

Function of MYST3 / MOZ: A Histone Acetyltransferase and Chromatin Regulator Implicated in Haematopoietic Stem Cell Development and Leukaemia

Cíntia Júnia Monteiro, MSc.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

January 2018

Dedicated to the memory of my mum and my dad, who always supported me, whatever path I took.

Abstract

MOZ (also known as KAT6A and MYST3) is a histone acetyltransferase that is implicated in the development of hematopoietic, heart and craniofacial tissues. Reciprocal translocations generating MOZ fusion proteins are associated with acute myeloid leukemia (AML), a heterogeneous hematopoietic stem cell disorder where the myeloid cells do not enter into differentiation process, resulting in accumulation of blast cells and absence of a fully functional granulocyte lineage. Also, mutations that truncate the C-terminus of MOZ are associated with global developmental delay. Mechanistically, it is known that MOZ acts as a cofactor for certain transcription factors, particularly those with haematopoietic or neuronal specificity, such as acute myeloid leukaemia/Runt-related transcription factor 1 (AML1/RUNX1) and PU.1. However, at the outset of this project it was relatively unknown which genes that are regulated by MOZ in leukaemia. Thus, this work investigates MOZ target genes in a leukemic cell line to fully understand its importance in cell function and disease conditions.

Firstly, we performed siRNA silencing in K562 cell line, coupled with transcription profiling by RNA sequencing. We identified 406 deregulated genes under MOZ knockdown, of which 76 were downregulated. Many of the MOZ-target genes identified are implicated in haematopoietic development and cell proliferation.

Secondly, we generated MOZ/KAT6A knockout K562 cells using CRISPR CAS9n nickase technology. It allowed further analysis of the function of MOZ in leukaemia. Our results confirmed the role of MOZ in maintaining levels of H3K14ac and H3K9ac and an apparent inverse relationship with the levels of H3K14 crotonylation. Furthermore, the results indicated that MOZ is important for leukemic cell proliferation and the gene dosage influences on the cell proliferation rate.

Finally, we found that genes requiring MOZ for their expression are sensitive to a novel class of bromodomain inhibitors called iBETs. Furthermore, we found evidence for the existence of MOZ in complex with BRD4 protein. This provides a hitherto unknown link between MOZ, MLL and BRD2-4 complexes and insight into how these epigenetic regulators cooperate to control gene expression.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. David Heery. He supported me at every step over the past four years. He always made time for me whenever I needed it, and was sympathetic during times of difficulty. I feel very fortunate to have been his student.

I would also like to thank Dr. Hilary Collins for the helpful discussions and for her patience and kindness.

My sincere gratitude goes to all the members of the Gene Regulation and RNA Biology Group (past and present members), specially Bismoy, Jonathan, Barbara, Chris, Brenda and Joel for all their help, friendship and for creating an enjoyable work environment.

I must also thank the Brazilian Government (Capes) for funding my PhD project.

I would like to give special thanks to my family and friends, especially my sisters Luciana and Patrícia, my nephew Vinícius and my niece Beatriz for their love and support.

Table of Contents

Abstract ..... **III**

Acknowledgements ..... **V**

Table of Contents ..... **VI**

List of Figures ..... **XIV**

List of Tables ..... **XVIII**

List of Abbreviations ..... **XIX**

Chapter 1: Introduction 1

1.1 Transcriptional Regulation in Eukaryotes 2

1.2 Chromatin Structure and Function 12

1.3 Histone Code Hypothesis 17

1.4 Regulation of Chromatin by Histone Modifications 19

1.4.1 Histone Lysine Acetylation 20

1.4.2 Histone Methylation 22

1.4.3 Phosphorylation 24

1.4.4 Ubiquitination and SUMOylation 25

1.4.5 ADP Ribosylation 26

1.4.6 Biotinylation 27

1.4.7 Proline Isomerisation 28

1.4.8 Crotonylation and other histone acyl post-translational modifications 29

1.4.9 Histone Tail Clipping 31

1.4.10 Chromatin Remodelling 32

1.5 The Superfamily of Lysine Acethyltransferase Enzymes ... 33

1.5.1 MYST family 34

1.5.2 CBP/p300 family 37

1.5.3 GNAT family 39

1.6 Structure and Function of MOZ / MYST3 Gene 41

1.6.1 Domain organization of MOZ 41

1.6.2 MOZ is part of multi-subunit protein complexes 43

1.6.3 MOZ functions as a transcriptional co-activator 46

1.6.4 MOZ in hematopoiesis and developmental process ... 47

1.6.5 Association of MOZ with acute myeloid leukaemia and other haematological disorders 53

1.7 Aims and Objectives 63

Chapter 2 : Materials and Methods 64

2.1 Materials 65

2.1.1 General suppliers 65

2.1.2 Bacterial reagents 65

2.1.3 Molecular biology reagents 65

2.1.4 Protein chemistry, western blotting reagents and equipment 66

2.1.5 Tissue culture reagents 66

2.1.6 Assay reagents 67

2.2 Propagation, purification and manipulation of plasmid DNA ... 67

2.2.1 Composition of solutions and media used in bacterial and molecular biology procedures 67

2.2.2 Culture of *E. coli* 68

2.2.3 Preparation of competent *E. coli* 68

2.2.4 Transformation of competent *E. coli* 69

2.2.5 Long-term storage of bacterial cultures 70

2.2.6 Small-scale purification from bacterial cells 70

2.2.7 Large-scale plasmid purification from bacterial cells 70

2.2.8 Spectrophotomeric quantification of DNA 71

2.2.9 Phenol/chloroform extraction and ethanol precipitation of DNA 71

2.2.10 Ethanol precipitation 72

2.2.11 Agarose gel electrophoresis 72

2.2.12 Purification of DNA fragments from agarose gels 72

2.2.13 Restriction digest of DNA 73

2.2.14 Removal of 5' terminal phosphate groups from cleaved plasmid DNA 73

2.2.15 Addition of 5' terminal phosphate groups to oligonucleotides 74
2.2.16 Annealing of single-stranded oligonucleotides 74
2.2.17 Ligation of DNA fragments 75
2.2.18 Polymerase chain reaction (PCR) 75
2.2.19 Colony PCR screening 76
2.2.20 Sequencing of plasmid DNA 77
2.2.21 Oligonucleotides 77
2.2.22 Genomic DNA purification 77
2.2.23 Reverse transcriptase and quantitative PCR analysis 77
2.3 Cell culture 78
2.3.1 Maintenance of cell lines 78
2.3.2 Maintenance of suspension cell lines 79
2.3.3 Cryopreservation of cell lines 79
2.3.4 Resuscitation of cell lines 80
2.3.5 Calcium phosphate-mediated transfection of adherent cells 80
2.3.6 Small interfering RNA and cell transfection 81
2.3.7 Electroporation 82
2.3.8 Cell proliferation assay 82
2.3.9 Apoptosis assays 83

2.4 Biochemical techniques 84

2.4.1 Composition of solutions used in biochemical methods ... 84

2.4.2 Preparation of whole-cell extracts 86

2.4.3 Protein concentration determination 86

2.4.4 Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) 87

2.4.5 Staining of SDS-polyacrylamide gels 88

2.4.6 Western blotting and immunodetection 88

2.4.7 Co-immunoprecipitation 90

2.4.8 Small-scale bacterial expression of GST-tagged proteins 91

2.4.9 Large-scale bacterial expression of GST-tagged proteins 91

2.4.10 GST-pulldown of histone extracts 92

2.4.11 RNA Seq and data analysis 93

2.4.12 Bioinformatic analysis 94

2.5 Statistical Analysis 95

Chapter 3 : Identification of KAT6A-regulated genes in K562 cells using siRNA Knockdown 96

3.1 Background 97

3.2 MOZ silencing in K562 cells using siRNA 99

3.3 Knockdown of MOZ in K562 cells reduces acetylation of H3K14 and H3K9 103

3.4 Transcription profiling of MOZ/KAT6A depleted K562 cells..105

3.5 Detection of alternatively spliced isoforms of the MOZ/KAT6A gene.....109

3.6 Identification of MOZ–regulated genes in K562 cells 113

3.7 Validation of DEGs in K562 after KAT6A knockdown.....115

3.8 Gene Ontology analyses: Functions of MOZ regulated genes 119

3.9 MOZ is required for expression of p53-regulated genes 122

3.10 MOZ regulates expression of transcription factors and is associated to development and cell growth.....128

3.10.1 CREM.....130

3.10.2 BACH2 130

3.10.3 IFI16 131

3.10.4 The nuclear receptor genes NR4A1, NR4A2 and NR2F2 132

3.10.5 The homeobox genes PROX1 and HHEX 133

3.11 Molecular and Biological process associated to MOZ expression 134

3.12 Genes down-regulated under MOZ-silencing are involved in coenzyme A metabolic process 138

3.13 Summary 141

Chapter 4 : CRISPR-CAS9 mediated knockout of MOZ in K562 cells 149

4.1 Background.....150

4.2 Design of synthetic guide RNA (sgRNA) for generating MOZ/KAT6A knockout using CRISPR CAS9 genome editing....151

4.3 Generation of KAT6A targeting Cas9n vectors.....153

4.4 KAT6A gene editing in K562 cells using CRISPR-CAS9n....157

4.5 Validation of KAT6A genome editing events by MOZ target genes expression analysis.....166

4.6 Effect of MOZ knockdown on H3K14 crotonylation levels....171

4.7 Impact of leukemogenic protein MOZ-TIF2 on expression of MOZ target genes.....174

4.8 Effect of KAT6A ablation on the K562 cells proliferation 176

4.9 Effect of MOZ ablation on K562 cells differentiation.....183

4.10 Summary 185

Chapter 5 : IBET-151 Inhibits Expression of MOZ-Dependent Genes: Evidence for MOZ-BRD4 complexes 193

5.1 Background.....194

5.2 I-BET 151 treatment affects expression of MOZ target genes 196

5.3 IBET-151 does not inhibit the binding of MOZ DPF to acetylated H3K14 200

5.4 Detection of MOZ / BRD4 complexes in K562 cells 201

5.5 I-BET151 inhibits K562 cells proliferation and causes cell cycle arrest.....205

5.6 K562 cells undego apoptosis upon treatment with I-BET151 208

5.7 Summary 212

Chapter 6 : Concluding remarks.....219

Appendix 224

References 235

List of Figures

Figure 1.1 - Step at which eukaryotic gene expression can be regulated from DNA to active protein. 3
Figure 1.2 - Structure of the gene showing the arrangement of exons and introns.....5
Figure 1.3 - Initiation of transcription of a eukaryotic gene by RNA polymerase II. 9
Figure 1.4 - Schematic representation DNA within the chromatin structure. 15
Figure 1.5 - X-Ray structure of the nucleosome core particle at 1.9Å resolution.. 20
Figure 1.6 - Modular structure of Lysine AcetylTransferases (KATs).. 36
Figure 1.7 - Modular structure of MOZ and its interacting partners. 45
Figure 1.8 - Schematics of leukaemia-associated MOZ fusion proteins.....58
Figure 3.1 - Knockdown of KAT6A/MOZ in K562 cells.. 102
Figure 3.2 - Knockdown of MOZ protein in K562 cells. 103
Figure 3.3 - Effect of MOZ knockdown in histone modification. 104
Figure 3.4 - RNA quality analysis.. 106
Figure 3.5 - Comparision of MOZ reads between non targeting siRNA (NT) and MOZ siRNA (siMOZ) transfected K562 cells across MOZ genomic region.. 109
Figure 3.6 - Sequences of MOZ/KAT6A variants showing alternative splicing at the 5' end of the transcript.. 111
Figure 3.7 - Sashimi Plots depicting the detection of KAT6A 5' Splice variants in K562 and HEK293 RNA Seq data. 112
Figure 3.8 - PCR detection of KAT6A 5' splice variants in K562 cells. PCR was performed using cDNA prepared from K562 cells.. 113
Figure 3.9 - Heatmap representing individual expression data of the genes differentially expressed in K562 MOZ silenced cells.. 114

Figure 3.10 - qPCR validation of selected down-regulated genes in MOZ knockdown K562 cells.. 116
Figure 3.11 - qPCR validation data of selected up-regulated genes in MOZ knockdown K562 cells from RNA-Seq data. 117
Figure 3.12 - Timecourse qPCR validation data of selected downregulated-regulated genes in MOZ knockdown K562 cells from RNA-Seq data. 118
Figure 3.13 - Western blot showing expression of MOZ, NR4A1/Nur77, HHEX, NR2F2/COUP-TFII, and CDKN2A/p16 proteins in K562 cells depleted for MOZ or control after 48, 72 and 96 hours siRNA transfection..... 119
Figure 3.14 - FunDO analysis of differentially expressed genes after MOZ knockdown in K562 cells.. 122
Figure 3.15 - Venn diagram showing the common genes regulated by MOZ in K562 cells and human p53 protein. 124
Figure 3.16 - KEGG/ ClueGO enrichment analysis of upregulated genes after MOZ knockdown in K562 cells.. 126
Figure 3.17 - qPCR validation data of up-regulated HDACs genes in MOZ knockdown K562 cells from RNA-Seq data.. 135
Figure 3.18 - Inhibition of cell proliferation by MOZ knockdown.....137
Figure 3.19 - ClueGO enrichment analysis of down-regulated genes under *MOZ* knockdown.. 139
Figure 4.1 - Synthetic guide RNA (sgRNA) for KAT6A knockout.. 153
Figure 4.2 - Schematic image of pSpCas9n(BB)-2A-GFP plasmid (PX461). 154
Figure 4.3 - Eletropherogram of the sgRNAs A, B and C targeting exon 3 of the KAT6A gene in K562 cells. 156
Figure 4.4 - PCR product of genomic DNA extracted from K562 cells clones tranfected with sgRNA targeting KAT6A.. 159
Figure 4.5 - Colony PCR of ACD6 and ACF6 targeted clones.. 161
Figure 4.6 - Alignment of MOZ alleles in clones ACD6 and ACF6, relative to the WT and A only sequence..... 162

Figure 4.7 - Scheme showing the region where MOZ primers bind to MOZ transcript. 164
Figure 4.8 - qRT-PCR analysis of MOZ expression in edited KAT6A gene. 165
Figure 4.9 - Knockdown of MOZ protein in K562 cells using CRISPR-CAS9 system. 166
Figure 4.10 - qPCR validation of selected down and up-regulated genes in A only, ACD6 and ACF6 clones. 168
Figure 4.11 - qRT-PCR of MOZ target genes after MOZ rescue expression in K562 MOZ knockout cells. 170
Figure 4.12 - Effect of KAT6A gene editing in histone modification. 173
Figure 4.13 - qRT-PCR of MOZ target genes after MOZ-TIF2 transfection in K562 MOZ knockout cells. 176
Figure 4.14 - Metabolic activity of ACD6 and ACF6 clones. 177
Figure 4.15 - Cell cycle analysis of ACF6 clone. 179
Figure 4.16 - ß-galactosidase assay to verify senescence in ACD6 and ACF6 clones. 180
Figure 4.17 - Apoptosis assay using annexin-V / PI assessed for the ACD6 and ACF6 clones. 182
Figure 4.18 - Caspase activity assessed for the ACD6 and ACF6 clones. 183
Figure 4.19 - Difference in K562 cell pellet colour between WT and treated cells. 184
Figure 5.1 - Cytotoxicity of I-BET151 in K562 cells and quantification by MTS assay. 197
Figure 5.2 - qRT-PCR analysis of MOZ target genes expression after treatment with I-BET151. 199
Figure 5.3 - *In vitro* binding assay of GST-MOZ DPF and H3K14ac. 201
Figure 5.4 - Expression of co-precipitation of MOZ/BRD4 complexes in the K562 cells. 203

Figure 5.5 - Attempted co-immunoprecipitation of YFP-FLAG-MOZ constructs with BRD4 proteins in the HEK293 cells.. 205
Figure 5.6 - MTS proliferation assay performed upon K562 cells treated with I-BET151. 206
Figure 5.7 - Cell cycle analysis of K562 cells treated with I-BET151.. 207
Figure 5.8 - Apoptosis assay using annexin-V / PI assessed for the K562 cells treated with I-BET151.. 209
Figure 5.9 - Caspase activity assessed for K562 cells treated with I-BET151. 211

List of Tables

Table 2.1 - Solution compositions for resolving and stacking gels for SDS-PAGE. The volumes of the components stated are for the production of one 5 ml or 2 ml volume of resolving or stacking gel respectively. 88

Table 2.2 - Origin and Dilutions of Antibodies. Antibodies utilised during the research described in this thesis were obtained as detailed. 90

Table 3.1 - Functional Disease Ontology database (FunDO) identified different cancer-associated to down regulated genes under MOZ knockdown in K562 cells 121

Table 3.2 - Transcriptional Regulators Identified by GO Analysis in MOZ Knockdown Cells 129

Table 3.3 - Top Biological Processes Identified by GO Analysis of Genes Whose Expression Is Down-Regulated by MOZ in K562 Cells. 136

