KAT2A inhibitor MB-3 vs vehicle DMSO. **(C)** Representative flow cytometry plot of data in (B).

Overall, these data indicate that medium-term survival of primary AML cells can be supported by the MS-5 co-culture method, which mimics the tumour microenvironment. Though, not all samples perform equally, with some expanding better than others, which likely reflects the heterogeneous nature of AML. Significantly, small-molecule inhibition of KAT2A resulted in decreased growth of CD34+ AML blasts over the of 2-3 weeks period. Impairment of L-GMP compartment, representative of leukaemia initiating /stem-like cells, was also verified for one of the patient samples upon treatment of MB-3, suggesting a role for KAT2A in the maintenance of primitive human AML.

Next, I tested whether *KAT2A* genetic inhibition reproduced the results observed with MB-3 treatment in co-culture.

5.3.3 *KAT2A* lentiviral-mediated knockdown mimics its chemical inhibition, reducing leukaemic expansion in the MS-5 co-culture system

When attempting to transduce human AML cells with shRNAs against *KAT2A*, few samples were able to transduce and I could observe GFP+ cells (transduced cells) in 33% of the samples tested, all of them were CD34+. Successfully transduced samples were kept in culture for 2-3 weeks in the presence of MS-5 stroma, while monitoring their expansion as well as the preservation of total GFP, CD34+ and L-GMP compartments, gated as per **Fig. 5.5**.

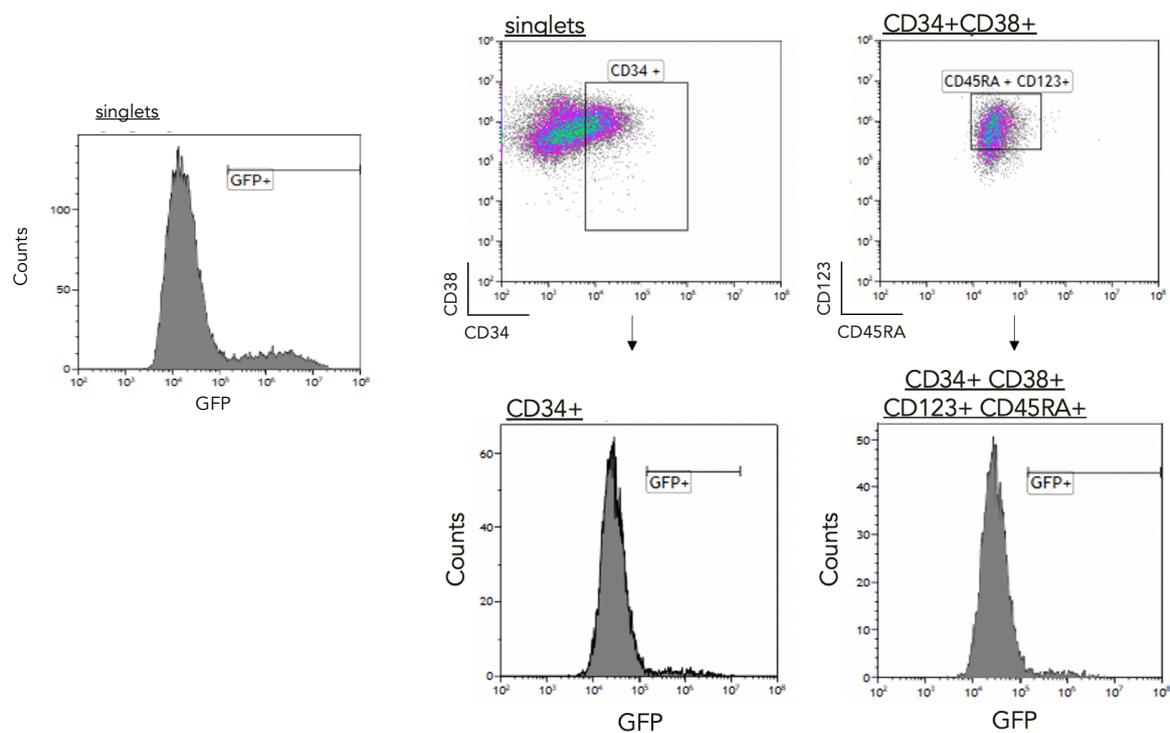


FIGURE 5.5 Flow cytometry gating strategy used for analysis of primitive AML CD34+ and GMP-like (L-GMPs) cell populations. Cells gated for live cells (SSC-A vs FSC-A), and singlets considered (SSC-A vs SSC-H). Singlets were then gated for GFP+ to establish the global level of GFP against which individual sub-populations were analysed. Sub-populations analysed were CD34+ (left), and L-GMPs (right), gated as CD34+CD38+CD123+CD45RA+ cells. GFP levels within CD34+ cells and L-GMP were obtained and compared with global GFP levels for each construct to determine relative preservation of each sub-population upon gene expression knockdown.

KAT2A knockdown (**Fig.5.6**) resulted in reduction in growth of transduced AML cells overtime (**Fig.5.7A-B**), as elicited by pharmacological inhibition with MB-3, thus validating the drug's specificity. Moreover, expansion of primitive CD34+ and L-GMP compartments was also reduced (**Fig.5.7C-D**).

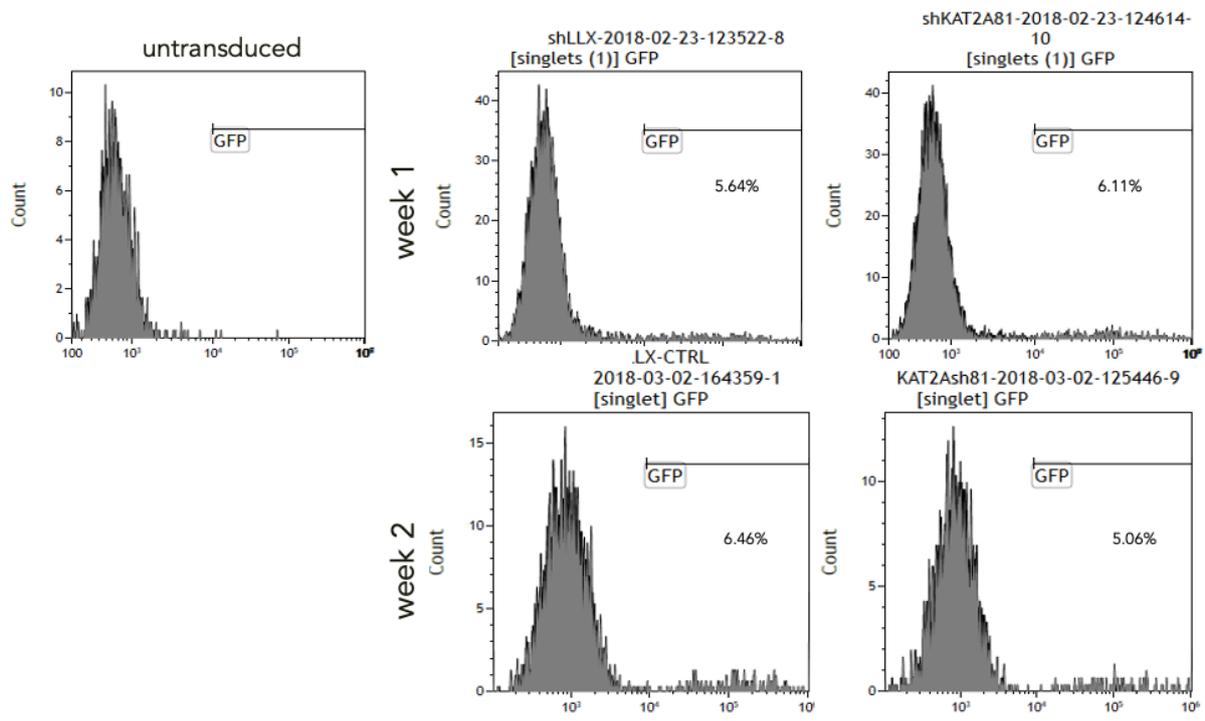


FIGURE 5.6 Flow cytometry profiling of *KAT2A* transduced (GFP+) cells over a 2-week co-culture. Representative plots for AML2.

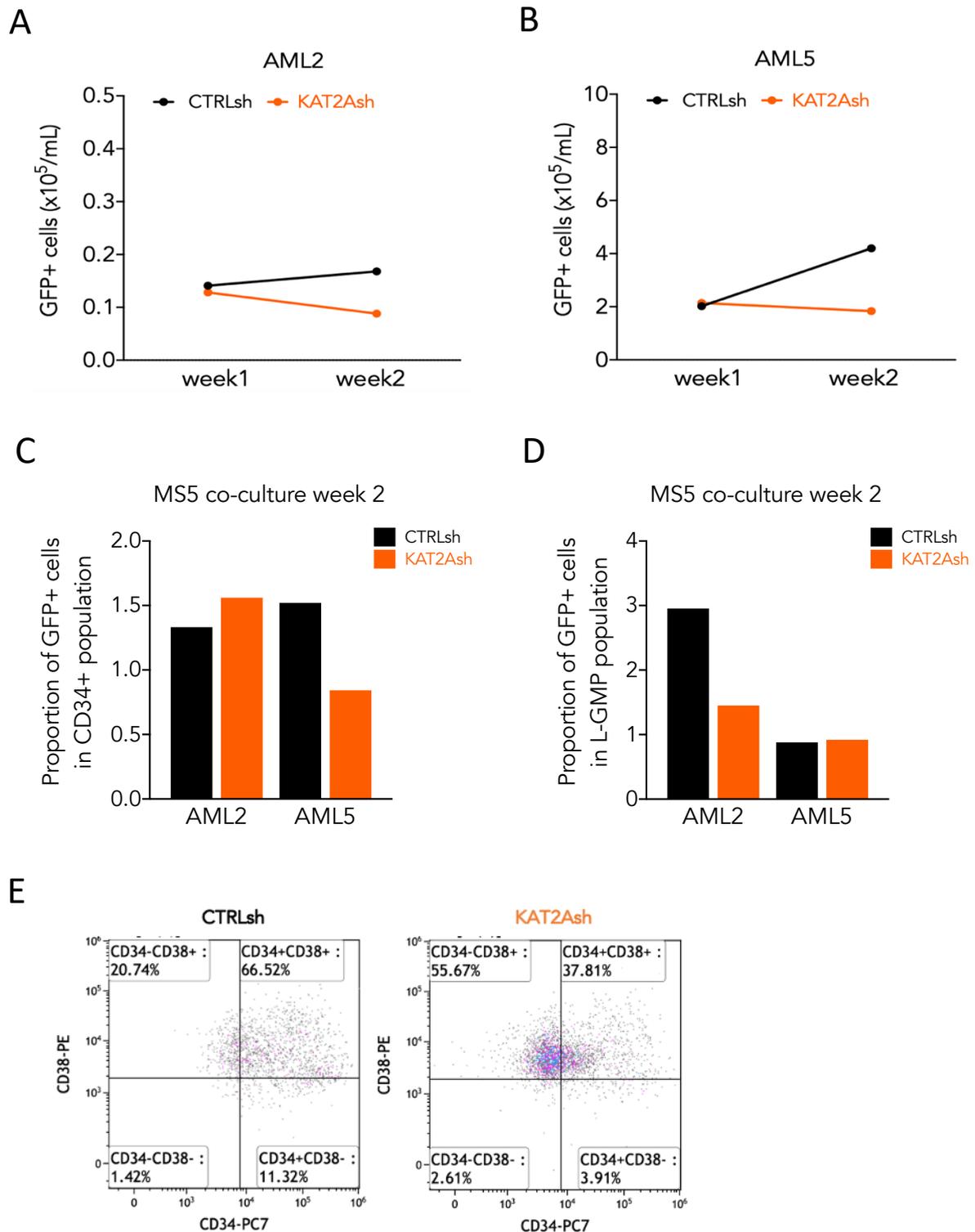


FIGURE 5.7 KAT2A knockdown reduces expansion of human primary AML cells in the MS-5 co-culture system. (A-B) Growth of human primary AML cells (CD34+ samples) transduced with *CTRLsh* and *KAT2Ash* and maintained in the MS-5 co-culture system for 2-weeks. **(C)** Percentage of GFP+ CD34+ cells in AML samples in (A-B) analysed at week 2 of the MS-5 co-culture. Analysis gates are presented in Fig. 5.5. Data are normalized to global GFP level, to correct for

unequal GFP transduction levels. Data are normalized to global GFP levels, to correct for unequal GFP transduction levels. **(D)** Percentage of GFP+ GMP-like (L-GMP) cells at week 2 of the MS-5 co-culture system in AML samples in (A-B). GFP transduction correction as in (C). **(E)** Representative flow cytometry plot of AML5 transduced with *CTRLsh* and *KAT2Ash*.

5.3.4 KAT2A complex activity maintains propagation of human primary AML cells in MS-5 co-culture

Finally, I attempted to transduce (**Fig. 5.8**) primary CD34+ AML patient blasts with the lentiviral vectors delivering shRNAs *SUPT20H* or *ZZZ3*.

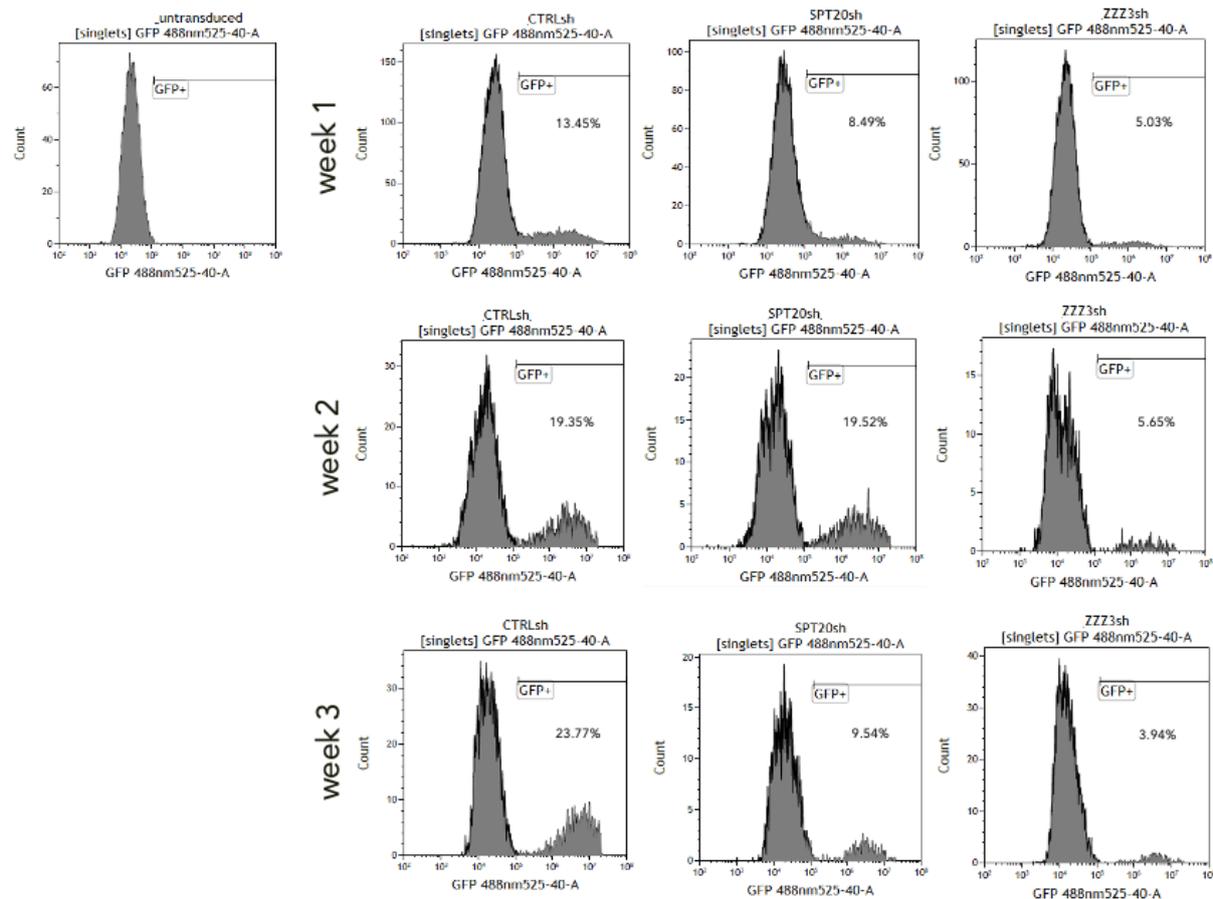


FIGURE 5.8 Flow cytometry profiling of shRNA-transduced AML7 sample – gating of total GFP+ cells in the whole sample over a 3-week co-culture period. Untransduced AML7 cells included for comparison.

In the sample analysed (AML7), downregulation of *ZZZ3*, but not of *SUPT20H*, resulted in reduced expansion of transduced AML blasts (**Fig. 5.9A**), which specifically affected CD34+ (**Fig. 5.9B**) and/or L-GMP (**Fig. 5.9C**), suggesting a loss of self-renewal potential.

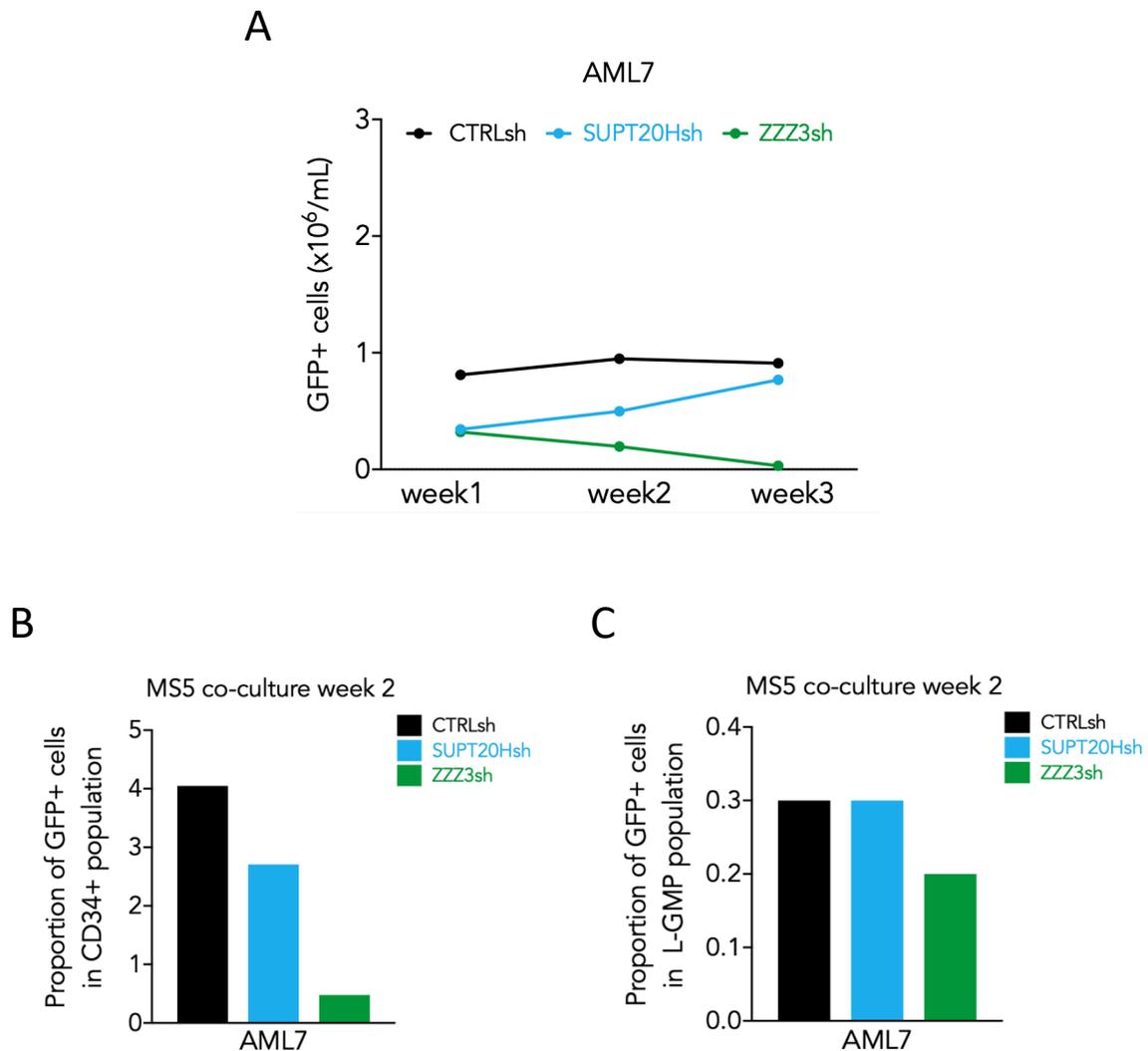


FIGURE 5.9 ATAC-ZZZ3 knockdown reduces expansion of human primary AML cells in the MS-5 co-culture system. (A) Growth of human CD34+ primary AML cells (AML7) transduced with *CTRLsh*, *SUPT20Hsh* or *ZZZ3sh* and maintained in the MS-5 co-culture system for 3 weeks. **(B)** Percentage of GFP+ CD34+ cells in AML7 (A) analysed at week 2 of the MS-5 co-culture. Analysis gates are presented in Fig. 5.5. Data are normalized to global GFP level, to correct for unequal GFP transduction levels. **(C)** Percentage of GFP+ GMP-like (L-GMP) cells at week 2 of the MS-5 co-culture system in AML7 (A).

5.4 Conclusions and future work

KAT2A has histone acetyltransferase activity through its catalytic HAT domain, exerted within SAGA or ATAC complexes, and which is druggable with a small molecule inhibitor, MB-3. These data shows that pharmacological inhibition of KAT2A with MB-3 is effective in reducing colony-initiation capacity of primary human leukaemias, confirming the disease's dependence on KAT2A function.

The MS-5 co-culture assay employed in this study, proved robust in maintaining AML cells for a few weeks, with leukaemic blast survival and proliferation benefiting from the culture on MS-5 layers. The main advantage of these assays to investigate stem/progenitor leukaemic cells is their accessibility when compared to mice xenotransplantation experiments, which typically take place over 8-20 weeks prior analysis. In turn, gene-perturbation analysis in candidate leukaemic stem cells can be performed on MS-5 stroma while regularly monitoring phenotypes. Similarly, with the development of new target therapies, this method seems an appropriate and predictive preclinical model to screen AML inhibitory compounds. By using the MS5-co culture strategy, I showed that as a result of both *KAT2A* loss or MB-3 inhibition, leukaemic cell expansion was decreased over a period of 2-3 weeks, and it was accompanied by a reduction of CD34+ and L-GMP phenotypic compartments.

Although I could only successfully transduce one patient sample with the ATAC and SAGA-delivering shRNAs, downmodulation of *ZZZ3* delayed AML cell expansion considerably, which was not observed with loss of SAGA-*SUPT20H*. This is compatible with a role for ATAC in sustaining leukaemic self-renewing properties as suggested in the previous Chapter in AML cell lines. It would have been meaningful to test by qRT-PCR if ribosomal protein genes were being downregulated upon loss of ATAC as observed in the context of cell line models. However, as the cell numbers were limiting and the population was being lost overtime time, this was not assessed. Furthermore, a more direct measure of self-renewal would be to harvest *ZZZ3sh* and *SUPT20Hsh* cultures and serial-replate onto new stroma. Indeed, a particular characteristic of MS-5 co-cultures is the formation of cobblestones, which are thought to correspond to leukaemic stem and progenitor cells [307], that when replated for a number of weeks would maintain the disease. Although, during the time frame of these experiments I failed to observe such structures, which may

require more than 3-weeks to form. Ideally, this would be done in parallel with xenotransplantation assessment in which the prediction would be that ZZZ3-depleted cultures would fail to re-capitulate the disease. In future, testing AML ATAC and SAGA dependencies in a larger cohort of AML patient samples is of considerable interest and requires further characterisation. Thus, it would be of great value to obtain information of a larger number of primary AMLs.

One other future aim would be to make use of the co-culture system to functionally validate plausible downstream effectors of KAT2A loss, yield by both RNA/ChIP-seq. Likewise, to validate unique targets of ATAC and SAGA complexes from our ChIP dataset, that could specifically act on leukaemic maintenance or differentiation.

Although the cohort of samples analysed was small, the capacity for medium-term *ex vivo* proliferation did not seem dependent on the genetic characteristics, as suggested by others [312]. However, it is recognised in the field that more aggressive leukaemias are more resistant to inhibitory compounds (and indeed engraft better into mice) [313]. Knowing this, it is possible to imagine that this could impose a bias towards more aggressive AML samples to self-propagate better in this co-culture *ex vivo* assays, than low risk leukaemias.

It has become apparent that malignant haematopoietic cells and their niche cells and their niche cells can affect each other in a bidirectional way [314]. However, little is known on how murine stromal cells stimulate human cell proliferation or mimic the BM microenvironment *ex vivo* [307]. One way, it has been shown, is that stromal cells prevent apoptosis of leukaemic blasts by up-regulating anti-apoptotic proteins in the AML cells [315]. In future, it would be interesting to characterise the phenotype of the blasts that are in direct contact with the feeder layer, for instance by immunostaining, and to investigate whether/how KAT2A and its complexes mediate AML-stromal cell interaction, as well as its potential niche dependencies across different types of AML. Also, it was not tested whether MB-3 treatment affected the viability and function of the MS-5 stromal cells themselves, a possibility that cannot be ignored.

In summary, while these experiments need reproduction using larger sample sizes, the co-culture experimental approach represents a good model for dissecting the differences in molecular mechanisms exerted by ATAC and SAGA complexes in AML.

It provided further evidence to suggest that KAT2A-mediated ATAC roles govern proliferation of human leukaemia. On the other hand, SAGA functions need to be better studied in primary AMLs. A recent study reported that MB-3, in combination with all-trans retinoic acid (ATRA), reduced the capacity of patient blasts to form colonies in non-APL AML [316]. This would be interesting to explore, particularly in the context of SUPT20H knockdown, as an attempt to enhance its pro-differentiation effects via targeting the Core module of SAGA. This is in keeping with the mild haematopoietic phenotype observed following shRNA treatment of SAGA DUB module.

Chapter 6

Conclusions

The unifying message of the work detailed in this thesis is the unique, differential contribution, of KAT2A chromatin complexes ATAC and SAGA, to normal and malignant haematopoiesis.

Both normal and leukaemic haematopoiesis are regulated by cytokines, lineage-specific Tfs, and epigenetic modifiers. The latter include 'readers', 'writers' or 'erasers', which all facilitate changes in chromatin state, influencing transcriptional programmes in both healthy and aberrant developmental scenarios. Indeed, epigenetic dysregulation is a widely recognised feature of AML, which remains a lethal disease for most patients. Therefore, examination of its chromatin susceptibilities is essential to gain a better understanding of the disease. Histone modifiers, central components of the epigenetic machinery, are often found in multiprotein complexes. Critically, while their enzymatic functions have been more extensively characterised and exploited as targets in AML therapeutics [120], little is known about the molecular context in which they function. *KAT2A* encodes for a HAT that is present within two multiprotein complexes, ATAC and SAGA, and which has been identified as a vulnerability in AML by our lab, but for which roles in normal blood were less well defined.

In this thesis, I investigated the roles of KAT2A in normal and malignant human haematopoiesis through analysis of the two co-activator complexes in which it exerts its HAT activity. Specific ATAC subunits *ZZZ3*, *TADA2A* and *YEATS2*, and SAGA-specific *SPT20*, *TADA2B* and *USP22* were used surrogates of one or the other complex. Through the use of CB, AML cell lines and primary samples and a shRNA-treatment approach, functional data addressing specificity of the two KAT2A complexes has been presented.

In normal blood, KAT2A was found to regulate erythroid progenitor specification and survival. While not required for granulocytic or lymphoid lineage development, the erythroid lineage was specifically dependent on KAT2A. RNA-seq in *KAT2A*-depleted CB cells provided a better understanding of the specific transcriptional programmes repressed in its absence. Those associated with oxygen transport and platelet function, agreeing with the defect in erythroid (and/or megakaryocytic) lineage development when removing KAT2A. Significantly, the defect along the erythropoietic lineage involved the contribution of ATAC and SAGA complexes, which were required in early and late developmental stages, respectively. Specifically, it was possible to

reconcile loss of the ATAC complex with a defect in fate commitment towards the erythroid lineage, suggesting that it is required for erythroid specification from CB. Conversely, while depletion of SAGA Core SPT20, did not affect normal blood cell function overall, SAGA DUB module USP22 was shown to impair erythroid maturation post-commitment. This suggested a role for SAGA in preservation – rather than initiation – of erythroid cell fate. Interestingly, the realisation that one other SAGA element with a different enzymatic activity than KAT2A regulates late erythroid fate could indicate combined activities of HAT and DUB modules, potentially explaining the more pronounced effect that seems to exceed loss of SUPT20H. On the other hand, there is also evidence in the literature for KAT2A-independent functions of the DUB module, which could potentially explain this disparity [278]. Similarly, Wang and colleagues recently elucidated the structure of yeast SAGA and reported a conformational change upon complex binding to DNA, in which the DUB module physically detaches from the Core [214], a mechanism potentially conserved in human SAGA complexes. It may also be possible that in late erythroid maturation, SAGA complexes rely on KAT2B, the KAT2A orthologue, mutually exclusive with KAT2A in ATAC and SAGA. In support of this hypothesis, expression of *KAT2B* increases in late erythroid differentiation (**Appendix C**). All these speculations need further investigation and could partially be answered by examining the effects of combined depletion of HAT and DUB subunits.

Regarding ATAC roles, its specific requirement in commitment to the erythroid lineage may relate to its association with protein synthesis and translation-affiliated genes, as revealed by the chromatin binding profile of *ZZZ3* in haematopoietic cells (**Appendix D**). In line with this view, studies of protein synthesis rates during hematopoiesis have indicated that dramatic changes occur during the early stages of lineage commitment [298]. More specifically, it has been shown that defects in the ribosome and in translation can selectively impact commitment to the erythroid, but not other hematopoietic lineages [317]. It would be interesting to consider this in the future with further investigation. It would also be important to investigate additional ATAC subunits, such as *YEATS2*. For example, *MLLT3*, which contains a *YEATS* domain, has previously been implicated in erythroid specification [318], and there is the possibility that the reader domain can play a role in the recruitment of this and other *YEATS*-containing proteins, such as ATAC-*YEATS2*, to promoters of erythroid regulators.

When addressing the transcriptional significance of unique ATAC and SAGA binding peaks in erythroid-affiliated K562 upon *ZZZ3* or *SUPT20H* knockdown, I observed loss of H3K9ac at unique ATAC-bound, as well as SAGA-bound genes - ribosomal protein genes and *HBB*, respectively. This was complex dependent and site specific. In other words, loss of H3K9ac at specific ATAC bound promoters coincided with down-regulation of its target genes, without affecting SAGA-bound regions, and *vice versa*. Given the identified KAT2A role in erythropoiesis, further assessment of H3K9ac status at other erythroid regulators, identified by the RNA/ChIP-seq datasets would provide better mechanistic insight into the exact maturation stages of erythroid development governed by the activities of either complex. I examined H3K9ac levels at *EPOR*, which expression was significantly reduced following *KAT2A* knockdown in CB HSCs, and which initially was thought to explain the defect in development of the erythroid lineage. In this experiment, however, I did not find H3K9ac levels to change in K562-*SUPT20Hsh* cells. Interestingly, in K562 cells, failure in downregulating *GATA2* upon loss of *SUPT20H* uniquely prevented erythroid differentiation. However, the exact mechanism is not clear at present. It has previously been shown that KAT2B interacts with, and directly acetylates TAL1, promoting erythroid differentiation in MEL cells [287]. One possibility is that failure of KAT2B in acetylation and activation of TAL1-dependent programmes (if SAGA indeed relies on KAT2B instead), would reflect the persistence of *GATA2* and differentiation arrest, which could be an indirect effect from an upstream dysregulation *via* KAT2B acetylation of TAL1. Given the observations the CB, it would be interesting to perform a similar time course of erythroid differentiation in K562s upon loss of SAGA DUB module *USP22*, to obtain more mechanistic insight on the SAGA contributions to erythroid differentiation.

I extended the investigation of ATAC and SAGA functions to malignant haematopoiesis, where the disparity in roles between the two KAT2A-containing complexes was also functionally observed. I made use of a panel of AML cell lines containing markedly different genetic mutations, allowing for a better representation of the disease. In investigating the functions of the two KAT2A complexes, I found that SAGA participates in preservation of the characteristic AML differentiation block, irrespectively of the leukaemic subtype, which may suggest general control of cell state/identity in line with its roles in normal haematopoiesis post-commitment. ATAC, on the other hand, regulates cell biosynthetic activities, *via* control of ribosomal protein and translation-associated genes. This molecular information,

together with the pervasive G1/G0 cell cycle arrest observed in most AML lines, indicates that the ATAC complex fulfils an essential role in maintaining leukaemic cells. Notably, the phenotypic changes associated with depletion of SAGA or ATAC elements were somewhat consistent across the different AML cell models, independently of their genetic background. Lack of cellular differentiation is a hallmark of AML. In addition, LSCs cells are dependent on tight translational activity [298], thus justifying a dependence on the activities of both complexes in the leukaemia biology. The data suggest that the pro-differentiation effects of KAT2A-inhibition are likely to be exerted through SAGA. From this point of view, one may think of targeting SAGA Core as an anti-leukaemic strategy as its effects in normal haematopoietic were minimal. This could, for example, be done in combination with chemotherapy, or ATRA which has recently been shown to have synergistic effects with KAT2A inhibitor MB-3, beyond APL leukaemias [316].

Mechanistically, in K562s, ChIP-qPCR analysis of H3K9ac, at self-renewal and ribosomal protein genes, showed reduction of the specific histone activating mark at these *loci*, upon knockdown of *SUPT20H* (SAGA) and *ZZZ3* (ATAC) elements. Nevertheless, ATAC and SAGA control of gene transcription may be more specific than its role as H3K9 acetyltransferase at gene promoters. In the leukaemic setting, loss of both SAGA and ATAC were accompanied by loss of H3K9ac at *Hoxa* genes as well as ribosomal protein-coding genes, suggesting a more complex epigenetic regulation at gene promoters, potentially involving additional complex enzymatic activities, as could also be the case in normal blood. Future studies investigating single and combined requirements of ATAC (for example *ZZZ3* and *KAT14*) and SAGA (HAT and DUB) enzymatic subunits in *locus* and cellular regulation will enhance our understanding of the 'crosstalk' between different chromatin marks and of their effects on transcriptional control in leukaemia and the normal blood system. Similarly, this work focused on the HAT activity of KAT2A, and does not address if other histone modifications catalysed by KAT2A such as histone succinylation of H3K79 [180], are involved normal or leukaemic processes. To my knowledge this newly described enzymatic function has not been linked to neither SAGA nor ATAC complexes, but speculation as to if this is the case requires investigation.