List of Abbreviations

Acetyl-CoA	Acetyl-CoenzymeA
AD	Activation domain
AdoMet	S-adenosylmethionine
ADP	Adenosine Di-Phosphate
ALL	Acute Lymphoblastic Leukaemia
AML	Acute Myeloid Leukaemia
AML1	Acute Myeloid Leukaemia Protein 1
ATP	Adenosine Tri-Phosphate
BD	Bromodomain
bp	Base Pairs
BPTF	Bromodomain and PHD domain Transcription Factor
BRD4	Bromodomain Protein 4
BRPF1	Bromodomain- and PHD Finger-containing Protein 1
BSA	Bovine Serum Albumin
Cas9	CRISPR associated protein 9
CBP	CREB Binding Protein
CGI	CpG Islands
CHD	Chromodomain, Helicase, DNA binding
ChIP	Chromatin Immunoprecipitation
CID	CBP Interacting Domain
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CREB	cAMP Response Element Binding Protein
DDR	DNA Damage Response
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribo Nucleic Acid
DPF	Double PHD Finger
DTT	Dithiothreitol
EB	Elution Buffer
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
FCS	Fetal Calf Serum
GFP	Green Fluorescent Protein
GNAT	Gcn5-related N-acetyltransferase
GST	Glutathione-S-Transferase
GTF	General Transcription Factor
H1,2A,2B,3,4	Histone 1,2A,2B,3,4
HAT	Histone Acetyltransferase
HBO1	Histone Acetyltransferase Binding to ORC-1

HBS	HEPES Buffered Saline
HDAC	Histone Deacetyltransferase
HEPES	4-(2-HydroxyEthyl)-1-PiperazineEthaneSulfonic acid
HMT	Histone MethylTransferase
Hox	Homeobox
HRP	Horseradish Peroxidase
HSC	Hematopoietic Stem Cells
ING	Inhibitor of Growth
IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISWI	Imitation SWI
KAT	Lysine Acetyltransferase
KDAC	Lysine Deacetylase
KIX	Kinase-Inducible Domain
KMT	Lysine Methyltransferase
LB	Luria-Bertani
MLL	Mixed Lineage Leukaemia
MOF	Males Absent On the First
MORF	Monocytic Leukaemia Zinc-finger Related Factor
MOZ	Monocytic Leukaemia Zinc-finger protein
mRNA	Messenger RNA
MYST	MOZ, Ybf2/Sas3, Sas2, TIP60
NAD	ADP-Ribosyl Nicotinamide
NCP	Nucleosome Core Particle
NEB	New England Biolabs
NELF	Negative Elongation Factor
NLS	Nuclear Localisation Signal
OD	Optical Density
PBS	Phosphate Buffer Saline
PCAF	p300/CBP Associated Factor
PCR	Polymerase Chain Reaction
PHD	Plant HomeoDomain
PIC	Pre-Initiation Complex
PML	Promyelocytic leukaemia nuclear bodies
PMSF	PhenylMethaneSulfonylFluoride
PTM	Post-Translational Modification
RARβ2	Retinoic Acid Receptor β2
RING	Really Interesting New Gene
RIPA	RadioImmunoprecipitation Assay buffer
RNA	Ribo Nucleic Acid
RNAP	RNA Polymerase
RPMI	Roswell Park Memorial Institute medium

rRNA	Ribosomal RNA
RT	Room Temperature
S phase	Synthesis phase
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SID	Steroid Receptor Coactivator 1 Interaction Domain
Sir	Sirtuin
siRNA	Short interfering RNA
snRNA	Small nuclear RNA
SUMO	Small Ubiquitin-related Modifier
SWI/SNF	Switching defective/Sucrose Nonfermenting
TAF	TBP Associated Factors
TAQ	Thermus Aquaticus
TBE	Tris-borate-EDTA
TBP	TATA-Binding Protein
TE	Tris- EDTA
TIF2	Transcriptional Intermediary Factor 2
TIP60	Tat Interacting Protein, 60kDa
tRNA	Transfer RNA
TSAP	Thermosensitive Alkaline Phosphatase
TSS	Transcription Start Site
UAS	Upstream Activating Sequence
UV	Ultraviolet

Chapter 1: Introduction

1.1 Transcriptional Regulation in Eukaryotes

In eukaryotic organisms, the regulation of gene expression controls cellular differentiation, proliferation and the maintenance of cellular function. The pathway leading from DNA to functional protein is regulated at many steps including transcription, RNA processing, RNA transport and localization (from nucleus to the cytoplasm), mRNA degradation, translation, and post-translational modifications (Figure 1.1). Other regulatory mechanisms such as microRNA, RNA interference and non-coding RNAs also contribute to the process of gene expression. Additionally, the accessibility of the RNA polymerase II (RNAPII) and transcription factors (TFs) to the DNA can be influenced through nucleosome positioning and chromatin modification expanding the complexity of this process [1].

Transcription is defined as the sequence of molecular events by which a gene sequence is copied into a complementary RNA copy, by an RNA polymerase enzyme. There are several RNA Polymerases in eukaryotes (described below), and protein coding genes are transcribed by RNA PolII. As the primary step in gene regulation, transcription is strictly coordinated and regulated at multiple stages. At transcription initiation, the PolII complex is recruited to the gene promoter region and begins to synthesise RNA. Following events that result in phosphorylation of C-terminal domain (CTD), the PolII complex then escapes the promoter and proceeds to transcribe the gene (transcription elongation), and finally when

transcription of the gene is complete, the PolII complex encounters the terminator region, when both polymerase and transcript are released from the DNA (transcription termination) [2].

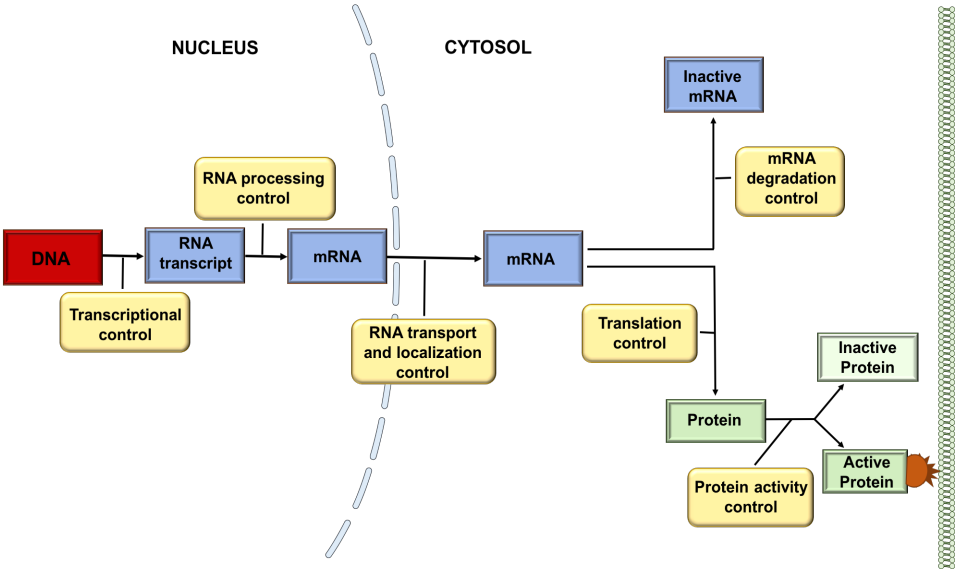


Figure 1.1 - Steps at which eukaryotic gene expression can be regulated from DNA to active protein. Transcriptional control regulates when a given gene is transcribed. RNA processing control regulates the splicing and processing of RNA transcripts. RNA transport and localization control selects which completed mRNA are exported from the nucleus to the cytosol also determines where in cytosol they are located. Translational control selects which mRNA in the cytosol are translated by ribosomes. mRNA degradation control selectively destabilizes some mRNA in the cytosol. Protein activity control selectively activates, inactivates or degrades specific protein molecules after they have been synthesized.

In higher eukaryotes, protein coding information is contained within gene exons, whereas intervening non-coding sequences separating the exons are called introns. To generate a mature messenger RNA (mRNA), intronic sequences are removed from the primary transcript by protein complexes and RNA referred to as spliceosomes (Figure 1.2). In addition, further processing events such as 5' capping of the transcript with 7-methyl guanylate and polyadenylation at the 3' end are performed.

The recruitment of PolII to gene promoters is coordinated by different regulatory sequences. The promoter region is located close to the start of the gene and may contain conserved regulatory elements such the Initiator (Inr) region, which usually overlaps the transcription start site (TSS) and downstream promoter elements (DPEs). These features help to ensure the correct positioning of the PolII complex so that transcription initiates from the correct start site. Other regulatory regions such as enhancers can be located upstream or downstream of the gene and can operate over very long distances. Experimental evidence (e.g. chromatin capture) shows that during gene expression there is likely to be juxtapositioning of the enhancers and promoters in 3D space within the nucleus. Sequence specific proteins (transcription factors- also called activators and repressor) can recognize and bind sequences within promoters and enhancers and help to regulate the accessibility and recruitment of PolII through protein-protein interactions.

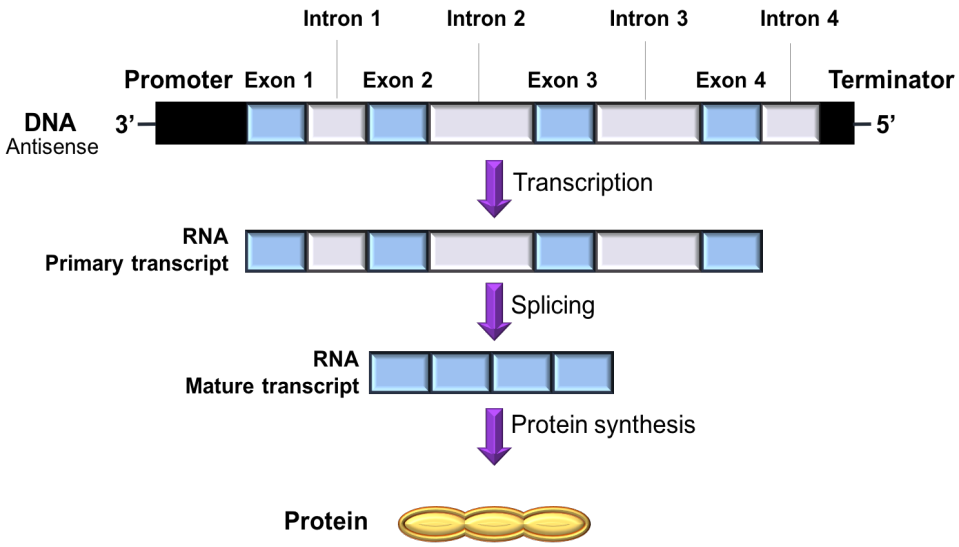


Figure 1.2 - Structure of the gene showing the arrangement of exons and introns. In eukaryotes cells, the mRNA molecule produced from transcription contains exons (coding sequences) (light blue) separated by introns (noncoding sequences) (grey). In the primary RNA transcript, the two ends of the RNA are modified, the introns are removed by splicing reaction and the resulting mature mRNA is transported from the nucleus to the cytoplasm to be translated into protein.

Following specific positive regulatory events due to the action of transcription factor (activator) complexes including histone modification and nucleosomal rearrangement, promoter DNA becomes transiently available for the interaction with the transcription machinery. The initiation starts with the binding of sequence-specific activator to the enhancer, a *cis*-acting element, which conduces successive recruitment of general transcription factors (GTFs) TFIIB, TFIID, TFIIE, TFIIF, TFIIH and the RNAPII to the target gene

promoter to form the pre-initiation complex (PIC) [3, 4]. Indeed, the classical model of stepwise PIC assembly starts with the binding of the TATA box-binding protein (TBP) of the TFIID complex to a consensus sequence TATAWAWR (W indicates A/T; R indicates A/G) of the TATA box, located around 30bp upstream of the transcription start site (TSS) in humans [5]. Studies show that the factor that binds to TATA box is functionally conserved from yeast to humans [6] and that the TBP of TFIID is enough for transcription *in vitro* and *in vivo* [7-9]. It is interesting note that some promoters are classified as TATA-less because they do not enclose TATA boxes, even though, TBP and other GTFs are placed genome-wide on the majority promoters in yeast suggesting that the general structure of the PIC is similar at TATA-containing and TATA-less promoters [10]. Although it is known that TBP can bind to TATA DNA variants [11], it is still unclear how transcription initiates at TATA-less promoters. Though all promoters require TBP for transcription, the other members of TFIID complex named TBP-associated factors (TAFs) have promoter-specific function as core promoter recognition (non-TATA elements) or coactivators [12].

Through the TAF subunits there are many possible sites for TFIID to interact with the core promoter – the TBP subunits bind the TATA box, TAF1 and TAF2 bind the Initiator (Inr) and TAF6 and TAF9 interact with the downstream core promoter element (DPE). TFIIA acts as auxiliary factor once it is not required for elemental transcription. Anyhow, TFIIA can stabilize the TBP–DNA complex by

binding to the upstream region of the TATA box and the underbody of the TBP saddle by its two conserved domains, a 4-helix bundle and a 12-stranded β-barrel [13]. TFIIB binds to TBP which facilitates the recruitment of the RNAPII to the core promoter. TFIIB recognition elements (BREu, BREd) can bind to TFIIB upstream and downstream, depending on the binding of TBP to the TATA box. Finally, the pre-transcription initiation complex (PIC) is completed by the association of TFIIE, TFIIF and TFIIH. TFIIH phosphorylates the C-terminal repeat domain (CTD) of RNA pol II facilitating the melting of DNA through helicase activity, promoter escape and the early steps of transcription [14]. RNA pol II clears the promoter and enters early elongation phase, after releasing contacts with the core promoter and the rest of the transcription machinery (Figure 1.3) [15].

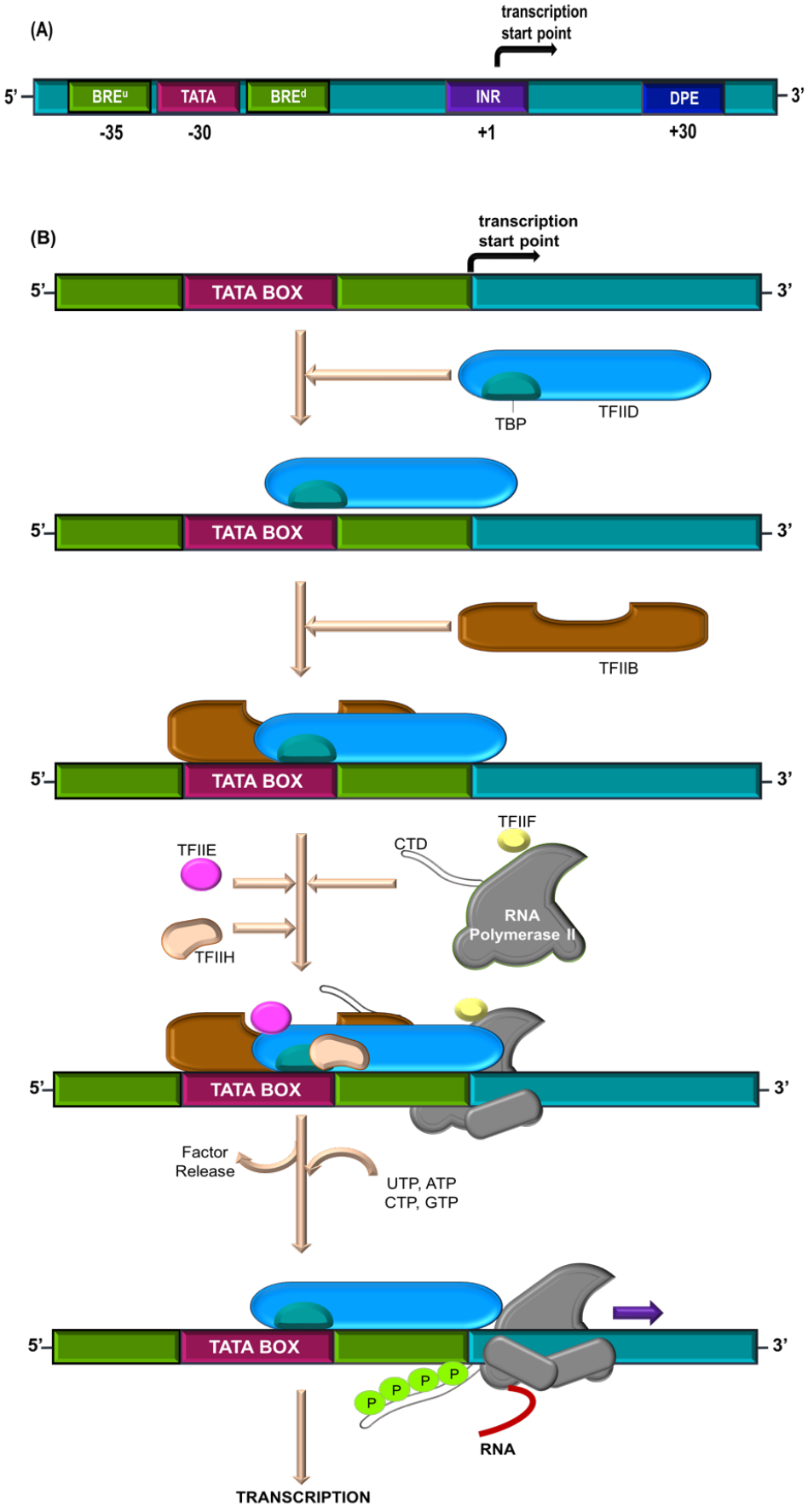

(A)
transcription start point
5'
BRE^u
TATA
BRE^d
INR
DPE
3'
-35
-30
+1
+30
(B)
transcription start point
5'
TATA BOX
3'
TBP
TFIID
TFIIB
TFIIE
TFIIH
CTD
TFIIF
RNA Polymerase II
Factor Release
UTP, ATP
CTP, GTP
P
RNA
TRANSCRIPTION

Figure 1.3 - Initiation of transcription of a eukaryotic gene by RNA pol II. (A) Consensus sequences found in the vicinity of eukaryotic RNA pol II start point. (B) Sequential recruitment of the basal transcription factors and RNA pol II are required for the initiation of transcription. Through its subunit TBP, TFIID recognizes and binds the TATA box, which then enables the adjacent binding of TFIIB. TAF1 and TAF2 binds to the Inr; and TAF6 and TAF9 to the DPE (not shown in the figure). TFIIE, TFIIH and the RNA pol II assemble at the promoter. Then, TFIIH uses energy from ATP hydrolysis to separate the DNA double helix at the transcription start site, exposing the template stand. TFIIH also phosphorylates RNA pol II, changing its conformation allowing the polymerase to release from the general factors and to starts the elongation phase of transcription. For simplicity, additional GTFs have been excluded.

In the course of elongation, nucleosomes are rearranged to allow the passage for RNAPII to transcribe throughout the entire gene sequence. Transcription terminates after the RNAPII complex has transcribed the gene, RNAPII removed from the DNA and the RNA released [16].

There are five functionally distinct RNA polymerases in the eukaryote cell nucleus: RNA polymerase I, II, III, IV and V. RNA pol I in combination with its transcription machinery acts in the nucleolus to synthesise ribosomal RNA (rRNA). RNA pol III is essential for the synthesis of short RNA molecules as such transfer RNA (tRNA) and rRNA, both of them are implicated in protein synthesis and RNA maturation [17, 18]. RNA pol IV and V synthesise short interfering RNA (siRNA) that are required for heterochromatin formation in plants [19, 20]. Lastly, RNA pol II is required for the synthesis of messenger RNA (mRNA) utilised as templates for protein synthesis and most small nuclear RNA (snRNA) used as templates for RNA

processing and microRNA required for regulation of gene activity [21].

Besides its capacity to synthesize RNA, RNA pol II is also able to induce denaturation and reannealing of double stranded DNA [22]. Regulation of RNA pol II activity underlies the control of cellular development and differentiation [23].

Regarding the transcription factors, they are known to play a key role in repressing or activating transcription throught the binding to specific sequences (enhancers and silencers) usually located in the 5' upstream region of target genes within the promoter, however, it can also bind at very distal 3' and 5' to transcriptional start sites or even within coding regions [24].

The structure of a transcription factor is important because it determines the specificity binding to DNA. The most common protein-DNA binding structure includes zinc finger motifs, helix-turn-helix, helix-loop-helix and basic zip domains [25, 26].

Transcription factors contain one or more DNA binding domains and a trans-activation domain to recruit cofactors [27]. Some transcription factors associate as homodimers or heterodimers resulting in inhibition or enhancement of transcriptional activity. This allows cross-talk between different signal transduction pathways at the level of gene expression. It also enables cells to overcome inhibition of one pathway and retain a capacity to activate specific transcription factors [24, 28].

Many transcription factors are ubiquitously expressed, whereas others are cell-specific and may determine the phenotypic characteristics of a cell [29, 30]. The activation of transcription factors is complex and involves multiple intracellular signal transduction pathways, including the kinases stimulated by cell-surface receptors [25, 26]. Some transcription factors, such as nuclear receptors, can also be directly activated by ligands as they usually have a signal-sensing domain [31]. Therefore, transcription factors may detect external signals and transmits them to the transcription complex acting as “nuclear messengers”. The activation of the transcription factors may occur within the nucleus, where usually they are already bound to DNA; or within the cytoplasm, resulting in exposure of nuclear localization signals and targeting to the nucleus [31]. It shows that post-translational modifications of these proteins by phosphorylation, acetylation, methylation have deep effects on both the DNA-binding capacity and the transcriptional activity of a transcription factor resulting in changes in gene expression profile [30].

As mentioned before, transcription is highly controlled and sequences close to protein coding genes are essential for this regulation. The CpG islands (CGIs), sequences are rich in G/C base composition and a have high density of CpG dinucleotides. CGIs surround the TSS and promoter regions of 60-70% of human genes [32]. These elements usually annotate active gene promoters, however, are refractory to epigenetic silencing by DNA methylation.

Furthermore, CGIs influence the chromatin modification status through the recruitment of the mediators CpG binding proteins [33]. In summary, genes in higher eukaryotes have a diverse and complex set of conserved *cis*-regulatory sequences which facilitate cell and context dependent regulation of their expression.

1.2 Chromatin Structure and Function

Human cells contain approximately 2 meters of DNA duplex that is packaged in an elegant way to allow it to fit in a typical cell nucleus which diameter is approximately 6 μm [34]. To accomplish this intricate function of packaging DNA, specialized nuclear proteins known as histones are needed to generate a series of coils and loops that fold the DNA tightly, however, in a way to allow it to be accessed by enzymes to replicate, repair, as well as to transcribe the DNA into an mRNA molecule. The complex consisting of DNA, histones and other chromosomal proteins is known as chromatin [35]. Chromatin consists of repeating units of nucleosome core particle so-called “beads-on-a string” that comprises 147bp of DNA wrapped around an octamer of core histones, comprising two molecules each of the histones H2A, H2B, H3 and H4 [36].

The core histones are synthesized during the S-phase of the cell cycle, their mRNAs lack introns and instead of the polyA tail, they have a special stem-loop structure which regulates their processing and stimulates translation [37]. Core histone proteins have a predominant globular structure containing a central domain that

mediates interactions within the histones, the nucleosomes and direct connects with the DNA to organize the two turns of nucleosomal DNA. Each histone also has an unstructured amino-terminal tail with 20–35 residue segment that protrudes from the surface of the nucleosome and is rich in basic amino-acids [38, 39]. These histone tails affect the structure and stability of the nucleosome, playing an essential role in the chromatin folding [40-42]. Crystal structures of mutant nucleosomes lacking the N-terminal tail of the histones H2A, H2B, H3 or H4 showed that removal of the H2B or H3 N-terminal tail affects the interaction between histone and DNA [43].

Histone tails are also common sites of post-translational modifications which alter nucleosomal interactions to disrupt the condensed chromatin architecture. Thus, the addition and reversal of histone tail modifications play essential role in gene access and, consequently, transcriptional control, as will be discussed in detail later.

Nucleosomes units are connected by a short DNA segment (approximately 20-80 nucleotides in length) with the help of the linker histone H1, that is present in the nucleosome in a half amount compared to the other four histones [44]. Histone H1 offers partial nuclease protection, stabilizes and improves higher order structures by turning nucleosomal arrays into stable 30nm structures with higher compaction and regularity [38, 45, 46]. The way in which the a "beads-on-a-string" is arranged in a condensed 30nm chromatin fiber

is still unclear due to the diverse properties of nucleosomes in native chromatin. However, two structural models, the one-start helix/solenoid model and the two-start helix model/zigzag model, were proposed based on the studies of native 30nm fibers in nuclei or isolated from nuclei [47-51]. In the solenoid model, nucleosomes are connected by linker DNA and arranged linearly in a one-start solenoid-type helix. In the zigzag model, adjacent nucleosomes are connected by a relatively straight DNA linker and are based on the zigzag back and forth arrangement in a two-start stack of nucleosomes [49, 52, 53].

In vivo, the 30nm chromatin fiber can be further condensed to form a 100-400nm interphase fiber [35]. The metaphase chromosome structures can be even more tightly compacted - the DNA is compacted 10000-20000 fold [44] (Figure 1.4).

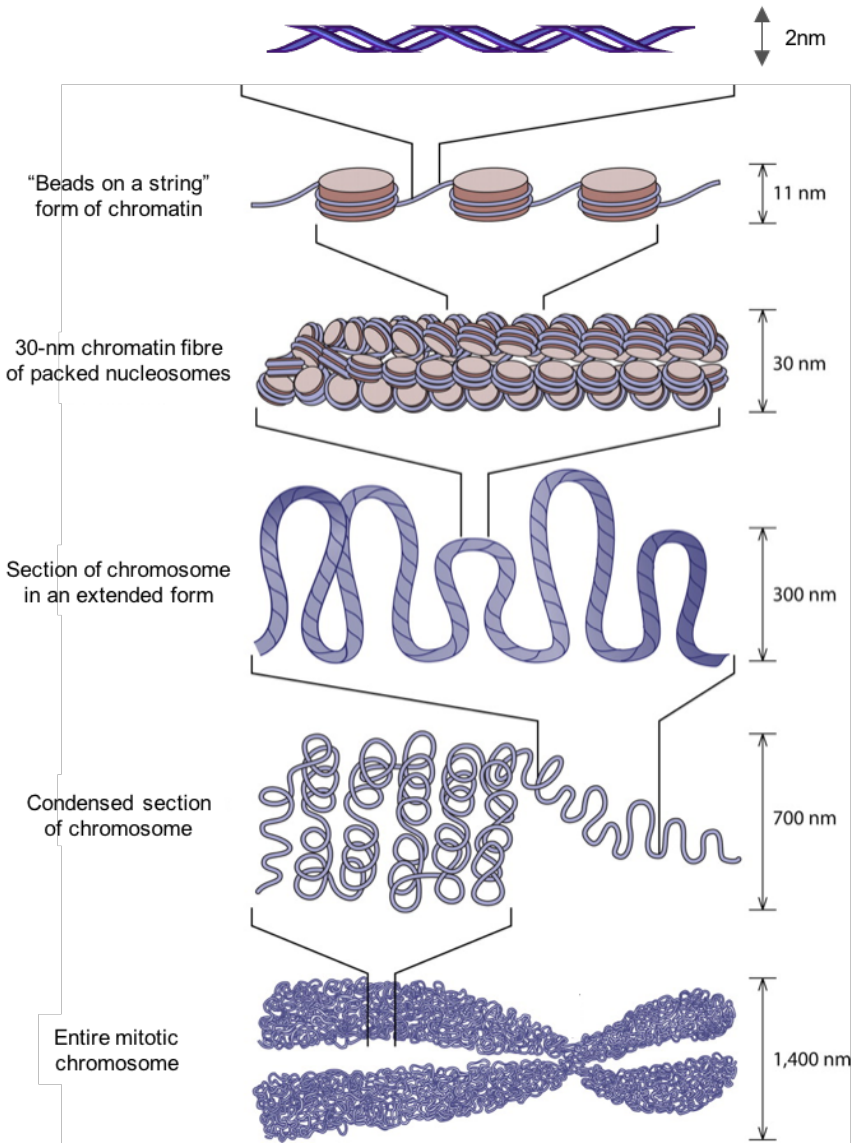


Figure 1.4 - Schematic representation DNA within the chromatin structure. The lowest level of organisation is the nucleosome, in which two superhelical turns of DNA (a total of 165 base pairs) are coiled around the outside of a histone octamer. Nucleosomes are joined to one another by short stretches of linker DNA. At the next level of organisation the string of nucleosomes is folded into a fibre about 30 nm in diameter, and these fibres are then further folded into higher-order structures. At levels of structure beyond the nucleosome the details of folding are uncertain. Figure was adapted [54].

The chromatin fiber is also composed of histone variants, the non-allelic isoforms of the canonical core histones, that play roles in transcription, chromosome segregation and DNA repair [55].

Structural studies suggested that these variants can contribute to the alteration of high-order chromatin structure by facilitating intranucleosomal interactions [36].

Differently from canonical histones, whose expression is coupled to DNA synthesis, histone variants are synthesized throughout the cell cycle and their transcripts usually have introns and polyA tails, and are typically not clustered [56]. Evolutionary analysis has suggested that the histone variants H3.3 and H2A.X are the ancestral forms of H3 and H2A respectively [57].

The variants are strategically distributed in chromatin, they are expressed in developmentally constrained or cell type-specific patterns [58]. The replacement of canonical histone with its variant changes the properties of nucleosomes and their interactions with chromatin remodellers and modifiers. For instance, the substitution of H3 with the centromeric H3 variant (or CENPA) into a nucleosome forms the foundation of centromeric chromatin [59, 60]. Many H2A variants can alter gene expression, as H2A.Z and H2A.B that are implicated in transcription initiation [61, 62] and macroH2A that seems to be associated with nucleosome immobility and transcriptional silencing [63, 64].

Histone folding, oligomerization, post-translational modification, nuclear import, stability, assembly and genomic localization are assisted by the association of histone with specific chaperone complexes [61]. Histone chaperones also prevent non-specific interactions between histones and DNA promoting specific

interactions that initiate nucleosome assembly through ATP-independent reactions [61, 65].

In eukaryotic cells, the accessibility of DNA is dependent on the packing density of chromatin fibers. Euchromatin is highly decondensed chromatin containing transcriptionally active genes that extend away from the compact chromosome region toward the interior of the interphase nucleus [66]. On the other hand, heterochromatin is tightly packaged with inactive genes located in the nuclear periphery region [66].

1.3 Histone Code Hypothesis

Histone modifications may be variously associated with gene expression or gene silencing. On the functional side, histone acetylation is usually associated with more relaxed chromatin state and active genes, whereas methylation is associated with both active and silent genes as will be discussed in detail in the next session. For example, tri-methylation of lysine 4 on histone H3 (H3K4me3) and acetylation of lysine 9 and 14 on histone H3 (H3K9/14ac) are correlated with gene activation, while deacetylation and tri-methylation of lysine 27 on histone 3 (H3K27me3) are associated with gene repression. In addition, either gene activation or repression can be affected by different modifications at the same amino acid position within a histone [67-69].

Besides the effect on nucleosomal structure, histone modifications also alter nucleosome surface which promotes binding

of specific factors and result in the recruitment or dissociation of proteins [70]. It is quite extensive the possibilities of histone modifications combinations that can lead to many functional outcomes. Thus, the “histone code” comprises the relationship between these combinations and the function they perform [67, 70, 71].

Molecular changes in normal physiological process and in specific diseases are already being determined by mass spectrometry quantification analysis of histone modifications and their combinatorial patterns [72, 73]. However, quantification of histone modification alone is not enough to determine its biological function. ChIP-seq is also an important technique used in combination to determine the localization of the changes in histone modification, although, this method relies on antibodies which is a drawback once it may not be applicable for rare or combinatorial marks. Furthermore, transcriptomic techniques as RNA-seq or qRT-PCR are also essential to assess if the gene downstream a histone modification, as H3K4me3 that is highly enriched on active promoter, is having its expression level altered [74].

Some aspects of the histone code make its interplay more similar to a language within the chromatin signaling pathway than a code: In some particular transcriptional readouts, histone modifications occur in a combination of context and time-dependent manner. Moreover, the order and mechanism by which histone modifications are added or removed are essential for the

transcription of a gene [75]. For this reason, it is crucial not only for mapping histone modification trends, but also to determine the context of the recruitment, regulation and interaction of the complexes that perform these marks to better understand the mechanisms that govern gene expression.

1.4 Regulation of Chromatin by Histone Modifications

Post-translational modification of histone was first described in 1964 by Allfrey, Faulkner [76]. The authors showed that histones are subjected to acetylation and methylation which modify their structure and affect RNA synthesis. Since then, many other histone modifications have been described, as discussed below. In 1997, Luger [36] solved the high-resolution X-ray structure of the nucleosome showing in atomic detail how the histone protein octamer is assembled and how histone modifications could affect chromatin structure. The X-ray crystal structure showed that histone amino terminal tails can pass over and between the gyres of DNA superhelix and make contact with adjacent nucleosomes (Figure 1.5). Thus, the modification of these tails can affect inter-nucleosomal interaction and, consequently, chromatin structure. Histone modifications also regulate chromatin structure by recruiting remodelling enzymes which consume energy from hydrolysis of ATP to reposition nucleosomes [68, 77]. Hence, histone modifications

can affect not only transcription but also others DNA process such as repair, replication and recombination.