On a more fundamental level, the data suggests diversification of SAGA complexes, or of KAT2A function itself, from yeast to human. In yeast, KAT2A had been found to promote global acetylation genome-wide. On the contrary, the ChIP data indicates

a relatively low number of SAGA-specific and ATAC-specific targets. This suggests a more restricted use in mammalian cells as compared to yeast. Although it is also arguable that precise identification of KAT2A complex targets may be limited by the tools available. When comparing between the categories of genes bound SAGA and ATAC, it seems that SAGA regulation may involve control of other transcriptional regulators (e.g. *MYC*, *STAT*, *ERG*). ATAC, in contrast, appears to control a more unique set of genes, which encompass the translational machinery, and are ubiquitous across multiple cell types. This was verified in context of haematopoiesis, in this work, but also by others in other cancer types [235–236]. In line with this view, and because different cell types and states (for example at different stages of lineage differentiation), may be differentially sensitive to loss of ‘identity’ control *vs* biosynthetic pathways, it is conceivable that the differential requirement would determine the cells’ relative dependency on ATAC or SAGA complexes. Similarly, the dependency on the roles - rather than the roles themselves - could direct the extent to which loss of *KAT2A* equates with loss of one or both complex activities. In the case of normal haematopoiesis, the fact that the requirement of either complex is restricted to erythropoiesis may reflect the distinctive nature of erythroid/megakaryocytic commitment and differentiation, which segregate from the stem cell root upstream of myeloid and lymphoid lineages [319–320] with *de novo* establishment of transcription programs [52]. In normal hematopoiesis, maintenance dominates pre-commitment, and identity post-commitment, with ATAC required early and SAGA late in lineage development. In light of this, a proposed molecular model of KAT2A function within ATAC and SAGA complexes is shown in **Fig 6.1**.

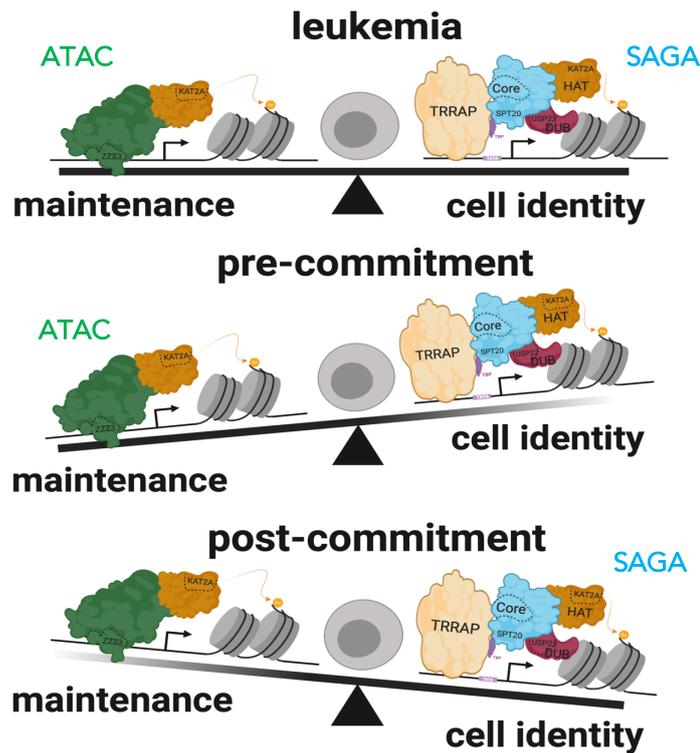


FIGURE 6.1 Working model of cell dependence on ATAC and SAGA complexes in normal and leukemic hematopoiesis. ATAC regulates leukaemia maintenance and SAGA cell identity/ state. In normal hematopoiesis, maintenance dominates pre-commitment, and identity post-commitment, with a stronger role of ATAC required early, and of SAGA in later stages of lineage development.

Overall, while mechanistic investigation is needed to understand exactly how KAT2A participation in individual ATAC or SAGA complexes balance cell fates in normal and malignant settings, KAT2A-containing complexes may operate in order to preserve the cellular *status quo*. This would be compatible with the role of KAT2A in sustaining, rather than initiating, gene transcription [179]. Similarly, it aligns with its function in stabilising promoter activity and minimising variability in transcriptional output as recently described by our lab in embryonic stem cells [187] and leukaemia [197]. In leukaemia, our lab demonstrated that *Kat2a* sustains LSCs in a *MLL-AF9* mouse model [197], and linked *Kat2a* loss to promoter dysregulation and enhanced transcriptional variability. In respect of control of transcriptional variability, *Kat2a*-dependent promoters that respond to *Kat2a* KO with loss of H3K9ac and increased transcriptional variability, were enriched in ATAC-dependent translation-associated genes. It is possible that *Kat2a* control of variability is exerted

in the context of its complexes [249], and this would be an interesting question to explore in future studies in response to ATAC and SAGA-specific depletions, using single-cell analysis.

Genetic and pharmacological perturbations have been performed to address the role of KAT2A and its complexes in human AML primary samples, as presented. By using the MS5 co-culture system, I showed that *KAT2A* loss or chemical inhibition reduced leukaemic cell expansion over a period of 2-3 weeks, confirming KAT2A roles in primary leukaemias. More significantly, there was reduction in CD34+ and L-GMP primitive leukaemic compartments. Though only a single patient sample was successfully transduced with the ATAC and SAGA-delivering shRNAs, down-modulation of *ZZZ3* impaired AML cell expansion, which was not verified upon loss of *SUPT20H*. While this warrant further investigation in larger patient cohorts and potentially optimisation of shRNA transduction conditions in primary AML cells, the result is compatible with a role for ATAC in sustaining leukaemic self-renewing properties. In future, the AML-MS-5 co-culture system could be used to functionally validate KAT2A downstream targets (yield by both RNA/ ATAC and SAGA ChIP datasets) that specifically act on leukaemic maintenance or differentiation. With this in mind, I have validated a CRISPR epigenetic-editing tool to directly deliver KAT2A activity to its putative target genes (**Appendix E**). In future this would involve: (a) redesign a more robust CRISPR-tool containing a fluorescent marker, thus avoiding the use of antibiotic selection on human samples; (b) design and test fusions between dCas9 and specific components of the KAT2A-containing complexes. The most obvious candidates would be DNA-binding *ZZZ3* in the case of ATAC, and for example DUB, in the case of SAGA.

To date no other knockouts of ATAC and SAGA subunits have been investigated in the hematopoietic system, at the exception of *Usp22*, which did not affect stem and progenitor compartments or normal myeloid differentiation [259], but had a surprising tumour suppressor effect in *Kras*-driven transformation. However, a recent screen of deubiquitinating enzymes required for 4-week bone marrow engraftment which identified *Usp15* as a regulator of HSC activity *in vitro* and *in vivo* [321], also identified *Usp22* as a possible hit, re-enforcing the notion that additional analysis of ATAC and SAGA-mediated roles in mouse haematopoiesis is necessary.

In summary, the data presented in this thesis provided functional *in vitro* assessment of ATAC and SAGA complexes in both normal and malignant blood development. It identified unique roles for ATAC and SAGA complexes which associate with maintenance of biosynthetic activity in the case of ATAC, and cell status/ identity in the case of SAGA [322]. It also prompted new questions that would be important to answer with future experiments:

-Is KAT2B the preferential HAT in SAGA complexes at late stages of erythroid differentiation?

-What are the consequences of combined knockdowns of SAGA-specific versus ATAC-specific depletions to normal and malignant haematopoiesis?

-Through which pathways/mechanisms does SUPT20H loss enforce cell differentiation?

-Is KAT2A control of transcriptional variability complex dependent, and if so, do both complexes play a role, or does it act uniquely through one of them?

-Would Cas9-fusions with molecules involved in large chromatin complexes pose a challenge to efficient epigenetic editing? Would this approach fail potentially because of inability of assembly of endogenous complexes?

Investigating these questions would further this work and shed new light into ATAC and SAGA chromatin and transcriptional regulation in normal haematopoiesis and AML.

Bibliography

1. S. Doulatov, F. Notta, E. Laurenti, & J. E. Dick, Hematopoiesis: A Human Perspective. *Cell Stem Cell*, **10** (2012) 120–136. <https://doi.org/https://doi.org/10.1016/j.stem.2012.01.006>.
2. L. Jacobson, M. Ek, & M. Robson, EFFECT OF SPLEEN PROTECTION ON MORTALITY FOLLOWING X-IRRADIATION. (1949).
3. E. Lorenz, C. Congdon, & D. Uphoff, Modification of Acute Irradiation Injury in Mice and Guinea-Pigs by Bone Marrow Injections. *Radiology*, **58** (1952). <https://doi.org/10.1148/58.6.863>.
4. J. E. Till & E. A. McCulloch, A Direct Measurement of the Radiation Sensitivity of Normal Mouse Bone Marrow Cells. *Radiation Research*, **14** (1961). <https://doi.org/10.2307/3570892>.
5. I. L. Weissman & J. A. Shizuru, The origins of the identification and isolation of hematopoietic stem cells, and their capability to induce donor-specific transplantation tolerance and treat autoimmune diseases. *Blood*, **112** (2008) 3543–3553. <https://doi.org/10.1182/blood-2008-08-078220>.
6. M. A. Goodell, K. Brose, G. Paradis, A. S. Conner, & R. C. Mulligan, Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *Journal of Experimental Medicine*, **183** (1996). <https://doi.org/10.1084/jem.183.4.1797>.
7. M. Kondo, A. J. Wagers, M. G. Manz, S. S. Prohaska, D. C. Scherer, G. F. Beilhack, J. A. Shizuru, & I. L. Weissman, Biology of Hematopoietic Stem Cells and Progenitors: Implications for Clinical Application. *Annual Review of Immunology*, **21** (2003) 759–806. <https://doi.org/10.1146/annurev.immunol.21.120601.141007>.
8. H. Iwasaki & K. Akashi, Myeloid Lineage Commitment from the Hematopoietic Stem Cell. *Immunity*, **26** (2007). <https://doi.org/10.1016/j.immuni.2007.06.004>.
9. G. J. Spangrude, S. Heimfeld, & I. L. Weissman, Purification and characterization of mouse hematopoietic stem cells. *Science*, **241** (1988) 58–62. <https://doi.org/10.1126/science.2898810>.
10. K. Ikuta & I. L. Weissman, Evidence that Hematopoietic Stem Cells Express Mouse c-kit but do not Depend on Steel Factor for Their Generation. *Proceedings of the National Academy of Sciences of the United States of America*, **89** (1992) 1502–1506.
11. C. M. Baum, I. L. Weissman, A. S. Tsukamoto, A. M. Buckle, & B. Peault, Isolation of a candidate human hematopoietic stem-cell population. *Proceedings of the National Academy of Sciences*, **89** (1992) 2804 LP – 2808. <https://doi.org/10.1073/pnas.89.7.2804>.
12. R. J. Berenson, R. G. Andrews, W. I. Bensinger, D. Kalamasz, G. Knitter, C. D. Buckner, & I. D. Bernstein, Antigen CD34+ marrow cells engraft lethally irradiated baboons. *Journal of Clinical Investigation*, **81** (1988). <https://doi.org/10.1172/JCI113409>.
13. M. Tuthill & Hatzimichael, Hematopoietic stem cell transplantation. *Stem Cells and Cloning: Advances and Applications*, (2010). <https://doi.org/10.2147/SCCAA.S6815>.
14. M. Kondo, I. L. Weissman, & K. Akashi, Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow. *Cell*, **91** (1997) 661–672. [https://doi.org/https://doi.org/10.1016/S0092-8674\(00\)80453-5](https://doi.org/https://doi.org/10.1016/S0092-8674(00)80453-5).
15. K. Akashi, D. Traver, T. Miyamoto, & I. L. Weissman, A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, **404** (2000) 193–197. <https://doi.org/10.1038/35004599>.
16. S. H. Orkin & L. I. Zon, Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell*, **132** (2008). <https://doi.org/10.1016/j.cell.2008.01.025>.

17. F. Shojaei, J. Trowbridge, L. Gallacher, L. Yuefei, D. Goodale, F. Karanu, K. Levac, & M. Bhatia, Hierarchical and Ontogenic Positions Serve to Define the Molecular Basis of Human Hematopoietic Stem Cell Behavior. *Developmental Cell*, **8** (2005) 651–663. <https://doi.org/https://doi.org/10.1016/j.devcel.2005.03.004>.
18. M. Jagannathan-Bogdan & L. I. Zon, Hematopoiesis. *Development*, **140** (2013) 2463 LP – 2467. <https://doi.org/10.1242/dev.083147>.
19. K. Ottersbach, Endothelial-to-haematopoietic transition: an update on the process of making blood. *Biochemical Society Transactions*, **47** (2019) 591–601. <https://doi.org/10.1042/BST20180320>.
20. A. Medvinsky, S. Rybtsov, & S. Taoudi, Embryonic origin of the adult hematopoietic system: advances and questions. *Development*, **138** (2011) 1017 LP – 1031. <https://doi.org/10.1242/dev.040998>.
21. H. K. A. Mikkola & S. H. Orkin, The journey of developing hematopoietic stem cells. *Development*, **133** (2006) 3733 LP – 3744. <https://doi.org/10.1242/dev.02568>.
22. C. N. Inra, B. O. Zhou, M. Acar, M. M. Murphy, J. Richardson, Z. Zhao, & S. J. Morrison, A perisinusoidal niche for extramedullary haematopoiesis in the spleen. *Nature*, **527** (2015) 466–471. <https://doi.org/10.1038/nature15530>.
23. E. Lefrançais, G. Ortiz-Muñoz, A. Caudrillier, B. Mallavia, F. Liu, D. M. Sayah, E. E. Thornton, M. B. Headley, T. David, S. R. Coughlin, M. F. Krummel, A. D. Leavitt, E. Passegué, & M. R. Looney, The lung is a site of platelet biogenesis and a reservoir for haematopoietic progenitors. *Nature*, **544** (2017) 105–109. <https://doi.org/10.1038/nature21706>.
24. S. Méndez-Ferrer, D. Lucas, M. Battista, & P. S. Frenette, Haematopoietic stem cell release is regulated by circadian oscillations. *Nature*, **452** (2008) 442–447. <https://doi.org/10.1038/nature06685>.
25. H. Geiger & G. van Zant, The aging of lympho-hematopoietic stem cells. *Nature Immunology*, **3** (2002) 329–333. <https://doi.org/10.1038/ni0402-329>.
26. E. Laurenti, C. Frelin, S. Xie, R. Ferrari, C. F. Dunant, S. Zandi, A. Neumann, I. Plumb, S. Doulatov, J. Chen, C. April, J.-B. Fan, N. Iscove, & J. E. Dick, CDK6 Levels Regulate Quiescence Exit in Human Hematopoietic Stem Cells. *Cell Stem Cell*, **16** (2015) 302–313. <https://doi.org/https://doi.org/10.1016/j.stem.2015.01.017>.
27. E. Laurenti & B. Göttgens, From haematopoietic stem cells to complex differentiation landscapes. *Nature*, **553** (2018) 418–426. <https://doi.org/10.1038/nature25022>.
28. S. J. Morrison & D. T. Scadden, The bone marrow niche for haematopoietic stem cells. *Nature*, **505** (2014) 327–334. <https://doi.org/10.1038/nature12984>.
29. A. Mendelson & P. S. Frenette, Hematopoietic stem cell niche maintenance during homeostasis and regeneration. *Nature Medicine*, **20** (2014) 833–846. <https://doi.org/10.1038/nm.3647>.
30. A. Wilson & A. Trumpp, Bone-marrow haematopoietic-stem-cell niches. *Nature Reviews Immunology*, **6** (2006) 93–106. <https://doi.org/10.1038/nri1779>.
31. J. Zhang, Q. Wu, C. B. Johnson, G. Pham, J. M. Kinder, A. Olsson, A. Slaughter, M. May, B. Weinhaus, A. D'Alessandro, J. D. Engel, J. X. Jiang, J. M. Kofron, L. F. Huang, V. B. S. Prasath, S. S. Way, N. Salomonis, H. L. Grimes, & D. Lucas, In situ mapping identifies distinct vascular niches for myelopoiesis. *Nature*, **590** (2021) 457–462. <https://doi.org/10.1038/s41586-021-03201-2>.
32. M. A. G. Essers, S. Offner, W. E. Blanco-Bose, Z. Waibler, U. Kalinke, M. A. Duchosal, & A. Trumpp, IFN α activates dormant haematopoietic stem cells in vivo. *Nature*, **458** (2009) 904–908. <https://doi.org/10.1038/nature07815>.
33. C. J. Watson, A. L. Papula, G. Y. P. Poon, W. H. Wong, A. L. Young, T. E. Druley, D. S. Fisher, & J. R. Blundell, The evolutionary dynamics and fitness landscape of clonal

- hematopoiesis. *Science*, **367** (2020) 1449 LP – 1454.
<https://doi.org/10.1126/science.aay9333>.
34. C. Benz, M. R. Copley, D. G. Kent, S. Wohrer, A. Cortes, N. Aghaeepour, E. Ma, H. Mader, K. Rowe, C. Day, D. Treloar, R. R. Brinkman, & C. J. Eaves, Hematopoietic Stem Cell Subtypes Expand Differentially during Development and Display Distinct Lymphopoietic Programs. *Cell Stem Cell*, **10** (2012) 273–283.
<https://doi.org/https://doi.org/10.1016/j.stem.2012.02.007>.
 35. A. L. Young, G. A. Challen, B. M. Birman, & T. E. Druley, Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nature Communications*, **7** (2016) 12484. <https://doi.org/10.1038/ncomms12484>.
 36. T. McKerrell, N. Park, T. Moreno, C. S. Grove, H. Ponstingl, J. Stephens, C. Crawley, J. Craig, M. A. Scott, C. Hodgkinson, J. Baxter, R. Rad, D. R. Forsyth, M. A. Quail, E. Zeggini, W. Ouwehand, I. Varela, & G. S. Vassiliou, Leukemia-Associated Somatic Mutations Drive Distinct Patterns of Age-Related Clonal Hemopoiesis. *Cell Reports*, **10** (2015) 1239–1245. <https://doi.org/https://doi.org/10.1016/j.celrep.2015.02.005>.
 37. S. Jaiswal, P. Fontanillas, J. Flannick, A. Manning, P. v Grauman, B. G. Mar, R. C. Lindsley, C. H. Mermel, N. Burt, A. Chavez, J. M. Higgins, V. Moltchanov, F. C. Kuo, M. J. Kluk, B. Henderson, L. Kinnunen, H. A. Koistinen, C. Ladenvall, G. Getz, A. Correa, B. F. Banahan, S. Gabriel, S. Kathiresan, H. M. Stringham, M. I. McCarthy, M. Boehnke, J. Tuomilehto, C. Haiman, L. Groop, G. Atzmon, J. G. Wilson, D. Neuberg, D. Altshuler, & B. L. Ebert, Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine*, **371** (2014) 2488–2498.
<https://doi.org/10.1056/NEJMoa1408617>.
 38. D. Metcalf, *Clonal culture of hemopoietic cells: techniques and applications* (Elsevier Publishing Company, 1984).
 39. T. Graf & T. Enver, Forcing cells to change lineages . *Nature*, **462** (2009) 587–594.
<https://doi.org/10.1038/nature08533>.
 40. M. Hu, D. Krause, M. Greaves, S. Sharkis, M. Dexter, C. Heyworth, & T. Enver, Multilineage gene expression precedes commitment in the hemopoietic system. *Genes & Development*, **11** (1997). <https://doi.org/10.1101/gad.11.6.774>.
 41. P. Laslo, C. J. Spooner, A. Warmflash, D. W. Lancki, H.-J. Lee, R. Sciammas, B. N. Gantner, A. R. Dinner, & H. Singh, Multilineage Transcriptional Priming and Determination of Alternate Hematopoietic Cell Fates. *Cell*, **126** (2006) 755–766.
<https://doi.org/https://doi.org/10.1016/j.cell.2006.06.052>.
 42. T. Enver, C. M. Heyworth, & T. M. Dexter, Do Stem Cells Play Dice? *Blood*, **92** (1998) 348–351. <https://doi.org/https://doi.org/10.1182/blood.V92.2.348>.
 43. C. Pina, C. Fugazza, A. J. Tipping, J. Brown, S. Soneji, J. Teles, C. Peterson, & T. Enver, Inferring rules of lineage commitment in haematopoiesis. *Nature Cell Biology*, **14** (2012) 287–294. <https://doi.org/10.1038/ncb2442>.
 44. L. Robb, Cytokine receptors and hematopoietic differentiation. *Oncogene*, **26** (2007) 6715–6723. <https://doi.org/10.1038/sj.onc.1210756>.
 45. J. E. Dick, M. C. Magli, D. Huszar, R. A. Phillips, & A. Bernstein, Introduction of a selectable gene into primitive stem cells capable of long-term reconstitution of the hemopoietic system of W/W^v mice. *Cell*, **42** (1985) 71–79.
[https://doi.org/https://doi.org/10.1016/S0092-8674\(85\)80102-1](https://doi.org/https://doi.org/10.1016/S0092-8674(85)80102-1).
 46. C. E. Müller-Sieburg, R. H. Cho, M. Thoman, B. Adkins, & H. B. Sieburg, Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. *Blood*, **100** (2002) 1302–1309.
https://doi.org/10.1182/blood.V100.4.1302.h81602001302_1302_1309.

47. R. Yamamoto, Y. Morita, J. Ooehara, S. Hamanaka, M. Onodera, K. L. Rudolph, H. Ema, & H. Nakauchi, Clonal Analysis Unveils Self-Renewing Lineage-Restricted Progenitors Generated Directly from Hematopoietic Stem Cells. *Cell*, **154** (2013) 1112–1126. <https://doi.org/https://doi.org/10.1016/j.cell.2013.08.007>.
48. A. Sanjuan-Pla, I. C. Macaulay, C. T. Jensen, P. S. Woll, T. C. Luis, A. Mead, S. Moore, C. Carella, S. Matsuoka, T. B. Jones, O. Chowdhury, L. Stenson, M. Lutteropp, J. C. A. Green, R. Facchini, H. Boukarabila, A. Grover, A. Gambardella, S. Thongjuea, J. Carrelha, P. Tarrant, D. Atkinson, S.-A. Clark, C. Nerlov, & S. E. W. Jacobsen, Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*, **502** (2013) 232–236. <https://doi.org/10.1038/nature12495>.
49. L. Velten, S. F. Haas, S. Raffel, S. Blaszkiewicz, S. Islam, B. P. Hennig, C. Hirche, C. Lutz, E. C. Buss, D. Nowak, T. Boch, W.-K. Hofmann, A. D. Ho, W. Huber, A. Trumpp, M. A. G. Essers, & L. M. Steinmetz, Human haematopoietic stem cell lineage commitment is a continuous process. *Nature Cell Biology*, **19** (2017) 271–281. <https://doi.org/10.1038/ncb3493>.
50. D. Karamitros, B. Stoilova, Z. Aboukhalil, F. Hamey, A. Reinisch, M. Samitsch, L. Quek, G. Otto, E. Repapi, J. Doondeea, B. Usukhbayar, J. Calvo, S. Taylor, N. Goardon, E. Six, F. Pflumio, C. Porcher, R. Majeti, B. Göttgens, & P. Vyas, Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nature Immunology*, **19** (2018) 85–97. <https://doi.org/10.1038/s41590-017-0001-2>.
51. J. Adolfsson, R. Månsson, N. Buza-Vidas, A. Hultquist, K. Liuba, C. T. Jensen, D. Bryder, L. Yang, O.-J. Borge, L. A. M. Thoren, K. Anderson, E. Sitnicka, Y. Sasaki, M. Sigvardsson, & S. E. W. Jacobsen, Identification of Flt3⁺ Lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential: A Revised Road Map for Adult Blood Lineage Commitment. *Cell*, **121** (2005) 295–306. <https://doi.org/https://doi.org/10.1016/j.cell.2005.02.013>.
52. F. Notta, S. Zandi, N. Takayama, S. Dobson, O. I. Gan, G. Wilson, K. B. Kaufmann, J. McLeod, E. Laurenti, C. F. Dunant, J. D. McPherson, L. D. Stein, Y. Dror, & J. E. Dick, Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*, **351** (2016) aab2116. <https://doi.org/10.1126/science.aab2116>.
53. J. Sun, A. Ramos, B. Chapman, J. B. Johnnidis, L. Le, Y.-J. Ho, A. Klein, O. Hofmann, & F. D. Camargo, Clonal dynamics of native haematopoiesis. *Nature*, **514** (2014) 322–327. <https://doi.org/10.1038/nature13824>.
54. K. J. Knudsen, M. Rehn, M. S. Hasemann, N. Rapin, F. O. Bagger, E. Ohlsson, A. Willer, A.-K. Frank, E. Søndergaard, J. Jendholm, L. Thorén, J. Lee, J. Rak, K. Theilgaard-Mönch, & B. T. Porse, ERG promotes the maintenance of hematopoietic stem cells by restricting their differentiation. *Genes & Development*, **29** (2015). <https://doi.org/10.1101/gad.268409.115>.
55. E. A. Kruse, S. J. Loughran, T. M. Baldwin, E. C. Josefsson, S. Ellis, D. K. Watson, P. Nurden, D. Metcalf, D. J. Hilton, W. S. Alexander, & B. T. Kile, Dual requirement for the ETS transcription factors Fli-1 and Erg in hematopoietic stem cells and the megakaryocyte lineage. *Proceedings of the National Academy of Sciences*, **106** (2009). <https://doi.org/10.1073/pnas.0906556106>.
56. K. Miyamoto, K. Y. Araki, K. Naka, F. Arai, K. Takubo, S. Yamazaki, S. Matsuoka, T. Miyamoto, K. Ito, M. Ohmura, C. Chen, K. Hosokawa, H. Nakauchi, K. Nakayama, K. I. Nakayama, M. Harada, N. Motoyama, T. Suda, & A. Hirao, Foxo3a Is Essential for Maintenance of the Hematopoietic Stem Cell Pool. *Cell Stem Cell*, **1** (2007). <https://doi.org/10.1016/j.stem.2007.02.001>.

57. J. Takai, T. Moriguchi, M. Suzuki, L. Yu, K. Ohneda, & M. Yamamoto, The Gata1 5' region harbors distinct cis-regulatory modules that direct gene activation in erythroid cells and gene inactivation in HSCs. *Blood*, **122** (2013). <https://doi.org/10.1182/blood-2013-01-476911>.
58. C. Perry & H. Soreq, Transcriptional regulation of erythropoiesis. *European Journal of Biochemistry*, **269** (2002) 3607–3618. <https://doi.org/https://doi.org/10.1046/j.1432-1033.2002.02999.x>.
59. N. P. Rodrigues, A. J. Tipping, Z. Wang, & T. Enver, GATA-2 mediated regulation of normal hematopoietic stem/progenitor cell function, myelodysplasia and myeloid leukemia. *The International Journal of Biochemistry & Cell Biology*, **44** (2012). <https://doi.org/10.1016/j.biocel.2011.12.004>.
60. C.-J. Ku, T. Hosoya, I. Maillard, & J. D. Engel, GATA-3 regulates hematopoietic stem cell maintenance and cell-cycle entry. *Blood*, **119** (2012). <https://doi.org/10.1182/blood-2011-07-366070>.
61. D. Yang, X. Zhang, Y. Dong, X. Liu, T. Wang, X. Wang, Y. Geng, S. Fang, Y. Zheng, X. Chen, J. Chen, G. Pan, & J. Wang, Enforced expression of Hoxa5 in haematopoietic stem cells leads to aberrant erythropoiesis in vivo. *Cell Cycle*, **14** (2015). <https://doi.org/10.4161/15384101.2014.992191>.
62. C. T. Collins & J. L. Hess, Role of HOXA9 in leukemia: dysregulation, cofactors and essential targets. *Oncogene*, **35** (2016) 1090–1098. <https://doi.org/10.1038/onc.2015.174>.
63. G. P. Souroullas, J. M. Salmon, F. Sablitzky, D. J. Curtis, & M. A. Goodell, Adult Hematopoietic Stem and Progenitor Cells Require Either Lyl1 or Scl for Survival. *Cell Stem Cell*, **4** (2009). <https://doi.org/10.1016/j.stem.2009.01.001>.
64. Y. K. Lieu & E. P. Reddy, Conditional c-myc knockout in adult hematopoietic stem cells leads to loss of self-renewal due to impaired proliferation and accelerated differentiation. *Proceedings of the National Academy of Sciences*, **106** (2009). <https://doi.org/10.1073/pnas.0907623106>.
65. N. K. Wilson, S. D. Foster, X. Wang, K. Knezevic, J. Schütte, P. Kaimakis, P. M. Chilarska, S. Kinston, W. H. Ouwehand, E. Dzierzak, J. E. Pimanda, M. F. T. R. de Bruijn, & B. Göttgens, Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators. *Cell Stem Cell*, **7** (2010). <https://doi.org/10.1016/j.stem.2010.07.016>.
66. H. Iwasaki, C. Somoza, H. Shigematsu, E. A. Duprez, J. Iwasaki-Arai, S. Mizuno, Y. Arinobu, K. Geary, P. Zhang, T. Dayaram, M. L. Fenyus, S. Elf, S. Chan, P. Kastner, C. S. Huettner, R. Murray, D. G. Tenen, & K. Akashi, Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. *Blood*, **106** (2005). <https://doi.org/10.1182/blood-2005-03-0860>.
67. J. L. Schultze & M. Beyer, Myelopoiesis Reloaded: Single-Cell Transcriptomics Leads the Way. *Immunity*, **44** (2016). <https://doi.org/10.1016/j.immuni.2015.12.019>.
68. F. Fazi, A. Rosa, A. Fatica, V. Gelmetti, M. L. de Marchis, C. Nervi, & I. Bozzoni, A Minicircuitry Comprised of MicroRNA-223 and Transcription Factors NFI-A and C/EBP α Regulates Human Granulopoiesis. *Cell*, **123** (2005). <https://doi.org/10.1016/j.cell.2005.09.023>.
69. F. Paul, Y. Arkin, A. Giladi, D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, D. Winter, D. Lara-Astiaso, M. Gury, A. Weiner, E. David, N. Cohen, F. K. B. Lauridsen, S. Haas, A. Schlitzer, A. Mildner, F. Ginhoux, S. Jung, A. Trumpp, B. T. Porse, A. Tanay, & I. Amit, Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell*, **163** (2015). <https://doi.org/10.1016/j.cell.2015.11.013>.