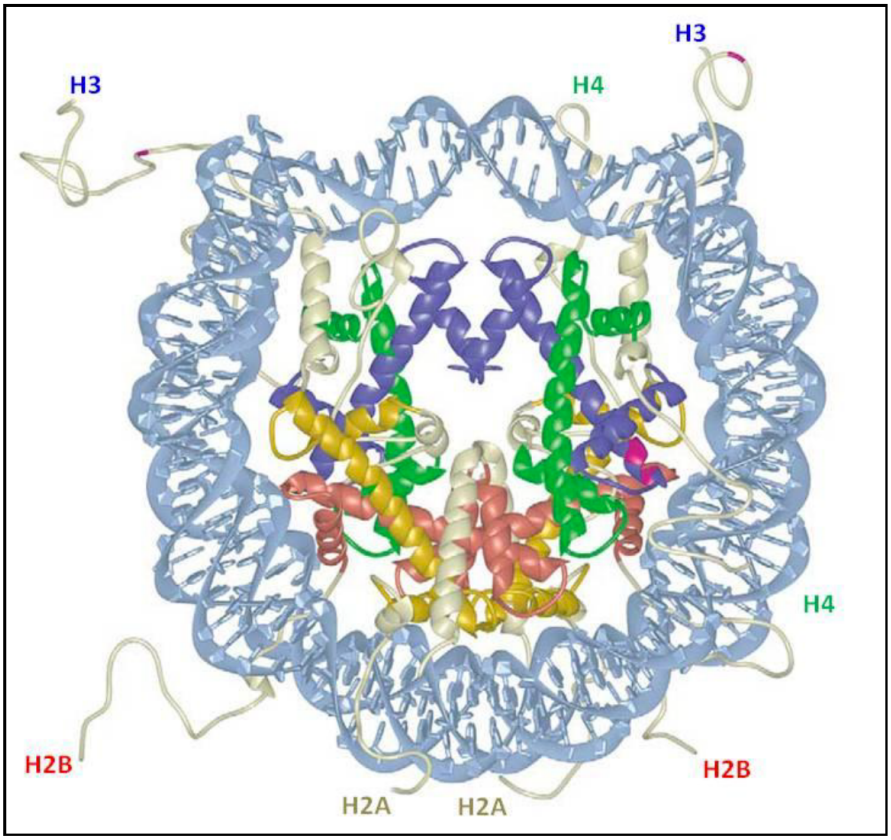


Figure 1.5 - X-Ray structure of the nucleosome core particle at 1.9Å resolution. The nucleosome core particle as viewed down the superhelical axis. The core histone octamer is composed of a central heterotetramer of histones H3 (blue) and H4 (green), flanked by two heterodimers of histones H2A (yellow) and H2B (red). The DNA helix is shown in light blue and protruding histone N-terminal tails in cream. Reprinted with permission (Elsevier publishing, licence number 2798250403120).

1.4.1 Histone Lysine Acetylation

Histone acetylation neutralizes the positive charge of lysine residues weakening of the interactions between histones and phosphate groups of DNA and facilitating the activation of

transcriptional events. Thus, histone acetylation promotes accessibility for transcriptional machinery through the reduction of internucleosomal contacts that prevent the compaction and folding of higher order chromatin structures [58, 78]. A genome-wide study demonstrated that all forms of histone acetylation are positively correlated with gene expression [79]. Despite the fact that the levels of histone acetylation are usually elevated in the promoters of active genes, H3K27ac was shown to be linked to active but not inactive enhancers [80]. Additional sites of acetylation are present within the globular histone core as in H3K56, which side chain points toward the DNA major groove, indicating the possibility of this acetylation also affect the interaction between histone and DNA [81].

Of the covalent modifications, acetylation of histones is one of the most well studied; its process is extremely dynamic and strictly regulated by histone acetyl transferases (HAT) and histone deacetylases (HDAC), also known as lysine deacetylase (KDAC).

HAT is also called Lysine (K) acetyltransferase (KAT) because its function is to transfer the acetyl-group of acetyl-CoA to the ε-amino group of an internal lysine residue [82]. Thus, the metabolic state of the cell regulates histone lysine acetylation levels and, consequently, the transcription process. Recently, it was shown that during metabolic stress when sources of acetyl-CoA are limited, which is common in tumours, nuclear acyl-CoA synthetase short-chain family member 2 (ACSS2) can form complexes with transcription activators at the promoter of lysosomal and

autophagy genes to locally support histone acetylation and activate gene expression programmes [83].

The repression of specific genes caused by HDAC activity is due not only to a direct compaction of chromatin structure but also to the binding of proteins, usually repressors, that recognize non-acetylated histones [84].

When normal activity of HATs or HDACs are disrupted, gene expression is affected and different types of cancer can be detected, such as acute myelogenous leukemia, gastric carcinomas and non-Hodgkin lymphoma [85-87].

1.4.2 Histone Methylation

Histone methylation is a reversible process that does not alter the charge of the histone protein and mainly occurs on the side chains of lysines and arginines. Lysines may be mono-, di- or tri-methylated, whereas arginines may be mono- or di-methylated. The methyl groups transferred from S-adenosylmethionine (also called AdoMet or SAM) to either lysine or arginine residues is catalyzed by histone methyltransferases (HMTs) and it is removed by histone demethylases (HDMs) [88-90].

The HMTs that catalyze lysine residues are grouped into the (Su(var)3-9, Enhancer of Zeste, Trithorax) SET domain-containing enzyme families (KMT1-3 and KMT5-7), the KMT4/DOT1 family, and others. The HDMs that catalyze lysine residues is composed of the flavin adenine dinucleotide- (FAD-) dependent monoamine oxidase

family (KDM1/LSD), the Jumonji C domain-containing demethylase families (KDM2-6), and others. The HMTs that catalyze arginine residues are protein arginine methyltransferases (PRMTs), which are categorized into types I, II, and III in mammalian cells. Type I PRMTs (PRMTs 1, 2, 3, 4, 6, and 8) catalyze the formation of monomethylarginine and asymmetric dimethylarginines. Type II PRMTs (PRMTs 5 and 7) catalyze the formation of monomethylarginine and symmetric dimethylarginines. PRMT7 also belongs to the type III PRMTs that solely catalyze monomethylarginine. HDMs for arginine residues still have not been reported [91]. However, the presence of lysine demethylases suggest it is likely arginine demethylases also exist. Lysine demethylase LSD1 acts via an amine oxidase reaction which is predicted to be compatible with arginine demethylation [92].

Histone methylation plays essential roles in many of nuclear processes, as such transcriptional regulation, heterochromatin formation and X-chromosome inactivation [93]. Depending on the sequence-specific methylation site in histones and the methylation state of the arginine or lysine residues, histone methylation can result in activation and repression of transcription [94]. For example, three methylation sites - H3K4, H3K36 and H3K79- are known to recruit activator proteins, whereas H3K9, H3K27 and H4K20 are connected to transcriptional repression. The functional effects of histone methylation are determined by both the position of the modified

residues and the number of methyl groups, also the type of protein it has evolved to recruit [95].

1.4.3 Phosphorylation

Phosphorylation of histones can occur on serine, tyrosine and threonine residues in the four nucleosomal histone tails. Histone phosphorylation can induce protein interaction [96] and also increase the electrostatic ion-pairing and hydrogen bonding capability by negatively charging phosphate moiety to the hydroxyl side chain [97]. The reversible addition of a phosphate group is performed by protein kinases and the dephosphorylation is catalyzed by phosphatases.

Histone phosphorylation has been associated with transcriptional regulation, apoptosis, cell cycle progression, DNA repair, chromosome condensation, and developmental gene regulation [68, 98-102]. Furthermore, it was suggested that phosphorylation of serine and threonine residues of the histone tails is involved in chromatin condensation during mitosis and meiosis. C-terminal phosphorylation of Thr119 in histone H2A is related to regulation of chromatin structure and function during mitosis [99], and H3 Ser10 (H3S10) phosphorylation is linked to chromatin compaction during mitosis. Nonetheless, H3S10 phosphorylation was shown to be involved in transcriptional activation of NF-κB pathway genes [68]. As for DNA repair, phosphorylation of histone H2AX at Ser139 (γ-H2AX) has been identified as one of the early events after a DNA double-strand break which assists recruitment of DNA damage repair proteins to the site affected [103, 104].

1.4.4 Ubiquitination and SUMOylation

Histone ubiquitination is a process of adding ubiquitin peptides to lysine residues [105]. Ubiquitin is a polypeptide composed of 76 amino acids and is covalently attached to lysine. The ubiquitination occurs via three-step enzymatic series: In an ATP-dependent reaction, ubiquitin activating enzymes (E1) attach ubiquitin to an E2 conjugating enzyme which acts in ubiquitin transfer. Then, E3 ligases acts as adaptors by binding to the substrate and to the E2-ubiquitin; thus they assists recognition of the substrates by facilitating isopeptide bond formation between the substrate and ubiquitin [96].

As the other modifications described before, protein ubiquitination is also a reversible process. Ubiquitin chains are removed by deubiquitinating enzymes [106]. Histones H2A and H2B are subject to monoubiquitination [107]. H2AK119 monoubiquitination is linked to transcriptional repression while H2BK120 monoubiquitination is associated with transcriptional activation, enhancement of transcriptional elongation, and induction of H3K4me2 and H3K4me3 [108].

Ubiquitin is well known to be a tag for protein turnover, thus ubiquitinated proteins usually undergo degradation via proteasomes. As research progressed, ubiquitination has been associated to other processes as such transcriptional control, DNA repair [68, 109, 110], endocytosis, cell signalling and trafficking, DNA-damage response, genomic integrity, maintainance of DNA methylation and the mediation of protein-protein interactions [96, 109, 111-113].

SUMO (small ubiquitin-like modifier) is an 11 kDa ubiquitin like protein that attaches covalently to histone lysine through the E1, E2 and E3 enzymes activity [114]. SUMO is structurally similar to ubiquitin, however, it plays different function. SUMOylation has been shown to mediate gene silencing through recruitment of histone deacetylase and heterochromatin protein 1 [115] and also to affect intracellular localization, stability and functional activity of the targets [116]. The SUMOylation process is implicated in the regulation of cell cycle and cellular differentiation. SUMOylation has been detected on all four core histones but was specifically identified on H2A, H2B and H4 [117, 118].

1.4.5 ADP Ribosylation

ADP ribosylation is a reversible post-translational modification involving the addition of the ADP-ribose moiety from β-NAD+ to an arginine/ glutamic acid residue on a target proteins by ADP-ribosyltransferases (ARTs). The majority of ARTs add single ADP-ribose by a process known as mono-ADP-ribosylation (MARylation). Nevertheless, many ARTs as such PARP1, PARP2, PARP5a and PARP5b catalyse ADP-ribose polymers [119-121].

ADP-ribosylation has been linked to cell growth and differentiation, transcriptional regulation and programmed cell death [121], however, its role in regulating DNA repair, specifically DNA strand breaks, is best-known [122-124]. The presence of double strand breaks leads to an induction of 10-500 fold in the PARP-1

activity, which induces local changes in chromatin architecture [121]. Indeed, poly-ADP ribosylated histones have been correlated with a relatively relaxed chromatin state, probably because this modification confers a negative charge to the histone [125, 126].

1.4.6 Biotinylation

Biotinylation is a reversible post-translational modification characterized by a covalent binding of the vitamin biotin to lysine residues in histones. This modification is mediated by biotinidase and holocarboxylase synthetase. More specifically, a mechanism by which this modification occurs involves cleavage of biocytin (biotin-ε-lysine) by biotinidase that leads to the formation of a biotinyl-thioester intermediate [127, 128]. Subsequently, the biotinyl moiety is transferred from the thioester to the ε-amino group of lysine in histones [129].

Biotinylation sites have been identified in histones H2A, H3 and H4 and their function role is linked to gene silencing, cellular responses to DNA damage, and cell proliferation [130]. H4K8bio and H4K12bio localize to alpha-satellite repeats in pericentromeric regions and to transcriptionally repressed chromatin loci. Thus, this modification has been shown to be associated with repression of a number of genes [131]. It may be due to the significant change in nucleosome structure caused by H4K12bio leading to a ~15% increase in the amount DNA wrapped around nucleosomes [132].

Crosstalk between biotinylation and methylation and phosphorylation was observed, for example, dimethylation of R2, R8 and R17 induces biotinylation of K4, K9 and K18 [133].

1.4.7 Proline Isomerisation

Proline isomerisation is a non-covalent modification that plays a role as a regulatory switch in signalling pathways. Peptidyl proline exists in either a *cis* or *trans* conformation, it dramatically affects the secondary structure of a polypeptide [134].

FPR4, a enzyme member of the FK506 binding proteins, has the ability to bind to the proline residues P30/38 of histone H3 in yeast. Crosstalk is also observed in transcription regulation between proline isomerization and lysine methylation, once isomerisation of P38 antagonises the methylation of H3K36 by SET2 [135].

Furthermore, it was shown that lysine acetylation influences proline isomerization and consequently local protein conformation. Acetylation of lysine 14 in the histone H3 N-terminal tail increases the proportion of neighbouring proline 16 (H3P16) in the *trans* conformation. This tail conformation was linked to reduced tri-methylation on histone H3 lysine 4 due to decreased methylation levels through the Set1 methyltransferase activity and increased demethylation levels through the demethylase Jhd2 activity.

The isomerization of peptidyl-prolyl bonds is catalyse by the parvulin (Pin1) family of enzymes in many transcription factors in a phosphorylation-dependent manner allowing the control of the

conformation, stability and activity of numerous proteins involved in transcriptional and cell cycle regulation [136].

1.4.8 Crotonylation and other histone acyl post-translational modifications

As with histone acetylation, histone lysine crotonylation has been associated with active transcription. This modification is enriched at the promoters and enhancers of active genes in human somatic cells. Crotonylation also occurs on the ε-amino group of lysine, however, it has planar orientation and crotonyl groups (C_4H_5O) have two more carbon atoms than acetyl groups (C_2H_3O) [137]. Although coactivator p300 has both crotonyltransferase and acetyltransferase activities, p300-catalysed histone crotonylation is a more potent transcriptional activator than histone acetylation [138]. The relative intracellular concentrations of crotonyl-CoA and acetyl-CoA determines whether the histone lysines will be crotonylated or acetylated [138].

A subgroup of bromodomains, as BRD9 and TAF1, can recognize lysine crotonylation but with reduced affinity compared to lysine acetylation recognition [139]. Lysine crotonylation is removed by the histone deacetylases Sirtuin 1-4 in a site-specific manner [140, 141].

Studies performed during spermatogenesis in mice to explore lysine crotonylation patterns showed that after meiosis this modification is enriched on sex chromosomes and marks testis-

specific genes, thus it was suggested that one of its roles may be to mark “escapee genes” during post-meiotic sex inactivation [142].

At least seven other types of short-chain lysine acylations have recently been identified on histones: propionylation, butyrylation, 2-hydroxyisobutyrylation, succinylation, malonylation, glutarylation, and β-hydroxybutyrylation. Studies have suggested that these histone modifications affect transcription and are structurally and functionally different from the widely studied histone lysine acetylation. They are classified into three groups based on their chemical structures in comparison to lysine acetylation: 1) hydrophobic acyl groups have extended hydrocarbon chains, which increase the hydrophobicity and bulk of the lysine residue compared to acetylated lysine. These groups include propionyl and butyryl. Crotonyl group is also included in these groups. 2) Polar acyl groups present a hydroxyl group that enables the modified lysine to form hydrogen bonds with other molecules. This group is composed of β-hydroxybutyryl and 2-hydroxyisobutyryl. 3). Negatively charged acidic acyl groups, which interestingly change the charge at the lysine residue from +1 to −1 at a physiological pH level. These groups include malonyl, succinyl, and glutaryl [143].

Specific non-acetyl acyls writers have not yet been identified; however, some histone acetyltransferases have been shown to use distinct acyl-CoAs as substrates for histone lysine acylation [144, 145].

1.4.9 Histone Tail Clipping

Histone tail clipping is an irreversible change involving proteolytic processing of histones that consists of the removal of the histone N-terminal tail by the histone H3 endopeptidase activity. It was first identified in *Tetrahymena* in 1980 [146], where the first six amino acids of H3 are removed. As research progressed, this modification has been observed in all histones in many organisms. In human cells, histone H3 clipping has been reported in some cells including embryonic stem cells [147], senescent fibroblasts and melanocytes [148], peripheral blood mononuclear cells [149] and hepatocytes [150]. N-terminal tail of H3 can be cleaved in distinct sites and in hESC cells it is mediated by a serine protease [147]. In models of oncogene-induced and replicative senescence, proteolytic clipping of histone H3 was performed by cathepsin L, a cysteine protease. Thus, these findings suggests that cleaved products of H3 may play a role in regulation of genes involved in cell cycle promotion, DNA replication, cellular proliferation, apoptosis and migration in senescence [148].

Clipping has also been involved in expulsion of repressive marks and proteins at promoters during the induction of gene expression [151].

1.4.10 Chromatin Remodelling

Apart from these chemical modifications involving covalent interactions described above, physical edits to the chromatin structure by alterations in the location or conformation of the chromosomes are also observed and are carried out by chromatin remodelers [152]. These chromatin remodelers are large multiprotein complexes that use the energy of ATP hydrolysis to alter the physical conformation of nucleosomes around the DNA [153, 154]. Remodelers are divided into four families: the Switching defective/Sucrose Nonfermenting (SWI/SNF), Imitation SWI (ISWI), Chromodomain, Helicase, DNA binding (CHD) and Inositol requiring 80 (INO80) remodelers, each with specialized biological roles. Nonetheless, all remodelers contain a subunit with a conserved ATPase domain. Evidence from the SWI/SNF and ISWI remodeler families revealed that remodeler ATPases are directional DNA translocases. The ATPase binds approximately 40 base pairs inside the nucleosome, then it pumps DNA around the histone-octamer surface which enables the movement of the nucleosome along the DNA. It permits the exposure of the DNA to regulatory factors [152]. The other domains that are attached to the ATPase are important for nucleosome selection and support the regulation of ATPase activity. Their binding to histones and nucleosomal DNA is affected by the histone modification state that determines whether the nucleosome is an appropriate substrate for a remodeler complex [152, 155].