70. N. Kraguljac, D. Marisavljevic, G. Jankovic, N. Radošević, M. Pantic, M. Donfrid, N. Miletic, D. Boškovic, & M. Colovic, Characterization of CD13 and CD33 Surface Antigen–Negative Acute Myeloid Leukemia. *American Journal of Clinical Pathology*, **114** (2000). <https://doi.org/10.1309/MFCP-7GMW-AQM4-ED3N>.
71. J. L. Schultze, E. Mass, & A. Schlitzer, Emerging Principles in Myelopoiesis at Homeostasis and during Infection and Inflammation. *Immunity*, **50** (2019). <https://doi.org/10.1016/j.immuni.2019.01.019>.
72. J. Palis, Primitive and definitive erythropoiesis in mammals . *Frontiers in Physiology* , **5** (2014) 3.
73. K. Chen, J. Liu, S. Heck, J. A. Chasis, X. An, & N. Mohandas, Resolving the distinct stages in erythroid differentiation based on dynamic changes in membrane protein expression during erythropoiesis. *Proceedings of the National Academy of Sciences*, **106** (2009) 17413 LP – 17418. <https://doi.org/10.1073/pnas.0909296106>.
74. S. M. Hattangadi, P. Wong, L. Zhang, J. Flygare, & H. F. Lodish, From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. *Blood*, **118** (2011) 6258–6268. <https://doi.org/10.1182/blood-2011-07-356006>.
75. V. G. Sankaran & S. H. Orkin, The Switch from Fetal to Adult Hemoglobin. *Cold Spring Harbor Perspectives in Medicine*, **3** (2013). <https://doi.org/10.1101/cshperspect.a011643>.
76. H. Lodish, J. Flygare, & S. Chou, From stem cell to erythroblast: Regulation of red cell production at multiple levels by multiple hormones. *IUBMB Life*, **62** (2010) 492–496. <https://doi.org/https://doi.org/10.1002/iub.322>.
77. E. Dzierzak & S. Philipsen, Erythropoiesis: Development and Differentiation. *Cold Spring Harbor Perspectives in Medicine*, **3** (2013). <https://doi.org/10.1101/cshperspect.a011601>.
78. W. Jelkmann, Regulation of erythropoietin production. *The Journal of Physiology*, **589** (2011). <https://doi.org/10.1113/jphysiol.2010.195057>.
79. H. Wu, X. Liu, R. Jaenisch, & H. F. Lodish, Generation of committed erythroid BFU-E and CFU-E progenitors does not require erythropoietin or the erythropoietin receptor. *Cell*, **83** (1995) 59–67. [https://doi.org/https://doi.org/10.1016/0092-8674\(95\)90234-1](https://doi.org/https://doi.org/10.1016/0092-8674(95)90234-1).
80. E. Liboi, M. Carroll, A. D. D’Andrea, & B. Mathey-Prevot, Erythropoietin receptor signals both proliferation and erythroid-specific differentiation. *Proceedings of the National Academy of Sciences*, **90** (1993). <https://doi.org/10.1073/pnas.90.23.11351>.
81. M. J. KOURY & M. C. BONDURANT, The molecular mechanism of erythropoietin action. *European Journal of Biochemistry*, **210** (1992). <https://doi.org/10.1111/j.1432-1033.1992.tb17466.x>.
82. A. Grover, E. Mancini, S. Moore, A. J. Mead, D. Atkinson, K. D. Rasmussen, D. O’Carroll, S. E. W. Jacobsen, & C. Nerlov, Erythropoietin guides multipotent hematopoietic progenitor cells toward an erythroid fate. *Journal of Experimental Medicine*, **211** (2014). <https://doi.org/10.1084/jem.20131189>.
83. R. Ferreira, K. Ohneda, M. Yamamoto, & S. Philipsen, GATA1 Function, a Paradigm for Transcription Factors in Hematopoiesis. *Molecular and Cellular Biology*, **25** (2005). <https://doi.org/10.1128/MCB.25.4.1215-1227.2005>.
84. T. Moriguchi & M. Yamamoto, A regulatory network governing Gata1 and Gata2 gene transcription orchestrates erythroid lineage differentiation. *International Journal of Hematology*, **100** (2014). <https://doi.org/10.1007/s12185-014-1568-0>.
85. L. Pevny, M. C. Simon, E. Robertson, W. H. Klein, S.-F. Tsai, V. D’Agati, S. H. Orkin, & F. Costantini, Erythroid differentiation in chimaeric mice blocked by a

- targeted mutation in the gene for transcription factor GATA-1. *Nature*, **349** (1991). <https://doi.org/10.1038/349257a0>.
86. Y. Fujiwara, C. P. Browne, K. Cunniff, S. C. Goff, & S. H. Orkin, Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proceedings of the National Academy of Sciences*, **93** (1996). <https://doi.org/10.1073/pnas.93.22.12355>.
 87. F.-Y. Tsai & S. H. Orkin, Transcription Factor GATA-2 Is Required for Proliferation/Survival of Early Hematopoietic Cells and Mast Cell Formation, But Not for Erythroid and Myeloid Terminal Differentiation. *Blood*, **89** (1997) 3636–3643. <https://doi.org/10.1182/blood.V89.10.3636>.
 88. J. Lacombe, G. Kroschl, M. Tremblay, B. Gerby, R. Martin, P. D. Aplan, S. Lemieux, & T. Hoang, Genetic interaction between Kit and Scl. *Blood*, **122** (2013) 1150–1161. <https://doi.org/10.1182/blood-2011-01-331819>.
 89. H. L. Hsu, J. T. Cheng, Q. Chen, & R. Baer, Enhancer-binding activity of the tal-1 oncoprotein in association with the E47/E12 helix-loop-helix proteins. *Molecular and Cellular Biology*, **11** (1991) 3037 LP – 3042. <https://doi.org/10.1128/MCB.11.6.3037>.
 90. M. A. Kerenyi & S. H. Orkin, Networking erythropoiesis. *Journal of Experimental Medicine*, **207** (2010). <https://doi.org/10.1084/jem.20102260>.
 91. P. Zhang, G. Behre, J. Pan, A. Iwama, N. Wara-aswapati, H. S. Radomska, P. E. Auron, D. G. Tenen, & Z. Sun, Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. *Proceedings of the National Academy of Sciences*, **96** (1999). <https://doi.org/10.1073/pnas.96.15.8705>.
 92. M. R. Tallack & A. C. Perkins, KLF1 directly coordinates almost all aspects of terminal erythroid differentiation. *IUBMB Life*, **62** (2010). <https://doi.org/10.1002/iub.404>.
 93. M. Siatecka & J. J. Bieker, The multifunctional role of EKLF/KLF1 during erythropoiesis. *Blood*, **118** (2011) 2044–2054. <https://doi.org/10.1182/blood-2011-03-331371>.
 94. A. B. Cantor & S. H. Orkin, Coregulation of GATA factors by the Friend of GATA (FOG) family of multitype zinc finger proteins. *Seminars in Cell & Developmental Biology*, **16** (2005). <https://doi.org/10.1016/j.semcdb.2004.10.006>.
 95. T. Hoang, M.-C. Sincennes, V. Lisi, D. Veiga, M. Humbert, F. Major, B. Affar, & A. Verreault, The dual function of LMO2 in driving erythroid cell fate. *Experimental Hematology*, **53** (2017). <https://doi.org/10.1016/j.exphem.2017.06.235>.
 96. A. Tamir, J. Howard, R. R. Higgins, Y.-J. Li, L. Berger, E. Zacksenhaus, M. Reis, & Y. Ben-David, Fli-1, an Ets-Related Transcription Factor, Regulates Erythropoietin-Induced Erythroid Proliferation and Differentiation: Evidence for Direct Transcriptional Repression of the Rb Gene during Differentiation. *Molecular and Cellular Biology*, **19** (1999). <https://doi.org/10.1128/MCB.19.6.4452>.
 97. M. Athanasiou, G. Mavrothalassitis, L. Sun-Hoffman, & D. G. Blair, FLI-1 is a suppressor of erythroid differentiation in human hematopoietic cells. *Leukemia*, **14** (2000) 439–445. <https://doi.org/10.1038/sj.leu.2401689>.
 98. A. P. Tsang, Y. Fujiwara, D. B. Hom, & S. H. Orkin, Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes & Development*, **12** (1998). <https://doi.org/10.1101/gad.12.8.1176>.
 99. A. J. Warren, W. H. Colledge, M. B. L. Carlton, M. J. Evans, A. J. H. Smith, & T. H. Rabbitts, The Oncogenic Cysteine-rich LIM domain protein Rbtn2 is essential for erythroid development. *Cell*, **78** (1994). [https://doi.org/10.1016/0092-8674\(94\)90571-1](https://doi.org/10.1016/0092-8674(94)90571-1).

100. R. A. Shivdasani, E. L. Mayer, & S. H. Orkin, Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature*, **373** (1995) 432–434. <https://doi.org/10.1038/373432a0>.
101. K. R. Kampen, The discovery and early understanding of leukemia. *Leukemia Research*, **36** (2012). <https://doi.org/10.1016/j.leukres.2011.09.028>.
102. E. Estey & H. Döhner, Acute myeloid leukaemia. *The Lancet*, **368** (2006). [https://doi.org/10.1016/S0140-6736\(06\)69780-8](https://doi.org/10.1016/S0140-6736(06)69780-8).
103. A. S. Roug, M. C. Hansen, L. Nederby, & P. Hokland, Diagnosing and following adult patients with acute myeloid leukaemia in the genomic age. *British Journal of Haematology*, **167** (2014). <https://doi.org/10.1111/bjh.13048>.
104. A. J. Dunbar, R. K. Rampal, & R. Levine, Leukemia secondary to myeloproliferative neoplasms. *Blood*, **136** (2020) 61–70. <https://doi.org/10.1182/blood.2019000943>.
105. T. Lapidot, C. Sirard, J. Vormoor, B. Murdoch, T. Hoang, J. Caceres-Cortes, M. Minden, B. Paterson, M. A. Caligiuri, & J. E. Dick, A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*, **367** (1994) 645–648. <https://doi.org/10.1038/367645a0>.
106. D. Bonnet & J. E. Dick, Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Medicine*, **3** (1997) 730–737. <https://doi.org/10.1038/nm0797-730>.
107. D. Thomas & R. Majeti, Biology and relevance of human acute myeloid leukemia stem cells. *Blood*, **129** (2017) 1577–1585. <https://doi.org/10.1182/blood-2016-10-696054>.
108. D. C. Taussig, J. Vargaftig, F. Miraki-Moud, E. Griessinger, K. Sharrock, T. Luke, D. Lillington, H. Oakervee, J. Cavenagh, S. G. Agrawal, T. A. Lister, J. G. Gribben, & D. Bonnet, Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34– fraction. *Blood*, **115** (2010). <https://doi.org/10.1182/blood-2009-02-206565>.
109. A. Cozzio, Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. *Genes & Development*, **17** (2003). <https://doi.org/10.1101/gad.1143403>.
110. A. v Krivtsov, D. Twomey, Z. Feng, M. C. Stubbs, Y. Wang, J. Faber, J. E. Levine, J. Wang, W. C. Hahn, D. G. Gilliland, T. R. Golub, & S. A. Armstrong, Transformation from committed progenitor to leukaemia stem cell initiated by MLL–AF9. *Nature*, **442** (2006) 818–822. <https://doi.org/10.1038/nature04980>.
111. J.-E. Sarry, K. Murphy, R. Perry, P. v. Sanchez, A. Secreto, C. Keefer, C. R. Swider, A.-C. Strzelecki, C. Cavelier, C. Récher, V. Mansat-De Mas, E. Delabesse, G. Danet-Desnoyers, & M. Carroll, Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2R γ c-deficient mice. *Journal of Clinical Investigation*, **121** (2011). <https://doi.org/10.1172/JCI41495>.
112. N. Goardon, E. Marchi, A. Atzberger, L. Quek, A. Schuh, S. Soneji, P. Woll, A. Mead, K. A. Alford, R. Rout, S. Chaudhury, A. Gilkes, S. Knapper, K. Beldjord, S. Begum, S. Rose, N. Geddes, M. Griffiths, G. Standen, A. Sternberg, J. Cavenagh, H. Hunter, D. Bowen, S. Killick, L. Robinson, A. Price, E. Macintyre, P. Virgo, A. Burnett, C. Craddock, T. Enver, S. E. W. Jacobsen, C. Porcher, & P. Vyas, Coexistence of LMPP-like and GMP-like Leukemia Stem Cells in Acute Myeloid Leukemia. *Cancer Cell*, **19** (2011) 138–152. <https://doi.org/10.1016/j.ccr.2010.12.012>.
113. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *New England Journal of Medicine*, **368** (2013) 2059–2074. <https://doi.org/10.1056/NEJMoa1301689>.

114. J. M. Bennett, D. Catovsky, M.-T. Daniel, G. Flandrin, D. A. G. Galton, H. R. Gralnick, & C. Sultan, Proposals for the Classification of the Acute Leukaemias French-American-British (FAB) Co-operative Group. *British Journal of Haematology*, **33** (1976). <https://doi.org/10.1111/j.1365-2141.1976.tb03563.x>.
115. D. A. Arber, A. Orazi, R. Hasserjian, J. Thiele, M. J. Borowitz, M. M. le Beau, C. D. Bloomfield, M. Cazzola, & J. W. Vardiman, The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*, **127** (2016). <https://doi.org/10.1182/blood-2016-03-643544>.
116. S.-J. Chen, Y. Shen, & Z. Chen, A panoramic view of acute myeloid leukemia. *Nature Genetics*, **45** (2013). <https://doi.org/10.1038/ng.2651>.
117. E. Papaemmanuil, M. Gerstung, L. Bullinger, V. I. Gaidzik, P. Paschka, N. D. Roberts, N. E. Potter, M. Heuser, F. Thol, N. Bolli, G. Gundem, P. van Loo, I. Martincorena, P. Ganly, L. Mudie, S. McLaren, S. O'Meara, K. Raine, D. R. Jones, J. W. Teague, A. P. Butler, M. F. Greaves, A. Ganser, K. Döhner, R. F. Schlenk, H. Döhner, & P. J. Campbell, Genomic Classification and Prognosis in Acute Myeloid Leukemia. *New England Journal of Medicine*, **374** (2016) 2209–2221. <https://doi.org/10.1056/NEJMoa1516192>.
118. Y. Sun, B.-R. Chen, & A. Deshpande, Epigenetic Regulators in the Development, Maintenance, and Therapeutic Targeting of Acute Myeloid Leukemia . *Frontiers in Oncology* , **8** (2018) 41.
119. P. Gallipoli, G. Giotopoulos, & B. J. P. Huntly, Epigenetic regulators as promising therapeutic targets in acute myeloid leukemia. *Therapeutic Advances in Hematology*, **6** (2015). <https://doi.org/10.1177/2040620715577614>.
120. K. A. Fennell, C. C. Bell, & M. A. Dawson, Epigenetic therapies in acute myeloid leukemia: where to from here? *Blood*, **134** (2019). <https://doi.org/10.1182/blood.2019003262>.
121. M. R. Corces, J. D. Buenrostro, B. Wu, P. G. Greenside, S. M. Chan, J. L. Koenig, M. P. Snyder, J. K. Pritchard, A. Kundaje, W. J. Greenleaf, R. Majeti, & H. Y. Chang, Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nature Genetics*, **48** (2016) 1193–1203. <https://doi.org/10.1038/ng.3646>.
122. G. L. Brien, D. G. Valerio, & S. A. Armstrong, Exploiting the Epigenome to Control Cancer-Promoting Gene-Expression Programs. *Cancer Cell*, **29** (2016). <https://doi.org/10.1016/j.ccell.2016.03.007>.
123. M. A. Dawson, R. K. Prinjha, A. Dittmann, G. Giotopoulos, M. Bantscheff, W.-I. Chan, S. C. Robson, C. Chung, C. Hopf, M. M. Savitski, C. Huthmacher, E. Gudgin, D. Lugo, S. Beinke, T. D. Chapman, E. J. Roberts, P. E. Soden, K. R. Auger, O. Mirguet, K. Doehner, R. Delwel, A. K. Burnett, P. Jeffrey, G. Drewes, K. Lee, B. J. P. Huntly, & T. Kouzarides, Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature*, **478** (2011). <https://doi.org/10.1038/nature10509>.
124. O. Abdel-Wahab & R. L. Levine, Mutations in epigenetic modifiers in the pathogenesis and therapy of acute myeloid leukemia. *Blood*, **121** (2013). <https://doi.org/10.1182/blood-2013-01-451781>.
125. M. Cusan, S. F. Cai, H. P. Mohammad, A. Krivtsov, A. Chramiec, E. Loizou, M. D. Witkin, K. N. Smitheman, D. G. Tenen, M. Ye, B. Will, U. Steidl, R. G. Kruger, R. L. Levine, H. Y. Rienhoff, R. P. Koche, & S. A. Armstrong, LSD1 inhibition exerts its antileukemic effect by recommissioning PU.1- and C/EBP α -dependent enhancers in AML. *Blood*, **131** (2018). <https://doi.org/10.1182/blood-2017-09-807024>.

126. E. M. Stein, G. Garcia-Manero, D. A. Rizzieri, R. Tibes, J. G. Berdeja, M. R. Savona, M. Jongen-Lavrenic, J. K. Altman, B. Thomson, S. J. Blakemore, S. R. Daigle, N. J. Waters, A. B. Suttle, A. Clawson, R. Pollock, A. Krivtsov, S. A. Armstrong, J. DiMartino, E. Hedrick, B. Löwenberg, & M. S. Tallman, The DOT1L inhibitor pinometostat reduces H3K79 methylation and has modest clinical activity in adult acute leukemia. *Blood*, **131** (2018). <https://doi.org/10.1182/blood-2017-12-818948>.
127. C. H. Waddington, The Epigenotype. *International Journal of Epidemiology*, **41** (2012). <https://doi.org/10.1093/ije/dyr184>.
128. C. H. WADDINGTON, Towards a Theoretical Biology. *Nature*, **218** (1968). <https://doi.org/10.1038/218525a0>.
129. C. D. Allis & T. Jenuwein, The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*, **17** (2016) 487–500. <https://doi.org/10.1038/nrg.2016.59>.
130. T. Kouzarides, Chromatin Modifications and Their Function. *Cell*, **128** (2007). <https://doi.org/10.1016/j.cell.2007.02.005>.
131. A. J. Bannister & T. Kouzarides, Regulation of chromatin by histone modifications. *Cell Research*, **21** (2011) 381–395. <https://doi.org/10.1038/cr.2011.22>.
132. K. K. Lee & J. L. Workman, Histone acetyltransferase complexes: one size doesn't fit all. *Nature Reviews Molecular Cell Biology*, **8** (2007). <https://doi.org/10.1038/nrm2145>.
133. V. G. Allfrey, R. Faulkner, & A. E. Mirsky, ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proceedings of the National Academy of Sciences*, **51** (1964). <https://doi.org/10.1073/pnas.51.5.786>.
134. M. Fournier & L. Tora, KAT2-mediated PLK4 acetylation contributes to genomic stability by preserving centrosome number. *Molecular and Cellular Oncology*, **4** (2017). <https://doi.org/10.1080/23723556.2016.1270391>.
135. K. Hyun, J. Jeon, K. Park, & J. Kim, Writing, erasing and reading histone lysine methylations. *Experimental & Molecular Medicine*, **49** (2017) e324–e324. <https://doi.org/10.1038/emm.2017.11>.
136. Y. Dou, T. A. Milne, A. J. Tackett, E. R. Smith, A. Fukuda, J. Wysocka, C. D. Allis, B. T. Chait, J. L. Hess, & R. G. Roeder, Physical Association and Coordinate Function of the H3 K4 Methyltransferase MLL1 and the H4 K16 Acetyltransferase MOF. *Cell*, **121** (2005). <https://doi.org/10.1016/j.cell.2005.04.031>.
137. E. C. Chittock, S. Latwiel, T. C. R. Miller, & C. W. Müller, Molecular architecture of polycomb repressive complexes. *Biochemical Society Transactions*, **45** (2017). <https://doi.org/10.1042/BST20160173>.
138. S. J. Geisler & R. Paro, Trithorax and Polycomb group-dependent regulation: a tale of opposing activities. *Development*, **142** (2015). <https://doi.org/10.1242/dev.120030>.
139. B. Jin, Y. Li, & K. D. Robertson, DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy? *Genes & Cancer*, **2** (2011). <https://doi.org/10.1177/1947601910393957>.
140. C. B. Schaefer, S. K. T. Ooi, T. H. Bestor, & D. Bourc'his, Epigenetic Decisions in Mammalian Germ Cells. *Science*, **316** (2007). <https://doi.org/10.1126/science.1137544>.
141. A. Avgustinova & S. A. Benitah, Epigenetic control of adult stem cell function. *Nature Reviews Molecular Cell Biology*, **17** (2016). <https://doi.org/10.1038/nrm.2016.76>.
142. I. L. de la Serna, Y. Ohkawa, & A. N. Imbalzano, Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. *Nature Reviews Genetics*, **7** (2006) 461–473. <https://doi.org/10.1038/nrg1882>.

143. E. Solary, O. A. Bernard, A. Tefferi, F. Fuks, & W. Vainchenker, The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia*, **28** (2014) 485–496. <https://doi.org/10.1038/leu.2013.337>.
144. K. L. Rice, I. Hormaeche, & J. D. Licht, Epigenetic regulation of normal and malignant hematopoiesis. *Oncogene*, **26** (2007) 6697–6714. <https://doi.org/10.1038/sj.onc.1210755>.
145. P. Jiang, H. Wang, J. Zheng, Y. Han, H. Huang, & P. Qian, Epigenetic regulation of hematopoietic stem cell homeostasis. *Blood Science*, **1** (2019). <https://doi.org/10.1097/BS9.0000000000000018>.
146. W.-I. Chan, R. L. Hannah, M. A. Dawson, C. Pridans, D. Foster, A. Joshi, B. Gottgens, J. M. van Deursen, & B. J. P. Huntly, The Transcriptional Coactivator Cbp Regulates Self-Renewal and Differentiation in Adult Hematopoietic Stem Cells. *Molecular and Cellular Biology*, **31** (2011). <https://doi.org/10.1128/MCB.05830-11>.
147. D. L. Letting, C. Rakowski, M. J. Weiss, & G. A. Blobel, Formation of a Tissue-Specific Histone Acetylation Pattern by the Hematopoietic Transcription Factor GATA-1. *Molecular and Cellular Biology*, **23** (2003). <https://doi.org/10.1128/MCB.23.4.1334-1340.2003>.
148. Y. Yamaguchi, M. Kurokawa, Y. Imai, K. Izutsu, T. Asai, M. Ichikawa, G. Yamamoto, E. Nitta, T. Yamagata, K. Sasaki, K. Mitani, S. Ogawa, S. Chiba, & H. Hirai, AML1 Is Functionally Regulated through p300-mediated Acetylation on Specific Lysine Residues. *Journal of Biological Chemistry*, **279** (2004). <https://doi.org/10.1074/jbc.M400355200>.
149. T. Katsumoto, MOZ is essential for maintenance of hematopoietic stem cells. *Genes & Development*, **20** (2006). <https://doi.org/10.1101/gad.1393106>.
150. H. Yin, J. Glass, & K. L. Blanchard, MOZ-TIF2 repression of nuclear receptor-mediated transcription requires multiple domains in MOZ and in the CID domain of TIF2. *Molecular Cancer*, **6** (2007). <https://doi.org/10.1186/1476-4598-6-51>.
151. Interaction of c-Myb with p300 is required for the induction of acute myeloid leukemia (AML) by human AML oncogenes. (n.d.).
152. P. Rimmelé, C. L. Bigarella, R. Liang, B. Izac, R. Dieguez-Gonzalez, G. Barbet, M. Donovan, C. Brugnara, J. M. Blander, D. A. Sinclair, & S. Ghaffari, Aging-like Phenotype and Defective Lineage Specification in SIRT1-Deleted Hematopoietic Stem and Progenitor Cells. *Stem Cell Reports*, **3** (2014). <https://doi.org/10.1016/j.stemcr.2014.04.015>.
153. C. Pessoa Rodrigues, J. S. Herman, B. Herquel, C. I. K. Valsecchi, T. Stehle, D. Grün, & A. Akhtar, Temporal expression of MOF acetyltransferase primes transcription factor networks for erythroid fate. *Science Advances*, **6** (2020) eaaz4815. <https://doi.org/10.1126/sciadv.aaz4815>.
154. P. Wang, Z. Wang, & J. Liu, Role of HDACs in normal and malignant hematopoiesis. *Molecular Cancer*, **19** (2020) 5. <https://doi.org/10.1186/s12943-019-1127-7>.
155. T. A. Milne, S. D. Briggs, H. W. Brock, M. E. Martin, D. Gibbs, C. D. Allis, & J. L. Hess, MLL Targets SET Domain Methyltransferase Activity to Hox Gene Promoters. *Molecular Cell*, **10** (2002). [https://doi.org/10.1016/S1097-2765\(02\)00741-4](https://doi.org/10.1016/S1097-2765(02)00741-4).
156. S. M. Cullen, A. Mayle, L. Rossi, & M. A. Goodell, Hematopoietic Stem Cell Development. (2014). <https://doi.org/10.1016/B978-0-12-416022-4.00002-0>.
157. C. M. McLean, I. D. Karemaker, & F. van Leeuwen, The emerging roles of DOT1L in leukemia and normal development. *Leukemia*, **28** (2014) 2131–2138. <https://doi.org/10.1038/leu.2014.169>.

158. Y. Feng, Y. Yang, M. M. Ortega, J. N. Copeland, M. Zhang, J. B. Jacob, T. A. Fields, J. L. Vivian, & P. E. Fields, Early mammalian erythropoiesis requires the Dot1L methyltransferase. *Blood*, **116** (2010). <https://doi.org/10.1182/blood-2010-03-276501>.
159. A. Daser & T. H. Rabbitts, The versatile mixed lineage leukaemia gene MLL and its many associations in leukaemogenesis. *Seminars in Cancer Biology*, **15** (2005). <https://doi.org/10.1016/j.semcancer.2005.01.007>.
160. Y. Okada, Q. Feng, Y. Lin, Q. Jiang, Y. Li, V. M. Coffield, L. Su, G. Xu, & Y. Zhang, hDOT1L Links Histone Methylation to Leukemogenesis. *Cell*, **121** (2005). <https://doi.org/10.1016/j.cell.2005.02.020>.
161. V. di Carlo, I. Mocavini, & L. di Croce, Polycomb complexes in normal and malignant hematopoiesis. *Journal of Cell Biology*, **218** (2019). <https://doi.org/10.1083/jcb.201808028>.
162. J. Jung, S. C. Buisman, E. Weersing, A. Dethmers-Ausema, E. Zwart, H. Schepers, M. R. Dekker, S. S. Lazare, F. Hammerl, Y. Skokova, S. M. Kooistra, K. Klauke, R. A. Poot, L. v. Bystrykh, & G. de Haan, CBX7 Induces Self-Renewal of Human Normal and Malignant Hematopoietic Stem and Progenitor Cells by Canonical and Non-canonical Interactions. *Cell Reports*, **26** (2019). <https://doi.org/10.1016/j.celrep.2019.01.050>.
163. A. C. Winters & K. M. Bernt, MLL-Rearranged Leukemias—An Update on Science and Clinical Approaches. *Frontiers in Pediatrics*, **5** (2017). <https://doi.org/10.3389/fped.2017.00004>.
164. F. Basheer, G. Giotopoulos, E. Meduri, H. Yun, M. Mazan, D. Sasca, P. Gallipoli, L. Marando, M. Gozdecka, R. Asby, O. Sheppard, M. Dudek, L. Bullinger, H. Döhner, R. Dillon, S. Freeman, O. Ottmann, A. Burnett, N. Russell, E. Papaemmanuil, R. Hills, P. Campbell, G. S. Vassiliou, & B. J. P. Huntly, Contrasting requirements during disease evolution identify EZH2 as a therapeutic target in AML. *Journal of Experimental Medicine*, **216** (2019). <https://doi.org/10.1084/jem.20181276>.
165. W. J. Harris, X. Huang, J. T. Lynch, G. J. Spencer, J. R. Hitchin, Y. Li, F. Ciceri, J. G. Blaser, B. F. Greystoke, A. M. Jordan, C. J. Miller, D. J. Ogilvie, & T. C. P. Somervaille, The Histone Demethylase KDM1A Sustains the Oncogenic Potential of MLL-AF9 Leukemia Stem Cells. *Cancer Cell*, **21** (2012). <https://doi.org/10.1016/j.ccr.2012.03.014>.
166. F. Izzo, S. C. Lee, A. Poran, R. Chaligne, F. Gaiti, B. Gross, R. R. Murali, S. D. Deochand, C. Ang, P. W. Jones, A. S. Nam, K.-T. Kim, S. Kothen-Hill, R. C. Schulman, M. Ki, P. Lhoumaud, J. A. Skok, A. D. Viny, R. L. Levine, E. Kenigsberg, O. Abdel-Wahab, & D. A. Landau, DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nature Genetics*, **52** (2020) 378–387. <https://doi.org/10.1038/s41588-020-0595-4>.
167. M. Farlik, F. Halbritter, F. Müller, F. A. Choudry, P. Ebert, J. Klughammer, S. Farrow, A. Santoro, V. Ciaurro, A. Mathur, R. Uppal, H. G. Stunnenberg, W. H. Ouwehand, E. Laurenti, T. Lengauer, M. Frontini, & C. Bock, DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell*, **19** (2016). <https://doi.org/10.1016/j.stem.2016.10.019>.
168. A.-M. Bröske, L. Vockentanz, S. Kharazi, M. R. Huska, E. Mancini, M. Scheller, C. Kuhl, A. Enns, M. Prinz, R. Jaenisch, C. Nerlov, A. Leutz, M. A. Andrade-Navarro, S. E. W. Jacobsen, & F. Rosenbauer, DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nature Genetics*, **41** (2009) 1207–1215. <https://doi.org/10.1038/ng.463>.
169. G. A. Challen, D. Sun, M. Jeong, M. Luo, J. Jelinek, J. S. Berg, C. Bock, A. Vasanthakumar, H. Gu, Y. Xi, S. Liang, Y. Lu, G. J. Darlington, A. Meissner, J.-P. J.

- Issa, L. A. Godley, W. Li, & M. A. Goodell, Dnmt3a is essential for hematopoietic stem cell differentiation. *Nature Genetics*, **44** (2012). <https://doi.org/10.1038/ng.1009>.
170. M. Jeong, H. J. Park, H. Celik, E. L. Ostrander, J. M. Reyes, A. Guzman, B. Rodriguez, Y. Lei, Y. Lee, L. Ding, O. A. Guryanova, W. Li, M. A. Goodell, & G. A. Challen, Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells In Vivo. *Cell Reports*, **23** (2018). <https://doi.org/10.1016/j.celrep.2018.03.025>.
171. M. Emperle, S. Adam, S. Kunert, M. Dukatz, A. Baude, C. Plass, P. Rathert, P. Bashtrykov, & A. Jeltsch, Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns. *Nucleic Acids Research*, **47** (2019). <https://doi.org/10.1093/nar/gkz911>.
172. L. Yang, R. Rau, & M. A. Goodell, DNMT3A in haematological malignancies. *Nature Reviews Cancer*, **15** (2015). <https://doi.org/10.1038/nrc3895>.
173. C. Gebhard, D. Glatz, L. Schwarzfischer, J. Wimmer, S. Stasik, M. Nuetzel, D. Heudobler, R. Andreesen, G. Ehninger, C. Thiede, & M. Rehli, Profiling of aberrant DNA methylation in acute myeloid leukemia reveals subclasses of CG-rich regions with epigenetic or genetic association. *Leukemia*, **33** (2019). <https://doi.org/10.1038/s41375-018-0165-2>.
174. K. Moran-Crusio, L. Reavie, A. Shih, O. Abdel-Wahab, D. Ndiaye-Lobry, C. Lobry, M. E. Figueroa, A. Vasanthakumar, J. Patel, X. Zhao, F. Perna, S. Pandey, J. Madzo, C. Song, Q. Dai, C. He, S. Ibrahim, M. Beran, J. Zavadil, S. D. Nimer, A. Melnick, L. A. Godley, I. Aifantis, & R. L. Levine, Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. *Cancer Cell*, **20** (2011). <https://doi.org/10.1016/j.ccr.2011.06.001>.
175. X. Zhang, J. Su, M. Jeong, M. Ko, Y. Huang, H. J. Park, A. Guzman, Y. Lei, Y.-H. Huang, A. Rao, W. Li, & M. A. Goodell, DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nature Genetics*, **48** (2016). <https://doi.org/10.1038/ng.3610>.
176. C. P. Rodrigues, M. Shvedunova, & A. Akhtar, Epigenetic Regulators as the Gatekeepers of Hematopoiesis. *Trends in Genetics*, **37** (2021). <https://doi.org/10.1016/j.tig.2020.09.015>.
177. J. E. Brownell, J. Zhou, T. Ranalli, R. Kobayashi, D. G. Edmondson, S. Y. Roth, & C. D. Allis, Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*, **84** (1996) 843–851. [https://doi.org/10.1016/S0092-8674\(00\)81063-6](https://doi.org/10.1016/S0092-8674(00)81063-6).
178. Z. Nagy & L. Tora, Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene*, **26** (2007). <https://doi.org/10.1038/sj.onc.1210604>.
179. Q. Jin, C. Wang, X. Kuang, X. Feng, V. Sartorelli, H. Ying, K. Ge, & S. Y. R. Dent, Gcn5 and PCAF Regulate PPAR and Prdm16 Expression To Facilitate Brown Adipogenesis. *Molecular and Cellular Biology*, **34** (2014). <https://doi.org/10.1128/MCB.00622-14>.
180. Y. Wang, Y. R. Guo, K. Liu, Z. Yin, R. Liu, Y. Xia, L. Tan, P. Yang, J. H. Lee, X. J. Li, D. Hawke, Y. Zheng, X. Qian, J. Lyu, J. He, D. Xing, Y. J. Tao, & Z. Lu, KAT2A coupled with the α -KGDH complex acts as a histone H3 succinyltransferase. *Nature*, **552** (2017) 273–277. <https://doi.org/10.1038/nature25003>.
181. T. Yamauchi, J. Yamauchi, T. Kuwata, T. Tamura, T. Yamashita, N. Bae, H. Westphal, K. Ozato, & Y. Nakatani, Distinct but overlapping roles of histone acetylase PCAF and of the closely related PCAF-B/GCN5 in mouse embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, **97** (2000) 11303–11306. <https://doi.org/10.1073/pnas.97.21.11303>.

182. W. Xu, D. G. Edmondson, Y. A. Evrard, M. Wakamiya, R. R. Behringer, & S. Y. Roth, Loss of Gcn512 leads to increased apoptosis and mesodermal defects during mouse development. *Nature Genetics*, **26** (2000) 229–232. <https://doi.org/10.1038/79973>.
183. C. Carré, D. Szymczak, J. Pidoux, & C. Antoniewski, The Histone H3 Acetylase dGcn5 Is a Key Player in Drosophila melanogaster Metamorphosis. *Molecular and Cellular Biology*, **25** (2005) 8228–8238. <https://doi.org/10.1128/mcb.25.18.8228-8238.2005>.
184. P. Bu, Y. A. Evrard, G. Lozano, & S. Y. R. Dent, Loss of Gcn5 Acetyltransferase Activity Leads to Neural Tube Closure Defects and Exencephaly in Mouse Embryos. *Molecular and Cellular Biology*, **27** (2007) 3405–3416. <https://doi.org/10.1128/mcb.00066-07>.
185. T. K. Ghosh, J. J. Aparicio-Sánchez, S. Buxton, A. Ketley, T. Mohamed, C. S. Rutland, S. Loughna, & J. D. Brook, Acetylation of TBX5 by KAT2B and KAT2A regulates heart and limb development. *Journal of Molecular and Cellular Cardiology*, **114** (2018) 185–198. <https://doi.org/10.1016/j.yjmcc.2017.11.013>.
186. W. Lin, G. Srajer, Y. A. Evrard, H. M. Phan, Y. Furuta, & S. Y. R. Dent, Developmental potential of Gcn5^{-/-} embryonic stem cells in vivo and in vitro. *Developmental Dynamics*, **236** (2007) 1547–1557. <https://doi.org/10.1002/dvdy.21160>.
187. N. Moris, S. Edri, D. Seyres, R. Kulkarni, A. F. Domingues, T. Balayo, M. Frontini, & C. Pina, Histone Acetyltransferase KAT2A Stabilizes Pluripotency with Control of Transcriptional Heterogeneity. *STEM CELLS*, **36** (2018) 1828–1838. <https://doi.org/10.1002/stem.2919>.
188. C. L. Hirsch, Z. C. Akdemir, L. Wang, G. Jayakumaran, D. Trcka, A. Weiss, J. J. Hernandez, Q. Pan, H. Han, X. Xu, Z. Xia, A. P. Salinger, M. Wilson, F. Vizeacoumar, A. Datti, W. Li, A. J. Cooney, M. C. Barton, B. J. Blencowe, J. L. Wrana, & S. Y. R. Dent, Myc and SAGA rewire an alternative splicing network during early somatic cell reprogramming. *Genes and Development*, **29** (2015) 803–816. <https://doi.org/10.1101/gad.255109.114>.
189. L. Wang, E. Koutelou, C. Hirsch, R. McCarthy, A. Schibler, K. Lin, Y. Lu, C. Jeter, J. Shen, M. C. Barton, & S. Y. R. Dent, GCN5 Regulates FGF Signaling and Activates Selective MYC Target Genes during Early Embryoid Body Differentiation. *Stem Cell Reports*, **10** (2018) 287–299. <https://doi.org/10.1016/j.stemcr.2017.11.009>.
190. B. R. Stanton, A. S. Perkins, L. Tessarollo, D. A. Sassoon, & L. F. Parada, Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop. *Genes and Development*, **6** (1992) 2235–2247. <https://doi.org/10.1101/gad.6.12a.2235>.
191. C. H. Lin, C. W. Lin, H. Tanaka, M. L. Fero, & R. N. Eisenman, Gene regulation and epigenetic remodeling in murine embryonic stem cells by c-Myc. *PLoS ONE*, **4** (2009). <https://doi.org/10.1371/journal.pone.0007839>.
192. J. H. Patel, Y. Du, P. G. Ard, C. Phillips, B. Carella, C.-J. Chen, C. Rakowski, C. Chatterjee, P. M. Lieberman, W. S. Lane, G. A. Blobel, & S. B. McMahon, The c-MYC Oncoprotein Is a Substrate of the Acetyltransferases hGCN5/PCAF and TIP60. *Molecular and Cellular Biology*, **24** (2004) 10826–10834. <https://doi.org/10.1128/mcb.24.24.10826-10834.2004>.
193. X. Liu, J. Tesfai, Y. A. Evrard, S. Y. R. Dent, & E. Martinez, c-Myc transformation domain recruits the human STAGA complex and requires TRRAP and GCN5 acetylase activity for transcription activation. *Journal of Biological Chemistry*, **278** (2003) 20405–20412. <https://doi.org/10.1074/jbc.M211795200>.

194. V. Martínez-Cerdeño, J. M. Lemen, V. Chan, A. Wey, W. Lin, S. R. Dent, & P. S. Knoepfler, N-Myc and GCN5 regulate significantly overlapping transcriptional programs in neural stem cells. *PLoS ONE*, **7** (2012) e39456. <https://doi.org/10.1371/journal.pone.0039456>.
195. B. Li, J. Sun, Z. Dong, P. Xue, X. He, L. Liao, L. Yuan, & Y. Jin, GCN5 modulates osteogenic differentiation of periodontal ligament stem cells through DKK1 acetylation in inflammatory microenvironment. *Scientific Reports*, **6** (2016). <https://doi.org/10.1038/srep26542>.
196. D. Bararia, H. S. Kwok, R. S. Welner, A. Numata, M. B. Sárosi, H. Yang, S. Wee, S. Tschuri, D. Ray, O. Weigert, E. Levantini, A. K. Ebralidze, J. Gunaratne, & D. G. Tenen, Acetylation of C/EBP α inhibits its granulopoietic function. *Nature Communications*, **7** (2016). <https://doi.org/10.1038/ncomms10968>.
197. A. F. Domingues, R. Kulkarni, G. Giotopoulos, S. Gupta, L. Vinnenberg, L. Arede, E. Foerner, M. Khalili, R. R. Adao, A. Johns, S. Tan, K. Zeka, B. J. Huntly, S. Prabakaran, & C. Pina, Loss of KAT2A enhances transcriptional noise and depletes acute myeloid leukemia stem-like cells. *eLife*, **9** (2020). <https://doi.org/10.7554/eLife.51754>.
198. H. Kikuchi, M. Nakayama, F. Kuribayashi, S. Imajoh-Ohmi, H. Nishitoh, Y. Takami, & T. Nakayama, GCN5 is essential for IRF-4 gene expression followed by transcriptional activation of Blimp-1 in immature B cells. *Journal of Leukocyte Biology*, **95** (2014) 399–404. <https://doi.org/10.1189/jlb.0413232>.
199. B. Gao, Q. Kong, Y. Zhang, C. Yun, S. Y. R. Dent, J. Song, D. D. Zhang, Y. Wang, X. Li, & D. Fang, The Histone Acetyltransferase Gcn5 Positively Regulates T Cell Activation. *The Journal of Immunology*, **198** (2017) 3927–3938. <https://doi.org/10.4049/jimmunol.1600312>.
200. Y. Wang, C. Yun, B. Gao, Y. Xu, Y. Zhang, Y. Wang, Q. Kong, F. Zhao, C. R. Wang, S. Y. R. Dent, J. Wang, X. Xu, H. bin Li, & D. Fang, The Lysine Acetyltransferase GCN5 Is Required for iNKT Cell Development through EGR2 Acetylation. *Cell Reports*, **20** (2017) 600–612. <https://doi.org/10.1016/j.celrep.2017.06.065>.
201. L. Li, B. Liu, X. Zhang, & L. Ye, The oncoprotein HBXIP promotes migration of breast cancer cells via GCN5-mediated microtubule acetylation. *Biochemical and Biophysical Research Communications*, **458** (2015) 720–725. <https://doi.org/10.1016/j.bbrc.2015.02.036>.
202. L. Zhao, A. Pang, & Y. Li, Function of GCN5 in the TGF β 1 induced epithelial to mesenchymal transition in breast cancer. *Oncology Letters*, **16** (2018) 3955–3963. <https://doi.org/10.3892/ol.2018.9134>.
203. L. Chen, T. Wei, X. Si, Q. Wang, Y. Li, Y. Leng, A. Deng, J. Chen, G. Wang, S. Zhu, & J. Kang, Lysine acetyltransferase GCN5 potentiates the growth of non-small cell lung cancer via promotion of E2F1, cyclin D1, and cyclin E1 expression. *Journal of Biological Chemistry*, **288** (2013) 14510–14521. <https://doi.org/10.1074/jbc.M113.458737>.
204. L. M. Mustachio, J. Roszik, A. T. Farria, K. Guerra, & S. Y. Dent, Repression of GCN5 expression or activity attenuates c-MYC expression in non-small cell lung cancer. *American journal of cancer research*, **9** (2019) 1830–1845.
205. Y. W. Yin, H. J. Jin, W. Zhao, B. Gao, J. Fang, J. Wei, D. D. Zhang, J. Zhang, & D. Fang, The histone acetyltransferase GCN5 expression is elevated and regulated by c-Myc and E2F1 transcription factors in human colon cancer. *Gene Expression*, **16** (2015) 187–196. <https://doi.org/10.3727/105221615X14399878166230>.

206. Y. Li, A. N. Jaramillo-Lambert, Y. Yang, R. Williams, N. H. Lee, & W. Zhu, And-1 is required for the stability of histone acetyltransferase Gcn5. *Oncogene*, **31** (2012) 643–652. <https://doi.org/10.1038/onc.2011.261>.
207. Y. Tong, D. Guo, D. Yan, C. Ma, F. Shao, Y. Wang, S. Luo, L. Lin, J. Tao, Y. Jiang, Z. Lu, & D. Xing, KAT2A succinyltransferase activity-mediated 14-3-3 ζ upregulation promotes β -catenin stabilization-dependent glycolysis and proliferation of pancreatic carcinoma cells. *Cancer Letters*, **469** (2020) 1–10. <https://doi.org/10.1016/j.canlet.2019.09.015>.
208. A. Riss, E. Scheer, M. Joint, S. Trowitzsch, I. Berger, & L. Tora, Subunits of ADA-two-A-containing (ATAC) or Spt-Ada-Gcn5-acetyltransferase (SAGA) coactivator complexes enhance the acetyltransferase activity of GCN5. *Journal of Biological Chemistry*, **290** (2015) 28997–29009. <https://doi.org/10.1074/jbc.M115.668533>.
209. D. Helmlinger & L. Tora, Sharing the SAGA. *Trends in Biochemical Sciences*, **42** (2017) 850–861. <https://doi.org/10.1016/j.tibs.2017.09.001>.
210. E. Koutelou, C. L. Hirsch, & S. Y. R. Dent, Multiple faces of the SAGA complex. *Current Opinion in Cell Biology*, **22** (2010) 374–382. <https://doi.org/10.1016/j.ceb.2010.03.005>.
211. T. Suganuma, J. L. Gutiérrez, B. Li, L. Florens, S. K. Swanson, M. P. Washburn, S. M. Abmayr, & J. L. Workman, ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. *Nature Structural and Molecular Biology*, **15** (2008) 364–372. <https://doi.org/10.1038/nsmb.1397>.
212. G. Spedale, H. T. M. Timmers, & W. W. M. P. Pijnappel, ATAC-king the complexity of SAGA during evolution. *Genes and Development*, **26** (2012) 527–541. <https://doi.org/10.1101/gad.184705.111>.
213. P. A. Grant, L. Duggan, J. Cote, S. M. Roberts, J. E. Brownell, R. Candau, R. Ohba, T. Owen-Hughes, C. D. Allis, F. Winston, S. L. Berger, & J. L. Workman, Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes & Development*, **11** (1997). <https://doi.org/10.1101/gad.11.13.1640>.
214. H. Wang, C. Dienemann, A. Stützer, H. Urlaub, A. C. M. Cheung, & P. Cramer, Structure of the transcription coactivator SAGA. *Nature*, **577** (2020) 717–720. <https://doi.org/10.1038/s41586-020-1933-5>.
215. G. Papai, A. Frechard, O. Kolesnikova, C. Crucifix, P. Schultz, & A. Ben-Shem, Structure of SAGA and mechanism of TBP deposition on gene promoters. *Nature*, **577** (2020) 711–716. <https://doi.org/10.1038/s41586-020-1944-2>.
216. C. Bian, C. Xu, J. Ruan, K. K. Lee, T. L. Burke, W. Tempel, D. Barsyte, J. Li, M. Wu, B. O. Zhou, B. E. Fleharty, A. Paulson, A. Allali-Hassani, J. Q. Zhou, G. Mer, P. A. Grant, J. L. Workman, J. Zang, & J. Min, Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO Journal*, **30** (2011) 2829–2842. <https://doi.org/10.1038/emboj.2011.193>.
217. Y. Zhang, W. Mi, Y. Xue, X. Shi, & T. G. Kutateladze, The ZZ domain as a new epigenetic reader and a degradation signal sensor. *Critical Reviews in Biochemistry and Molecular Biology*, **54** (2019) 1–10. <https://doi.org/10.1080/10409238.2018.1564730>.
218. Z. Nagy, A. Riss, C. Romier, X. le Guezennec, A. R. Dongre, M. Orpinell, J. Han, H. Stunnenberg, & L. Tora, The Human SPT20-Containing SAGA Complex Plays a Direct Role in the Regulation of Endoplasmic Reticulum Stress-Induced Genes. *Molecular and Cellular Biology*, **29** (2009) 1649–1660. <https://doi.org/10.1128/mcb.01076-08>.

219. S. B. McMahon, H. A. van Buskirk, K. A. Dugan, T. D. Copeland, & M. D. Cole, The novel ATM-related protein TRRAP is an essential cofactor for the c- Myc and E2F oncoproteins. *Cell*, **94** (1998) 363–374. [https://doi.org/10.1016/S0092-8674\(00\)81479-8](https://doi.org/10.1016/S0092-8674(00)81479-8).
220. S. Allard, NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. *The EMBO Journal*, **18** (1999). <https://doi.org/10.1093/emboj/18.18.5108>.
221. M. Fuchs, J. Gerber, R. Drapkin, S. Sif, T. Ikura, V. Ogryzko, W. S. Lane, Y. Nakatani, & D. M. Livingston, The p400 complex is an essential E1A transformation target. *Cell*, **106** (2001) 297–307. [https://doi.org/10.1016/S0092-8674\(01\)00450-0](https://doi.org/10.1016/S0092-8674(01)00450-0).
222. E. Martinez, V. B. Palhan, A. Tjernberg, E. S. Lyman, A. M. Gamper, T. K. Kundu, B. T. Chait, & R. G. Roeder, Human STAGA Complex Is a Chromatin-Acetylating Transcription Coactivator That Interacts with Pre-mRNA Splicing and DNA Damage-Binding Factors In Vivo. *Molecular and Cellular Biology*, **21** (2001) 6782–6795. <https://doi.org/10.1128/mcb.21.20.6782-6795.2001>.
223. R. Stegeman, P. J. Spreacker, S. K. Swanson, R. Stephenson, L. Florens, M. P. Washburn, & V. M. Weake, The Spliceosomal Protein SF3B5 is a Novel Component of Drosophila SAGA that Functions in Gene Expression Independent of Splicing. *Journal of Molecular Biology*, **428** (2016) 3632–3649. <https://doi.org/10.1016/j.jmb.2016.05.009>.
224. B. K. Das, L. Xia, L. Palandjian, O. Gozani, Y. Chyung, & R. Reed, Characterization of a Protein Complex Containing Spliceosomal Proteins SAPs 49, 130, 145, and 155. *Molecular and Cellular Biology*, **19** (1999) 6796–6802. <https://doi.org/10.1128/mcb.19.10.6796>.
225. S. M. Armour, E. J. Bennett, C. R. Braun, X.-Y. Zhang, S. B. McMahon, S. P. Gygi, J. W. Harper, & D. A. Sinclair, A High-Confidence Interaction Map Identifies SIRT1 as a Mediator of Acetylation of USP22 and the SAGA Coactivator Complex. *Molecular and Cellular Biology*, **33** (2013) 1487–1502. <https://doi.org/10.1128/mcb.00971-12>.
226. R. T. Sussman, T. J. Stanek, P. Estes, J. D. Gearhart, K. E. Knudsen, & S. B. McMahon, The epigenetic modifier ubiquitin-specific protease 22 (USP22) regulates embryonic stem cell differentiation via transcriptional repression of sex-determining region Y-box 2 (SOX2). *Journal of Biological Chemistry*, **288** (2013) 24234–24246. <https://doi.org/10.1074/jbc.M113.469783>.
227. A. Köhler, E. Zimmerman, M. Schneider, E. Hurt, & N. Zheng, Structural Basis for Assembly and Activation of the Heterotetrameric SAGA Histone H2B Deubiquitinase Module. *Cell*, **141** (2010) 606–617. <https://doi.org/10.1016/j.cell.2010.04.026>.
228. R. D. Mohan, G. Dialynas, V. M. Weake, J. Liu, S. Martin-Brown, L. Florens, M. P. Washburn, J. L. Workman, & S. M. Abmayr, Loss of Drosophila Ataxin-7, a SAGA subunit, reduces H2B ubiquitination and leads to neural and retinal degeneration. *Genes and Development*, **28** (2014) 259–272. <https://doi.org/10.1101/gad.225151.113>.
229. D. Umlauf, J. Bonnet, F. Waharte, M. Fournier, M. Stierle, B. Fischer, L. Brino, D. Devys, & L. Tora, The human TREX-2 complex is stably associated with the nuclear pore basket. *Journal of Cell Science*, **126** (2013) 2656–2667. <https://doi.org/10.1242/jcs.118000>.
230. V. M. Weake, K. K. Lee, S. Guelman, C. H. Lin, C. Seidel, S. M. Abmayr, & J. L. Workman, SAGA-mediated H2B deubiquitination controls the development of neuronal connectivity in the Drosophila visual system. *EMBO Journal*, **27** (2008) 394–405. <https://doi.org/10.1038/sj.emboj.7601966>.
231. K. K. Lee, M. E. Sardi, S. K. Swanson, J. M. Gilmore, M. Torok, P. A. Grant, L. Florens, J. L. Workman, & M. P. Washburn, Combinatorial depletion analysis to

- assemble the network architecture of the SAGA and ADA chromatin remodeling complexes. *Molecular Systems Biology*, **7** (2011). <https://doi.org/10.1038/msb.2011.40>.
232. J. H. M. Soffers, X. Li, A. Saraf, C. W. Seidel, L. Florens, M. P. Washburn, S. M. Abmayr, & J. L. Workman, Characterization of a metazoan ADA acetyltransferase complex. *Nucleic Acids Research*, **47** (2019). <https://doi.org/10.1093/nar/gkz042>.
 233. A. Shukla, P. Bajwa, & S. R. Bhaumik, SAGA-associated Sgf73p facilitates formation of the preinitiation complex assembly at the promoters either in a HAT-dependent or independent manner in vivo. *Nucleic Acids Research*, **34** (2006). <https://doi.org/10.1093/nar/gkl844>.
 234. Y. L. Wang, F. Faiola, M. Xu, S. Pan, & E. Martinez, Human ATAC is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein. *Journal of Biological Chemistry*, **283** (2008) 33808–33815. <https://doi.org/10.1074/jbc.M806936200>.
 235. W. Mi, H. Guan, J. Lyu, D. Zhao, Y. Xi, S. Jiang, F. H. Andrews, X. Wang, M. Gagea, H. Wen, L. Tora, S. Y. R. Dent, T. G. Kutateladze, W. Li, H. Li, & X. Shi, YEATS2 links histone acetylation to tumorigenesis of non-small cell lung cancer. *Nature Communications*, **8** (2017). <https://doi.org/10.1038/s41467-017-01173-4>.
 236. W. Mi, Y. Zhang, J. Lyu, X. Wang, Q. Tong, D. Peng, Y. Xue, A. H. Tencer, H. Wen, W. Li, T. G. Kutateladze, & X. Shi, The ZZ-type zinc finger of ZZZ3 modulates the ATAC complex-mediated histone acetylation and gene activation. *Nature Communications*, **9** (2018). <https://doi.org/10.1038/s41467-018-06247-5>.
 237. F. Gori, P. Divieti, & M. B. Demay, Cloning and Characterization of a Novel WD-40 Repeat Protein That Dramatically Accelerates Osteoblastic Differentiation. *Journal of Biological Chemistry*, **276** (2001) 46515–46522. <https://doi.org/10.1074/jbc.M105757200>.
 238. J. Chen, Q. Luo, Y. Yuan, X. Huang, W. Cai, C. Li, T. Wei, L. Zhang, M. Yang, Q. Liu, G. Ye, X. Dai, & B. Li, Pygo2 Associates with MLL2 Histone Methyltransferase and GCN5 Histone Acetyltransferase Complexes To Augment Wnt Target Gene Expression and Breast Cancer Stem-Like Cell Expansion. *Molecular and Cellular Biology*, **30** (2010) 5621–5635. <https://doi.org/10.1128/mcb.00465-10>.
 239. F. Grebien, M. Vedadi, M. Getlik, R. Giambruno, A. Grover, R. Avellino, A. Skucha, S. Vittori, E. Kuznetsova, D. Smil, D. Barsyte-Lovejoy, F. Li, G. Poda, M. Schapira, H. Wu, A. Dong, G. Senisterra, A. Stukalov, K. V. M. Huber, A. Schönegger, R. Marcellus, M. Bilban, C. Bock, P. J. Brown, J. Zuber, K. L. Bennett, R. Al-awar, R. Delwel, C. Nerlov, C. H. Arrowsmith, & G. Superti-Furga, Pharmacological targeting of the Wdr5-MLL interaction in C/EBP α N-terminal leukemia. *Nature Chemical Biology*, **11** (2015) 571–578. <https://doi.org/10.1038/nchembio.1859>.
 240. M. Ma, Y. Zhang, M. Weng, Y. Hu, Y. Xuan, Y. Hu, & K. Lv, lncRNA GCAWKR Promotes Gastric Cancer Development by Scaffolding the Chromatin Modification Factors WDR5 and KAT2A. *Molecular Therapy*, **26** (2018) 2658–2668. <https://doi.org/10.1016/j.ymthe.2018.09.002>.
 241. T.-T. Sun, J. He, Q. Liang, L.-L. Ren, T.-T. Yan, T.-C. Yu, J.-Y. Tang, Y.-J. Bao, Y. Hu, Y. Lin, D. Sun, Y.-X. Chen, J. Hong, H. Chen, W. Zou, & J.-Y. Fang, lncRNA GCIncl Promotes Gastric Carcinogenesis and May Act as a Modular Scaffold of WDR5 and KAT2A Complexes to Specify the Histone Modification Pattern. *Cancer Discovery*, **6** (2016) 784–801. <https://doi.org/10.1158/2159-8290.CD-15-0921>.
 242. S. Guelman, K. Kozuka, Y. Mao, V. Pham, M. J. Solloway, J. Wang, J. Wu, J. R. Lill, & J. Zha, The Double-Histone-Acetyltransferase Complex ATAC Is Essential for

- Mammalian Development. *Molecular and Cellular Biology*, **29** (2009) 1176–1188. <https://doi.org/10.1128/mcb.01599-08>.
243. C. Carré, A. Ciurciu, O. Komonyi, C. Jacquier, D. Fagegaltier, J. Pidoux, H. Tricoire, L. Tora, I. M. Boros, & C. Antoniewski, The Drosophila NURF remodelling and the ATAC histone acetylase complexes functionally interact and are required for global chromosome organization. *EMBO Reports*, **9** (2008) 187–192. <https://doi.org/10.1038/sj.embor.7401141>.
 244. J. L. Vogel & T. M. Kristie, The dynamics of HCF-1 modulation of herpes simplex virus chromatin during initiation of infection. *Viruses*, **5** (2013) 1272–1291. <https://doi.org/10.3390/v5051272>.
 245. A. R. Krebs, K. Karmodiya, M. Lindahl-Allen, K. Struhl, & L. Tora, SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. *Molecular Cell*, **44** (2011) 410–423. <https://doi.org/10.1016/j.molcel.2011.08.037>.
 246. S. C. Dolfi, L. L.-Y. Chan, J. Qiu, P. M. Tedeschi, J. R. Bertino, K. M. Hirshfield, Z. N. Oltvai, & A. Vazquez, The metabolic demands of cancer cells are coupled to their size and protein synthesis rates. *Cancer & Metabolism*, **1** (2013) 20. <https://doi.org/10.1186/2049-3002-1-20>.
 247. Y. Martineau, D. Müller, & S. Pyronnet, Targeting protein synthesis in cancer cells. *Oncoscience*, **1** (2014) 484–485. <https://doi.org/10.18632/oncoscience.63>.
 248. L. M. Mustachio, J. Roszik, A. Farria, & S. Y. R. Dent, Targeting the SAGA and ATAC Transcriptional Coactivator Complexes in MYC-Driven Cancers. *Cancer research*, **80** (2020) 1905–1911. <https://doi.org/10.1158/0008-5472.CAN-19-3652>.
 249. L. Arede & C. Pina, Buffering noise: KAT2A modular contributions to stabilization of transcription and cell identity in cancer and development. *Experimental Hematology*, **93** (2021). <https://doi.org/10.1016/j.exphem.2020.10.003>.
 250. M. Gotoh, H. Ichikawa, E. Arai, S. Chiku, H. Sakamoto, H. Fujimoto, M. Hiramoto, T. Nammo, K. Yasuda, T. Yoshida, & Y. Kanai, Comprehensive exploration of novel chimeric transcripts in clear cell renal cell carcinomas using whole transcriptome analysis. *Genes, Chromosomes and Cancer*, **53** (2014) 1018–1032. <https://doi.org/10.1002/gcc.22211>.
 251. R. L. Milne, J. Lorenzo-Bermejo, B. Burwinkel, N. Malats, J. I. Arias, M. Pilar Zamora, J. Benítez, M. K. Humphreys, M. García-Closas, S. J. Chanock, J. Lissowska, M. E. Sherman, A. Mannermaa, V. Kataja, V. M. Kosma, H. Nevanlinna, T. Heikkinen, K. Aittomäki, C. Blomqvist, H. Anton-Culver, A. Ziogas, P. Devilee, C. J. van Asperen, R. A. E. M. Tollenaar, C. Seynaeve, P. Hall, K. Czene, J. Liu, A. K. Irwanto, D. Kang, K. Y. Yoo, D. Y. Noh, F. J. Couch, J. E. Olson, X. Wang, Z. Fredericksen, B. G. Nordestgaard, S. E. Bojesen, H. Flyger, S. Margolin, A. Lindblom, P. A. Fasching, R. Schulz-Wendtland, A. B. Ekici, M. W. Beckmann, S. Wang-Gohrke, C. Y. Shen, J. C. Yu, H. M. Hsu, P. E. Wu, G. G. Giles, G. Severi, L. Baglietto, D. R. English, A. Cox, I. Brock, G. Elliott, M. W. R. Reed, J. Beesley, X. Chen, O. Fletcher, L. Gibson, I. dos S. Silva, J. Peto, B. Frank, J. Heil, A. Meindl, J. Chang-Claude, R. Hein, A. Vrieling, D. Flesch-Janys, M. C. Southey, L. Smith, C. Apicella, J. L. Hopper, A. M. Dunning, K. A. Pooley, P. D. P. Pharoah, U. Hamann, B. Pesch, Y. D. Ko, D. F. Easton, & G. Chenevix-Trench, 7q21-rs6964587 and breast cancer risk: An extended case-control study by the Breast Cancer Association Consortium. *Journal of Medical Genetics*, **48** (2011) 698–702. <https://doi.org/10.1136/jmedgenet-2011-100303>.
 252. C. Han, L. Yu, X. Liu, T. Yu, W. Qin, X. Liao, Z. Liu, S. Lu, Z. Chen, H. Su, G. Zhu, X. Qin, Y. Gui, J. Li, K. Xiao, X. Chen, X. Ye, M. Peng, J. Dong, & T. Peng, ATXN7

- Gene Variants and Expression Predict Post-Operative Clinical Outcomes in Hepatitis B Virus-Related Hepatocellular Carcinoma. *Cellular Physiology and Biochemistry*, **39** (2016). <https://doi.org/10.1159/000452511>.
253. S. Majaz, Z. Tong, K. Peng, W. Wang, W. Ren, M. Li, K. Liu, P. Mo, W. Li, & C. Yu, Histone acetyl transferase GCN5 promotes human hepatocellular carcinoma progression by enhancing AIB1 expression. *Cell and Bioscience*, **6** (2016). <https://doi.org/10.1186/s13578-016-0114-6>.
 254. R. S. Schrecengost, J. L. Dean, J. F. Goodwin, M. J. Schiewer, M. W. Urban, T. J. Stanek, R. T. Sussman, J. L. Hicks, R. C. Birbe, R. A. Draganova-Tacheva, T. Visakorpi, A. M. DeMarzo, S. B. McMahon, & K. E. Knudsen, USP22 regulates oncogenic signaling pathways to drive lethal cancer progression. *Cancer Research*, **74** (2014) 272–286. <https://doi.org/10.1158/0008-5472.CAN-13-1954>.
 255. J. J. McCann, I. A. Vasilevskaya, N. P. Neupane, A. A. Shafi, C. McNair, E. Dylgjeri, A. C. Mandigo, M. J. Schiewer, R. S. Schrecengost, P. Gallagher, T. J. Stanek, S. B. McMahon, L. D. Berman-Booty, W. F. Ostrander, & K. E. Knudsen, USP22 functions as an oncogenic driver in prostate cancer by regulating cell proliferation and DNA repair. *Cancer Research*, **80** (2020) 430–443. <https://doi.org/10.1158/0008-5472.CAN-19-1033>.
 256. K. Zhang, L. Yang, J. Wang, T. Sun, Y. Guo, R. Nelson, T. R. Tong, R. Pangen, R. Salgia, & D. J. Raz, Ubiquitin-specific protease 22 is critical to in vivo angiogenesis, growth and metastasis of non-small cell lung cancer. *Cell Communication and Signaling*, **17** (2019) 167. <https://doi.org/10.1186/s12964-019-0480-x>.
 257. C. C. Lim, J. C. Xu, T. Y. Chen, J. X. Xu, W. F. Chen, J. W. Hu, Q. L. Li, & Y. Q. Zhang, Ubiquitin-specific peptide 22 acts as an oncogene in gastric cancer in a son of sevenless 1-dependent manner. *Cancer Cell International*, **20** (2020). <https://doi.org/10.1186/s12935-020-1137-y>.
 258. R. L. Kosinsky, M. Helms, M. Zerche, L. Wohn, A. Dyas, E. Prokakis, Z. B. Kazerouni, U. Bedi, F. Wegwitz, & S. A. Johnsen, USP22-dependent HSP90AB1 expression promotes resistance to HSP90 inhibition in mammary and colorectal cancer. *Cell Death and Disease*, **10** (2019) 1–11. <https://doi.org/10.1038/s41419-019-2141-9>.
 259. J. Melo-Cardenas, Y. Xu, J. Wei, C. Tan, S. Kong, B. Gao, E. Montauti, G. Kirsammer, J. D. Licht, J. Yu, P. Ji, J. D. Crispino, & D. Fang, USP22 deficiency leads to myeloid leukemia upon oncogenic Kras activation through a PU.1-dependent mechanism. *Blood*, **132** (2018) 423–434. <https://doi.org/10.1182/blood-2017-10-811760>.
 260. K. Tzelepis, H. Koike-Yusa, E. de Braekeleer, Y. Li, E. Metzakopian, O. M. Dovey, A. Mupo, V. Grinkevich, M. Li, M. Mazan, M. Gozdecka, S. Ohnishi, J. Cooper, M. Patel, T. McKerrell, B. Chen, A. F. Domingues, P. Gallipoli, S. Teichmann, H. Ponstingl, U. McDermott, J. Saez-Rodriguez, B. J. P. Huntly, F. Iorio, C. Pina, G. S. Vassiliou, & K. Yusa, A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Reports*, **17** (2016) 1193–1205. <https://doi.org/10.1016/j.celrep.2016.09.079>.
 261. Z. Lin, H. Yang, Q. Kong, J. Li, S. M. Lee, B. Gao, H. Dong, J. Wei, J. Song, D. D. Zhang, & D. Fang, USP22 Antagonizes p53 Transcriptional Activation by Deubiquitinating Sirt1 to Suppress Cell Apoptosis and Is Required for Mouse Embryonic Development. *Molecular Cell*, **46** (2012) 484–494. <https://doi.org/10.1016/j.molcel.2012.03.024>.
 262. W. Lin, Z. Zhang, G. Srajer, C. C. Yi, M. Huang, H. M. Phan, & S. Y. R. Dent, Proper expression of the Gcn5 histone acetyltransferase is required for neural tube closure in

- mouse embryos. *Developmental Dynamics*, **237** (2008) 928–940. <https://doi.org/10.1002/dvdy.21479>.
263. S. Warriar, S. Nuwayhid, J. A. Sabatino, K. F. Sugrue, & I. E. Zohn, Supt20 is required for development of the axial skeleton. *Developmental Biology*, **421** (2017) 245–257. <https://doi.org/10.1016/j.ydbio.2016.11.009>.
 264. D. Seruggia, M. Oti, P. Tripathi, M. C. Canver, L. LeBlanc, D. C. di Giammartino, M. J. Bullen, C. M. Nefzger, Y. B. Y. Sun, R. Farouni, J. M. Polo, L. Pinello, E. Apostolou, J. Kim, S. H. Orkin, & P. P. Das, TAF5L and TAF6L Maintain Self-Renewal of Embryonic Stem Cells via the MYC Regulatory Network. *Molecular Cell*, **74** (2019) 1148–1163.e7. <https://doi.org/10.1016/j.molcel.2019.03.025>.
 265. T. Kusch, S. Guelman, S. M. Abmayr, & J. L. Workman, Two Drosophila Ada2 Homologues Function in Different Multiprotein Complexes. *Molecular and Cellular Biology*, **23** (2003) 3305–3319. <https://doi.org/10.1128/mcb.23.9.3305-3319.2003>.
 266. E. Griessinger, F. Anjos-Afonso, I. Pizzitola, K. Rouault-Pierre, J. Vargaftig, D. Taussig, J. Gribben, F. Lassailly, & D. Bonnet, A Niche-Like Culture System Allowing the Maintenance of Primary Human Acute Myeloid Leukemia-Initiating Cells: A New Tool to Decipher Their Chemoresistance and Self-Renewal Mechanisms. *STEM CELLS Translational Medicine*, **3** (2014). <https://doi.org/10.5966/sctm.2013-0166>.
 267. M. Biel, A. Kretsovali, E. Karatzali, J. Papamatheakis, & A. Giannis, Design, Synthesis, and Biological Evaluation of a Small-Molecule Inhibitor of the Histone Acetyltransferase Gcn5. *Angewandte Chemie International Edition*, **43** (2004). <https://doi.org/10.1002/anie.200453879>.
 268. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, & J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, **102** (2005). <https://doi.org/10.1073/pnas.0506580102>.
 269. J. W. Lau, E. Lehnert, A. Sethi, R. Malhotra, G. Kaushik, Z. Onder, N. Groves-Kirkby, A. Mihajlovic, J. DiGiovanna, M. Srdic, D. Bajcic, J. Radenkovic, V. Mladenovic, D. Krstanovic, V. Arsenijevic, D. Klisic, M. Mitrovic, I. Bogicevic, D. Kural, & B. Davis-Dusenbery, The Cancer Genomics Cloud: Collaborative, Reproducible, and Democratized—A New Paradigm in Large-Scale Computational Research. *Cancer Research*, **77** (2017). <https://doi.org/10.1158/0008-5472.CAN-17-0387>.
 270. C. Lozzio & B. Lozzio, Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood*, **45** (1975). <https://doi.org/10.1182/blood.V45.3.321.321>.
 271. J. Sutherland, A. Turner, P. Mannoni, L. McGann, & J. Turc, Differentiation of K562 leukemia cells along erythroid, macrophage, and megakaryocyte lineages. *Journal of biological response modifiers*, **5** **3** (1986) 250–262.
 272. J. F. Leary, B. M. Ohlsson-Wilhelm, R. Giuliano, S. Labella, B. Farley, & P. T. Rowley, Multipotent human hematopoietic cell line K562: Lineage-specific constitutive and inducible antigens. *Leukemia Research*, **11** (1987). [https://doi.org/10.1016/0145-2126\(87\)90065-8](https://doi.org/10.1016/0145-2126(87)90065-8).
 273. S. T. Chou, E. Khandros, L. C. Bailey, K. E. Nichols, C. R. Vakoc, Y. Yao, Z. Huang, J. D. Crispino, R. C. Hardison, G. A. Blobel, & M. J. Weiss, Graded repression of PU.1/Sfp1 gene transcription by GATA factors regulates hematopoietic cell fate. *Blood*, **114** (2009). <https://doi.org/10.1182/blood-2009-03-207944>.
 274. A. T. Merryweather-Clarke, A. Atzberger, S. Soneji, N. Gray, K. Clark, C. Waugh, S. J. McGowan, S. Taylor, A. K. Nandi, W. G. Wood, D. J. Roberts, D. R. Higgs, V. J.