1.5 The Superfamily of Lysine Acethyltransferase Enzymes

Lysine acetyltransferases (KATs), also known as histone acetyltransferases (HATs), are a large group of enzymes evolutionarily conserved across eukaryotes. They are found in the nucleus, where they function to regulate gene expression but also are located in the cytoplasm and mitochondria where they play an important role regulating protein activity in processes including metabolism, cell migration, cancer and neurodegenerative disorders [144, 156].

The KAT family consist of 17 known members, as defined by the HUGO gene nomenclature committee and are classified in 3 main families based on structural homologies: Gcn5 N-acetyltransferases or GNATs; MYST; and a class of HAT enzymes, including p300/CBP and Taf1, which do not possess a consensus HAT domain. Variation between specific members is due to the presence of further structural features, such as zinc fingers, PHD fingers and chromodomains. These additional domains provides an insight into the substrate specificity of the family members [144].

KATs catalyse the addition of an acetyl moiety from an acetyl-CoA molecule to the ε-amino group of a protein lysine residue. The enzymatic activity is regulated in many different ways including post-translational modification, degradation or alteration of subcellular localisation [157]. Most KATs are physiologically found as part of a multisubunit complex, which are generally more active compared to

their counterpart enzymes. Additionally, the complexes have distinctive substrate targeting, demonstrating the importance of the non-catalytic proteins in the regulation of activity and substrate specificity [144, 157].

It has been shown that multiple KATs co-occur simultaneously to a similar set of loci in the genome, suggesting a low degree of specificity in their function [158].

1.5.1 MYST family

The MYST family is one of the most studied groups of HATs protein because of their broad conservation, varied functions and association with disease. They play key roles in post-translational modification of histones and are deeply involved in chromatin structure, transcriptional activation, DNA repair, development, transcriptional silencing and cell cycle progression [159]. The name MYST (MOZ/Morf, Ybf2, Sas2, and Tip60) is due to the presence of the conserved MYST domain that consists of an acetyl-CoA binding motif and a zinc finger. Plant homeodomain-linked (PHD) zinc fingers, chromodomains, and others are also present in some proteins of this family [157, 160] (Figure 1.6).

Although MYST proteins share a structurally conserved acetyl coenzyme A (acetyl CoA) core-binding region with other HATs, they use unique structural and catalytic properties within their conserved HAT domain. The MYST HAT domains contain N- and C-terminal regions that flank the core-binding region that are structurally

different compared to other HAT families. These include motifs required for chromatin regulatory activities as N-terminal C_2HC zinc-binding region and C-terminal helix-turn-helix [159, 161]. To perform the catalysis activity, the MYST proteins have been shown to employ a ping-pong catalytic mechanism in which the general base glutamate first deprotonates the active site cysteine to allow the transfer of acetyl group from acetyl CoA to the cysteine. The glutamate then protonates the leaving cofactor and deprotonates the substrate lysine so that the cysteine can transfer the acetyl group to the lysine. To return to their initial state, glutamate and cysteine react with water [161]. Recently, it was found that MYST catalytic activity requires active lysine autoacetylation site for cognate substrate acetylation [162].

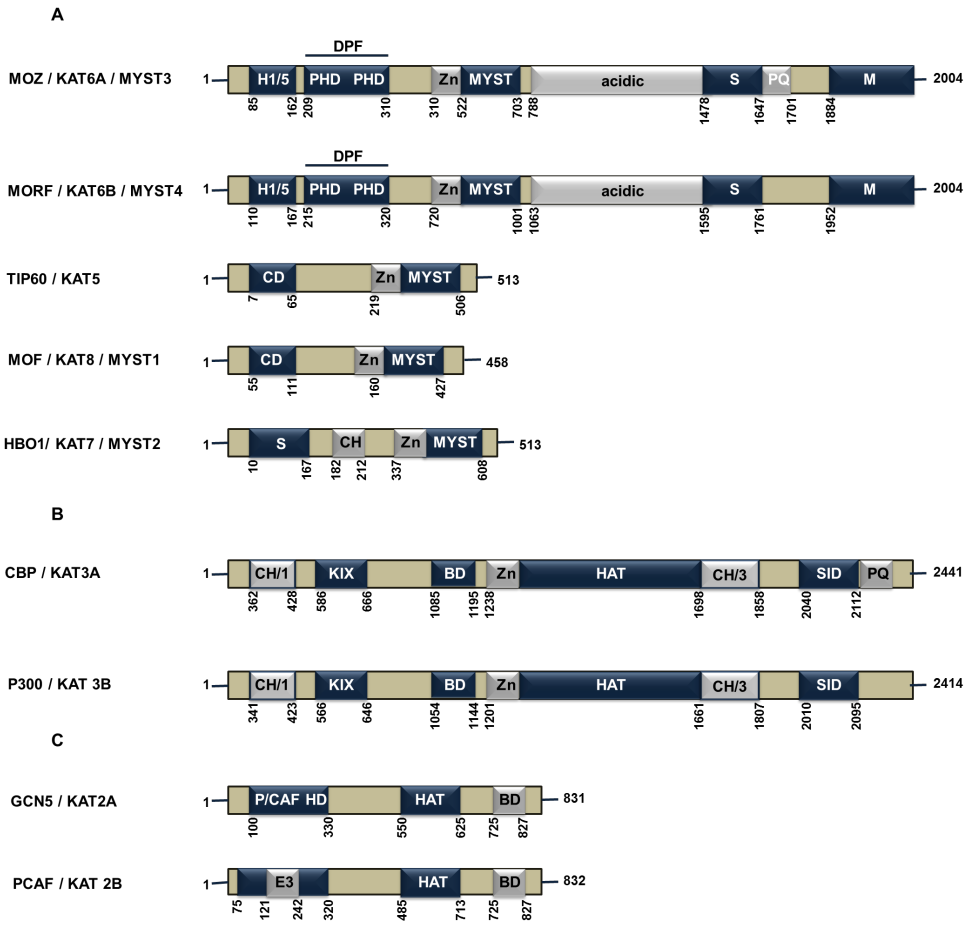


Figure 1.6 - **Modular structure of Lysine AcetylTransferases (KATs).** Domain organisation of lysine acetyltransferase enzyme families (A) MYST (B) CBP/p300 (C) GCN5/ PCAF from humans, indicating highly conserved regions and functional domains. Domains are labelled as follows: H1/5, histones H1- and H5-like domain; PHD, plant homeodomain; DPF, double PHD finger; Zn, C2HC zinc finger; MYST, (MOZ, Ybf2/Sas3, Sas2, TIP60) homologous acetyltransferase region; acidic, glutamate/aspartate-rich region; S/M, serine/ methionine-rich region; PQ, proline/glutamine-stretch; CD, chromodomain; CH, cysteine/histidine rich motif; KIX, CREB binding region; BD, bromodomain; SID, steroid receptor coactivator 1 interaction domain; P/CAF HD, P/CAF homology domain; E3, E3 ligase domain.

Because MYST family are involved in many important nuclear processes, mutations of genes encoding members of this family are associated with cancer and others diseases [159, 163, 164]. A more

complete understanding of chromatin regulation is therefore of fundamental importance to allow development of assays to predict or detect disease at earlier stages, as well as new efficacious treatments.

The development of some MYST inhibitors, as such a bisubstrate CoA-peptide has been reported. It has the advantage to be more potent than the natural product HAT inhibitors as anacardic acid and curcumin, however, it exhibited only micromolar IC50 values and poor selectivity for the MYST proteins vs other families of acetyltransferases [165]. Therefore, it is still important to identify a MYST inhibitor with higher selectivity and more potent pharmacokinetic properties.

1.5.2 CBP/p300 family

CBP/p300 family of acetyltransferases is composed of the two members: the human protein p300 and its homologue CREB-Binding Protein (CBP). The nuclear CBP was first identified as an interacting partner of the activated phosphorylated form of the CREB (cAMP response element binding protein) transcription factor [166]. Whereas p300 was isolated as a cellular factor bound by the adenovirus E1A protein [167]. Both proteins are ubiquitously expressed in all higher eukaryotes but not in lower eukaryotes such as yeast [168]. CBP and p300 share 91% sequence identity and are thought to be functionally equivalent [169].

The p300/CBP acetyltransferase is an important transcriptional coactivator involved in several biological processes

including DNA transcription, DNA repair development, innate immune response and cell cycle regulation [170, 171]. It also have been implicated in other cellular processes not directly linked to chromatin, such as the regulation of p53 turnover in cycling cells [172] and nuclear import [173].

Interestingly, as a result of the high versatility and the broad specificity of p300/CBP, these coactivator are usually used in a competitive way by different pathways and represents a limiting factor for transcriptional activity because the cellular demand for this acetyltransferase is very high [107].

With regard to transcriptional activity, p300/CBP can activate transcription through acetylation of histone or non-histone proteins. p300/CBP can function as a multivalent scaffold to recruit other cofactors or to allow the assembly of multiprotein complexes or it can also act as a bridge to connect sequence-specific transcription factors to the components of the basal transcriptional machinery [170].

CBP and p300 functions are executed through different conserved domains, particularly the N- and C-terminal regions act as transactivation domains and contain modules for protein interaction [168, 170]. With regard to the p300/CBP HAT domain, it is comprised of the coenzyme A (CoA) binding and catalytic site, and zinc finger. Although it has structural conservation within the central core region associated with CoA binding compared to other KATs, the structural elements flanking this central region are divergent from other HATs

(Figure 1.6). For instance, p300 present a long loop intimately associated with the CoA portion of the LysCoA inhibitor and contributes to one-third of the protein-cofactor interactions. Also, p300 has a shallow and more electronegative substrate binding site [174]. This is consistent with the binding properties of p300 HAT that prefers substrate with multiple positive charges, compared to other HATs [169]. Furthermore, p300 HAT is able to acetylate the four histones (*in vivo* and *in vitro*) and many non-histone proteins, such as transcription factors [3, 106].

1.5.3 GNAT family

The term GNAT was first used to describe *N*-acetyltransferases derived from the yeast GCN5, which was shown to be a histone acetyltransferase [175, 176]. The GNAT superfamily is found in eukaryotes, prokaryotes, and archea [177].

The related GCN5 acetyltransferase subclasses GCN5 and PCAF were described in mammals and are considered global coativators [58]. Both KATs are considered indispensable global coactivators for normal cellular function. In vertebrate, GCN5 and PCAF proteins are very similar with approximately 73% similiarity, most of it focused into three highly homologous domain structures. The N-terminal half presents the PCAF homology domain, and the C-terminal contains the bromo- and acetyltransferase domains [58].

GCN5 and PCAF play a role in transcriptional activation, cell-cycle, differentiation and apoptosis. They are also associated to nuclear-receptor-mediated activation [178] and growth-factor-

signalled activation [179]. GCN5 and PCAF are known to have several substrates, however, their major targets are histones and nucleosomes [180].

Members of the GNAT superfamily have a particularly conserved fold structure [181, 182]. The core secondary elements of all GNAT structures present six or seven *β*-strands and four *α*-helices. The pyrophosphate binding loop is considered as a characteristic structural feature of the GNAT superfamily and it plays an important role in acetyl-CoA binding [181-183].

These enzymes exist as part of multisubunit complexes. For example, GCN5 forms complex with ySAGA, hTFTC, hSTAGA and the PCAF forms complex with hPCAF. Acetyl-CoA binds firstly with a high affinity for the free form of the enzyme, followed successively by the H3 histone substrate. Within the ternary complex, the ε-amino group of the lysine substrate directly attacks the carbonyl carbon of acetyl-CoA, transferring the acetyl group to the acceptor peptide. Before the catalysis, a conserved glutamate residue (hPCAF-E570, yGCN5-E173) acts as a general base catalyst, deprotonating the ε-amino group and facilitating nucleophilic attack. Finally, the resulting products are released sequentially, with the CoA following the ejection of the acetylated H3 product. PCAF can regulate itself through its own post-translation modification [184].

1.6 Structure and Function of MOZ / MYST3 Gene

The human KAT6A gene located on chromosome 8p11 encodes the monocytic leukaemia zinc finger protein, also referred to as MOZ, MYST3 or KAT6A. KAT6A gene is composed of 18 exons, whereof 16 are coding. Additionally, there are three major splice forms that present differences among them in the 5' untranslated region (UTR), however, all of the splice forms produce protein that contains 2004 amino acids. MOZ protein shares a conserved MYST acetyltransferase family in common with other human proteins such as MYST1/MOF/KAT8, MYST2/HBO1/KAT7, MYST4/MORF/KAT6B and TIP60/KAT5 [185] (Figure 1.6).

MOZ was first identified as part of a fusion gene generated by a recurrent chromosomal translocation t(8;16)(p11;p13) that creates a fusion between *KAT6A*, and *CREBBP*. It is considered a cytogenetic hallmark of myeloblasts of some patients of the M4 and M5 subtypes of acute myeloid leukaemia (AML) [185]. Later, more translocations and inversions involving *KAT6A* were discovered, including those encoding HATs p300 and TIF2 [186]. Since then, diverse biological roles of MOZ have been discovered, and its critical requirement has been illuminated; however, complete detail of the structure and function of MOZ are not yet fully understood.

1.6.1 Domain organization of MOZ

MOZ is a multidomain protein that has, in addition to a catalytic MYST module, a NEMM (N-terminal part of Enok, MOZ or

MORF) domain, a double PHD finger (DPF) and acid domains. The C-terminus of MOZ contains regions rich in serine (S), proline (P), glutamine (Q) and methionine (M) (Figure 1.6A).

The catalytic MYST domain of MOZ is composed of eight α-helices and nine β-strands folded into two β-sheets enclosed by α-helical bundles [187]. The central core region, conserved in other KATs, is implicated in the acetyl-CoA cofactor binding, once it accommodates E680 and C646 that are the key enzymatic residues. On the other hand, it contains an additional C2HC-type zinc finger at the N-terminus and a α6-helix-turn-α7-helix motif at the C-terminus, unlike other acetyltransferase domains. Both of these regions interact tightly with DNA. It makes MOZ MYST domain an unique module that combines a non-mutually exclusive enzymatic and DNA-binding activities [187].

In regard to NEMM (also named H1/5) domain, it has similar sequence homology to histone H1 and H5 motifs and its function has been associated to nuclear localization and transcriptional coactivation [188]. Although its biological role still remains to be understood.

The adjacent DPF domain, a double plant homeodomain (PHD) finger, possesses ability to recognise and bind to methylated, acetylated or the unmodified histone lysine residues [189]. Recently, our research group showed evidence that MOZ DPF operates in recognition and binding of histones, especially H3; however, this binding is dependent on specific post-translational modifications

(PTMs), including acetylated H3K14, which has been correlated with transcriptional activation. On the other hand, trimethylated H3K4, a chromatin modification known to mark the transcription start sites of active genes, prevents the binding of H3 to the DPF [190]. Furthermore, it was shown for the first time that MOZ DPF induces conformational change in the sequence of H3K4-T11 to an α-helix in order to facilitate sampling of H3K4 methylation status, and present H3K9 and other residues for modification [190]. Thus, DPF plays important role in determining if H3 will be subjected to HAT activity. More recently, it was showed that DPF domain of MOZ is also a specific reader of the histone lysine crotonylation, presenting four fold higher binding compared to H3K14 acetylation. MOZ DPF is also responsive to H3K14pr and H3K14bu, however, with lower affinity [191].

1.6.2 MOZ is part of multi-subunit protein complexes

MOZ, as the others histone acetyltransferases of the MYST family as such TIP60, HBO1, MORF, and MOF, functions in multi-subunit protein complexes [192-194]. Beside the role that protein complexes associated with HAT proteins play to enable them to acetylate chromatin substrates, these proteins also select which histone tail will be targeted.

MOZ is known to be a catalytic subunit in many nuclear multi-subunit complexes. Indeed, MOZ was first isolated as part of the AML1-CBFβ transcription factor complex, by which MOZ plays a role in stimulation of AML1-mediated transcription [195].

Since then, a number of interacting partners have been identified, such as the tetrameric ING5 (inhibitor of growth 5) complex *in vivo*, which specifically acetylates nucleosomal histone H3K14 [192].

ING5 is a member of the ING family of putative tumour suppressors which do not possess intrinsic enzymatic activity but it has been shown to be essential for the KAT acetylation of chromatin substrates, possibly playing a role in the initial binding to the chromatin through their PHD fingers [192]. Other subunits of the complexes are EAF6 (homolog of Esa1-associated factor 6) and BRPF1 (bromodomain-PHD finger protein). BRPF family of protein also have multiple domains which include PHD fingers, bromodomain and a proline-tryptophan-tryptophan-proline (PWWP) domain. BRPF1 interacts with MOZ through the MYST domain, and bridges the interaction between MOZ and ING5 proteins and the fourth partner Esa1-Associated Factor 6 (hEAF6) [196, 197] (Figure 1.7).

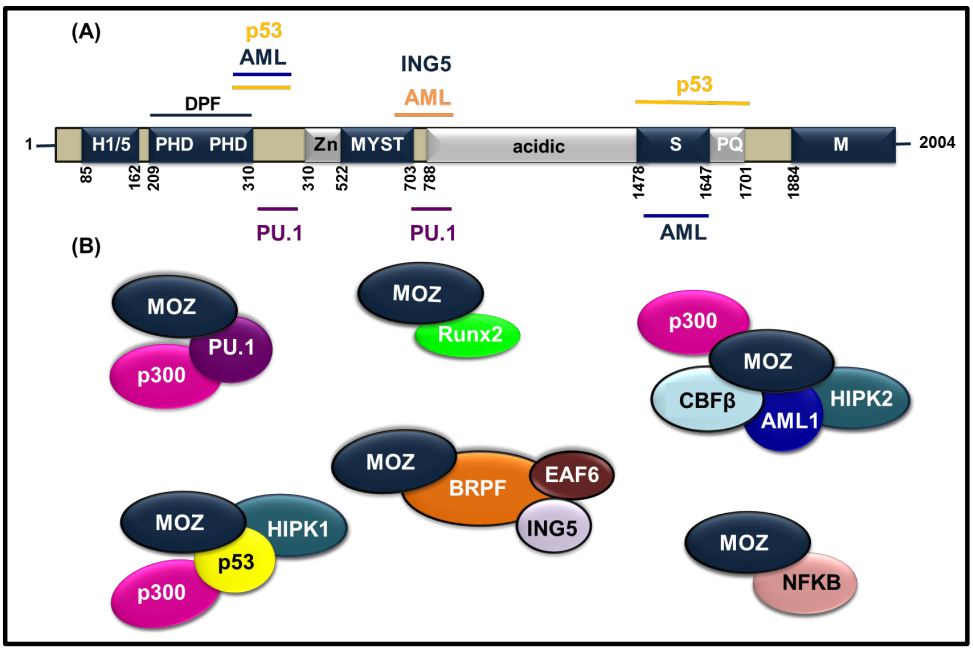


Figure 1.7 - Modular structure of MOZ and its interacting partners. (A) Schematic representation of MOZ Domain organisation, indicating highly conserved regions and functional domains. Domains are labelled as follows: H1/5, histones H1- and H5-like domain; PHD, plant homeodomain; DPF, double PHD finger; Zn, C2HC zinc finger; MYST homologous acetyltransferase region; acidic, glutamate/aspartate-rich region; S/M, serine/ methionine-rich region; PQ, proline/glutamine-stretch. The mapped binding sites of published interacting partners are also detailed on the schematic. **(B)** MOZ functions as the catalytic subunit of a variety of multi-subunit complexes.

Functionally, this complex stimulates MOZ MYST domain acetyltransferase activity and influences its choice of target, which substrate specificity is primarily acetylating H3 and H4 free histones and nucleosomal histone H3 [196]. BRPF1 was also seen to amplify the transcriptional activation potential of MOZ towards RUNX2 through their interaction. Additionally, it was recently suggested that BRPF2-MOZ complex plays an important role in the differentiation of ESCs via H3K14 acetylation [198]. Furthermore, BRPF3 was shown

to preferentially forms a tetrameric complex with HBO1 and two other subunits but not with MOZ [199]

MOZ and ING5 complex plays an important role in the *in vivo* regulation of T-box DNA binding Transcription Factor 1 gene (*Tbx1*), responsible for heart and aortic arch development, in the 22q11 deletion. In KAT6A knockout mouse the embryos have similar heart and aorta defects to those seen in children with DiGeorge/velo-cardio-facial syndrome. MOZ is crucial for normal activity of Tbx1, and the level of MOZ activity may contribute to determining how severe the defects are in children with DiGeorge syndrome [200].

1.6.3 MOZ functions as a transcriptional co-activator

Many research groups have demonstrated that MOZ acts as a co-activator for selected DNA binding transcription factors, particularly with haematopoietic specificity, such as acute myeloid leukaemia/Runt-related transcription factor 1 (AML1/RUNX1) [195, 201, 202] and the ETS family factor PU.1 [203] (Figure 1.7). In addition, MOZ appears to enhance the transcriptional activity of RUNX2 [204], p53 [205], NRF2/MAFK [206], NF-κB [207], and others.

The interaction of MOZ with the p65 subunit of NF-kappaB enhances expression of NF-kappaB-dependent promoters and may play an important role in the pathogenesis of certain acute myeloid leukemia [207].

Regarding MOZ-p53 interaction, MOZ acts as a coactivator for p53-mediated transactivation of specific target genes. It was showed

that MOZ acetylates p53 at K120 and K382 and colocalizes with p53 in promyelocytic leukemia (PML) nuclear bodies under cellular stress. MOZ-PML-p53 interaction enhances MOZ-mediated acetylation of p53, and this ternary complex enhances p53-dependent p21 expression and cell arrest [205, 208]. In fact, p53 leads damaged cells to apoptosis or initiate G1 cell-cycle arrest for damage repair [209-211].

Although it is known that MOZ cooperates to activate transcription factor-dependent expression of genes as such AML-1 dependent transcription of myeloperoxidase [195, 202], it is still unclear the role that HAT plays in MOZ coactivator function once the MYST domain is required for NF-κB-dependent transcription [207] but not for AML1-dependent transcription [195]. It indicates that the promoter context is important to determine the role of HAT activity in gene regulation.

1.6.4 MOZ in hematopoiesis and developmental process

As mentioned above, MOZ is a coactivator of the transcription factors AML1 and PU.1 and they are known to play fundamental roles in the establishment of definitive hematopoiesis, it is not surprising that MOZ function is essential in blood formation.

MOZ-deficient mice are embryonic lethal at stage of development E15 with an undersized liver, organ responsible for the production of hematopoietic cells during this developmental stage [203]. MOZ was also implicated in the generation and maintenance of E14.5 fetal liver cells extracted from the MOZ null mice as it was

observed reductions in the number of short-term hematopoietic stem cells (HSCs) and lineage-committed progenitors. Furthermore, the number of B cells was critically reduced in lineage-committed cells, also an increased population of myeloid lineage cells and maturation arrest were observed in erythroid cells by an accumulation of erythroblasts [203]. Finally, expansion in the committed myeloid cells indicates that differentiation is interrupted at the progenitor stage in the $MOZ^{-/-}$ fetal livers. Therefore, all these findings show that MOZ plays a key role for correct development and differentiation of lymphoid and myeloid cells [212].

To verify whether $MOZ^{-/-}$ embryos possess functional hematopoietic stem cells, these embryos were injected into irradiated normal recipient mice by competitive repopulation assay. The results indicated that MOZ is essential for sustaining the long-term repopulation of stem cells because no mice that received cells from MOZ-deficient embryos survived after 28 days. Thus, MOZ is involved in the regulation of hematopoietic stem cells self-renewal activity to reconstitute the hematopoietic system in recipients following transplantation [203, 213].

To better understand the function of MOZ acetyltransferase activity in hematopoiesis, a study using a mouse strain carrying a single amino acid mutation that inactivates HAT activity was performed [214]. The result showed a reduction in the number of hematopoietic stem cells and lineage-committed precursors, also a B-cell development defect. Severe decrease in the hematopoietic

precursors proliferation was also observed [214]. In conclusion, these analysis indicates an essential role for MOZ HAT activity in the proliferation and maintenance of hematopoietic precursors at all stages of development.

1.6.4.1 MOZ regulates B-cells progenitor proliferation

More recently, a role for MOZ in regulating B-cells progenitor proliferation and hematopoietic malignancy was reported. The Eμ-Myc model of aggressive pre-B/B-cell lymphoma model was used and showed that the loss of one Moz allele increased by 3.9 fold the survival of mice. Thus, MOZ maintained the proliferative capacity of B-cell progenitors even when c-MYC was up-regulated through the directly maintenance of the transcriptional activity of genes necessary for normal B-cell development. A reduction of B-cell progenitor number was also observed in MOZ haploinsufficient mice. Furthermore, an overlap in genes regulated by MOZ, mixed lineage leukemia 1, and mixed lineage leukemia 1 cofactor, menin, was also found [215]. The authors highlighted that the homeobox gene required for B-cell development, Meis1, was among the overlap genes. It is an interesting observation because this gene is characteristically upregulated as a result of MLL1 translocations in leukemia. They also demonstrated that MOZ is localised to the Meis1 locus in pre-B-cells and maintains MEIS1 expression. Thus, the results suggest that even partial inhibition of MOZ is able to reduce the proliferative capacity of MEIS1, in HOX-driven lymphoma and leukemia cells [215].

1.6.4.2 MOZ regulates Hox genes expression

MOZ has been associated with developmental processes through the regulation of Homeobox gene expression, the gene clusters responsible for body segment identity during development. For example, a zebrafish MOZ mutant phenotype presented developmental defects in the head region, exactly where MOZ mutation causes the abolition of HOX gene expression in skeletogenic cranial neural crest cells [216]. However, an extensive anterior homeotic transformation of the axial skeleton and nervous system is observed in MOZ-deficient mice [217]. For this reason, MOZ was implicated to regulate many HOX proteins. Other studies confirmed that MOZ deletion leads to a global downregulation of HOX gene expression via ChIP analysis using antibodies against specific histone post-translation modifications followed by quantitative genomic PCR. The results showed a reduction of H3K9 acetylation at the TSS of HOX *loci* in MOZ-deficient mice. However, H3K14 acetylation levels were not affected and H4K16 acetylation and H3K9 trimethylation were increased in the $MOZ^{-/-}$ mutant compared to wild-type controls, contradicting previous *in vitro* studies [218]. Therefore, it was concluded that MOZ is needed to maintain normal levels of H3K9 acetylation at HOX loci *in vivo.* In additon, it was found that retinoic acid signaling, a potent activator of HOX gene expression, rescues H3K9 hypoacetylation, HOX gene repression and the anterior homeotic transformation caused by a lack of MOZ [217].

Furthermore, it was the scaffold protein Symplekin, that supports the assembly of polyadenylation machinery, interacts and co-localizes with both MOZ and MLL in immature hematopoietic cells. Its inhibition caused a decrease of the HOXA9 protein level but not of HOXA9 mRNA also to an over-recruitment of MOZ and MLL onto the HOXA9 promoter [219].

The aberrant HOX transcription factors expression has been associated to MLL-rearranged mixed lineage leukaemia [220] and MOZ/CBP-associated leukemogenesis [221]. MOZ/MLL-targeted knockdown in multipotent CD34+ cells alters the recruitment of the enzymes to HOX promoters, thus the levels of H3K4me2/3 decrease and cause repression of HOX gene which affects differentiation into myeloid progenitors [222].

MOZ and BMI1 have opposing effects on the initiation of HOX gene expression during embryonic development. BMI1 is a member of the polycomb repressive complex 1 (PRC1) which ubiquitinates (ub) H2AK119 via its catalytic units RING1A/RING1B [223]. Through the interrogation of temporal HOX gene expression dynamics, it was shown that MOZ and BMI1 control the balance between activation and repression in the period before and during HOX activation. It was concluded that MOZ but not BMI1 was critical for the maintenance of normal transcription levels [224]. However, a recent study showed that MOZ and BMI1 act synergistically to preserve adult hematopoietic stem cells by maintaining quiescence and repressing senescence, respectively [225].

1.6.4.3 MOZ is implicated in brain development

One study indicates that mouse MOZ regulates neural stem cell proliferation by repressing p16/INK4A expression [226]. Furthermore, it was reported that some individuals with intellectual disabilities had mutation in MOZ. They also presented feeding difficulties, cardiac defects and distinct facial features. The patients presented different mutations in the KAT6A gene that resulted in a truncated transcript. Some of them had small insertions and deletions that caused a frameshift in the coding sequencing while one of them had deletion of the whole reading frame of MOZ. During the first year of life, the patients suffered microcephaly, poor vision, seizures and global developmental delay. Verbal communication was also affected afterwards [227]. Interestingly, all of the patients investigated had the exon 17/18 affected by these mutations, more specifically, the acidic glutamate/aspartate-rich region, which includes the transactivation domain of MOZ [227].

MOZ is also required in the mesoderm for heart development as lack of MOZ in the mesoderm leads to cardiac septal defects and causes conotruncal and aortic arch defects [228]. Finally, dentinogenesis is also affect by MOZ expression. It was found that the microRNA 665 (miR-665), known to repress early and late odontoblast marker genes and proteases involved in dentin maturation, targets MOZ. Thus, it inhibits MOZ acetylation at H3K9 which maintains the tri-methylation (H3K9me3) state and, consequently, inhibits dentinogenesis by repressing the expression

of the proteins dentin DSPP and DMP1 which form primary dentin via mineralization [229].

1.6.5 Association of MOZ with acute myeloid leukaemia and other haematological disorders

As mentioned earlier, MOZ was first identified in recurrent chromosomal translocations which involve the fusion of KAT6A to the gene encoding CBP t(8;16)(p11.2;p13.3) [185]. Since then, others MOZ fusions have been identified such as incorporating the nuclear coactivators p300 t(8;22)(p11;q13) [230], TIF2 (transcriptional intermediary factor) inv(8)(p11q13) [231] and *NCOA3* t(8;20)(p11;q13) [232]. More recently, it was found in therapy-related AML that MOZ is also fused to the LEUTX (leucine twenty homeobox) t(8;19)(p11;q13), however, the mechanism of how this translocation contributes to leukemogenesis is not yet known [233] (Figure 1.8).

Dysfunction of MOZ activity, caused by reciprocal translocations generating MOZ fusion proteins has been shown to play a causative role in acute myeloid leukaemia (AML), a heterogeneous hematopoietic stem cell disorder where the myeloid cells do not enter into differentiation process, resulting in accumulation of blast cells and absence of a fully functional granulocyte lineage [234, 235]. AML can arise in patients with an underlying hematological disorder, or as a consequence of prior therapy such as exposure to topoisomerases II, alkylating agents

or radiation [236]. Most commonly, it appears as a *de novo* malignancy in previously healthy individuals.

The majority of clinical manifestations of AML are caused by accumulation of malignant, poorly differentiated myeloid cells within the bone marrow, peripheral blood and less frequenty in other organs. Most patients have a combination of leukocytosis and signs of bone marrow failure, such as anemia and thrombocytopenia. Fatigue, anorexia and weight loss are usually observed in these patients. Non-treated AML leads to death within months of diagnosis usually secondary to infection or bleeding. The diagnosis of acute leukemia is established, among other analysis, by the presence of 20% or more blasts in the bone marrow or peripheral blood [237]. AML is further diagnosed by demonstrating the myeloid origin of these cells by testing for myeloperoxidase activity, immunophenotyping or documenting the presence of Auer rods.

Well-characterized chromosomal translocations, such as t(8:21) in core-binding factor AML or t(15:17) in acute promyelocytic leukemia result in the formation of chimeric proteins (RUNX1-RUNX1T1 and PML-RARA, respectively), which affect the normal maturation process of myeloid precursor cells. Furthermore, molecular changes have also been linked to the development of AML. In fact, genetic mutations are identified in more than 97% of cases [238], often combined to a large chromosomal abnormality [239]. Based on the study performed using animal models a two-hit

model that offers a conceptual framework for classifying the various mutations associated with AML was developed [240]. According to this model, class I mutations which result in the activation of pro-proliferative pathways must occur in conjunction with class II mutations which impair normal hematopoietic differentiation in order for leukemia to develop [241, 242]. Common class I mutations, such as FLT3 (internal tandem duplications, ITD, and tyrosine kinase domain mutations, TKD), K/NRAS, TP53 and c-KIT are found in ~28, 12, 8 and 4% of cases, respectively [239]. The role of signal transducer and activator of transcription 3 (STAT3) in the stimulation of cellular proliferation and survival was also accessed in solid and hematological malignancies [243-245]. Increased tyrosine phosphorylation of STAT3 whether due to increased secretion of cytokines, such as IL-6 [246] or mutations in receptor tyrosine kinases [247, 248] is seen in up to 50% of AML cases and is associated to a worse prognosis. Notable class II mutations confer a better prognosis and include NPM1 and CEBPA, which are found in ~27% and 6% of cases, respectively [239]. A third class of mutations that includes genes involved in epigenetic regulation was included in this classification. These gene alterations causes downstream effects on both cellular differentiation and proliferation [238].

Recently, a unique case of infantile leukemia in which chromothripsis in chromosome 8 completely altered the G-band structure and resulted in concurrent changes in MOZ, FGFR1,

RUNX1T1, and RUNX1 was reported. These multiple-hit abnormalities in AML have not been described previously [249].

AML expressing MOZ-related translocations are generally monocytic leukaemias classified under the M4/M5 French-American-British subtypes [250] and are particularly aggressive; the median survival post-diagnosis for these patients ranges from 2 – 10 months [233].