- Buckle, & K. J. H. Robson, Global gene expression analysis of human erythroid progenitors. *Blood*, **117** (2011). <https://doi.org/10.1182/blood-2010-07-290825>.
275. M. v. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, M. G. McDermott, C. D. Monteiro, G. W. Gundersen, & A. Ma'ayan, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, **44** (2016). <https://doi.org/10.1093/nar/gkw377>.
276. B. K. Tusi, S. L. Wolock, C. Weinreb, Y. Hwang, D. Hidalgo, R. Zilionis, A. Waisman, J. R. Huh, A. M. Klein, & M. Socolovsky, Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature*, **555** (2018). <https://doi.org/10.1038/nature25741>.
277. X.-Y. Zhang, M. Varthi, S. M. Sykes, C. Phillips, C. Warzecha, W. Zhu, A. Wyce, A. W. Thorne, S. L. Berger, & S. B. McMahon, The Putative Cancer Stem Cell Marker USP22 Is a Subunit of the Human SAGA Complex Required for Activated Transcription and Cell-Cycle Progression. *Molecular Cell*, **29** (2008). <https://doi.org/10.1016/j.molcel.2007.12.015>.
278. Y. Zhao, G. Lang, S. Ito, J. Bonnet, E. Metzger, S. Sawatsubashi, E. Suzuki, X. le Guezennec, H. G. Stunnenberg, A. Krasnov, S. G. Georgieva, R. Schüle, K.-I. Takeyama, S. Kato, L. Tora, & D. Devys, A TFTC/STAGA Module Mediates Histone H2A and H2B Deubiquitination, Coactivates Nuclear Receptors, and Counteracts Heterochromatin Silencing. *Molecular Cell*, **29** (2008). <https://doi.org/10.1016/j.molcel.2007.12.011>.
279. S. Ghosh & B. F. Pugh, Sequential Recruitment of SAGA and TFIID in a Genomic Response to DNA Damage in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, **31** (2011). <https://doi.org/10.1128/MCB.00317-10>.
280. S. J. Zanton, Full and partial genome-wide assembly and disassembly of the yeast transcription machinery in response to heat shock. *Genes & Development*, **20** (2006). <https://doi.org/10.1101/gad.1437506>.
281. T. Baptista, S. Grünberg, N. Minoungou, M. J. E. Koster, H. T. M. Timmers, S. Hahn, D. Devys, & L. Tora, SAGA Is a General Cofactor for RNA Polymerase II Transcription. *Molecular Cell*, **68** (2017). <https://doi.org/10.1016/j.molcel.2017.08.016>.
282. L. T. Lam, C. Ronchini, J. Norton, A. J. Capobianco, & E. H. Bresnick, Suppression of Erythroid but Not Megakaryocytic Differentiation of Human K562 Erythroleukemic Cells by Notch-1. *Journal of Biological Chemistry*, **275** (2000). <https://doi.org/10.1074/jbc.M002866200>.
283. A. de Thonel, J. Vandekerckhove, D. Lanneau, S. Selvakumar, G. Courtois, A. Hazoume, M. Brunet, S. Maurel, A. Hammann, J. A. Ribeil, Y. Zermati, A. S. Gabet, J. Boyes, E. Solary, O. Hermine, & C. Garrido, HSP27 controls GATA-1 protein level during erythroid cell differentiation. *Blood*, **116** (2010). <https://doi.org/10.1182/blood-2009-09-241778>.
284. O. N. Kuvardina, J. Herglotz, S. Kolodziej, N. Kohrs, S. Herkt, B. Wojcik, T. Oellerich, J. Corso, K. Behrens, A. Kumar, H. Hussong, H. Urlaub, J. Koch, H. Serve, H. Bonig, C. Stocking, M. A. Rieger, & J. Lausen, RUNX1 represses the erythroid gene expression program during megakaryocytic differentiation. *Blood*, **125** (2015). <https://doi.org/10.1182/blood-2014-11-610519>.
285. Y. Xie, L. Gao, C. Xu, L. Chu, L. Gao, R. Wu, Y. Liu, T. Liu, X. Sun, R. Ren, J. Tang, Y. Zheng, Y. Zhou, & S. Shen, ARHGEF12 regulates erythropoiesis and is involved in erythroid regeneration after chemotherapy in acute lymphoblastic leukemia patients. *Haematologica*, **105** (2020). <https://doi.org/10.3324/haematol.2018.210286>.

286. K. Anamika, A. R. Krebs, J. Thompson, O. Poch, D. Devys, & L. Tora, Lessons from genome-wide studies: an integrated definition of the coactivator function of histone acetyl transferases. *Epigenetics & Chromatin*, **3** (2010). <https://doi.org/10.1186/1756-8935-3-18>.
287. S. Huang, P/CAF-mediated acetylation regulates the function of the basic helix-loop-helix transcription factor TAL1/SCL. *The EMBO Journal*, **19** (2000). <https://doi.org/10.1093/emboj/19.24.6792>.
288. Y. Matsuo, R. MacLeod, C. Uphoff, H. Drexler, C. Nishizaki, Y. Katayama, G. Kimura, N. Fujii, E. Omoto, M. Harada, & K. Orita, Two acute monocytic leukemia (AML-M5a) cell lines (MOLM-13 and MOLM-14) with interclonal phenotypic heterogeneity showing MLL-AF9 fusion resulting from an occult chromosome insertion, ins(11;9)(q23;p22p23). *Leukemia*, **11** (1997). <https://doi.org/10.1038/sj.leu.2400768>.
289. S. Tsuchiya, M. Yamabe, Y. Yamaguchi, Y. Kobayashi, T. Konno, & K. Tada, Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *International Journal of Cancer*, **26** (1980). <https://doi.org/10.1002/ijc.2910260208>.
290. H. Quentmeier, J. Reinhardt, M. Zaborski, & H. G. Drexler, FLT3 mutations in acute myeloid leukemia cell lines. *Leukemia*, **17** (2003). <https://doi.org/10.1038/sj.leu.2402740>.
291. L. Larizza, I. Magnani, & A. Beghini, The Kasumi-1 cell line: a t(8;21)-kit mutant model for acute myeloid leukemia. *Leukemia & Lymphoma*, **46** (2005). <https://doi.org/10.1080/10428190400007565>.
292. M. She, X. Niu, X. Chen, J. Li, M. Zhou, Y. He, Y. Le, & K. Guo, Resistance of leukemic stem-like cells in AML cell line KG1a to natural killer cell-mediated cytotoxicity. *Cancer Letters*, **318** (2012). <https://doi.org/10.1016/j.canlet.2011.12.017>.
293. A. Seifert, Role of *Hox* genes in stem cell differentiation. *World Journal of Stem Cells*, **7** (2015). <https://doi.org/10.4252/wjsc.v7.i3.583>.
294. M. Orpinell, M. Fournier, A. Riss, Z. Nagy, A. R. Krebs, M. Frontini, & L. Tora, The ATAC acetyl transferase complex controls mitotic progression by targeting non-histone substrates. *EMBO Journal*, **29** (2010) 2381–2394. <https://doi.org/10.1038/emboj.2010.125>.
295. D. L. Myster & R. J. Duronio, Cell cycle: To differentiate or not to differentiate? *Current Biology*, **10** (2000). [https://doi.org/10.1016/S0960-9822\(00\)00435-8](https://doi.org/10.1016/S0960-9822(00)00435-8).
296. L. Larizza, I. Magnani, & A. Beghini, The Kasumi-1 cell line: a t(8;21)-kit mutant model for acute myeloid leukemia. *Leukemia & Lymphoma*, **46** (2005). <https://doi.org/10.1080/10428190400007565>.
297. M. Fournier, M. Orpinell, C. Grauffel, E. Scheer, J. M. Garnier, T. Ye, V. Chavant, M. Joint, F. Esashi, A. Dejaegere, P. Gönczy, & L. Tora, KAT2A/KAT2B-targeted acetylome reveals a role for PLK4 acetylation in preventing centrosome amplification. *Nature Communications*, **7** (2016). <https://doi.org/10.1038/ncomms13227>.
298. R. A. J. Signer, J. A. Magee, A. Salic, & S. J. Morrison, Haematopoietic stem cells require a highly regulated protein synthesis rate. *Nature*, **509** (2014). <https://doi.org/10.1038/nature13035>.
299. Z.-F. Yang, H. Zhang, L. Ma, C. Peng, Y. Chen, J. Wang, M. R. Green, S. Li, & A. G. Rosmarin, GABP transcription factor is required for development of chronic myelogenous leukemia via its control of PRKD2. *Proceedings of the National Academy of Sciences*, **110** (2013). <https://doi.org/10.1073/pnas.1212904110>.
300. S. Yu, K. Cui, R. Jothi, D.-M. Zhao, X. Jing, K. Zhao, & H.-H. Xue, GABP controls a critical transcription regulatory module that is essential for maintenance and

- differentiation of hematopoietic stem/progenitor cells. *Blood*, **117** (2011). <https://doi.org/10.1182/blood-2010-09-306563>.
301. C. L. Hirsch, Z. C. Akdemir, L. Wang, G. Jayakumaran, D. Trcka, A. Weiss, J. J. Hernandez, Q. Pan, H. Han, X. Xu, Z. Xia, A. P. Salinger, M. Wilson, F. Vizeacoumar, A. Datti, W. Li, A. J. Cooney, M. C. Barton, B. J. Blencowe, J. L. Wrana, & S. Y. R. Dent, Myc and SAGA rewire an alternative splicing network during early somatic cell reprogramming. *Genes and Development*, **29** (2015) 803–816. <https://doi.org/10.1101/gad.255109.114>.
 302. B. Hoffman, A. Amanullah, M. Shafarenko, & D. A. Liebermann, The proto-oncogene c-myc in hematopoietic development and leukemogenesis. *Oncogene*, **21** (2002). <https://doi.org/10.1038/sj.onc.1205400>.
 303. T. M. Dexter, T. D. Allen, & L. G. Lajtha, Conditions controlling the proliferation of haemopoietic stem cells in vitro. *Journal of Cellular Physiology*, **91** (1977). <https://doi.org/10.1002/jcp.1040910303>.
 304. G. de Haan & R. Ploemacher, The Cobblestone-Area-Forming Cell Assay. *Hematopoietic Stem Cell Protocols* (New Jersey: Humana Press). <https://doi.org/10.1385/1-59259-140-X:143>.
 305. W. J. C. Rombouts, A. Broyl, A. C. M. Martens, R. Slater, & R. E. Ploemacher, Human acute myeloid leukemia cells with internal tandem duplications in the Flt3 gene show reduced proliferative ability in stroma supported long-term cultures. *Leukemia*, **13** (1999) 1071–1078. <https://doi.org/10.1038/sj.leu.2401446>.
 306. S. P. S. Dhama, S. S. Kappala, A. Thompson, & E. Szegezdi, Three-dimensional ex vivo co-culture models of the leukaemic bone marrow niche for functional drug testing. *Drug Discovery Today*, **21** (2016). <https://doi.org/10.1016/j.drudis.2016.04.019>.
 307. D. van Gosliga, H. Schepers, A. Rizo, D. van der Kolk, E. Vellenga, & J. J. Schuringa, Establishing long-term cultures with self-renewing acute myeloid leukemia stem/progenitor cells. *Experimental Hematology*, **35** (2007). <https://doi.org/10.1016/j.exphem.2007.07.001>.
 308. D. G. J. Cucchi, R. W. J. Groen, J. J. W. M. Janssen, & J. Cloos, Ex vivo cultures and drug testing of primary acute myeloid leukemia samples: Current techniques and implications for experimental design and outcome. *Drug Resistance Updates*, **53** (2020). <https://doi.org/10.1016/j.drup.2020.100730>.
 309. C. R. Marlein, L. Zaitseva, R. E. Piddock, S. D. Robinson, D. R. Edwards, M. S. Shafat, Z. Zhou, M. Lawes, K. M. Bowles, & S. A. Rushworth, NADPH oxidase-2 derived superoxide drives mitochondrial transfer from bone marrow stromal cells to leukemic blasts. *Blood*, **130** (2017). <https://doi.org/10.1182/blood-2017-03-772939>.
 310. M. Azadniv, J. R. Myers, H. R. McMurray, N. Guo, P. Rock, M. L. Coppage, J. Ashton, M. W. Becker, L. M. Calvi, & J. L. Liesveld, Bone marrow mesenchymal stromal cells from acute myelogenous leukemia patients demonstrate adipogenic differentiation propensity with implications for leukemia cell support. *Leukemia*, **34** (2020). <https://doi.org/10.1038/s41375-019-0568-8>.
 311. L. Quek, G. W. Otto, C. Garnett, L. Lhermitte, D. Karamitros, B. Stoilova, I.-J. Lau, J. Doondeea, B. Usukhbayar, A. Kennedy, M. Metzner, N. Goardon, A. Ivey, C. Allen, R. Gale, B. Davies, A. Sternberg, S. Killick, H. Hunter, P. Cahalin, A. Price, A. Carr, M. Griffiths, P. Virgo, S. Mackinnon, D. Grimwade, S. Freeman, N. Russell, C. Craddock, A. Mead, A. Peniket, C. Porcher, & P. Vyas, Genetically distinct leukemic stem cells in human CD34⁺ acute myeloid leukemia are arrested at a hemopoietic precursor-like stage. *Journal of Experimental Medicine*, **213** (2016) 1513–1535. <https://doi.org/10.1084/jem.20151775>.

312. A. K. Brenner, E. Aasebø, M. Hernandez-Valladares, F. Selheim, F. Berven, I.-S. Grønningsæter, S. Bartaula-Brevik, & Ø. Bruserud, The Capacity of Long-Term in Vitro Proliferation of Acute Myeloid Leukemia Cells Supported Only by Exogenous Cytokines Is Associated with a Patient Subset with Adverse Outcome. *Cancers*, **11** (2019). <https://doi.org/10.3390/cancers11010073>.
313. D. J. Pearce, D. Taussig, K. Zibara, L.-L. Smith, C. M. Ridler, C. Preudhomme, B. D. Young, A. Z. Rohatiner, T. A. Lister, & D. Bonnet, AML engraftment in the NOD/SCID assay reflects the outcome of AML: implications for our understanding of the heterogeneity of AML. *Blood*, **107** (2006). <https://doi.org/10.1182/blood-2005-06-2325>.
314. A. Sánchez-Aguilera & S. Méndez-Ferrer, The hematopoietic stem-cell niche in health and leukemia. *Cellular and Molecular Life Sciences*, **74** (2017). <https://doi.org/10.1007/s00018-016-2306-y>.
315. M. Konopleva, S. Konoplev, W. Hu, A. Zaritskey, B. Afanasiev, & M. Andreeff, Stromal cells prevent apoptosis of AML cells by up-regulation of anti-apoptotic proteins. *Leukemia*, **16** (2002). <https://doi.org/10.1038/sj.leu.2402608>.
316. M. Kahl, A. Brioli, M. Bens, F. Perner, A. Kresinsky, U. Schnetzke, A. Hinze, Y. Sbirkov, S. Stengel, G. Simonetti, G. Martinelli, K. Petrie, A. Zelent, F.-D. Böhmer, M. Groth, T. Ernst, F. H. Heidel, S. Scholl, A. Hochhaus, & T. Schenk, The acetyltransferase GCN5 maintains ATRA-resistance in non-APL AML. *Leukemia*, **33** (2019). <https://doi.org/10.1038/s41375-019-0581-y>.
317. R. K. Khajuria, M. Munschauer, J. C. Ulirsch, C. Fiorini, L. S. Ludwig, S. K. McFarland, N. J. Abdulhay, H. Specht, H. Keshishian, D. R. Mani, M. Jovanovic, S. R. Ellis, C. P. Fulco, J. M. Engreitz, S. Schütz, J. Lian, K. W. Gripp, O. K. Weinberg, G. S. Pinkus, L. Gehrke, A. Regev, E. S. Lander, H. T. Gazda, W. Y. Lee, V. G. Panse, S. A. Carr, & V. G. Sankaran, Ribosome Levels Selectively Regulate Translation and Lineage Commitment in Human Hematopoiesis. *Cell*, **173** (2018). <https://doi.org/10.1016/j.cell.2018.02.036>.
318. C. Pina, G. May, S. Soneji, D. Hong, & T. Enver, MLLT3 Regulates Early Human Erythroid and Megakaryocytic Cell Fate. *Cell Stem Cell*, **2** (2008). <https://doi.org/10.1016/j.stem.2008.01.013>.
319. S. Belluschi, E. F. Calderbank, V. Ciaurro, B. Pijuan-Sala, A. Santoro, N. Mende, E. Diamanti, K. Y. C. Sham, X. Wang, W. W. Y. Lau, W. Jawaid, B. Göttgens, & E. Laurenti, Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. *Nature Communications*, **9** (2018). <https://doi.org/10.1038/s41467-018-06442-4>.
320. J. Carrelha, Y. Meng, L. M. Kettyle, T. C. Luis, R. Norfo, V. Alcolea, H. Boukarabila, F. Grasso, A. Gambardella, A. Grover, K. Höglstrand, A. M. Lord, A. Sanjuan-Pla, P. S. Woll, C. Nerlov, & S. E. W. Jacobsen, Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature*, **554** (2018). <https://doi.org/10.1038/nature25455>.
321. P. van den Berk, C. Lancini, C. Company, M. Serresi, M. P. Sanchez-Bailon, D. Hulsman, C. Pritchard, J.-Y. Song, M. J. Schmitt, E. Tanger, O. Popp, P. Mertins, I. J. Huijbers, H. Jacobs, M. van Lohuizen, G. Gargiulo, & E. Citterio, USP15 Deubiquitinase Safeguards Hematopoiesis and Genome Integrity in Hematopoietic Stem Cells and Leukemia Cells. *Cell Reports*, **33** (2020). <https://doi.org/10.1016/j.celrep.2020.108533>.
322. L. Arede, E. Foerner, S. Wind, R. Kulkarni, A. F. Domingues, S. Kleinwaechter, S. Gupta, E. Scheer, L. Tora, & C. Pina, Unique roles of ATAC and SAGA - KAT2A

- complexes in normal and malignant hematopoiesis. *bioRxiv*, (2020) 2020.05.14.096057. <https://doi.org/10.1101/2020.05.14.096057>.
323. R. Ahrends, A. Ota, K. M. Kovary, T. Kudo, B. O. Park, & M. N. Teruel, Controlling low rates of cell differentiation through noise and ultrahigh feedback. *Science*, **344** (2014) 1384–1389. <https://doi.org/10.1126/science.1252079>.
324. J. A. Doudna & E. Charpentier, The new frontier of genome engineering with CRISPR-Cas9. *Science*, **346** (2014). <https://doi.org/10.1126/science.1258096>.
325. P. D. Hsu, E. S. Lander, & F. Zhang, Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, **157** (2014). <https://doi.org/10.1016/j.cell.2014.05.010>.
326. L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman, & L. S. Qi, CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*, **154** (2013). <https://doi.org/10.1016/j.cell.2013.06.044>.
327. I. B. Hilton, A. M. D’Ippolito, C. M. Vockley, P. I. Thakore, G. E. Crawford, T. E. Reddy, & C. A. Gersbach, Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology*, **33** (2015). <https://doi.org/10.1038/nbt.3199>.

Appendices

Appendix A

RNA-seq differentially expressed genes in CTRLsh vs KAT2Ash HSCs (10% FDR).

FC- log fold change; CPM- log counts per million; FDR- false discovery rate

Gene ID	logFC	logCPM	F	PValue	FDR
FAM178B	-2.7919339	4.57662182	209.804103	6.36E-08	0.00055831
NDUFA7	-10.65197	1.56035749	273.732958	6.52E-08	0.00055831
RP11.442H21.2	-10.718431	1.59202772	240.327631	1.13E-07	0.00055831
AC026703.1	7.97918721	1.91624787	182.503811	1.22E-07	0.00055831
ITGB3.1	-3.3532307	3.66336714	176.048598	1.44E-07	0.00055831
GATS	-9.8349675	0.76338795	168.500593	5.06E-07	0.00163249
PPBP	-2.3129192	4.47036984	129.031556	6.04E-07	0.00167001
IGFBP4	-2.0176159	3.9944502	107.138753	1.41E-06	0.00339818
PKHD1L1	-2.1180676	5.62881016	96.497307	2.25E-06	0.00483162
CCND1	-1.5451196	4.9786278	91.951013	2.79E-06	0.00539332
AC068831.15	-9.2135714	0.21358427	107.191964	3.29E-06	0.00563785
SELP	-2.3323822	5.01591171	87.3775137	3.50E-06	0.00563785
VWF	-1.5753351	6.29388924	84.4974527	4.06E-06	0.00603634
RP11.1415C14.2	7.12949796	1.12423216	81.0659378	4.87E-06	0.00626421
RP11.1415C14.1	7.12949796	1.12423216	81.0659378	4.87E-06	0.00626421
SYP	2.25769994	3.15817232	78.3308404	5.66E-06	0.00626421
CD36	-1.326918	7.01035101	78.0046359	5.77E-06	0.00626421
GRAP2	-1.2923771	5.2741455	76.9619779	6.12E-06	0.00626421
HBB	-1.6668311	5.87745836	75.25629	6.75E-06	0.00626421
FOXJ1	-1.5969441	4.42706646	74.5606195	7.03E-06	0.00626421
CDH1	-1.8230216	3.76533043	74.4864825	7.06E-06	0.00626421
YWHAB	-1.2845636	6.9949001	74.0013034	7.26E-06	0.00626421
CTTN	-1.822755	4.0674422	73.5734519	7.45E-06	0.00626421
ABCC3	-2.394305	3.25978991	71.3241095	8.52E-06	0.00680041
PRG2	-1.9734162	7.53643255	70.8261119	8.79E-06	0.00680041
PKLR	-1.4229324	4.42151634	70.154519	9.16E-06	0.00681476
GABRE	-2.1886179	4.24873927	68.8742229	9.92E-06	0.0071073
ITM2B	-1.557922	6.61457459	66.8155077	1.13E-05	0.0078133
GNG4	2.95563226	1.24601719	64.9897665	1.27E-05	0.00849883
YBX3	-1.4253536	6.6826618	64.4799722	1.32E-05	0.00849883
SLC2A14	-2.1432289	2.8145318	63.1737887	1.44E-05	0.00897912

BSN	1.91960277	3.42469595	60.9870835	1.67E-05	0.01011188
RAB39A	3.35323633	0.69731341	60.0074013	1.79E-05	0.01050583
RHAG	-1.3504209	4.35702225	59.2009561	1.90E-05	0.0108007
CTAGE4	8.6059153	-0.3081708	64.0312481	2.58E-05	0.01426574
CA2	-2.1260744	2.12866322	53.3948398	2.93E-05	0.01576866
TOP1MT	-1.307755	4.38732688	51.9844273	3.28E-05	0.0170132
F2R	-1.5511334	6.09105276	51.7597565	3.34E-05	0.0170132
NSMAF	-0.9951983	6.01800862	50.1991702	3.80E-05	0.0188299
RIC3	2.02759409	2.18867936	49.6014567	3.99E-05	0.01901569
LTBP1	-1.6018366	7.41311926	49.4804437	4.03E-05	0.01901569
C5orf15	-1.0839755	4.76990089	49.0924172	4.16E-05	0.01917822
GDAP1	1.30514176	4.07378424	48.4088614	4.41E-05	0.01985042
ATPAF1	-1.1499677	4.48521769	48.0974605	4.53E-05	0.0199231
ZDHHC9	-1.1517861	4.47687887	47.3933614	4.81E-05	0.02025234
VWA5A	-0.9983599	5.11952474	47.392728	4.82E-05	0.02025234
UBXN10	-2.4949024	1.30463104	46.9555644	5.00E-05	0.02059178
SLC25A1	-0.9251074	5.84309404	45.8367069	5.52E-05	0.02214392
EPX	-1.2447312	4.18530108	45.6648631	5.61E-05	0.02214392
CBLL1	-1.0071338	6.56940098	45.3804639	5.75E-05	0.02225436
LEPR	-1.038541	5.00235499	44.9792063	5.97E-05	0.02225436
AHSP	-2.3857605	1.26399934	44.5915816	6.18E-05	0.02225436
AC107982.4	8.31512407	-0.5415332	50.7194788	6.33E-05	0.02225436
GOLGA7B	4.04593555	0.3217727	44.284004	6.36E-05	0.02225436
CERS6	-0.9272692	6.41153102	44.2112845	6.40E-05	0.02225436
SLC38A5	-0.9433713	5.52936971	44.1414792	6.44E-05	0.02225436
NAA50	-0.9482909	6.7311562	43.7335918	6.69E-05	0.02270539
ARHGEF40	0.94868749	5.65523475	43.361784	6.93E-05	0.02310135
TNS4	-5.2103863	-0.0832531	43.1433841	7.07E-05	0.02317986
NEO1	-2.0606482	3.74001808	42.0458542	7.85E-05	0.02495992
PF4	-1.8572984	3.42942765	42.0141851	7.87E-05	0.02495992
RP11.676J12.4	-5.8285494	-0.720811	41.7191008	8.10E-05	0.02498902
PROS1	-0.9298074	5.26689604	41.6675479	8.14E-05	0.02498902
PTGER3	-1.2232757	4.45593947	41.3488967	8.39E-05	0.02537133
PTGDR2	1.13971918	4.79593733	40.3128622	9.29E-05	0.02766214
LRRC32	-1.7191298	3.16815555	39.9482605	9.64E-05	0.02779905
TRIM29	-2.6631628	0.62281202	39.8853808	9.70E-05	0.02779905
ASIC4	1.09577074	4.48802608	39.8120555	9.77E-05	0.02779905
ITGB3	-1.1972307	7.40461172	39.3024809	0.00010286	0.02866704
RAD51L3.RFFL	-8.1705926	-0.7568719	44.1904411	0.00010628	0.02866704
APOE	-1.9903519	1.98727918	38.3016778	0.00011399	0.02866704
RNU1.28P	0.88579753	6.24989788	38.1689679	0.00011557	0.02866704
RNU1.27P	0.88579753	6.24989788	38.1689679	0.00011557	0.02866704

RNU1.1	0.88579753	6.24989788	38.1689679	0.00011557	0.02866704
RNVU1.18	0.88579753	6.24989788	38.1689679	0.00011557	0.02866704
RNU1.2	0.88579753	6.24989788	38.1689679	0.00011557	0.02866704
RNU1.4	0.88579753	6.24989788	38.1689679	0.00011557	0.02866704
RNU1.3	0.88579753	6.24989788	38.1689679	0.00011557	0.02866704
CD9	-1.9467042	3.64928251	37.3473395	0.00012597	0.03007184
ARL6IP5	-0.8337588	5.8761186	37.3375136	0.00012611	0.03007184
MGAT5B	1.73447744	2.37236315	37.3222539	0.00012631	0.03007184
TMEM56	-1.6718854	2.69502293	37.2374884	0.00012745	0.03007184
HBG2	-0.8282934	5.94648663	36.7330561	0.0001345	0.03135243
TMCO3	-0.9437675	4.80969303	36.5031394	0.00013787	0.03175491
GHITM	-0.9183674	6.01344833	36.3460458	0.00014022	0.03191844
MANEAL	1.96595002	1.48162798	35.8582543	0.00014787	0.03281923
TRAP1	-0.8838828	6.13952359	35.8115418	0.00014863	0.03281923
PDE3A	-0.9356703	4.77743718	35.4844034	0.00015407	0.03281923
LGMN	-1.77024	2.12904195	35.4594272	0.00015449	0.03281923
UCA1	-2.0350359	1.93501915	35.3996797	0.00015552	0.03281923
CHKB.CPT1B	1.99840445	2.10248956	35.3847193	0.00015577	0.03281923
SARM1	1.24214122	4.35948296	35.3684057	0.00015606	0.03281923
ID2	-1.0669651	4.39141285	35.1214586	0.00016039	0.03336809
PCSK9	-1.3372199	3.97783467	34.8629172	0.00016508	0.03397938
SLC2A3	-1.0086984	6.7553463	34.5535491	0.00017092	0.03481049
BLVRB	-0.8682192	5.41037738	34.3484159	0.00017493	0.03525562
TUBB2A	-1.7278508	2.24568264	33.8049665	0.00018611	0.03712182
ADD2	-0.8998515	4.75469628	33.657413	0.00018929	0.03737134
AC087392.1	7.97840136	-0.8923142	37.3754379	0.00019667	0.03843671
AK4	-0.795312	5.74617402	33.2172086	0.00019918	0.03845034
NPDC1	1.13703402	4.24048107	33.0978855	0.00020197	0.03845034
CANX	-1.1322802	8.74190948	33.0296228	0.00020359	0.03845034
STXBP6	1.39834956	3.03612252	32.9832994	0.00020469	0.03845034
PVRIG2P	-1.78422	1.66888481	32.7054642	0.00021148	0.03933609
CLN3.1	-7.9868634	-0.9843453	36.5381172	0.00021347	0.03933609
TPSB2	-1.2373947	7.08879588	32.0307471	0.00022912	0.04153233
CSF3R	0.78123683	5.55651326	32.0100722	0.00022969	0.04153233
BNIP3	-0.8138449	5.1524408	31.8735765	0.00023348	0.04182725
TP53RK	-0.7768699	6.3211129	31.6150208	0.00024087	0.04275594
STC2	-2.2428256	0.94733728	31.3566473	0.00024854	0.04371677
FAM43A	1.18372714	4.15838417	31.2073312	0.00025311	0.04411922
TPSAB1	-1.2030019	8.19963128	30.9484953	0.00026127	0.04449602
EPCAM	-1.5339833	2.76771534	30.8414242	0.00026474	0.04449602
C3orf58	-0.7969384	6.23164933	30.7944966	0.00026628	0.04449602
RP11.407N17.3	-4.1358966	0.70791621	30.78418	0.00026662	0.04449602

C3orf14	1.95600349	1.57611997	30.7795367	0.00026677	0.04449602
ENPP3	-0.9862373	4.32312895	30.6248766	0.00027193	0.04483512
EMP2	-1.7056126	1.73977288	30.5379116	0.00027488	0.04483512
SERP1	-0.7894917	6.71876662	30.5123182	0.00027576	0.04483512
AMMECR1	-0.8031819	5.82654732	30.4408996	0.00027822	0.04485859
ARPP21	-1.6987374	2.71608016	30.2255962	0.00028581	0.04542714
PHLDA3	2.04345096	1.05355279	30.2078871	0.00028644	0.04542714
CFH	1.04028486	4.04377359	30.0904145	0.0002907	0.04572731
MAT2B	-0.8163782	5.42627552	29.9404309	0.00029625	0.046085
EPOR	-0.8176704	5.16392446	29.9006506	0.00029774	0.046085
PRAM1	1.24923531	3.79806198	29.7722213	0.00030262	0.04617265
TPSD1	-1.2369253	4.26435643	29.6810532	0.00030614	0.04617265
PMP22	-0.9559369	4.26586747	29.6706337	0.00030655	0.04617265
BTF3L4	-0.7748487	5.21217322	29.6372609	0.00030785	0.04617265
HLA.DOA	1.28331849	3.22561627	29.5707343	0.00031047	0.0462073
SCAMP5	1.40545233	2.83120183	29.4633752	0.00031475	0.0463338
BASP1	-0.8003496	5.63455362	29.3968746	0.00031744	0.0463338
AP001266.1	2.21428311	1.14298939	29.3708237	0.0003185	0.0463338
LRRC16B	2.08687342	0.85782887	29.1502071	0.00032766	0.04731052
EIF5B	-0.8287479	7.20088314	29.064146	0.00033132	0.04748447
AKAP2	-0.9017449	7.8219992	28.9669432	0.00033551	0.04754467
PVRL1	-0.7814922	5.41328982	28.9407013	0.00033666	0.04754467
SDPR	-0.8974509	6.92870613	28.7706565	0.00034418	0.04825438
LTC4S	1.46920778	2.41047034	28.6794954	0.00034829	0.04829741
RNF7	-1.0065562	4.02338962	28.6383457	0.00035017	0.04829741
ALAS2	-1.3153514	3.42050156	28.5626644	0.00035365	0.04829741
MAPK11	1.87368602	3.39885673	28.4521791	0.00035881	0.04829741
MORF4L2	-0.7627965	6.49246579	28.4248837	0.0003601	0.04829741
FAM175A	-0.815534	4.81136455	28.4081103	0.00036089	0.04829741
AC138123.2	-1.0987862	4.01386774	28.3759497	0.00036242	0.04829741
SHPK.1	2.01459595	1.74908342	28.3179362	0.0003652	0.04829741
PFKFB4	-0.7520225	5.50115389	28.2571484	0.00036814	0.04829741
CTD.2031P19.4	-1.1367786	4.0513881	28.2304554	0.00036944	0.04829741
EDA2R	1.7680313	1.62496642	28.0244169	0.00037967	0.04930088
ANXA5	0.78343047	5.19469237	27.7668226	0.00039294	0.05047524
FAM45A	-0.7444832	5.47672151	27.7479327	0.00039393	0.05047524
PRSS23	0.74256901	5.2823292	27.5410165	0.00040503	0.05155573
MBOAT2	-0.7126237	6.04089307	26.9133373	0.00044108	0.05577731
WSB2	-0.9982382	4.40503946	26.7451262	0.00045139	0.05671047
ADAM28	0.7746216	4.88769837	26.6472454	0.00045752	0.05711019
RPL13AP5	-0.7030387	6.07815373	26.5385448	0.00046445	0.05726508
LDHA	-1.0378328	8.5129856	26.5349582	0.00046468	0.05726508