The recurrent breakpoints of MOZ associated with inv8 AMLs are located within the region encoding the acidic –rich domain (Figure 1.9). Thus, the N-terminal region of MOZ, containing the H1/5 like-, DPF-, MYST- and basic-domains is maintained and one of the fusion partners listed above replaces the MOZ C-terminus [212]. The chimeric products disturb gene expression programs necessary for normal myeloid differentiation because these proteins can cause a reduction of enzymatic activity at intended genomic *loci*, aberrant global and promoter-specific acetylation, sequestration of essential nuclear regulators [251] or the instability of co-regulator complexes [252].

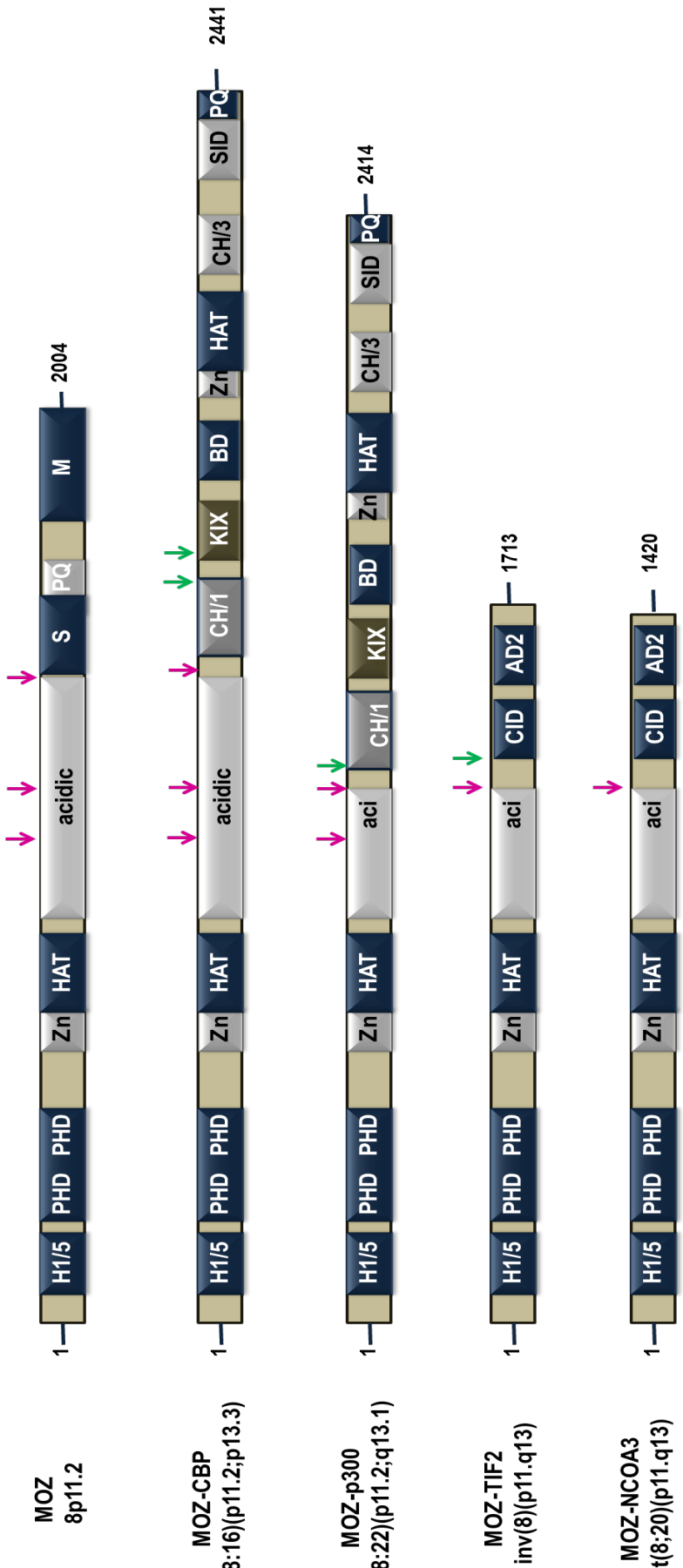

MOZ
8p11.2
1
2004
H1/5
PHD
PHD
Zn
HAT
acidic
S
PQ
M
MOZ-CBP
t(8:16)(p11.2;p13.3)
1
2441
H1/5
PHD
PHD
Zn
HAT
acidic
CH/1
KIX
BD
Zn
HAT
CH/3
SID
PQ
MOZ-p300
t(8:22)(p11.2;q13.1)
1
2414
H1/5
PHD
PHD
Zn
HAT
aci
CH/1
KIX
BD
Zn
HAT
CH/3
SID
PQ
MOZ-TIF2
inv(8)(p11.q13)
1
1713
H1/5
PHD
PHD
Zn
HAT
aci
CID
AD2
MOZ-NCOA3
t(8;20)(p11.q13)
1
1420
H1/5
PHD
PHD
Zn
HAT
aci
CID
AD2

Figure 1.8 - Schematics of leukaemia-associated MOZ fusion proteins. MOZ contains several functional domains including the H1/H5 (Histone1/5-like), PHD (plant homeodomain), Zn (zinco finger), HAT (histone acetyltransferase), acidic domains and S- (serine), PQ- (proline/glutamine), M- (methionine) rich domais. MOZ fuses with four known partners: CBP, p300, TIF2 and the recently identified NCOA3, MOZ fusion partners contain CH (cysteine/histidine rich), KIX (kinase-inducible), BD (bromodomain), SID (SRD1 interacting domain), CID (CBP interacting domain) and AD (activation domain). Chromosomal translocation breakpoints in the MOZ gene and its fusion partners described are described as chromosome loci and mapped as pink arrows and green arrows, respectively.

Our group showed that expression of MOZ-TIF2 in different cell lines causes broad changes in histone modifications and disturbs transcription by CBP-dependent transcription factors such as p53 and retinoic acid receptors [252, 253].

MOZ-TIF2 and MOZ-CBP alter cofactor recruitment and chromatin modification at target promoters. The transcriptional activity of RUNX factors, involved in regulation of blood and bone differentiation, is enhanced by MOZ-TIF2 [202]. On the other hand, MOZ-TIF2 inhibits transcriptional activities of CBP-dependent activators such as p53 and nuclear receptors. This is due to the direct binding of CBP to the C-terminal activation domain (AD1) of MOZ-TIF2, as a consequence, CBP is depleted from PML bodies [252]. Cells treated with calpain/cathepsin inhibitor (CPI-2) blocks the effect of MOZ-TIF2-mediated destruction of CBP/p300 and restor CBP expression to normal levels. It also rescues MOZ-TIF2 mediated gene repression [253].

MOZ-TIF2 associates with the *RARβ2* promoter disturbing the ligand-dependent recruitment of CBP and causing down-regulation of the RARβ2 gene. On the other hand, MOZ-TIF2 enhanced AML1/RUNX1 reporter activation [202]. The MOZ-CBP fusion protein shows similar transcriptional regulation functions at the RARβ2 and AML1 promoters as compared to MOZ-TIF2 [202].

Binding of MOZ-TIF2 to the CBP cofactor increases the proliferative capacity of hematopoietic progenitors *in vitro* and induces AML *in vivo* [252]. Irradiated mice depleted for normal bone marrow and rescued by bone marrow transplant, showed that retroviral delivery of MOZ-TIF2 can immortalise committed myeloid progenitor cells and cause AML [254]. The transforming properties of MOZ-TIF2 is dependent of the C2HC nucleosome recognition motif of MOZ, and the AD1 domain of TIF2, whereas the HAT activity of the MYST domain is not essential [255].

Studies using PU.1-deficient mice showed that PU.1 is essential for the ability of MOZ-TIF2 to establish and maintain AML stem cells. Both MOZ-TIF2/CBP fusion proteins can interact with the PU.1 transcription factor to stimulate the expression of macrophage colony-stimulating factor receptor [256]. A recent study using murine MOZ-TIF2 AML models showed that myeloid differentiation to granulocyte macrophage progenitors (GMP) is critical for leukemia stem cell generation and AML development. Disrupting GMP formation by deleting the lineage-restricted transcription factor

blocked normal granulocyte formation and prevented initiation of AML [257].

1.6.5.1 MOZ in other cancer types

MOZ is associated with other types of cancer also due to its association with p53 protein, the most frequently mutated gene among all human cancers. As mentioned before, MOZ directly interacts with p53 and is essential for p53-dependent selective regulation of p21 expression [205]. Also, MOZ increases p53 acetylation and premature senescence through its complex formation with promyelocytic leukemia nuclear bodies [258].

Prevalent p53 missense mutations abrogate its tumour suppressive function and lead to a 'gain-of-function' (GOF) that promotes cancer. A study revealed a novel chromatin mechanism underlying the progression of tumours with GOF p53 where p53 GOF mutants bind to and upregulate chromatin regulatory genes, including the methyltransferases MLL1, MLL2 and MOZ, resulting in genome-wide increases of histone methylation and acetylation [259]. Analysis of The Cancer Genome Atlas shows specific upregulation of MLL1, MLL2, and MOZ/KAT6A in p53 GOF patient-derived tumours, but not in wild-type p53 or p53 null tumours [239, 259].

MOZ/KAT6A is commonly found in the chromosome amplicon consisting of loci 8p11-p12, which contains approximately 55 genes and is present in 12-15% of breast cancers. It is associated with poor prognosis in primary breast tumours and a lower 5-year metastasis-

free survival rate [260]. Also, analysis of The Cancer Genome Atlas data [239] from 1,000 primary breast cancers revealed that in 22% of cases, the expression of MOZ is altered and it is overexpressed in 8.7% of invasive breast cancers at the mRNA level. Moreover, a high level of MOZ expression was linked to reduction in the overall survival rate [260].

Recently MOZ was found to have a role in breast cancer development as novel histone acetyltransferase that activates ERα promoter. It was suggested that targeting MOZ may serve as a novel therapeutic strategy to block ERα gene expression in MOZ-high breast cancers [261].

Apart from cancer, MOZ is also involved in the DiGeorge/velo-cardio-facial syndrome, caused by a 22q11 microdeletion or mutation of the TBX1 gene, the most common deletion syndrome in humans [262]. It is characterized by severe cardiac and aortic arch defects, thymus dysplasia, cleft palate, and craniofacial abnormalities [263]. As mentioned before, MOZ has been shown to be essential for the normal activation of TBX1 *in vivo* [264]. TBX1 recruits chromatin modifiers to their target genes, enabling gene expression. Importantly, DiGeorge syndrome-like anomalies are displayed in mice with homozygous mutation of MOZ and in heterozygous MOZ mutants when combined with Tbx1 haploinsufficiency or oversupply of retinoic acid. MOZ complex occupies the TBX1 locus, promoting H3K9 acetylation and its expression; transgenic expression of TBX1

rescues the velo-cardio-facial syndrome phenotype in homozygous MOZ mutant mice [264].

1.7 Aims and Objectives

MOZ is an essential coactivator for genes that are involved in haematopoietic and developmental processes. Dysfunction of MOZ activity, caused by reciprocal translocations generating MOZ fusion proteins, has been shown to play a causative role in acute myeloid leukaemia (AML). Additionally, it is known that MOZ acts as a co-activator for many transcription factors, particularly with haematopoietic specificity. However, it is also important to investigate which genes have their expression dependent of MOZ activity to fully understand MOZ role in the physiological and disease conditions. Thereby, the primary aim of this project was to explore the role of MOZ in acute myeloid leukaemia by identifying genes that are regulated by MOZ in K562, a human erythroleukemic cell line which expresses robust levels of MOZ. Secondly, we aimed to generate a MOZ/KAT6A knockout K562 cells using CRISPR CAS9n nickase technology to allow further analysis of the function of MOZ in leukaemia. The final objective was to explore a possible effect of I-BET151 on the expression of MOZ target genes.

Chapter 2 : Materials and Methods

2.1 Materials

2.1.1 General suppliers

All general laboratory chemicals were of analytical grade and supplied by Fisher Chemicals or Sigma-Aldrich, unless otherwise stated. Phosphate buffered saline (PBS) was prepared using PBS tablets (1 tablet/100 ml distilled water) supplied by OXOID Ltd. Double deionised water was used to make all solutions, unless RNase free water was required in which case deionised water was purified by passage through a Purite select/Neptune water purification system. The pH of solutions was measured using a Jenway 3510 pH meter. Where appropriate, sterilisation was achieved by filtration through a 0.22 micron filter or by autoclaving for 20 minutes at 120℃ and 25 lb/in2 pressure.

2.1.2 Bacterial reagents

Tryptone and yeast extract for bacterial growth medium were purchased from OXOID Ltd. The *Escherichia coli* strain DH5α [265] was used for all molecular cloning applications in this project and was purchased from Stratagene.

2.1.3 Molecular biology reagents

All enzymes and their buffers were purchased from Roche Applied Science, New England Biolabs (NEB) or Promega. 1 kb Plus DNA ladder was purchased from Invitrogen Life Sciences. Polymerase enzymes and their respective buffers and supplements

were purchased from NEB, Finnzymes or Invitrogen Life Sciences. dATP, dCTP, dGTP and dTTP were purchased from Invitrogen Life Sciences. Gel extraction kits were acquired from York Biosiences, Nucleospin® plasmid kits from Macherey-Nagel, and maxi-DNA kits from QIAGEN Ltd. Oligonucleotides for sequencing and PCR applications were purchased from Sigma-Aldrich.

2.1.4 Protein chemistry, western blotting reagents and equipment

Acrylamide mix 30% (w/v) was obtained from National Diagnostics. SDS-PAGE electrophoresis apparatus, Precision Plus Protein Standards, wet transfer apparatus and nitrocellulose transfer membrane were all purchased from Bio-Rad Laboratories. Complete Protease Inhibitor cocktail tablets were acquired from Roche. All primary antibodies, Horseradish peroxidase (HRP) conjugated secondary antibodies, Protein A or G sepharose beads were purchased from Abcam, Sigma-Aldrich, Roche and Millipore. Glutathione Sepharose™ 4Fastflow beads and Amplify solution were obtained from GE Healthcare and X-ray film was obtained from FUJI-medical. Isopropyl-β-D-Thiogalactopyranoside (IPTG) and Dithiothreitol (DTT) were purchased from Merck chemicals.

2.1.5 Tissue culture reagents

K562 and HEK293 cell lines were purchased from the European Collection of Cell Cultures (ECACC). All tissue culture

supplements were purchased from Lonza BioWhittaker® and plasticware was obtained from Helena Biosciences.

2.1.6 Assay reagents

Protein assay reagent was purchased from Bio-rad and bovine serum albumin (BSA) for protein standards was purchased from First Link (UK) Ltd.

2.2 Propagation, purification and manipulation of plasmid DNA

2.2.1 Composition of solutions and media used in bacterial and molecular biology procedures

Luria-Bertani (LB) medium: 1% (w/v) tryptone, 0.5% (w/v) yeast extract 1% (w/v) NaCl (pH 7.0). The medium was solidified by the addition of 2% (w/v) bacteriological agar where required.

Solid growth medium: For LB plates 1 L of LB medium was supplemented with 15 g bacto-agar prior to sterilisation. The appropriate antibiotic was added to the molten agar medium and 20-25 ml was poured into each petri-dish and allowed to solidify at room temperature. Plates were stored at 4℃ for up to 1 month.

Ampicillin 1000× stock solution: 100 mg/ml in 50% ethanol.

Cloramphenicol 1000× stock solution: 34 mg/ml 100% ethanol.

Kanamycin 1000× stock solution: 50 mg/ml in H2O.

Gel loading buffer (5×): 15 g Ficoll400 dissolved in 10 ml 10 mM Tris-HCl-pH 7.5 and 50 mM EDTA-pH 8.0, containing 0.4% orange G, 0.03% bromophenol blue and 0.03% xylene cyanol FF.

Tris-borate-EDTA (TBE): 40 mM Tris-base, 40 mM boric acid, 0.5 mM EDTA (pH 8.0).

2.2.2 Culture of *E. coli*

The bacterial strain *E. coli* DH5α was used for all DNA manipulations while the strain Rosetta was employed for optimised protein expression. Single colonies were obtained by streaking the bacteria onto agar plates containing the appropriate antibiotic (100 μg/ml ampicillin, 50 μg/ml kanamycin or 50 μg/ml chloramphenicol) and grown at 37℃ which can be stored for several weeks at 4℃. Liquid cultures were grown by inoculating a single colony into liquid LB (containing appropriate antibiotic) and incubated in a universal screw-top tube (Bibby Sterlin) or conical flasks at 37℃ with shaking at 220 rpm.