KRT13	-2.7745848	-0.0672338	26.4132031	0.0004726	0.05787218
AL139819.1	-1.4321284	4.19845397	29.0921665	0.00047788	0.05815113
RYR3	-0.8103062	7.47668867	26.1931394	0.00048732	0.05892968
DISP2	1.48135078	2.18861542	25.7348167	0.00051981	0.06246813
SCAMP1	-0.7362454	5.08817676	25.6082556	0.00052924	0.06287205
GP6	-1.3526847	3.14022592	25.5241435	0.00053562	0.06287205
CPNE2	-1.0194098	3.83230925	25.5203829	0.00053591	0.06287205
FCGR2A	-0.9061804	4.41477167	25.5169386	0.00053617	0.06287205
ATP11C	-0.6942706	5.54854695	25.4426789	0.00054189	0.06315968
RP11.46D6.1	1.87153375	1.03619075	25.3348793	0.00055032	0.06375826
SEMA3C	-0.7469942	5.12045685	25.1482586	0.00056529	0.06483875
EPAS1	1.60624714	1.8186077	25.0956176	0.00056961	0.06483875
FST	-1.0687936	3.57486032	25.0944674	0.0005697	0.06483875
AP000350.10	7.6728791	-1.2001501	27.5795353	0.0005742	0.06494734
SEL1L	-0.7148802	6.54762541	24.991435	0.00057826	0.06494734
WSCD2	-2.7346026	0.7409234	24.9620664	0.00058073	0.06494734
LY6G6D	-3.9627172	-0.5807463	24.7803104	0.00059628	0.06600066
SLC18A2	-0.7128584	6.36427728	24.7724465	0.00059697	0.06600066
STMN3	1.35399173	5.79589784	24.5132719	0.00062006	0.06789482
EIF3J	-0.7212653	5.71019568	24.4801888	0.00062309	0.06789482
BMS1P1	0.96707672	3.8743946	24.4634103	0.00062463	0.06789482
PTPRJ	-1.7443619	3.28185298	24.3951825	0.00063094	0.06819794
CECR1	0.86571777	4.57861688	24.3000011	0.00063988	0.06873636
ARPC1A	-0.7675712	4.79342371	24.2668375	0.00064303	0.06873636
ODC1	-0.7008897	6.63176562	24.198468	0.00064958	0.06905527
MTFP1	-1.1123866	3.35157625	23.9960313	0.00066946	0.07077971
TSPAN7	1.3950794	2.10073699	23.798262	0.00068959	0.07229403
PRKAA1	-0.7319216	5.76108611	23.7822533	0.00069125	0.07229403
TOMM20	-0.7096446	6.86970144	23.7096101	0.00069886	0.07237508
HPSE	-1.4908599	3.3433987	23.7033881	0.00069951	0.07237508
FN3K	-0.8259704	4.41538903	23.4903277	0.00072242	0.07417593
BLOC1S6	-0.6698255	5.72460557	23.4706108	0.00072458	0.07417593
ANKRD36BP1	-0.9917558	3.7318181	23.3488568	0.00073813	0.07516495
AP000295.9	-1.8267503	1.03520702	23.1695836	0.00075864	0.07668658
ETV5	-0.7601633	4.70321533	23.149299	0.000761	0.07668658
SLFN5	-1.1064563	3.91599488	23.0544978	0.00077216	0.07740839
KRT19	-1.4654521	1.78739353	22.9773267	0.0007814	0.07793028
TRIB2	-0.6801247	6.18045808	22.6833799	0.00081781	0.08114375
CLEC1B	-1.7555943	1.71975901	22.6497139	0.00082211	0.08115423
SOS1	-0.9795413	8.45725912	22.5641488	0.00083316	0.08147846
GEMIN5	-0.6496575	5.71098756	22.5076845	0.00084056	0.08147846
RAB8B	-0.6579424	5.87062131	22.4889479	0.00084303	0.08147846

SH3TC1	1.25077137	4.67893985	22.4595957	0.00084691	0.08147846
AC011558.5	5.3271911	-0.7538758	22.4529686	0.0008478	0.08147846
RGS6	-1.0646422	3.99516341	22.43144	0.00085066	0.08147846
CHML	-0.64272	5.82534996	22.3127432	0.00086669	0.08260478
ANP32E	-0.6822003	6.82690423	22.1963674	0.00088276	0.08274363
FAM136A	-0.7214768	5.16170751	22.1921695	0.00088335	0.08274363
MICALL2	-0.7335035	4.81988097	22.1829178	0.00088464	0.08274363
SELENBP1	-1.4292228	2.10522309	22.1785452	0.00088526	0.08274363
ARHGAP18	-0.7174464	5.9672185	22.1099423	0.00089493	0.08301663
SEZ6L	3.56941383	0.36849377	22.0970864	0.00089676	0.08301663
RP11.512M8.5	1.38325472	3.02174272	22.0430399	0.00090449	0.0833336
ZFPM2	-2.8317646	1.18586688	21.9652103	0.00091576	0.08397254
HIF1A	-0.7826863	6.36404852	21.884628	0.00092762	0.0846582
HSP90AB1	-1.0449005	9.83087933	21.8105577	0.00093868	0.08505706
RASGRP3	-0.9267197	5.05454515	21.7965861	0.00094078	0.08505706
CYP1B1	0.68642082	5.23351094	21.7658897	0.00094542	0.0850792
KCNE1	-1.7829089	0.99369599	21.6632025	0.00096115	0.08520657
PRICKLE2	-1.8392286	1.22267361	21.6269857	0.00096678	0.08520657
RASSF4	1.78838528	1.10412488	21.6260212	0.00096693	0.08520657
ICK	-0.8644244	4.13343114	21.5850528	0.00097334	0.08520657
PLD4	1.21746499	4.32928614	21.553292	0.00097834	0.08520657
FAM98A	-0.7348393	4.99416749	21.5489105	0.00097904	0.08520657
SERPINE2	-0.8448244	4.85528786	21.5386295	0.00098066	0.08520657
LPXN	0.82316008	4.56609367	21.5207727	0.0009835	0.08520657
ARHGEF17	0.65242592	5.36809812	21.5021234	0.00098647	0.08520657
UBE2L5P	-1.9673131	1.17867075	21.4152467	0.00100046	0.08603087
SLC16A9	-1.5640968	1.28257914	21.2519576	0.00102741	0.08784995
MPO	1.69480594	3.90681474	21.2091112	0.00103463	0.08784995
KDELC1	-1.10165	3.07349912	21.1919322	0.00103754	0.08784995
SLC6A8	-1.0328807	3.42313539	21.174008	0.00104058	0.08784995
CACNA1I	-1.7482538	1.27206738	21.1521196	0.00104432	0.08784995
RP11.618P17.4	-1.3882373	1.97481629	21.1093208	0.00105167	0.08808536
RPS17	-0.6573157	5.29603309	21.0797522	0.00105679	0.08810133
KAZN	-0.974455	3.61019998	21.0557045	0.00106097	0.08810133
PDK1	-0.6870707	6.26762363	20.8205075	0.00110294	0.09074915
EFCAB12	-1.4560328	1.46056569	20.8035261	0.00110605	0.09074915
KLHL5	-0.6594703	5.11697821	20.7987653	0.00110693	0.09074915
RAB11FIP4	1.03622119	3.43676387	20.7490204	0.0011161	0.09111513
MCFD2	-0.6264449	5.86844186	20.7018124	0.00112489	0.09144692
C10orf54	0.6731067	4.9702224	20.6355187	0.00113738	0.09178737
IFITM2	-0.6844192	5.39916135	20.6292467	0.00113857	0.09178737
TYW3	-0.7344454	4.6107657	20.512051	0.00116108	0.09321395

NPW	-1.1658232	2.75862878	20.4812785	0.00116708	0.0933085
TET1	-0.7882876	5.75960703	20.4421707	0.00117476	0.09353599
PPAP2B	-1.6064195	1.07058221	20.3876471	0.00118557	0.09400976
BCL2L1	-0.6552427	6.07704532	20.3265286	0.00119783	0.09459421
S100A4	0.62059892	5.83953456	20.2195066	0.00121967	0.09511883
RP11.113K21.4	-7.3313967	-1.5568306	21.985637	0.00122237	0.09511883
SMIM3	-2.6157678	-0.0636175	20.1996258	0.00122378	0.09511883
ANKRD36B	-0.6256711	5.85247061	20.1885333	0.00122608	0.09511883
NCS1	-1.2608174	2.51106657	20.1659098	0.00123078	0.09511883
RAC3	-0.9084007	3.90240005	20.1506621	0.00123397	0.09511883
TMOD2	1.14626111	3.23821838	20.0829097	0.00124824	0.09583726
UBE2Q2	-0.7216157	4.81964366	19.9643019	0.00127371	0.09690211
ERG	0.63718607	6.28535264	19.9339021	0.00128034	0.09690211
C1orf186	-0.7590097	4.71743077	19.9270683	0.00128184	0.09690211
BIN1	0.95547765	4.22704575	19.925666	0.00128214	0.09690211
CHD7	-0.6443556	6.68937852	19.886201	0.00129083	0.09717873
RP11.1415C14.4	-1.4671649	4.48257937	19.8524624	0.00129831	0.09736295
ACE	1.24420367	3.79514181	19.7865621	0.00131307	0.09808967
SGPL1	0.96750681	3.60354119	19.7087999	0.00133075	0.09902798
CHGB	2.83209038	-0.1141612	19.6182886	0.00135169	0.09970443
VGF	1.95736732	0.58544287	19.5804203	0.00136057	0.09970443
CTC.273B12.8	-1.9726701	0.38204147	19.5558751	0.00136636	0.09970443
KCNH8	1.54689242	1.12519072	19.5485975	0.00136808	0.09970443
GFI1B	-0.7888666	6.9236806	19.5297529	0.00137256	0.09970443
COX7A2L	-0.7116701	4.93033386	19.5050119	0.00137846	0.09970443
LINC01122	1.50532819	1.32053011	19.4976759	0.00138021	0.09970443
SH3GL1P3	1.52428129	1.14383303	19.4941346	0.00138106	0.09970443
HBG1	-1.0645057	3.12242422	19.4592863	0.00138945	0.09993677
UBE2G1	-0.6043346	5.63717823	19.436631	0.00139493	0.09995955

Appendix B

Enriched genes in erythroid-basophil-megakaryocyte-biased progenitors (EBMP) as per detailed single-cell profiling of erythroid development by Merryweather Clarke (2011).

EBMP region selected using SPRING tool:

https://kleintools.hms.harvard.edu/paper_websites/tusi_et_al/

GENE ID		GENE ID		GENE ID		GENE ID	
ApoE	1.638	Fermt3	0.334	Fam102a	0.269	Galnt6	0.233
Ctla2a	1.391	Pla2g4a	0.333	Gm11114	0.269	Gns	0.232
Muc13	1.112	Tbc1d14	0.333	Sptssa	0.269	Pkd1l3	0.232
Gata2	1.101	mt-Atp6	0.332	Spock2	0.268	Tor4a	0.232
H1fx	1.012	Mecom	0.332	Proser3	0.268	Rai1	0.232
Ctla2b	1.012	Vwc2l	0.331	Serinc1	0.268	Mettl14	0.232
Pbx1	0.994	Git2	0.331	Gm14735	0.268	Gm13552	0.232
F2r	0.958	2610034B18Rik	0.331	Fam60a	0.268	Ndst1	0.232
Nrgn	0.939	Bcr	0.33	Rhoh	0.268	Spint2	0.232
Itga2b	0.921	Zmynd11	0.33	Bex1	0.267	Runx2	0.232
Slc22a3	0.864	Rpl13a	0.33	Glul	0.267	Gna15	0.232
Pafah2	0.852	Rps14	0.328	Ankef1	0.267	Gm8522	0.231
Diap1	0.848	Pnir	0.328	Jam3	0.266	Abcc3	0.231
Adgrg1	0.845	Gnmt	0.328	Ski	0.266	Cep55	0.231
Cdk6	0.835	Sfi1	0.327	Zfp13	0.266	Mtus2	0.231
Meis1	0.766	Zdhhc24	0.327	Klk8	0.266	Gm14607	0.231
H2-K1	0.764	Tnfrsf18	0.327	Heg1	0.266	Mbtps2	0.231
Eef1a1	0.759	Tnrc18	0.327	Neur11b	0.266	Ncor2	0.231
Slc18a2	0.757	Zfp664	0.327	Nrxn1	0.266	Rhobtb3	0.231
Zfp385a	0.756	Atf7ip	0.327	Usp40	0.265	Col1a2	0.231
Mpl	0.751	Clec11a	0.326	Mzt1	0.265	Tet1	0.231
Csgalnact1	0.748	Tjp2	0.326	Notch2	0.265	Ptprc	0.231
Vezf1	0.746	Unk	0.326	Tmem59	0.265	Gm4613	0.231
Txnip	0.744	Fam189b	0.326	Stxbp1	0.265	Gm355	0.231
Fam65a	0.738	Tmem40	0.326	Gm14975	0.265	Gm13247	0.23
Mdga1	0.729	Rbms3	0.325	Nr2e1	0.264	Fbxl19	0.23
Angpt1	0.703	Dock8	0.324	Ddx52	0.264	Hsd3b2	0.23
Spns2	0.703	Gm18291	0.324	Hoxa10	0.264	Tuba8	0.23

Serpina3g	0.701	Stat3	0.323	Dennd1c	0.264	Rpl17-ps8	0.23
Rab37	0.698	Fam199x	0.323	Plgrkt	0.264	Creb1	0.23
Itga6	0.691	Rftn2	0.323	Rps11-ps1	0.263	Clec4d	0.23
Atp13a2	0.667	Trip6	0.323	Rps16	0.263	Tnip3	0.23
Ptrf	0.663	Sesn3	0.321	Pgpep1	0.263	5430427019Rik	0.23
Sox4	0.663	Rps21	0.321	Trappc6a	0.263	Kat6a	0.23
Ikzf2	0.652	Klc3	0.321	Bin1	0.263	Rpl23a-ps14	0.23
Rgs18	0.643	Aplnr	0.32	Stbd1	0.263	Plbd2	0.23
Dapp1	0.639	Usp11	0.319	RbmX	0.263	Tmem47	0.229
Slc7a8	0.634	Mrm1	0.319	4930555F03Ri	0.263	Ndn	0.229
Rpl15	0.632	Rbms2	0.319	Laptm4a	0.262	Srrm2	0.229
NdrG1	0.628	Zfp82	0.318	Taok3	0.262	Oit3	0.229
Gnb4	0.624	Plek	0.317	Fam64a	0.262	Cul7	0.229
Vim	0.602	Gpr161	0.317	Nedd4	0.262	Rfc2	0.229
Mycn	0.597	Glud1	0.317	Gm6368	0.262	Homer3	0.228
Plec	0.59	Afap111	0.317	Gm5356	0.262	Prl8a8	0.228
Gse1	0.59	Lgals1	0.317	Fnbp11	0.262	Rpl26	0.228
B2m	0.585	Smad5	0.317	Epsti1	0.261	Hint2	0.228
Inafm1	0.581	Zxdc	0.316	Atp6v0a2	0.261	Dtx3	0.228
Dach1	0.58	Neil1	0.316	Tap1	0.261	Rell2	0.228
Bin2	0.58	Snx10	0.315	Psat1	0.261	Zfhx3	0.228
Rps19	0.577	Ago1	0.315	Rplp2	0.261	Eef2	0.228
H2-D1	0.577	Tctn3	0.315	Ctnnbip1	0.261	Ube2e1	0.228
Tmem123	0.569	Slamf1	0.315	Rai14	0.261	Rtfdc1	0.227
Zfp36l2	0.566	Pdcd4	0.315	Ier3	0.26	Kcnt1	0.227
Tmem176b	0.565	Pold4	0.314	Eml2	0.26	Orai2	0.227
S100a10	0.565	Socs7	0.314	4933415A04Ri	0.26	Rpl38	0.227
Pitpnc1	0.564	Fam212a	0.314	Eml1	0.26	Aamdc	0.227
Gcnt2	0.56	Cobll1	0.314	Zfp54	0.26	Arhgef18	0.227
Rps15	0.555	Erg	0.313	Cpsf1	0.26	Gm38356	0.227
Rps9	0.553	Fam222a	0.313	Cstf2t	0.259	A930004D18Rik	0.227
Mef2c	0.551	Zbed3	0.312	Gm7329	0.259	Gm5493	0.227
Myct1	0.547	Mme	0.311	Ubald2	0.259	Gtf2i	0.227
H1f0	0.544	Fam3c	0.31	Cyp4f16	0.258	Fam65b	0.227
Trem12	0.536	Tmco4	0.31	Nav1	0.258	Ltbr	0.226
Rpl10a	0.535	AW551984	0.31	Sdr39u1	0.258	Ilf3	0.226
Rps18	0.534	Hsd17b10	0.309	Nek9	0.258	Svil	0.226
Plxnc1	0.527	Dag1	0.309	Arsk	0.257	Otos	0.226
Stmn1	0.526	Pdcd2	0.309	Tfap4	0.257	Nlgn2	0.226
Kit	0.526	Prkcq	0.309	Armc10	0.257	Best3	0.226
Rps5	0.519	Gm8420	0.309	Smad4	0.257	Gm21092	0.226
Cbfa2t3	0.518	Rps13	0.308	Cdk17	0.257	Ctr9	0.226

Lat	0.516	Ifitm1	0.308	Slc9b2	0.257	Gm38280	0.226
Zyx	0.515	Il17ra	0.307	Kansl1	0.256	Rnf2	0.226
Ccdc8	0.514	Utrn	0.307	Xlr4c	0.256	Phldb3	0.225
Gnb2l1	0.513	Rpl18-ps1	0.306	Camk1	0.256	1810030O07Rik	0.225
Tmem176a	0.513	Prkcdbp	0.306	Cd274	0.256	Rasip1	0.225
Macf1	0.508	Slit1	0.306	Cep68	0.255	Arl6ip1	0.225
Fbxl20	0.508	Crim1	0.306	RP24-461L18.3	0.255	Nap1l3	0.225
Tmem98	0.504	Ctsk	0.305	Ldlr	0.255	Rpl27a	0.225
Lgals9	0.503	Grb2	0.305	Ttc21b	0.254	Akt1	0.225
Smarca2	0.5	Samd14	0.305	Kctd3	0.254	Gm16411	0.225
Sfxn3	0.499	Itgb3	0.305	Prkar2b	0.254	Olfr132	0.225
Esam	0.497	Mlph	0.304	Tmem240	0.254	Gm12268	0.225
Soga1	0.495	Pcbp3	0.304	Lmo4	0.254	Gm12518	0.225
Ankrd13a	0.494	Actg1	0.304	Map4k4	0.253	2410004B18Rik	0.225
Tbxas1	0.491	Dcaf8	0.304	Gm14007	0.253	Echdc3	0.225
Ifitm3	0.49	Ppif	0.304	Cirbp	0.253	Vmn2r106	0.225
Numa1	0.489	Bahcc1	0.304	Clec4e	0.253	Gm26914	0.225
Msi2	0.488	Zfp191	0.304	Eif2ak3	0.252	Gm13195	0.224
Gp5	0.487	Dock2	0.304	Cuzd1	0.252	Acap1	0.224
Mtfr2	0.484	Ctf2	0.303	Pon3	0.252	Gm6136	0.224
H2-Q7	0.479	Bicd1	0.303	Hspb8	0.251	Lrrc4c	0.224
Rpl18	0.478	Ttc28	0.302	Ccdc112	0.251	Setd2	0.224
Ptms	0.476	Tbx20	0.302	D17H6S53E	0.251	Stx4a	0.224
Rps3	0.474	Ttc3	0.301	Slit2	0.251	Vmn1r52	0.224
Smim5	0.473	Rpl22	0.301	Myo1b	0.251	Spaca7	0.224
Arid1b	0.472	Dgkz	0.301	Ino80b	0.251	Adamts15	0.224
Rabgap1l	0.466	Rest	0.301	Gemin8	0.251	Kng1	0.224
Arid1a	0.466	Tom1l2	0.301	Fam198b	0.251	Pcmt2	0.224
Tgfbr2	0.462	Acp5	0.3	Lypla1	0.251	Sord	0.224
Por	0.459	Spry3	0.3	Calr3	0.25	Anks1	0.223
Sgce	0.459	Ric8	0.3	Lrrc9	0.25	Dlx1	0.223
Samsn1	0.458	Zer1	0.3	Ccdc121	0.25	Rbpj	0.223
Hsd3b7	0.455	Nckap5l	0.3	Kat2a	0.25	Gm2308	0.223
Tgfb1	0.454	Bcl2	0.3	Fhod1	0.25	Mlycd	0.223
Kctd1	0.451	Map1lc3b	0.299	Phf14	0.25	Elk1	0.223
Ms4a2	0.451	Ltc4s	0.299	Tnni3	0.25	Sptbn1	0.223
Prkca	0.45	As3mt	0.299	Rsrc1	0.249	Wfikkn1	0.223
Ehd3	0.449	Med12l	0.299	Olfr729	0.249	Npas3	0.223
Ccng2	0.448	Cd84	0.299	RP23-183O3.1	0.249	Gm6829	0.223
Csrp3	0.445	Cldn34-ps	0.298	Il10rb	0.249	Socs4	0.223
Spn	0.445	Cbln2	0.298	Rps11	0.249	Gprasp2	0.223
Crlf2	0.442	Grcc10	0.298	Odf3b	0.249	Gm7776	0.223

Fyn	0.441	Gm7206	0.298	Vsig10l	0.249	Gm28239	0.223
Ptprcap	0.441	Atf7	0.298	Zfp282	0.249	Ttc41	0.222
Dok2	0.439	Rtn4	0.297	Sncb	0.248	Col16a1	0.222
Tgfb3	0.439	Ankrd16	0.297	Zcchc11	0.248	Dxo	0.222
Phf21a	0.433	Rapgef3	0.296	Gnpda2	0.248	Med15	0.222
Cuedc1	0.432	Btbd8	0.296	Prkar1a	0.248	Foxo1	0.222
Mprip	0.428	Dusp23	0.296	Zbtb9	0.248	RP24-188B2.2	0.222
Vamp5	0.426	Ggcx	0.295	Col14a1	0.248	Wdr63	0.222
Sigirr	0.426	Myl12b	0.295	Rcor2	0.248	Tle2	0.222
Zbtb20	0.425	Adgr1	0.295	Rpl34	0.247	Fam161b	0.221
Pafah1b3	0.425	H2-T22	0.295	Kcnab2	0.247	Podxl2	0.221
Cox14	0.424	Cnbp	0.295	Rpl37a	0.247	AI314180	0.221
Myb	0.423	Gimap9	0.295	Mapk10	0.247	Pcdhb5	0.221
Celf2	0.423	Glms-ps1	0.295	Ikbkb	0.247	Mt3	0.221
Anxa5	0.421	Zfp831	0.295	Slc39a1	0.247	Vmn1r5	0.221
Pear1	0.42	Ick	0.295	Gm2830	0.247	Gm6771	0.221
Slc45a3	0.42	Tagln2	0.294	Anxa4	0.247	Gng2	0.221
Rab27b	0.419	Nr1d1	0.294	Gm13478	0.247	Ankrd7	0.221
St8sia6	0.418	Cenpb	0.293	Skp1a	0.247	Cdkn2c	0.22
Cnot6l	0.413	Gem	0.293	Cyb5r4	0.246	RP24-335I21.2	0.22
Csnk1e	0.412	Commd2	0.293	Gm8909	0.246	Adamts13	0.22
Gmpr	0.411	Cbfb	0.293	Sapcd1	0.246	Ddx58	0.22
Tsc22d1	0.409	Srgap3	0.293	Bcor	0.246	Vkorc1	0.22
Rpl6	0.408	Calhm2	0.292	Cacul1	0.246	Gucy1b3	0.22
Tnfsf12	0.407	Dhrs1	0.292	Wdpcp	0.246	Arhgap25	0.22
Emp3	0.407	Polr3k	0.292	Gm11914	0.246	Cnpy4	0.22
Sdsl	0.406	Lrba	0.292	Glb1	0.246	Itm2c	0.219
Kifc3	0.405	Vsig10	0.292	Gprasp1	0.245	2810403A07Rik	0.219
F2rl2	0.405	Ywhag	0.291	Hhex	0.245	Pip5k1c	0.219
Vwa5a	0.405	Chd9	0.291	Ppp3r2	0.245	Mndal	0.219
Pecr	0.404	Atxn7	0.29	mt-Co2	0.245	Htr2a	0.219
Ctnnd1	0.403	Hnrnph1	0.29	Ncor1	0.245	Gm9795	0.219
Pard3b	0.402	Zfp157	0.289	Dennd6a	0.245	Itgal	0.219
Rpl23	0.401	Sepw1	0.289	Gm7142	0.245	Rnase6	0.219
Trim47	0.4	Gna14	0.289	Runx1	0.245	Ajuba	0.219
Abca2	0.4	Nynrin	0.289	Hic2	0.245	Gbas	0.219
Ssbp2	0.4	Magi2	0.289	Ntrk2	0.245	Akr1b3	0.219
Kcnj5	0.4	Pwwp2a	0.289	D430041D05Ri	0.244	1700021K19Rik	0.218
Rps28	0.399	Casp2	0.289	Krba1	0.244	Ccdc94	0.218
Clec4f	0.398	R3hdm4	0.288	Cers4	0.244	Mro	0.218
Cux1	0.397	Map3k12	0.287	Pla2g4b	0.244	Qrfp	0.218
H3f3b	0.397	Zfp113	0.287	Kctd10	0.244	Wdfy3	0.218

Tpt1	0.397	Akr1c13	0.287	Sik3	0.244	Uvssa	0.218
Madd	0.395	Map7	0.287	Suv420h1	0.244	Olfr961	0.218
Stx3	0.395	Olfr609	0.287	Btk	0.244	Copg1	0.218
Bcl11a	0.394	Eif3h	0.287	Slc9a3r1	0.244	Lrrc3	0.218
Ctdsp2	0.394	Stat5a	0.286	Fam180a	0.243	Mfap3	0.218
Ankrd33b	0.394	Slc14a1	0.286	Zzz3	0.243	Zbtb12	0.218
Tpst2	0.394	Ocr1	0.286	Cabp4	0.243	Tmem201	0.218
Zfp358	0.392	Fgf13	0.285	4732471J01Rik	0.243	Ldlrad4	0.217
Serinc3	0.392	Pcbp4	0.285	Gm6472	0.243	Mapk14	0.217
Rab38	0.391	Rhof	0.285	Prdx4	0.243	Panx3	0.217
Morf4l1	0.391	Ilvbl	0.285	Ubr2	0.243	Ebpl	0.217
Fgf3	0.389	Kmt2c	0.284	Znhit6	0.243	Ddit4l	0.217
Sipa1l1	0.388	Ptov1	0.284	Pou5f2	0.243	Lrig3	0.217
Pts	0.388	Gm15427	0.284	Gp9	0.243	Rpl32	0.217
Clstn3	0.388	Gm10638	0.284	Ttc14	0.243	Smc5	0.217
Gabarapl1	0.387	Rpl18a	0.284	Cep89	0.242	2010012O05Rik	0.217
Klhl22	0.387	Adipor2	0.283	Hk1	0.242	Zfp580	0.217
Mon1b	0.387	Chrn4	0.283	Gm7589	0.242	Il6	0.217
Ncoa1	0.386	Hoxa7	0.283	Bcl7c	0.242	C1ql2	0.217
Phc1	0.386	Ccdc148	0.283	Cndp2	0.242	Syt13	0.217
Fam171a2	0.385	Dennd5a	0.283	Ift74	0.242	Pdgfb	0.216
Samd13	0.383	Lim2	0.283	Best2	0.241	Gm13038	0.216
Gramd1a	0.383	Copz2	0.283	Zfpm1	0.241	Gm14880	0.216
Fut8	0.382	Hlf	0.282	Sulf1	0.241	Trhr2	0.216
Psd3	0.38	Fut7	0.282	Msl3l2	0.241	Gm5224	0.216
Rps15a	0.379	Polr3gl	0.282	Itga5	0.241	2810039B14Rik	0.216
Zc4h2	0.379	Mink1	0.282	Mmrn1	0.241	Cggbp1	0.216
Akap13	0.375	Dnmt3b	0.282	Rusc2	0.241	Ddx17	0.216
Aipl1	0.374	Ube2d2a	0.281	Cdh16	0.241	Gm11197	0.216
Jak3	0.372	Ap5z1	0.281	Trp53	0.241	Cd109	0.216
Tnfsf10	0.371	Crebzf	0.281	Gm10123	0.241	Plscr3	0.215
St3gal1	0.37	Tbc1d31	0.281	Ociad2	0.241	Fbxl8	0.215
Prdx1	0.369	Tigd5	0.28	Ddx19b	0.241	Nomo1	0.215
Rpl31	0.369	Tsc22d3	0.28	RP24-383C4.1	0.241	Tmem219	0.215
Ptk2b	0.369	Cxhc5	0.28	Sp6	0.24	Fam159b	0.215
Tia1	0.368	Gng11	0.28	Usp17la	0.24	Hnrnph3	0.215
Ccnd2	0.368	Tm9sf3	0.28	Ahdc1	0.24	Acbd6	0.215
Ifi47	0.368	Ptpn18	0.28	Rps18-ps1	0.24	Tmem57	0.215
Ctnn	0.368	Gm20699	0.28	Sumo3	0.24	Chst3	0.215
Ecm2	0.367	Rps17	0.28	Ppp1r35	0.24	Hfe2	0.215
Gm14805	0.366	Eri3	0.279	Rsf1	0.24	Zfp677	0.215
Ormdl1	0.365	Map3k4	0.279	Rsbnl1	0.239	Fam168b	0.215

Dapk1	0.36	Tes	0.279	Sdr42e1	0.239	Gatad1	0.214
Prkacb	0.359	Tspan14	0.279	Fam117a	0.239	Gstk1	0.214
Sepp1	0.357	2610002J02Rik	0.279	Nat1	0.239	Tek	0.214
Rpl7	0.355	Rsph3b	0.279	Snw1	0.239	Rps4x	0.214
Tmod3	0.355	Tns1	0.278	Gtpbp10	0.239	Rps13-ps6	0.214
Nfe2	0.354	Cys1	0.278	Wdr72	0.239	RP23-246L14.4	0.214
H2-Q4	0.353	Fabp5	0.278	Pcp4l1	0.238	P2rx3	0.214
Cox7a2l	0.353	Apaf1	0.278	Rps26	0.238	Mfsd6	0.214
Fto	0.352	Ppfibp1	0.277	Oxct1	0.238	Dusp4	0.214
Tbrg1	0.351	Sde2	0.277	Tubd1	0.238	Rbbp6	0.214
Ddah2	0.351	Rps29	0.277	Pttg1ip	0.238	Atxn3	0.214
Tmsb10	0.351	Parp12	0.277	Aebp2	0.238	Dyx1c1	0.214
Hdac7	0.35	Rftn1	0.277	Nfya	0.238	Ermn	0.214
Smim10l1	0.35	Aktip	0.277	Zfp764	0.238	Fpr-rs3	0.214
Oplah	0.349	Iqgap1	0.276	Gm14931	0.238	Zdhhc9	0.214
Deptor	0.349	Marcksl1	0.276	Cabin1	0.238	Mbd6	0.214
Zeb2	0.347	Neto2	0.276	Klhdc9	0.238	Nmb	0.214
Igf1r	0.347	Terf2	0.275	Gm9320	0.238	Gm14867	0.213
Trappc9	0.346	3110062M04Rik	0.275	Nras	0.238	Rxb	0.213
Sdpr	0.346	Uba7	0.275	Abhd14a	0.237	Ldha	0.213
Kmt2a	0.345	Psme1	0.275	Zfx	0.237	Serpina3i	0.213
Zbtb17	0.345	Rnh1	0.275	Prrt2	0.237	Gm16210	0.213
Eif4a2	0.345	Yap1	0.275	Gm10540	0.237	Trp53bp1	0.213
Ak1	0.345	Rpl28	0.275	Csf2rb	0.237	Gm10300	0.213
Emid1	0.345	Gm6546	0.275	Ifitm2	0.237	Ift46	0.213
Srebf2	0.344	Gpatch8	0.274	Trim33	0.237	Trip10	0.213
Spred2	0.343	Mex3b	0.274	Zfp949	0.237	Pnrc1	0.212
Nfkb1	0.343	Ric8b	0.274	Nkain3	0.237	Cers2	0.212
Acss2	0.343	Fzd6	0.274	Olfr389	0.237	Cnot4	0.212
1500012F01Rik	0.343	Coro1b	0.273	Gltscr2	0.236	Klhl6	0.212
Gbp2b	0.343	Prpf40b	0.273	Trim35	0.236	Gm15382	0.212
Obsl1	0.342	Guca1a	0.273	Slc35b3	0.236	Pkd1	0.212
Senp6	0.342	Prmt2	0.273	Tmem222	0.236	Gigyf2	0.212
Renbp	0.342	Chd3	0.272	Gsap	0.236	Pik3cd	0.212
Skor1	0.342	Prrc2c	0.272	Rassf3	0.236	Olig2	0.212
Rpl7a	0.341	Dgkq	0.272	Flot1	0.236	Cage1	0.212
Nucb1	0.34	Appbp2	0.271	Gm17727	0.235	Cib3	0.212
Osbpl1a	0.34	Fkbp9	0.271	Obfc1	0.235	Zbtb33	0.211
Arl6ip5	0.339	Ctcf1	0.271	Tsr2	0.235	2410089E03Rik	0.211
Zfp457	0.339	Mfge8	0.271	Arap3	0.235	Ltbp3	0.211
Gp1bb	0.338	Dynll2	0.271	Cacnb4	0.235	Vmn2r95	0.211
Cd48	0.338	Pcdhb14	0.271	Zfyve21	0.235	Gm12302	0.211

PISD	0.338	Rpl13	0.27	Pnrc2	0.235	Fam49a	0.211
Csrp1	0.338	Fam174b	0.27	Rpl10a-ps1	0.234	Pdp2	0.211
Jsrp1	0.337	Fxyd5	0.27	Rabac1	0.234	Gm7593	0.211
Cd47	0.337	Dnm3	0.27	Lrwd1	0.234	Gpx2	0.211
Rpl14	0.337	Rpl37rt	0.27	Acvr1b	0.234	Pde4b	0.211
Foxp1	0.337	Ptger3	0.269	1110037F02Ri	0.234	B930041F14Rik	0.211
Timp2	0.337	Slc45a4	0.269	Add3	0.234	Bace2	0.211
Eef1b2	0.336	Rab11b	0.269	Tef	0.233	Tbc1d22a	0.21
Cpne2	0.336	Pomgnt1	0.269	Chrm3	0.233	Mterf1a	0.21
Kmt2e	0.335	Emcn	0.269	Ccdc120	0.233	Ly6g6f	0.21
Ptprs	0.334	Tcf25	0.269	Msn	0.233	Sash1	0.21
Tmco1	0.334	Khk	0.269	Maf1	0.233	Brca2	0.21

Appendix C

Genes enriched in Intermediate Erythroblasts (IntE) as defined in a microarray study of human erythroid differentiation by Tusi et al. (2018).

Cluster 17 obtained from the Human Erythroblast Maturation database:
<https://cellline.molbiol.ox.ac.uk/eryth/index.html>

Gene ID	Probes	Description
ABI2	225112_at	Entrez: 10152
ACADM	202502_at	Entrez: 34
ACER3	227776_at, 222688_at, 222689_at	Entrez: 55331
ACSBG1	206466_at	Entrez: 23205
ACSL3	201660_at	Entrez: 2181
ACSL4	202422_s_at	Entrez: 2182
ADAL	239711_at	Entrez: 161823
ADCY7	203741_s_at	Entrez: 113
ADSSL1	226325_at	Entrez: 122622
AHCTF1	226115_at	Entrez: 25909
AK3	224655_at	Entrez: 50808
AKR1C3	209160_at	Entrez: 8644
ALKBH6	225969_at	Entrez: 84964
ALMS1	214221_at	Entrez: 7840
AMMECR1	204976_s_at	Entrez: 9949
AMPD3	207992_s_at	Entrez: 272
AMY2B	228023_x_at	Entrez: 280
ANGEL2	221825_at	Entrez: 90806
ANKRD10	223251_s_at	Entrez: 55608
ANKRD49	219069_at	Entrez: 54851
ANP32E	221505_at	Entrez: 81611
AP3M2	203410_at	Entrez: 10947
APOBEC3A	210873_x_at	Entrez: 200315
APOBEC3B	206632_s_at	Entrez: 9582
APOBEC3G	204205_at	Entrez: 60489
ARF6	224788_at	Entrez: 382

ARGLU1	227448_at	Entrez: 55082
ARID2	1553349_at	Entrez: 196528
ARL6IP6	225711_at	Entrez: 151188
ASB3	224524_s_at	Entrez: 51130
ASPM	219918_s_at	Entrez: 259266
ATF7IP2	228381_at, 219870_at	Entrez: 80063
ATL3	224893_at	Entrez: 25923
ATP13A3	212297_at	Entrez: 79572
ATP2C1	211137_s_at	Entrez: 27032
AURKA	208079_s_at	Entrez: 6790
AURKB	239219_at, 209464_at	Entrez: 9212
AZI2	227904_at	Entrez: 64343
BCL2L2	209311_at	Entrez: 599
BIRC6	233093_s_at	Entrez: 57448
BLM	205733_at	Entrez: 641
BRCA1	204531_s_at, 211851_x_at	Entrez: 672
BRCA2	208368_s_at	Entrez: 675
BRWD2	229694_at, 218090_s_at	Entrez: 55717
BTAF1	209430_at	Entrez: 9044
BTF3L4	238675_x_at, 225976_at	Entrez: 91408
BUB1	209642_at	Entrez: 699
BUB3	209974_s_at	Entrez: 9184
C11orf30	242847_at, 222807_at, 222806_s_at	Entrez: 56946
C11orf80	238593_at	Entrez: 79703
C12orf69	237484_at	Entrez: 440087
C14orf106	241816_at	Entrez: 55320
C15orf15	222465_at	Entrez:
C16orf52	230296_at	Entrez: 730094
C17orf85	218896_s_at	Entrez: 55421
C18orf1	209574_s_at	Entrez: 753
C18orf18	229829_at	Entrez: 147525
C18orf54	244324_at, 229442_at	Entrez: 162681
C1orf25	223404_s_at	Entrez: 81627
C21orf66	218515_at	Entrez: 94104
C22orf30	1558097_at	Entrez: 253143
C2CD3	36552_at	Entrez: 26005
C2orf30	224628_at	Entrez: 27248
C2orf49	226951_at	Entrez: 79074
C2orf64	225409_at	Entrez: 493753
C3orf23	1555906_s_at	Entrez: 285343
C3orf38	229174_at	Entrez: 285237
C4orf41	218179_s_at	Entrez: 60684

C5orf22	203738_at	Entrez: 55322
C5orf24	224876_at	Entrez: 134553
C5orf28	1557828_a_at	Entrez: 64417
C5orf34	229886_at	Entrez: 375444
C6orf153	235223_at	Entrez: 88745
C6orf89	1556359_at	Entrez: 221477
C8orf83	225603_s_at	Entrez: 286144
C9orf102	228211_at	Entrez: 375748
C9orf64	235940_at	Entrez: 84267
CAMSAP1	212710_at	Entrez: 157922
CARD8	204950_at	Entrez: 22900
CARHSP1	224910_at	Entrez: 23589
CASC4	224619_at	Entrez: 113201
CASC5	1552680_a_at, 228323_at	Entrez: 57082
CBX5	226085_at	Entrez: 23468
CCDC14	225017_at	Entrez: 64770
CCDC82	223300_s_at, 223301_s_at	Entrez: 79780
CCNE2	211814_s_at, 205034_at	Entrez: 9134
CD47	226016_at	Entrez: 961
CDAN1	228516_at	Entrez: 146059
CDC23	223651_x_at	Entrez: 8697
CDCA3	223307_at	Entrez: 83461
CDK2	204252_at	Entrez: 1017
CENPF	209172_s_at	Entrez: 1063
CENPM	218741_at	Entrez: 79019
CENPQ	219294_at	Entrez: 55166
CEP120	226449_at	Entrez: 153241
CEP152	238535_at	Entrez: 22995
CEP192	218827_s_at	Entrez: 55125
CEP72	219531_at	Entrez: 55722
CGGBP1	224599_at	Entrez: 8545
CHD1L	238070_at	Entrez: 9557
CHD6	225031_at	Entrez: 84181
CHMP2B	202536_at	Entrez: 25978
CHUK	209666_s_at	Entrez: 1147
CKAP2	1554264_at	Entrez: 26586
CNOT6	222476_at	Entrez: 57472
COMMD10	222637_at	Entrez: 51397
COPS8	202141_s_at	Entrez: 10920
COQ7	209745_at	Entrez: 10229
COX15	223281_s_at	Entrez: 1355
CP110	204662_at	Entrez: 9738

CRLS1	237502_at, 225324_at	Entrez: 54675
CRYL1	220753_s_at	Entrez: 51084
CRYZL1	1552347_at	Entrez: 9946
CSE1L	210765_at	Entrez: 1434
CSNK1G1	231920_s_at, 226888_at	Entrez: 53944
CSTF2T	212905_at	Entrez: 23283
CTTN	201059_at	Entrez: 2017
CWF19L2	228916_at	Entrez: 143884
CYTL1	219837_s_at	Entrez: 54360
DCLRE1A	209804_at	Entrez: 9937
DCLRE1C	222233_s_at	Entrez: 64421
DCUN1D4	212855_at	Entrez: 23142
DDX59	228385_at	Entrez: 83479
DEK	200934_at	Entrez: 7913
DEPDC1B	226980_at	Entrez: 55789
DNAJC6	204720_s_at	Entrez: 9829
DNAJC9	213092_x_at	Entrez: 23234
DTL	222680_s_at	Entrez: 51514
DUT	208956_x_at	Entrez: 1854
DYNC1LI1	217976_s_at	Entrez: 51143
E2F7	228033_at	Entrez: 144455
E2F8	219990_at	Entrez: 79733
EFHC1	225656_at	Entrez: 114327
EID1	211698_at, 208670_s_at	Entrez: 23741
EIF2AK2	213294_at	Entrez: 5610
EIF3F	200023_s_at	Entrez: 8665
ELAVL1	244660_at	Entrez: 1994
ELL2	226982_at	Entrez: 22936
EME1	234465_at	Entrez: 146956
ENO2	201313_at	Entrez: 2026
EPC2	242960_at	Entrez: 26122
EPS15	217886_at	Entrez: 2060
ERAP1	214012_at	Entrez: 51752
ERCC4	235215_at	Entrez: 2072
ERGIC2	226422_at	Entrez: 51290
ERI2	240603_s_at	Entrez: 112479
ESCO2	235178_x_at	Entrez: 157570
EXOC4	224926_at	Entrez: 60412
EXOSC8	215136_s_at	Entrez: 11340
FAM119A	1553743_at	Entrez: 151194
FAM122B	222673_x_at	Entrez: 159090
FAM126A	227239_at	Entrez: 84668

FAM178A	203482_at, 235590_at	Entrez: 55719
FAM29A	222685_at	Entrez: 54801
FAM40B	231880_at, 1555292_at	Entrez: 57464
FANCA	203805_s_at	Entrez: 2175
FANCB	1557218_s_at	Entrez: 2187
FANCD2	242560_at	Entrez: 2177
FANCI	213007_at, 213008_at	Entrez: 55215
FAR1	224866_at	Entrez: 84188
FBXO3	238686_at	Entrez: 26273
FBXO5	218875_s_at	Entrez: 26271
FBXW2	209630_s_at	Entrez: 26190
FEN1	204767_s_at	Entrez: 2237
FLJ11151	239135_at	Entrez:
FN3KRP	218210_at	Entrez: 79672
FUSIP1	204299_at	Entrez: 10772
GARNL1	214855_s_at, 213049_at	Entrez: 253959
GCOM1	212244_at	Entrez: 145781
GINS3	218719_s_at	Entrez: 64785
GINS4	1554356_at	Entrez: 84296
GLCCI1	225706_at	Entrez: 113263
GLTSCR2	217807_s_at	Entrez: 29997
GNA11	213766_x_at	Entrez: 2767
GNAI1	227692_at	Entrez: 2770
GNG10	201921_at	Entrez: 2790
GOLGA8E	222149_x_at	Entrez: 390535
GOLGA8F	222149_x_at	Entrez: 100132565
GOLGA8G	222149_x_at	Entrez: 283768
GOLT1B	222552_at	Entrez: 51026
GPR89A	222140_s_at	Entrez: 653519
GPR89B	222140_s_at	Entrez: 51463
GPR89C	222140_s_at	Entrez: 728932
GRINL1A	212244_at	Entrez: 81488
GTSE1	211040_x_at, 204318_s_at	Entrez: 51512
GUSBL2	232207_at	Entrez: 375513
GYPE	1561136_at	Entrez: 2996
H2AFV	212205_at	Entrez: 94239
HCCS	203745_at	Entrez: 3052
HCFC2	235264_at	Entrez: 29915
HCG4P6	215973_at	Entrez: 80868
HDAC4	228813_at	Entrez: 9759
HELLS	220085_at, 227350_at	Entrez: 3070
HIGD1A	221896_s_at, 217845_x_at	Entrez: 25994

HINT1	1555960_at	Entrez: 3094
HIRIP3	204504_s_at	Entrez: 8479
HMGA2	208025_s_at	Entrez: 8091
HMGB1	224734_at	Entrez: 3146
HMGXB4	212596_s_at, 212597_s_at	Entrez: 10042
HNRNPA2B1	225107_at	Entrez: 3181
HNRNPA3	211929_at	Entrez: 220988
HNRNPD	213359_at	Entrez: 3184
HNRNPU	235603_at	Entrez: 3192
HSD17B6	37512_at	Entrez: 8630
IFFO1	36030_at	Entrez: 25900
IFI27L1	230172_at	Entrez: 122509
IKZF1	227344_at, 227346_at	Entrez: 10320
INPP5B	213804_at	Entrez: 3633
INPP5F	203607_at	Entrez: 22876
INSIG2	209566_at	Entrez: 51141
INTS7	218783_at	Entrez: 25896
IPO7	200994_at	Entrez: 10527
IPO9	213785_at	Entrez: 55705
IREB2	225892_at	Entrez: 3658
ITFG1	1556151_at	Entrez: 81533
ITGB3BP	205176_s_at	Entrez: 23421
ITPKB	203723_at	Entrez: 3707
JAZF1	225798_at	Entrez: 221895
KIAA0101	211713_x_at	Entrez: 9768
KIAA0841	213054_at, 36888_at	Entrez: 23354
KIAA1143	226816_s_at	Entrez: 57456
KIAA1429	243927_x_at	Entrez: 25962
KIAA1524	231855_at	Entrez: 57650
KIAA1704	226429_at	Entrez: 55425
KIAA1731	215018_at	Entrez: 85459
KIAA1841	243539_at	Entrez: 84542
KIF11	204444_at	Entrez: 3832
KIF21A	226003_at	Entrez: 55605
KIF9	231319_x_at, 228429_x_at	Entrez: 64147
KLHDC5	225961_at	Entrez: 57542
KLHL2	219157_at	Entrez: 11275
KPNA2	201088_at	Entrez: 3838
KRAS	204009_s_at	Entrez: 3845
LCORL	240592_at	Entrez: 254251
LGALS8	210731_s_at, 208933_s_at, 208935_s_at	Entrez: 3964

LIG1	202726_at	Entrez: 3978
LIMK2	202193_at	Entrez: 3985
LIN7C	221568_s_at	Entrez: 55327
LIN9	235039_x_at	Entrez: 286826
LOC100128661	230293_at	Entrez: 100128661
LOC149832	228456_s_at	Entrez: 149832
LOC158402	236769_at	Entrez: 158402
LOC284926	1556064_at	Entrez: 284926
LOC285147	236166_at	Entrez: 285147
LOC642558	204299_at	Entrez: 642558
LOC645030	232932_at	Entrez: 645030
LOC728179	234512_x_at	Entrez:
LOC728860	201088_at	Entrez: 728860
LOC729082	225225_at, 225332_at	Entrez: 729082
LOC91431	1565935_at	Entrez: 91431
LOC92270	228816_at	Entrez:
LRRCC1	231872_at	Entrez: 85444
LSM8	219119_at	Entrez: 51691
LYST	210943_s_at	Entrez: 1130
MAD2L1	203362_s_at	Entrez: 4085
MALT1	210017_at	Entrez: 10892
MANEA	222805_at, 219003_s_at	Entrez: 79694
MAP3K7	206853_s_at	Entrez: 6885
MAPK14	202530_at	Entrez: 1432
MASTL	228468_at	Entrez: 84930
MATR3	228012_at	Entrez: 9782
MDM1	213761_at	Entrez: 56890
MED4	222438_at	Entrez: 29079
MED6	207078_at	Entrez: 10001
MELK	204825_at	Entrez: 9833
METTL4	232194_at	Entrez: 64863
MGC21881	228040_at	Entrez: 389741
MIB1	224720_at	Entrez: 57534
MIER3	231975_s_at, 228961_at	Entrez: 166968
MIS12	221559_s_at	Entrez: 79003
MITF	226066_at	Entrez: 4286
MKLN1	235067_at, 1560145_at	Entrez: 4289
MOBK13	202919_at	Entrez: 25843
MON2	212755_at	Entrez: 23041
MOSPD2	64883_at	Entrez: 158747
MPHOSPH9	215731_s_at	Entrez: 10198
MPPE1	213727_x_at	Entrez: 65258

MRPL39	236910_at	Entrez: 54148
MSL3	236165_at	Entrez: 10943
MSTP9	213380_x_at	Entrez: 11223
MT1X	204326_x_at	Entrez: 4501
MTBP	233436_at	Entrez: 27085
MTERFD2	226486_at	Entrez: 130916
MTR	203774_at	Entrez: 4548
MUM1	221290_s_at	Entrez: 84939
MYEF2	229464_at	Entrez: 50804
N4BP2L1	229718_at	Entrez: 90634
N4BP2L2	202259_s_at	Entrez: 10443
NAF1	229757_at	Entrez: 92345
NAG	240579_at	Entrez:
NAP1L1	208752_x_at, 204528_s_at, 212967_x_at	Entrez: 4673
NAPG	225448_at	Entrez: 8774
NCAPD3	212789_at	Entrez: 23310
NCAPG2	219588_s_at	Entrez: 54892
NCAPH	212949_at	Entrez: 23397
NCKAP1	207738_s_at	Entrez: 10787
NDC80	204162_at	Entrez: 10403
NDFIP2	224801_at, 224802_at	Entrez: 54602
NDUFA8	218160_at	Entrez: 4702
NEDD1	1560116_a_at	Entrez: 121441
NEK9	212299_at	Entrez: 91754
NFATC2IP	217527_s_at, 217526_at	Entrez: 84901
NIP30	217896_s_at	Entrez: 80011
NIPA1	225752_at	Entrez: 123606
NIPSNAP3A	224436_s_at	Entrez: 25934
NPAT	209798_at	Entrez: 4863
NR2C1	210530_s_at	Entrez: 7181
NRF1	204651_at	Entrez: 4899
NRIP1	202599_s_at	Entrez: 8204
NSL1	209484_s_at	Entrez: 25936
NUCKS1	217802_s_at, 224582_s_at	Entrez: 64710
NUDCD1	225438_at	Entrez: 84955
NUDT21	224830_at	Entrez: 11051
NXT2	209628_at	Entrez: 55916
OPA1	214306_at	Entrez: 4976
OSBP2	1569617_at	Entrez: 23762
OSBPL8	212582_at, 212585_at	Entrez: 114882
PBK	219148_at	Entrez: 55872

PCBD2	223712_at, 240846_at	Entrez: 84105
PCGF3	212753_at	Entrez: 10336
PCM1	228905_at	Entrez: 5108
PCNA	201202_at	Entrez: 5111
PCSK7	232521_at	Entrez: 9159
PDCL	204449_at	Entrez: 5082
PDE7A	223358_s_at	Entrez: 5150
PDS5A	212140_at	Entrez: 23244
PDS5B	204742_s_at	Entrez: 23047
PGM2L1	229256_at	Entrez: 283209
PGRMC2	213227_at, 201701_s_at	Entrez: 10424
PHC3	226508_at	Entrez: 80012
PHF17	225816_at	Entrez: 79960
PHF2	212726_at	Entrez: 5253
PHF20L1	230098_at	Entrez: 51105
PIAS1	217863_at	Entrez: 8554
PIGK	227639_at, 209707_at	Entrez: 10026
PIGM	235532_at	Entrez: 93183
PJA2	201133_s_at	Entrez: 9867
PLK4	204886_at	Entrez: 10733
PLRG1	227246_at	Entrez: 5356
PNKD	225298_at	Entrez: 25953
POLH	231115_at	Entrez: 5429
POLI	238992_at	Entrez: 11201
POLQ	207746_at	Entrez: 10721
POLR3GL	223269_at	Entrez: 84265
POT1	204354_at	Entrez: 25913
POU2F1	227254_at	Entrez: 5451
PPP2R1B	202883_s_at	Entrez: 5519
PPP6C	225429_at	Entrez: 5537
PRKAA1	225985_at	Entrez: 5562
PRKX	204060_s_at, 204061_at	Entrez: 5613
PRKY	204060_s_at	Entrez: 5616
PROS1	207808_s_at	Entrez: 5627
PSIP1	205961_s_at	Entrez: 11168
PSMC2	201067_at	Entrez: 5701
PTBP2	218683_at	Entrez: 58155
PTK2	241453_at	Entrez: 5747
PTPLB	212640_at	Entrez: 201562
PTS	209694_at	Entrez: 5805
PURA	204020_at	Entrez: 5813
PYCR2	224855_at	Entrez: 29920

PYROXD1	219802_at	Entrez: 79912
QKI	212265_at	Entrez: 9444
R3HDM2	203831_at	Entrez: 22864
RAB11FIP2	203884_s_at	Entrez: 22841
RAB3GAP2	202372_at	Entrez: 25782
RAB3IP	223471_at, 231399_at, 238526_at	Entrez: 117177
RAB6B	221792_at	Entrez: 51560
RAD51	205023_at	Entrez: 5888
RAD54B	219494_at	Entrez: 25788
RAD54L	204558_at	Entrez: 8438
RAI1	226143_at	Entrez: 10743
RAP1GDS1	237856_at	Entrez: 5910
RB1	203132_at	Entrez: 5925
RBBP4	244872_at, 210371_s_at, 225396_at	Entrez: 5928
RBL1	1555004_a_at	Entrez: 5933
RBL2	212331_at	Entrez: 5934
RBM17	224780_at	Entrez: 84991
RBM25	212033_at	Entrez: 58517
RBPJ	211974_x_at, 207785_s_at	Entrez: 3516
REEP5	208873_s_at	Entrez: 7905
RFC1	208021_s_at, 209085_x_at	Entrez: 5981
RFC3	204127_at	Entrez: 5983
RFC5	203209_at	Entrez: 5985
RFPL2	207227_x_at	Entrez: 10739
RFWD3	218564_at	Entrez: 55159
RMI1	218979_at	Entrez: 80010
RNASEH2B	219056_at	Entrez: 79621
RNF34	236288_at	Entrez: 80196
 RNGTT	211387_x_at	Entrez: 8732
ROCK1	213044_at	Entrez: 6093
RPA1	201528_at	Entrez: 6117
RPL10	200725_x_at	Entrez: 6134
RPRD1A	225953_at	Entrez: 55197
RPS21	200834_s_at	Entrez: 6227
RPS3A	200099_s_at	Entrez: 6189
RPS4X	216342_x_at	Entrez: 6191
RRM1	201477_s_at	Entrez: 6240
RSPH10B2	1555272_at	Entrez: 728194
RSU1	201980_s_at	Entrez: 6251
S100PBP	218370_s_at	Entrez: 64766
SASS6	231895_at	Entrez: 163786
SCLT1	1569495_at, 236487_at	Entrez: 132320

SCML2	206147_x_at	Entrez: 10389
SCYL3	205607_s_at	Entrez: 57147
SDCCAG1	218649_x_at	Entrez: 9147
SEC63	229969_at	Entrez: 11231
SENP1	226619_at	Entrez: 29843
39692	226627_at	Entrez: 23176
SESTD1	227041_at	Entrez: 91404
SFI1	36545_s_at	Entrez: 9814
SFRS1	227164_at	Entrez: 6426
SFRS11	200685_at	Entrez: 9295
SFRS12IP1	227288_at	Entrez: 285672
SFRS14	64371_at, 212000_at, 214092_x_at, 213505_s_at	Entrez: 10147
SFRS15	233753_at	Entrez: 57466
SGPP1	223391_at	Entrez: 81537
SIK2	213221_s_at	Entrez: 23235
SLC11A2	210047_at	Entrez: 4891
SLC25A21	220474_at	Entrez: 89874
SLC25A40	227012_at	Entrez: 55972
SLC25A46	226831_at	Entrez: 91137
SLC2A13	227176_at	Entrez: 114134
SLC2A3P1	221751_at	Entrez: 100128062
SLC30A7	226601_at, 226217_at	Entrez: 148867
SLC30A9	202614_at	Entrez: 10463
SLC36A1	213119_at	Entrez: 206358
SLC39A10	225295_at	Entrez: 57181
SLC7A1	212295_s_at	Entrez: 6541
SLMO2	217851_s_at	Entrez: 51012
SMC1A	201589_at	Entrez: 8243
SMC2	204240_s_at	Entrez: 10592
SMC4	201664_at	Entrez: 10051
SMC5	212927_at	Entrez: 23137
SMCHD1	212569_at	Entrez: 23347
SMEK2	226230_at, 222270_at	Entrez: 57223
SMYD4	229175_at	Entrez: 114826
SNRNP200	200058_s_at	Entrez: 23020
SNRNP48	226263_at	Entrez: 154007
SPC24	235572_at	Entrez: 147841
SPC25	209891_at	Entrez: 57405
SRBD1	219055_at	Entrez: 55133
SS18L1	213140_s_at	Entrez: 26039
STAMPB	202811_at, 227515_at	Entrez: 10617

STK11	231017_at	Entrez: 6794
STX16	221499_s_at	Entrez: 8675
STX7	212631_at	Entrez: 8417
STX8	204690_at	Entrez: 9482
SUDS3	224974_at	Entrez: 64426
SUMO1	208762_at	Entrez: 7341
SUMO2	213881_x_at, 200740_s_at	Entrez: 6613
SUMO3	200740_s_at	Entrez: 6612
SUPT7L	201838_s_at	Entrez: 9913
SVIP	230005_at	Entrez: 258010
TADA1L	225455_at	Entrez: 117143
TADA2B	236248_x_at	Entrez: 93624
TAF2	209523_at	Entrez: 6873
TAF5L	1554415_at	Entrez: 27097
TAL1	206283_s_at	Entrez: 6886
TAPT1	227407_at	Entrez: 202018
TBC1D1	227945_at	Entrez: 23216
TBC1D24	227632_at	Entrez: 57465
TBRG1	230320_at	Entrez: 84897
TCEB3	213604_at	Entrez: 6924
TCERG1	202396_at	Entrez: 10915
TCF19	223274_at	Entrez: 6941
TGOLN2	1554608_at	Entrez: 10618
TIA1	201449_at, 201448_at	Entrez: 7072
TIPRL	214773_x_at	Entrez: 261726
TLE3	212770_at	Entrez: 7090
TLK1	244389_at, 202606_s_at	Entrez: 9874
TMEM161B	238783_at	Entrez: 153396
TMEM167A	224702_at	Entrez: 153339
TMEM188	235812_at	Entrez: 255919
TMEM194A	212621_at	Entrez: 23306
TMEM41B	212623_at, 212622_at	Entrez: 440026
TMPO	203432_at	Entrez: 7112
TOP2A	237469_at	Entrez: 7153
TOP3A	214299_at	Entrez: 7156
TOPBP1	202633_at	Entrez: 11073
TPM1	210986_s_at	Entrez: 7168
TPR	228709_at	Entrez: 7175
TRA2A	213593_s_at	Entrez: 29896
TRIM10	56748_at	Entrez: 10107
TRIM33	214815_at	Entrez: 51592
TRIM45	242056_at	Entrez: 80263

TRIM73	1554250_s_at	Entrez: 375593
TRNAU1AP	222800_at	Entrez: 54952
TSNAX	203983_at	Entrez: 7257
TSPYL4	212928_at	Entrez: 23270
TTC13	219481_at	Entrez: 79573
TTC14	225180_at	Entrez: 151613
TTC17	224852_at	Entrez: 55761
TTC31	218838_s_at	Entrez: 64427
TTF2	204407_at	Entrez: 8458
TUBD1	231853_at	Entrez: 51174
TUBGCP3	215739_s_at, 1554086_at	Entrez: 10426
TXNDC10	225302_at	Entrez:
TYMS	202589_at	Entrez: 7298
UBA2	229587_at	Entrez: 10054
UBE2K	225179_at	Entrez: 3093
UBR1	226921_at	Entrez: 197131
UBR7	218108_at	Entrez: 55148
UGCGL2	235749_at	Entrez: 55757
USP1	202413_s_at	Entrez: 7398
USP14	201671_x_at, 226567_at	Entrez: 9097
USP22	216957_at	Entrez: 23326
USP32	244871_s_at	Entrez: 84669
USP37	226730_s_at, 226729_at	Entrez: 57695
UXS1	225583_at	Entrez: 80146
VANGL1	229492_at	Entrez: 81839
VDAC3	208845_at	Entrez: 7419
VEGFA	210512_s_at	Entrez: 7422
VKORC1L1	224881_at	Entrez: 154807
VPS13C	218396_at, 235023_at	Entrez: 54832
VPS13D	1557571_at, 212324_s_at	Entrez: 55187
VPS33B	218415_at, 44111_at	Entrez: 26276
VRK1	203856_at	Entrez: 7443
WAPAL	212267_at	Entrez: 23063
WDR53	227814_at	Entrez: 348793
WDR67	214061_at	Entrez: 93594
WDR68	221745_at	Entrez: 10238
WHSC1	223472_at, 222777_s_at	Entrez: 7468
WSB1	201295_s_at	Entrez: 26118
WSB2	213734_at	Entrez: 55884
XPO1	235927_at	Entrez: 7514
YEATS4	218911_at	Entrez: 8089
YIPF5	224934_at	Entrez: 81555

YME1L1	201352_at	Entrez: 10730
YWHAQ	212426_s_at	Entrez: 10971
ZBED5	218263_s_at	Entrez: 58486
ZBTB3	220391_at	Entrez: 79842
ZBTB44	225845_at	Entrez: 29068
ZCCHC10	221193_s_at	Entrez: 54819
ZFP1	226807_at	Entrez: 162239
ZFX	229022_at	Entrez: 7543
ZMYM6	227595_at	Entrez: 9204
ZNF137	207394_at	Entrez: 7696
ZNF22	218005_at	Entrez: 7570
ZNF223	207128_s_at	Entrez: 7766
ZNF238	212774_at	Entrez: 10472
ZNF25	228185_at	Entrez: 219749
ZNF264	230063_at	Entrez: 9422
ZNF274	232436_at	Entrez: 10782
ZNF302	228392_at	Entrez: 55900
ZNF33A	231864_at	Entrez: 7581
ZNF345	207236_at	Entrez: 25850
ZNF451	1556060_a_at	Entrez: 26036
ZNF468	214751_at	Entrez: 90333
ZNF519	1568873_at	Entrez: 162655
ZNF623	206188_at	Entrez: 9831
ZNF638	211257_x_at	Entrez: 27332
ZNF721	228029_at	Entrez: 170960
ZNF828	226194_at	Entrez: 283489
ZNF84	228630_at	Entrez: 7637
ZRSR2	208174_x_at, 213876_x_at	Entrez: 8233
unknown_217019_at	217019_at	Entrez:
unknown_227921_at	227921_at	Entrez:

Genes in **red** – this probe can be mapped to other genes.

Appendix D

ZZZ3 and SPT20 ChIP-seq peaks in human K562 cells.

Peaks and analysis obtained from EnrichR database are shown below.

GO Biological processes for ZZZ3 binding

Term	Overlap	P-Value	Adjusted P-Value	Z-score	Combined Score	Genes
translation (GO:0006412)	67/233	1.4E-30	1.5E-27	-1.2E+00	8.4E+01	RPL9; RPL6; RPL7; RPS15; RPS4X; MRPL3; RPL7A; RPS14; MRPL1; RPS18; RARS; RPL35; RPL38; RPL37; RPL39; RPS10; RPS7; RPS8; RPL23; RPS6; RPS3A; MRPS7; MRPS6; MRPS18C; TUFM; EEF1A1; EEF1G; MRPL51; MRTO4; RPL27; TARS2; RPL26; RPL29; RPL12; MRPS34; RPL11; IGHMBP2; MRPL16; MRPL13; RPS15A; RPS3; RPL13; RPS2; RPL18; RPL17; ABCF1; EEFSEC; RPL41; MRPL27; RPL35A; EEF2; RPL17-C18ORF32; HARS2; RPS25; RPS20; RSL24D1; RPS24; EIF4G1; RPS23
rRNA processing (GO:0006364)	62/203	5.3E-30	2.7E-27	-1.6E+00	1.1E+02	RPL5; RPL3; RPL32; RPL31; RPL34; FCF1; RPL10A; RPL8; RPL9; RPL6; RPL7; RPS15; RPS4X; FBL; RPL7A; RPS14; MRPL1; RPS18; RPL35; RPL38; RIOK1; RPL37; RPL39; RPS10; RRP15; RPS7; DIS3; RPS8; PIH1D1; RPL23; RPS6; CSNK1D; RPS3A; NSA2; MRTO4; RPL27; RPL26; RPL29; RPL12; RPL11; ISG20L2; NOL8; EXOSC4; RPS15A; BMS1; RPS3; RPL13; RPS2; EXOSC3; RPL18; RPL17; MDN1; RPL41; UTP3; RPL35A; BYSL; WDR55; RPS25; LAS1L; RPS20; RPS24; RPS23
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (GO:0000377)	30/237	0.0000	0.0013	-1.2412	13.8052	ISY1; SF3B3; DHX9; CSTF2; CSTF2T; CWC27; SNRPD2; POLR2C; RALY; SNRPD3; POLR2K; HNRNPA0; AQR; PPIL1; CPSF1; NCBP2; CPSF3; PRPF40A; LSM5; CHERP; SFPO; PRPF3; SNRNP27; RNPS1; GEMIN7; PPIL3; SNRNP200; SLU7; RBM22; SNRPB

GO Cellular component for ZZZ3 binding

Term	Overlap	P-Value	Adjusted P-Value	Z-score	Combined Score	Genes
cytosolic ribosome (GO:0022626)	52/125	6.3E-33	2.0E-30	-1.3E+00	9.9E+01	RPL5; DDX3X; RPL3; RPL32; RPL31; RPL34; RPL10A; RPL8; RPL9; RPL6; RPL7; RPS15; RPS4X; RPL7A; RPS14; MRPL1; RPS18; RPL35; EIF2D; RPL38; RPL37; RPL39; RPS10; RPS7; RPS8; RPL23; RPS6; RPS3A; MRPS18C; MRTO4; RPL27; RPL26; RPL29; RPL12; RPL11; RPS15A; RPS3; RPL13; RPS2; ZNF622; RPL18; RPL17; RPL41; RPL35A; LARP4; RPL17-C18ORF32; RPS25; RPS20; RSL24D1; RPS24; EIF4G1; RPS23
large ribosomal subunit (GO:0015934)	31/73	1.8E-20	1.4E-18	-1.5E+00	6.9E+01	RPL5; RPL3; RPL32; RPL31; RPL12; RPL34; RPL11; RPL10A; RPL8; RPL9; RPL6; RPL7; RPL7A; MRPL1; RPL13; RPL35; RPL38; RPL37; ZNF622; RPL18; RPL17; RPL39; RPL41; RPL23; RPL35A; RPL17-C18ORF32; MRTO4; RPL27; RPL26; RPL29; RSL24D1
preribosome large subunit precursor (GO:0030687)	7/27	4.7E-04	9.2E-03	-2.1E+00	1.6E+01	NSA2; PPAN-P2RY11; LAS1L; PPAN; ZNF622; MDN1; RRP15

GO Molecular function for ZZZ3 binding

Term	Overlap	P-Value	Adjusted P-Value	Z-score	Combined Score	Genes
mRNA binding (GO:0003729)	33/180	6.9E-10	2.3E-07	-1.2E+00	2.5E+01	EIF4A1; RPL5; SLBP; DDX3X; SLC4A1AP; DHX9; CSTF2; CSTF2T; MRPL13; EDC3; RPL7; PURB; RPS14; DHX33; RPS3; RPL35; RPS2; RPL41; CPSF1; RPS7; NCBP2; RPS3A; MRPS7; PUM1; THOC5; SLPI; LUC7L3; TUT1; RNPS1; RPL26; DCP1A; EIF3D; EIF4G1
rRNA binding (GO:0019843)	12/44	2.6E-06	4.7E-04	-2.1E+00	2.7E+01	RPL5; RPS14; PPAN-P2RY11; RPL12; RPL23; RPL11; MDM2; MRPL16; PPAN; RPS3; MRPS7; MRPS6
translation factor activity. RNA binding (GO:0008135)	15/68	2.9E-06	4.7E-04	-1.2E+00	1.6E+01	EEFSEC; EIF4A1; GFM1; GFM2; MTIF2; GTF2H3; TUFM; EEF1A1; EEF1G; EIF2D; EIF2B1; EIF3D; EIF4G3; ABCF1; EIF4G1

ENCODE TF ChIP-seq 2015 for combined ZZZ3 targets (top 3 hits)

Term	Overlap	P-Value	Adjusted P-Value	Z-score	Combined Score	Genes
GABPA_HL-60_hg19	594/2724	7.2E-243	5.9E-240	-1.7E+00	9.4E+02	... _
ZZZ3_GM12878_hg19	173/293	4.3E-143	8.7E-141	-1.9E+00	6.2E+02	RPL5; NUP107; TMEM41A; RPL3; SLC4A1AP; GMFB; RPL32; TACO1; RPL34; CASC4; ZDHHC6; UBE3B; PHB2; NDUFC2-KCTD14; IPO4; RPL6; RPL7; TATDN3; ALKBH3; MTBP; RPS14; FGF7; STX16-NPEPL1; ALKBH1; CDK5RAP1; RPL35; RPL38; BANF1; MAP3K8; RPL37; ZNF165; RPS7; GIN1; TYW5; RPL23; RFX3; NOSIP; CSNK1D; SUPT7L; KCTD10; ZNF837; SNRNP27; TARS2; ZNF799; RPL27; RPL26; RPL29; VPS25; ALG10B; SNRPB; RTEL1; COX16; GSTCD; SH2D2A; SHOX2; MRPL16; COPS7B; MRPL13; PHF21A; TOR1AIP1; ZMPSTE24; TMEM242; NPAS3; EIF1AD; DHX33; ZNF827; VT1A; MLEC; DMAP1; GRIA3; FZD1; RPL41; NR2F1; PUM1; RPL17-C18ORF32; MRPL21; KLHL20; MRPL30; LAS1L; COPS2; INTS5; DUS1L; NOL12; MNAT1; HNRNPA1L2; EIF4G3; BRWD3; ISY1; INTS12; C17ORF75; C11ORF31; C2ORF47; GFM2; ARHGAP1; EFCAB7; CHD2; SMG7; MRPL3; RPL7A; DSTYK; SPRED1; MRPL40; MRPL1; ZMYM3; ZNF408; AIFM2; NUF2; AP3S2; USP30; MAN2C1; TNPO3; TOP3B; COMMD2; VPS13B; NDUFC2; RPS3A; FAM126B; REXO4; MRPL44; NME1; SENP1; MED22; LSG1; NSA2; NDUFS7; NDUFS3; TUT1; ITGB3BP; SNAP47; GTF3C3; STX16; TIMM9; SEMA3A; GBA; NDUFB3; RPL11; MRPS31; IGHMBP2; CSTF2T; NOL8; HDAC8; CXXC4; NSL1; HIRA; FOXS1; RBBP5; C7ORF10; BMS1; PRRG2; KIAA1324; C11ORF49; STX6; RPS3; DCAF10; RPL17; KCNJ5; BRF2; SPAG8; NR1H2; RPL35A; KLK10; PEX13; MEIS2; SNX16; KIAA0586; FAM98B; MITD1; SSBP1; CENPP; MDGA1; PFKM; MBTPS2; FRA10AC1
BCLAF1_K562_hg19	322/1007	3.0E-169	8.2E-167	-1.8E+00	6.8E+02	... _

GO Biological processes for SPT20 binding

Term	Overlap	P-Value	Adjusted P-Value	Z-score	Combined Score	Genes
regulation of transcription. DNA-templated (GO:0006355)	26/1599	6.5E-04	1.4E-01	-1.8E+00	1.3E+01	KMT2A; RELA; HIRA; PPP3CB; SHH; IFI16; MYC; TTC21B; RPS6KA1; DMAP1; JARID2; FZD1; ZNF561; BRPF1; UBE2I; NFYB; STAT1; STAT3; ZNF76; ELF2; RYBP; KAT6A; BCOR; PKN1; RBAK; SSBP2
histone acetylation (GO:0016573)	5/72	3.2E-04	1.2E-01	-1.5E+00	1.2E+01	BRPF1; KAT6A; DMAP1; SUPT7L; TAF5L
positive regulation of nucleic acid-templated transcription (GO:1903508)	12/503	9.8E-04	1.5E-01	-2.0E+00	1.4E+01	FZD1; PPP3CB; SHH; BRPF1; ELF2; RYBP; KMT2A; STAT1; KAT6A; MYC; STAT3; RELA

GO Cellular component for SPT20 binding

Term	Overlap	P-Value	Adjusted P-Value	Z-score	Combined Score	Genes
AP-3 adaptor complex (GO:0030123)	2/7	1.4E-03	2.7E-02	-3.8E+00	2.5E+01	AP3S2; AP3S1
H3 histone acetyltransferase complex (GO:0070775)	2/8	1.8E-03	2.7E-02	-2.9E+00	1.8E+01	BRPF1; KAT6A
chromatin (GO:0000785)	8/297	3.3E-03	4.3E-02	-1.6E+00	9.1E+00	MBD3; HIRA; STAT1; MYC; STAT3; E2F4; IPO4; RELA

GO Molecular function for SPT20 binding

Term	Overlap	P-Value	Adjusted P-Value	Z-score	Combined Score	Genes
repressing transcription factor binding (GO:0070491)	6/54	5.4E-06	1.3E-03	-1.8E+00	2.1E+01	NFYB; STAT1; MYC; STAT3; DMAP1; RELA
RNA polymerase II core promoter proximal region sequence-specific DNA binding (GO:0000978)	9/263	3.5E-04	2.8E-02	-1.2E+00	9.4E+00	MBD3; IFI16; STAT1; MYC; STAT3; E2F4; SSBP2; RELA; SSBP4

core promoter proximal region sequence-specific DNA binding (GO:0000987)	9/279	5.3E-04	3.2E-02	-1.6E+00	1.2E+01	MBD3; IFI16; STAT1; MYC; STAT3; E2F4; SSBP2; RELA; SSBP4
--	-------	---------	---------	----------	---------	--

ENCODE TF ChIP-seq 2015 for combined SPT20 targets (top 3 hits)

Term	Overlap	P-Value	Adjusted P-Value	Z-score	Combined Score	Genes
SUPT20H_HeLa-S3_hg19	45/172	6.8E-56	5.5E-53	-2.0E+00	2.5E+02	SLC4A1AP; CELF3; IGHMBP2; NOC4L; UBR1; CCDC106; NT5C2; RELA; PREX1; PPP3CB; SYNGR2; OGFOD2; CTNNA1; BMF; E2F4; AP3S1; DCAF11; PUSL1; JARID2; ARID2; HNRNPA0; FZD1; PIAS3; TIGD5; TPRA1; UBE2I; RHBDD3; NFYB; STAT3; VWA5A; MRPL28; ZNF76; DDX51; ARID3A; MRPL21; ATG12; SUPT7L; EEF1A1; ACAP3; EWSR1; EEF1D; KAT6A; SNX15; RNF121; B4GALT5
SUPT20H_GM12878_hg19	64/1132	5.1E-37	2.1E-34	-1.8E+00	1.5E+02	FAM47E-STBD1; SLC4A1AP; CELF3; IPO4; PREX1; PPP3CB; NRD1; OGFOD2; AP3S1; JARID2; ARID2; TOP3B; PIAS3; TIGD5; TPRA1; PCYT1A; RPS7
STAT1_HeLa-S3_hg19	46/945	3.0E-23	8.1E-21	-1.8E+00	9.4E+01	VWA5A; DDX51; SYP; ATG12; SUPT7L; MRPL44; EEF1A1; ACAP3; AIM2; EWSR1; EEF1D; KAT6A; TUT1; RNF121; B4GALT5; IGHMBP2; NOC4L; UBR1; CCDC106; NT5C2; RELA; SYNGR2; CTNNA1; BMF; DMAP1; E2F4; DCAF11; PUSL1; HNRNPA0; FZD1; UBE2I; RHBDD3; NFYB; STAT1; KLRC4; STAT3; MRPL28; ZNF76; ARID3A; HSPE1; MRPL21; VANGL2; DPY19L4; SNX15; FAM98B; DUS1L; SSBP1

Appendix E

Validation of a CRISPR/Cas9-KAT2A epigenetic editing tool

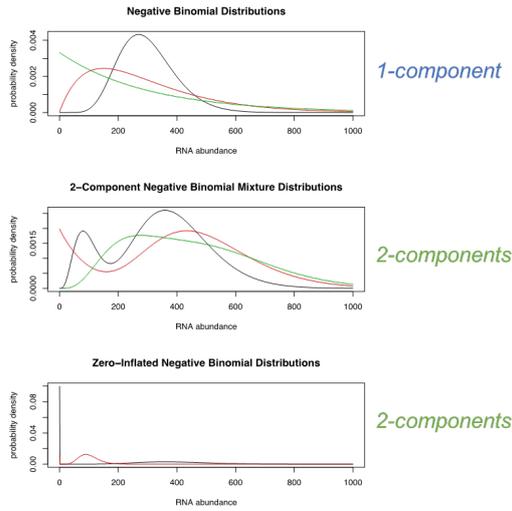
Introduction

Gene expression heterogeneity has the potential to alter fate decision probabilities in mammalian cells [323]. In the haematopoietic system, cell fate decisions associate with increased cell-to-cell transcriptional variability [43]. Our lab has shown that loss of *Kat2a* associates with enhanced transcriptional variability in a *MLL-AF9* mouse model, with reduction of leukaemia stem-like cells [260]. Furthermore, a single cell RNA-seq pilot study from the lab in Kasumi-1 cells upon KAT2A inhibition with MB-3 (unpublished) (**Fig. E-1**), revealed genes that are heterogeneously expressed following loss of KAT2A activity. Specifically, single-gene contribution to overall heterogeneity was either by the acquisition of an additional component (change from a unimodal to a bimodal distribution) in the case of *CHEK1*, *EP300* and *CEBPβ* genes, or by increased variance in a single-component, in the case of *SOX4*. Thus, it is possible that these genes play a role in leukaemic phenotypes mediated by the activity of KAT2A. However, assigning causal relationships between chromatin regulation to changes in gene expression, and ultimately, cell behaviour, remains a challenge.

The CRISPR/Cas9 gRNA system adapted from the genomes of bacteria has been widely used as a gene editing tool due to its robust, versatile and easily programmable nature [324–325]. The utility of Cas9 is further expanded with the engineering of the nuclease-null Cas9 (dCas9), which can be fused to epigenetic-effector domains for site-specific epigenome modifications [326–327]. Thus, by enabling precise manipulation of specific regulatory gene *loci*, CRISPR/ dCas9 tools may offer direct assessment of the impact of chromatin regulation and transcriptional activity to leukaemic phenotypes. In line with this view, our lab had cloned in a dCas9-KAT2A epigenetic tool that can direct KAT2A activity to putative target genes, and which I have preliminarily validated.

Single-cell RNA-seq data fits to 1 and 2-component distributions

Single-gene contribution to overall heterogeneity:



- acquisition of additional component (e.g. CEBPB, EP300, CHEK1)
- increased variability in single-component (e.g. SOX4)

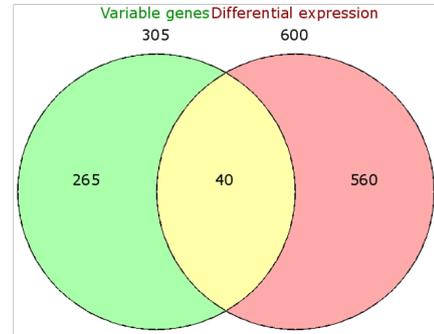


FIGURE E-1 Single-cell transcriptomics identifies a KAT2A-dependent signature in Kasumi-1 AML cells. Analysis performed by Sonya Ridden and Ben MacArthur, University of Southampton. Single-cell RNA-seq experiment performed by Cristina Pina and Mark Lynch, Fluidigm Inc.

Methods

dCas9-KAT2A-core epigenetic tool

I started by validating the editing capacity of the dCas-KAT2A tool, by directing its activity to one of the candidate target genes – *CHEK1*. To this end, I made use of a lentiviral vector previously cloned in the lab (**Fig. E-2**), in which dCas9 is fused to an expressed variant of *KAT2A* that includes its catalytic HAT domain and its reader bromodomain but that lacks the N-terminal PCAF-HD domain – here designated “KAT2A-core”.

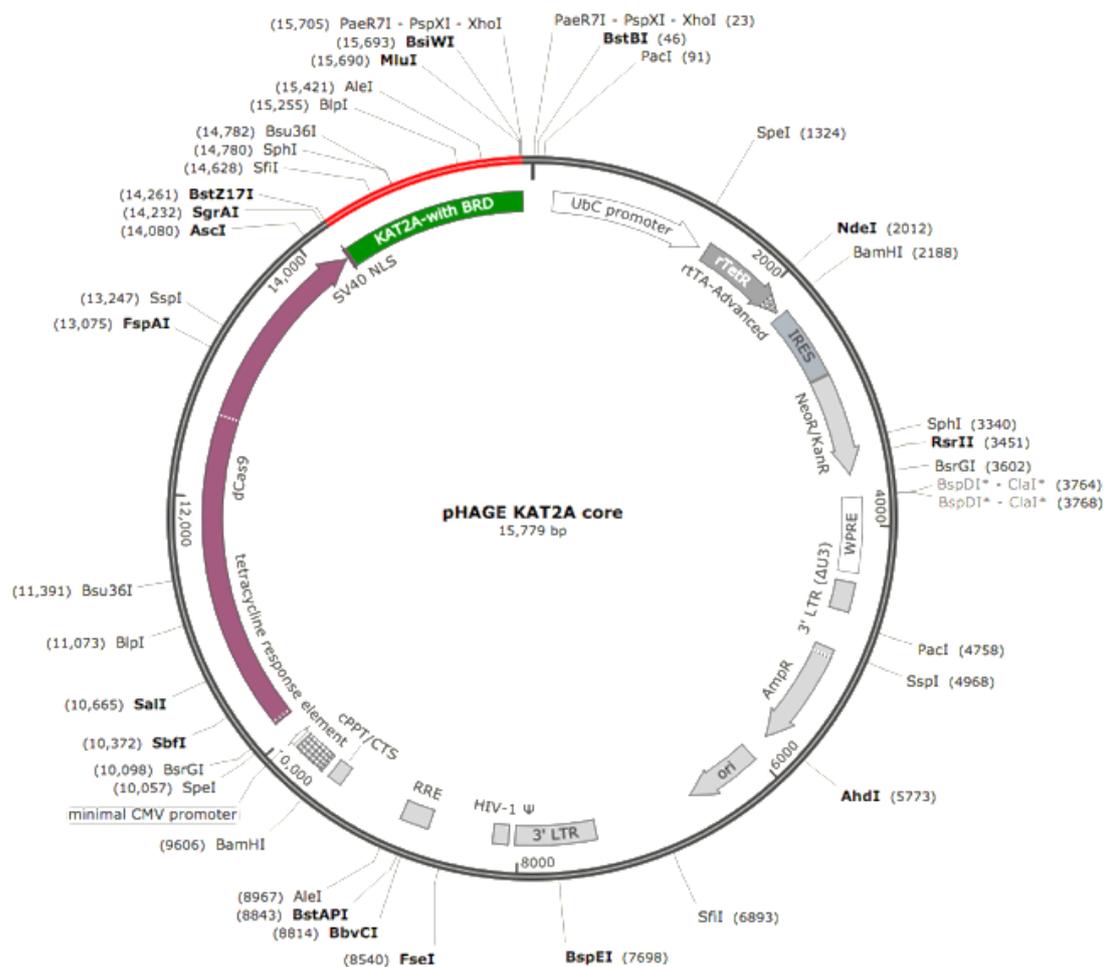


FIGURE E-2 pHAGE vector containing a dCas9-KAT2A-core fusion.

The fusion between Cas9 and KAT2A proteins was validated by WB, as shown in **Fig E-3**.

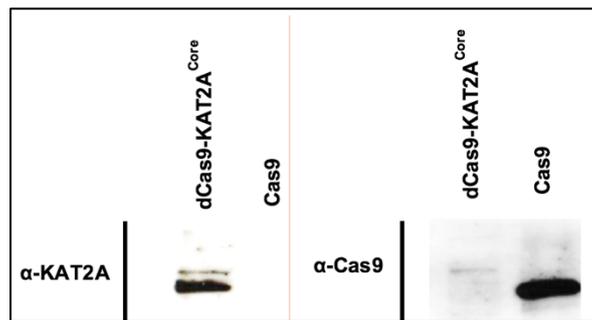


FIGURE E-3 Western blot analysis of dCas9- KAT2A-Core fusion. A band of ~230 kDa corresponding to the fusion between Cas9 and KAT2A was detected with an anti-Cas9 and an anti-KAT2A antibody in 293T cells transduced with the lentiviral vector shown in Fig. E-2 (left). A WB for detection of Cas9 in MOLM-13 cells expressing WT Cas9 (MW ~160 kDa) is shown (right). Performed by a member of the Pina lab.

Another member of the Pina lab had designed 2 sgRNA directed against the *CHEK1* promoter, which are represented in **Fig. E-4**. Single guide RNAs (sgRNAs) had been cloned into a second lentiviral vector, pKLV2 [260], represented in (**Fig. E-5**).

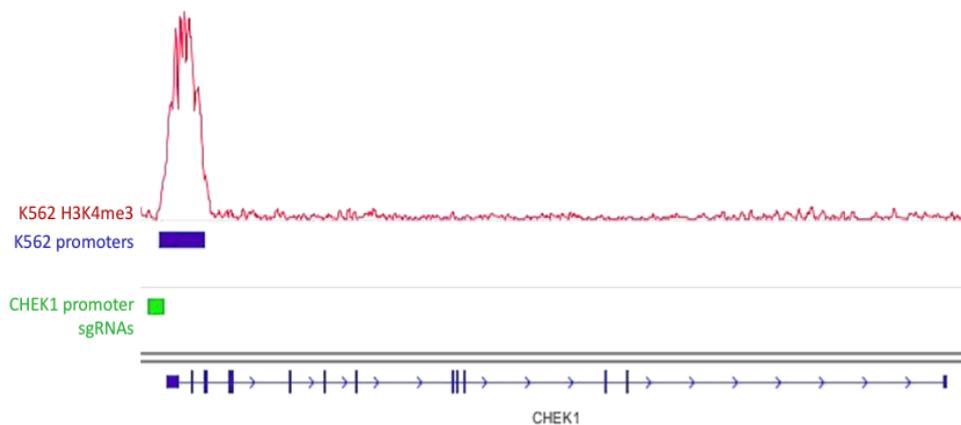


FIGURE E-4 Region encompassing the human *CHEK1* locus on chromosome 11 (125,586,692-125,701,556; GRCh38/hg38 assembly). A ChIP-seq region for H3K4me3 enrichment signal in K562 is shown in red and a known promoter in K562 cells shown in blue for mapping of the promoter region. Data generated in IGV, retrieved from ENCODE/Broad Institute. *CHEK1* promoter sgRNA target locations are indicated in green. Guides were designed using

CRISPR algorithms DNA 2.0, Zhang Lab and CHOP-CHOP. sgRNA sequences are *CTTAAATGACGTACGCAGCT* for *CHEK1* guide 115 and *GTAGAACTAAGCTAAGCAGA* for *CHEK1* guide 210.

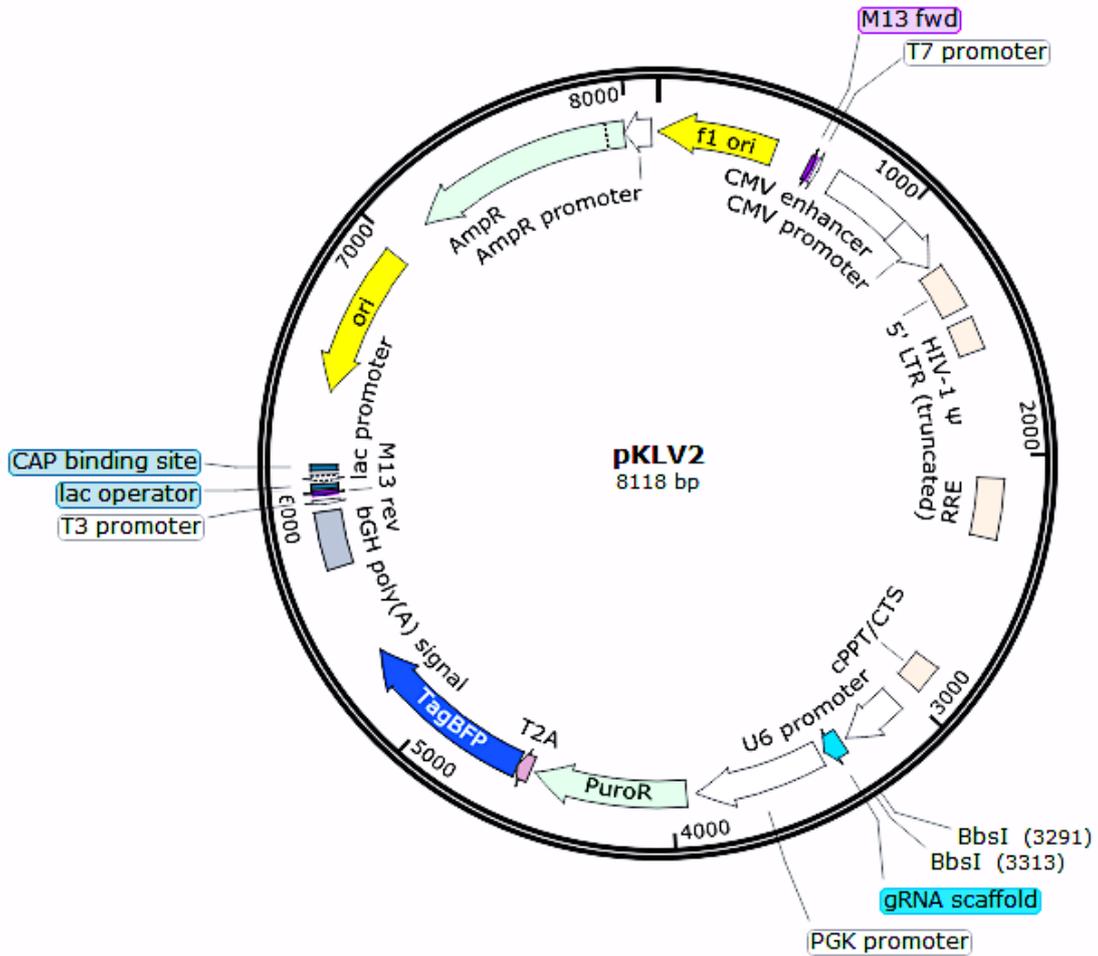


FIGURE E-5 CRISPR gRNA expression vector.

Experimental strategy

As detailed in (**Fig E-6**), for establishing Kasumi-1 cell lines expressing dCas9 and dCas9-KAT2A, cells were lentiviral transduced with the plasmid in (**Fig E-2**), as well as with the same vector without the KAT2A-core part (dCas9 only), as control. Selection for transduced cells was initiated after three days post-transduction with Geneticin (750 μ g/mL). After stable Kasumi-1 lines were established, cells were subjected to a second round of lentiviral transduction with the vector containing sgRNAs (**Fig E-6**) directed at the promoter region of *CHEK1*. As control for the presence of the guides, cells were also transduced with the vector without sgRNAs, here designated 'empty vector'. Cells were sorted for BFP positivity (successfully transduced cells), expanded, and Doxycycline (2 μ g/mL) daily administered to induce expression of the dCas9-KAT2A construct, from the start of the experiment (Day0). Cells were followed in culture for two weeks and collected at different time points for ChIP-qPCR analysis.

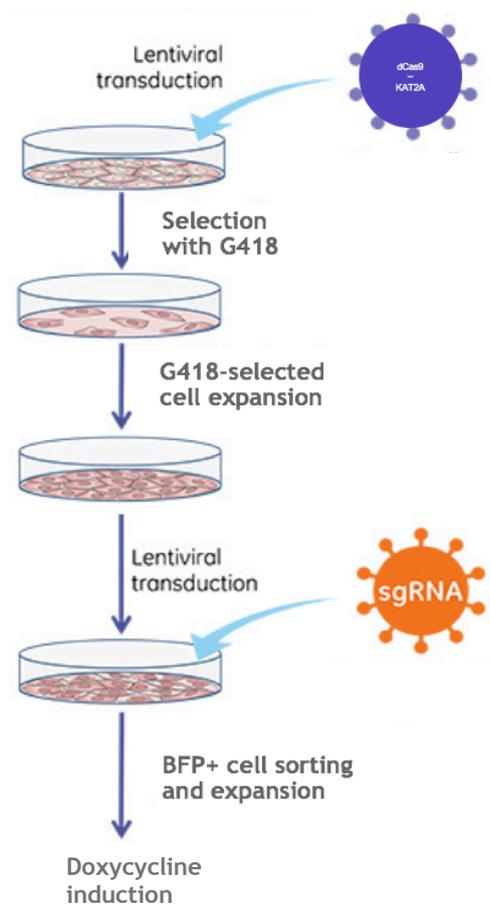


FIGURE E-6 Experimental strategy.

Results

Kasumi-1 cells transduced with lentiviral vector delivering the dCas9-KAT2A-Core fusion proliferate normally in culture

When assessing cell growth (**Fig. E-7**), no difference was observed in cells transduced with dCas9-containing vector alone, or with the dCas9-KAT2A-Core fusion vector. Similarly, the presence of the sgRNAs (*CHEK1* g115 and g210) did not seem to affect cell growth substantially, when compared to control, empty vector.

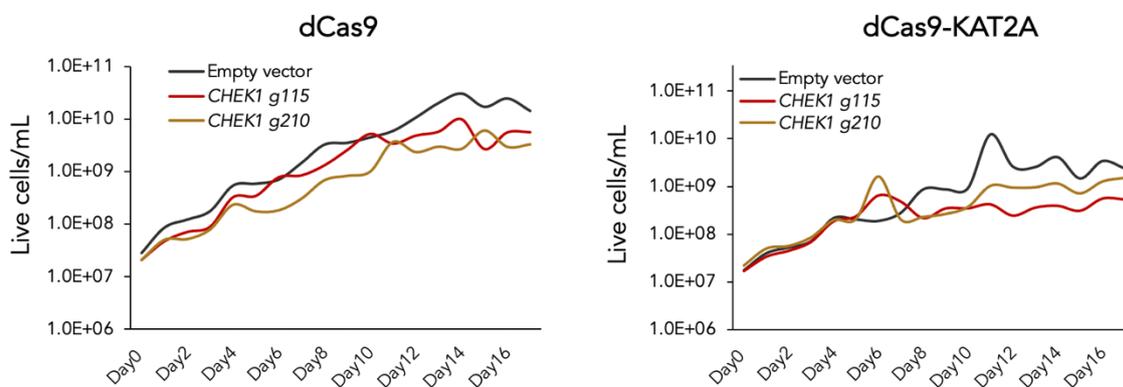


FIGURE E-7 Cell proliferation curves of *Kasumi-1* cells transduced with dCas9-KAT2A-Core fusion and with dCas9 only, together with *CHEK1* gRNAs or empty vector.

No differences in cell viability were noted in cells expressing dCas9-KAT2A, compared to dCas9 or between cells expressing *CHEK1* sgRNAs and empty vector (**Fig. E-8**).

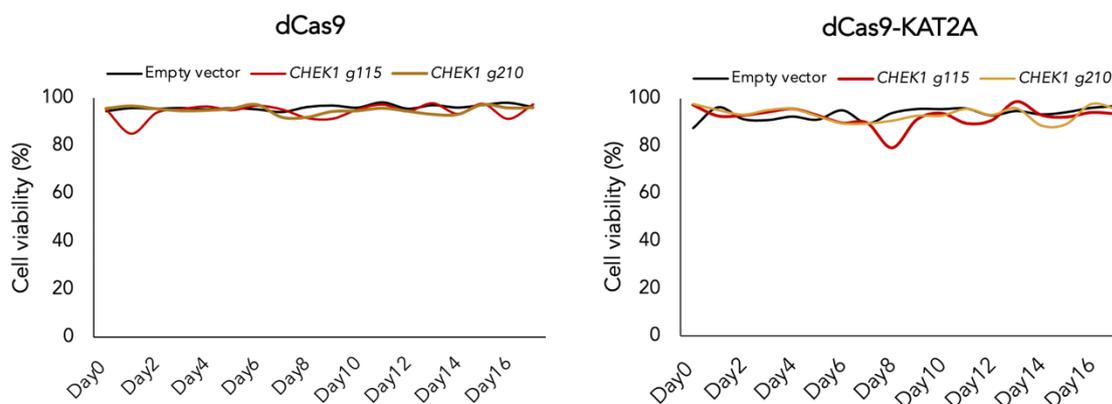


FIGURE E-8 Cell viability of *Kasumi-1* cells transduced with dCas9-KAT2A-Core fusion and with dCas9 only, together with *CHEK1* gRNAs or empty vector.

This suggests that dCas9 expression has no detrimental effect on cell proliferation or viability. Thus, any phenotype observed in sgRNA-expressing cells is likely caused by genetic perturbation of the target gene, *CHEK1*, but at least for proliferation and viability I did not observe one.

dCas9-KAT2A-Core deposits H3K9 and H3K27 at the CHEK1 promoter

In order to understand if KAT2A was indeed being directed to the *CHEK1* promoter, I assessed H3K9ac changes by ChIP-qPCR (**Fig. E-9, left**), using *E2F3* as a positive control for KAT2A histone acetylation activity based on ENCODE data. I verified an increase in H3K9ac at the *CHEK1* promoter region with *CHEK1* gRNA 115 overtime. This was not observed at the *E2F3* promoter until Day 10 (**Fig. E-9, right**). Thus, is possible that Day10 corresponds to maximal KAT2A expression, but this should be confirmed by WB at different time points. As for *CHEK1* sgRNA 210, the levels of H3K9ac mark are no different from empty vector. This may be because the sgRNA is not working, thus not directing KAT2A to the target location, or because that region is not responsive to KAT2A activity. Both hypotheses need to be studied further.

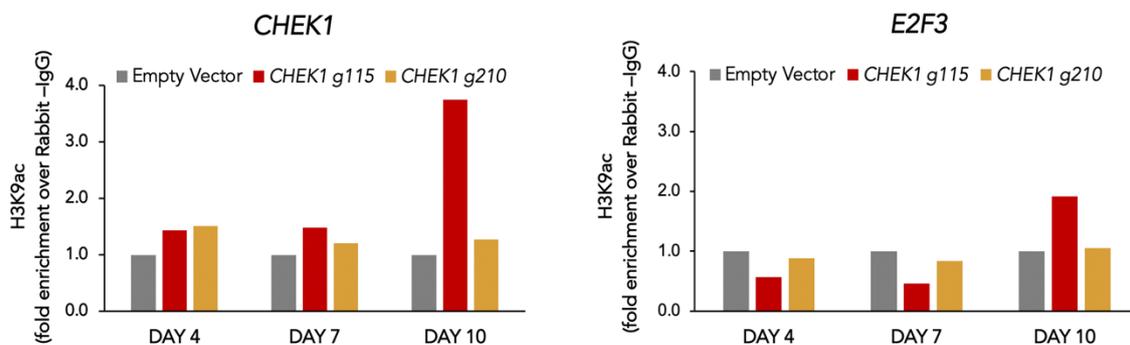


FIGURE E-9 ChIP-qPCR analysis of H3K9ac at the *CHEK1* and *E2F3* promoters. H3K9ac is shown as enrichment over Rabbit-IgG and normalised against GATA1 ($n=1$).

H3K27ac, a well-documented mark of enhancer activity, was also investigated. It was shown to be enriched at the *CHEK1* target region 115 (**Fig. E-10, left**). Therefore, it is possible that the acetylation driven by KAT2A at this location may promote proximal/facultative enhancer activity at promoters. Although, it more likely shows

promiscuity of mark deposition. Interestingly, a more robust enrichment of both H3K9ac and H3K27ac marks was observed overtime with the sgRNA at location 115.

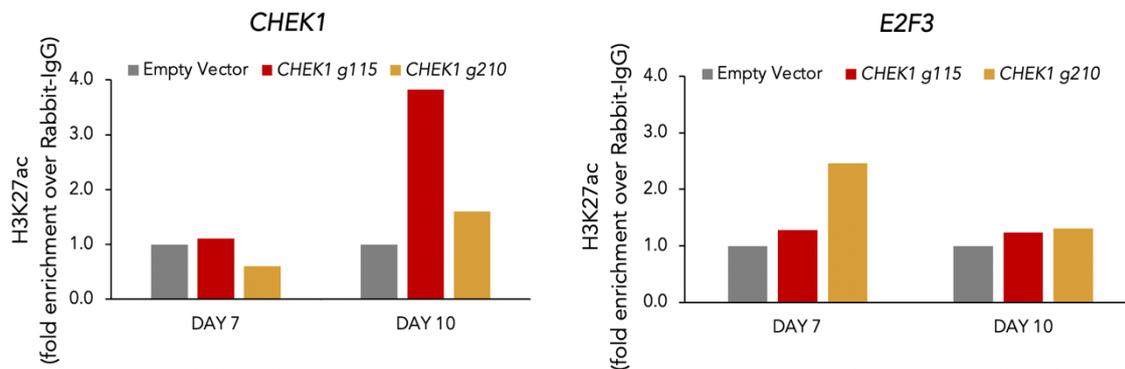


FIGURE E-10 ChIP-qPCR analysis of H3K27ac at the *CHEK1* and *E2F3* promoters.

H3K27ac is shown as enrichment over Rabbit-IgG and normalised against GATA1 ($n=1$).

These preliminary results show increased deposition of H3K9ac mark at the *CHEK1* promoter locus, deposited by KAT2A. The peak of H3K9ac should, in principle, correspond to the time point where I observed maximum d-Cas9-KAT2A expression, which may be at Day10, but this should be quantitatively measured to make firm conclusions. Failure of epigenetic remodelling at CHEK1 region 210 could be explained by (1) inefficiency of gRNA in directing KAT2A activity; (2) absence of HAT activity from the truncated KAT2A moiety of the fusion protein; (3) incorrect assembly of the endogenous KAT2A complexes with failure to recruit additional components that may be required to its acetyl-transferase activity. These points should be addressed with further experiments. In future, to employ the CRISPR tool in primary human AML samples, and screen KAT2A activity at putative target genes implicated in primary samples, including leukaemia, antibiotic-resistance cassettes should be avoided. Therefore, a vector encoding for the expression of a fluorescent reporter, instead of a neomycin resistance cassette would be desirable. Similarly, a new cloning strategy consisting of a single vector that contains both the dCas9-KAT2A fusion and the sgRNAs would avoid two consecutive rounds of transduction and selection, making the tool more versatile.