2.2.3 Preparation of competent *E. coli*

E. coli were made competent for DNA transformation using the calcium chloride-magnesium chloride method. *E. coli* were streaked from a glycerol stock stored at -80℃, onto Luria Bertani (LB) plates and incubated at 37℃ overnight. A single colony from the plate was used to inoculate a 2.5 ml LB medium and grown overnight at 37℃. The culture was diluted 1:100 in LB medium and grown at 37℃ to an optical density (OD595) of 0.3. The cells were then chilled

on ice for 20 minutes and consequently harvested by centrifugation at 4℃, 3500 rpm for 5 minutes. On ice, cell pellets were resuspended in 20 ml of ice cold 100 mM MgCl2 and incubated on ice for 5 minutes. After centrifuging as above, cells were resuspended in 10 ml ice cold 100 mM CaCl2 and incubated on ice for 30 minutes. The cells were then harvested by centrifugation and resuspended in 5 ml CaCl2 solution (70 mM CaCl2, 15% v/v glycerol) prior to being aliquoted into 200 µl samples in microfuge tubes and snap-frozen in liquid nitrogen. Competent cells were then stored at -80℃. The transformation efficiency was tested using the pcDNA3.1+ plasmid vector using serial dilutions of the transformation to give a minimum of ~1 × 106 transformants/1 µg DNA.

2.2.4 Transformation of competent *E. coli*

DNA products were transformed into competent DH5α or Rosetta *coli* strains. The competent host cells were thawed on ice. For each transformation 50 ng of plasmid DNA was gently mixed with 50 µl cell suspension and incubated on ice for 30 minutes. The cells were heat-shocked in a water bath at 42℃ for 2 minutes and returned to ice for a further 2 minutes. 500 µl of LB medium was added to each reaction and incubated at 37℃ with shaking at 220 rpm for 1 hour. Transformations were spun at 13,000 rpm for 1 minute to pellet the cells and 350 µl of supernatant was discarded. Cells were resuspended in remaining LB medium and spread onto

LB agar plates containing the appropriate antibiotic. The plates were incubated overnight at 37ºC.

2.2.5 Long-term storage of bacterial cultures

Bacterial cultures were maintained for long periods of time by transferring 0.8 ml of a turbid bacterial culture to a 1.5 ml eppendorf. 0.2 ml sterile glycerol was then added to the culture. The contents were then vortexed and stored at -80℃.

2.2.6 Small-scale purification from bacterial cells

E. coli cultures of 5-10 ml were grown overnight in an orbital shaker at 37℃, 220 rpm. Cells were harvested by centrifugation and the plasmid DNA was recovered from the pellet using the NucleoSpin® Plasmid kit (Machery-Nagel) according to the manufacturer's protocol. DNA was eluted from the NucleoSpin column with the supplied Elution Buffer (EB).

2.2.7 Large-scale plasmid purification from bacterial cells

To obtain a large-scale preparation of plasmid DNA, 250 ml LB media with the appropriate antibiotic was inoculated with suspended cells to make a 1/100 dilution of and left to grow overnight at 37℃, 220 rpm. To isolate plasmid DNA from these cells the QIAGEN Plasmid Maxi Kit protocol was used, involving alkaline lysis of the cells, followed by immobilisation of plasmid DNA on a QIAGEN anion-exchange resin. Elution and precipitation of the DNA was achieved in isopropanol, prior to being washed in 75% ethanol and dried in a category II microbiological safety cabinet under

laminar flow. The resulting DNA pellet was resuspended in sterile water. Details of this protocol and the composition of buffers used can be found in the QIAGEN maxiprep handbook.

2.2.8 Spectrophotomeric quantification of DNA

The quantification and quality of DNA were determined by optical density measurement of 260 nm using the Nanodrop® ND-1000 UV-Vis spectrophotometer. The ratio of OD260:OD280 was used to estimate purity, with pure DNA having an OD260/OD280 of between 1.7 and 1.9. The measurement at 280 nm indicates the presence of protein contamination in the sample.

2.2.9 Phenol/chloroform extraction and ethanol precipitation of DNA

To separate nucleic acids from contaminating proteins a phenol/chloroform extraction was carried out. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 saturated with 10 mM Tris pH 8.0, 1 mM EDTA) solution was added to the DNA sample to extract protein contaminants. The sample was vigorously mixed until an emulsion formed, then spun at 13,000 rpm for 2 minutes in a desk-top centrifuge (Heraeus Biofuge *pico*). The aqueous layer was transferred to a new microfuge tube and a further round of phenol:chloroform extraction carried out until all protein had been removed from the interphase. Following the preceding protocol, several rounds of chloroform extraction were used to remove residual phenol. The DNA was then collected by ethanol precipitation.

2.2.10 Ethanol precipitation

DNA was precipitated using 0.1 volumes of 3 M NaOAc (pH 5.2) and 2 volumes of 100% ethanol. The precipitate was recovered after vigorous mixing and centrifugation for 30 minutes at 13,000 rpm, 4℃ (ALC PK131R, rotor T528). To remove excess salt ions, the pellet was washed with 75% ethanol and the centrifugation repeated. The pellet was dried under vacuum in a heated speed-vac (Savant DNA120 speed-vac) for 5 minutes before resuspending in ddH2O.

2.2.11 Agarose gel electrophoresis

DNA was resolved according to size on neutral agarose gels of between 0.7% and 2% agarose (w/v) depending on the size of fragments being resolved. Gels containing 0.5 µg/ml ethidium bromide were set and run in 0.5× TAE buffer as described in standard protocols (Sambrook J, 1989). DNA samples were loaded in 5× gel loading buffer, in conjunction with Invitrogen standard 1 kb DNA ladder (0.2 µg/µl). Bands were then visualised using an ultraviolet transilluminator (BioRad Gel Doc 2000) and BioRad Quality One software. The approximate size (kb) and concentration (ng/µl) of DNA were determined by comparison with standard molecular weight markers.

2.2.12 Purification of DNA fragments from agarose gels

DNA bands to be purified were visualised using an ultraviolet light box and excised from the gel after electrophoresis. DNA was

purified using the York Biosciences Gel/PCR purification kit according to manufacturer's instructions.

2.2.13 Restriction digest of DNA

Plasmid DNA was digested using restriction endonucleases to generate compatible ends for cloning and to verify the insertion of the gene fragment. The restriction enzymes chosen were based on the primer design used in the cloning protocol and digests were carried out in the buffers and at the temperature recommended by the manufacturer (Roche, NEB or Promega). Digests were carried out in a 20 µl reaction volume consisting of 1× SuRE/cut reaction buffer, 1 units restriction endonuclease, template DNA and made up to volume with sterile water. Digests were carried out at 37℃ for 3-4 hours.

2.2.14 Removal of 5' terminal phosphate groups from cleaved plasmid DNA

To reduce the efficiency with which plasmid DNA cleaved by a single restriction endonuclease would re-ligate to itself in the absence of insert DNA, the 5' terminal phosphate groups were removed by treatment with Thermosensitive Alkaline Phosphatase (TSAP), acquired from Promega. Restriction digests were then set up and incubated as in Section 2.2.13, followed by incubation with Thermosensitive Alkaline Phosphatase (1 unit/µl) in 1x Multicore™ buffer for 30 minutes at 37℃, and incubation at 74℃ for 20 minutes to inactivate this enzyme.

2.2.15 Addition of 5' terminal phosphate groups to oligonucleotides

Short oligonucleotides (27-39mers) which were required to generate short double stranded DNA inserts, first had to have a phosphate group added to their 5' terminus to make them compatible with cut, TSAP treated vector. Therefore 5 µl DNA was phosphorylated in a 50 µl reaction volume using 2 µl Polynucleotide Kinase and 10 µl 5× T4 DNA ligase buffer (Promega) at 37℃ for 30 minutes. The ligase buffer served as a source of ATP. After this reaction kinased oligonucleotides were then ethanol precipitated (Section 2.2.10) and resuspended in 10 µl ddH2O.

2.2.16 Annealing of single-stranded oligonucleotides

2 µL of each oligonucleotide at 100 µM were placed at 37 ºC for 1 hour along with 2 µL of 10X T4 DNA ligase buffer, 10 units of polynucleotide kinase and 15 µL of diH2O to make a total volume of 20 µL. This mixture produces the phosphorylation of the oligonucleotide followed by incubation at 65 ºC for 20 minutes to inactivate the enzyme. Afterwards, 5 µL of each of the two relevant phosphorylated oligos (top and bottom strands) were mixed in a PCR tube with 40 µL of diH2O and placed in a thermocycler to heat the mixture to 95 ºC for 3 minutes, then reduce by 1 ºC each minute until 26 ºC, followed by 5 minutes at 25ºC.

2.2.17 Ligation of DNA fragments

Recombinant plasmids were created by annealing cut fragments with T4 DNA ligase (Promega). Relative quantities of vector and insert were determined by agarose gel electrophoresis (Section 2.2.11) and spectrophotomeric quantification (Section 2.2.8). Ligation reactions were composed using a vector:insert ratio of 1:2 and in accordance with the manufacturer's instructions. Vector only control ligations were set up to represent self-ligation of the vector in the absence of insert DNA. Ligated DNA was transformed in DH5α *E. coli* (Section 2.2.4), plasmid preparations made (Section 2.2.5) and verified for the correct integration of gene fragment (2.2.13, 2.2.20).

2.2.18 Polymerase chain reaction (PCR)

PCR was used to generate DNA fragments needed for construction of recombinant plasmids and also to screen newly generated recombinant plasmids for the correct DNA inserts (colony PCR). For the synthesis of DNA to be used in cloning, a DNA polymerase with proofreading (3' to 5' exonuclease) activity, Phusion® High Fidelity DNA polymerase (NEB) was used. A typical reaction mixture consisted of 20 ng of DNA template and a final concentration of 1× Phusion HF reaction buffer, 0.5 μM of forward and reverse primers, 0.2 μM of each dNTP, and 1.25 units of Phusion® High Fidelity DNA polymerase, made up to final volume of 50 μl with nuclease free sterile water. Reactions were incubated in

an Applied Biosystems 2720 thermal cycler, a 'hot start' PCR method was utilised to prevent misannealing of primers. Samples were initially denatured by incubating at 98℃ for 30 seconds before typical thermal cycling parameters as follows: denaturation, 98℃ for 10 seconds, annealing, 55℃ for 30 seconds and extension, 72℃ for 30 seconds per kb of target DNA, for a total of 25-35 cycles. As a final step, the samples were incubated at 72℃ for 10 minutes before being held at 10℃ until collection. Upon completion, 10% of PCR reaction was separated by agarose gel electrophoresis to confirm that appropriately sized DNA fragments had been produced.

2.2.19 Colony PCR screening

After transformation, ligation colonies were picked from LB agar plates containing selective antibiotic and inoculated in 400 µl of LB (+antibiotic) cultures. These were grown at 37 ºC for 2 hours or until the culture was turbid. Using the bacterial culture as the template, a PCR reaction was set up using TAQ™ DNA polymerase (Invitrogen) in order to screen for the appropriate DNA insert. Primers were chosen where one was complementary for the vector DNA and the other was complementary for insert DNA; this ensured specificity of PCR product. Reactions consisted of the same reagents as previously stated with the following amendments: 1.25 units of TAQ DNA polymerase and its corresponding buffer (1×), 1.5 mM $MgCl_2$ and 2.5 µl of bacterial culture instead of specific DNA template.

2.2.20 Sequencing of plasmid DNA

Plasmid DNA samples were dispatched to Geneservice Ltd (Source Bioscience Ltd Nottingham, UK), and sequenced using Big Dye v3.1 chemistry and an AB13730xl automated sequencer.

2.2.21 Oligonucleotides

Oligonucleotides were purchased from Sigma-Aldrich as lyophilised and desalted pellets. These were resuspended in sterile water to 100 μM concentration and stored at -20ºC. Oligonucleotides used in this project are listed in the Appendix (A1).

2.2.22 Genomic DNA purification

Genomic DNA was purified following the manufacturer's protocol. After harvesting cells, they were placed in a 2 mL microcentrifuge tube and centrifuged at 2000 rpm for 5 minutes. Cell pellet was then resuspended in 1 mL of 1X PBS and centrifugation step was repeated. Then, supernatant was removed and the genomic DNA purification took place.

2.2.23 Reverse transcriptase and quantitative PCR analysis

Total RNA was extracted with RNeasy Mini Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. Reverse transcription PCR (RT-PCR) was carried out with 500 ng of total RNA using QuantiTect Reverse Transcription Kit (Qiagen). Genomic DNA elimination was carried out twice, during RNA extraction and before the complementary DNA (cDNA) synthesis. RNA concentration and integrity were evaluated using an Agilent's 2100 Bioanalyzer.

Following cDNA conversion, real time quantitative PCR (RT-qPCR) reactions were performed in 96-well white-walled plates using 2x Brilliant II SYBR® Green QPCR Master Mix (Agilent). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as a reference control and ΔΔCq method was used to calculate relative gene expression. Primer sequences are listed in the Appendix (A2).

2.3 Cell culture

2.3.1 Maintenance of cell lines

The embryonic kidney cell line, HEK293 was routinely maintained in Roswell Park Memorial Institute medium (RPMI) containing phenol red and supplemented with 1% (v/v) 0.2 M L-glutamine, 1% Penicillin/Streptomycin antibiotics and 10% (v/v) heat inactivated fetal calf serum (FCS). Cultures were grown in sterile plasticware at 37℃ in a humidified atmosphere containing 5% CO2. Adherent cells were passaged at 90% confluency for a maximum of 30 passages, at which point the cells were discarded. To split HEK 293 cells, RPMI was aspirated and the cells washed twice with sterile 1× PBS. 1 ml trypsin/EDTA was added and the cell incubated for 2 minutes at 37℃, 5% CO2. HEK293 cells were passaged at 90% confluency for a maximum of 30 passages, at which point the cells were discarded. To split cells, RPMI was aspirated and the cells washed twice with sterile 1x PBS. Cells were resuspended and subsequently diluted in fresh DMEM in 10 cm plates (for maintenance) or 6-well plates for transient transfections.

2.3.2 Maintenance of suspension cell lines

The human erythroleukemia, K562 cell line was routinely maintained in Roswell Park Memorial Institute medium (RPMI) containing phenol red, supplemented with 1 % (v/v) 0.2 M L-glutamine, 1 % Penicillin/Streptomycin antibiotics and 10 % (v/v) heat inactivated FCS. Cultures were grown in sterile plastic flasks at 37 °C in a humidified atmosphere containing 5 % CO2. Suspension cells were passaged every four days for a maximum of 30 passages, at which point the cells were discarded. To passage, cells were first counted. Cells were diluted in a 1:1 ratio and mixed thoroughly with trypan blue (Autogen Bioclear), loaded onto a hepatocytometer (Hawksley, BS.748) and counted under a light microscope (Nikon eclipse TS100). Cells were then diluted in fresh RPMI media to a concentration of 100,000 cells/ml in a sterile tissue culture flask and incubated at 37 ºC, 5 % CO2 (for maintenance). To harvest suspension cells for whole cell extract preparation, cells were collected at 1250 rpm for 10 minutes at room temperature and the RPMI media aspirated.

2.3.3 Cryopreservation of cell lines

Cells were incubated with trypsin for 1 minute and after removal were returned to the incubator for 5 minutes. Cells were then resuspended in 2 ml of appropriate media (for a 10cm2 dish), counted (section 2.4.2) and diluted to the appropriate concentration. 400 µl (40 %) of cell suspension in complete media was transferred

to labelled cryovials containing 100 µl (10 %) DMSO and a supplement of 500 µl (50 %) FCS. The cell suspension was then gently pipetted to ensure even mixing of DMSO, FCS and media. The vials were then stored at -20 °C until suspensions were frozen solid. The cryovials were then placed at -80 °C overnight, after which the vials were stored in liquid nitrogen.

2.3.4 Resuscitation of cell lines

To recover cells from liquid nitrogen storage, cells were thawed at 37℃ and transferred into 10 ml of pre-warmed media. The suspension was mixed thoroughly and the cells pelleted by centrifugation at 1100 rpm for 5 minutes to remove residual DMSO which can impede growth. Cells were then resuspended in 10 ml of fresh media and incubated at 37℃, 5% CO2 with humidity. A few hours after cell plating the media was changed to remove dead and unattached cells followed by another change of media after 24 hours. Cells were passaged at least three times (approximately 12 days) before use in experiments.

2.3.5 Calcium phosphate-mediated transfection of adherent cells

Adherent cells were seeded at 20% confluency in 10 cm plates 24 hours prior to transfection. The following day DMEM was replaced with fresh media 3-4 hours prior to transfection. For transient transfections, 50-80% subconfluent cells were transfected using the calcium phosphate co-precipitation method. Cells were

transfected with the appropriate amount of plasmid DNA (quantity requires optimisation for each plasmid) and empty vector to standardise the quantity of each sample. To each sample 12.5% (v/v) 2 M calcium chloride solution was added to make a final volume of 800 µl with sterile nuclease-free water pH 7.0. This mixture was added dropwise to an equivalent volume of 2x hepes buffered saline (HBS) solution while vortexing gently. After a 20 minute incubation at room temperature, 1.5 ml was added dropwise to the appropriate cell plate. After 16 hours of incubation at 37℃, 5% CO2, the cells were washed three times with 1x PBS and maintained in RPMI for a further 24 hours. Cells were lysed after 48 hours.

2.3.6 Small interfering RNA and cell transfection

Pools of 4 small interfering RNA oligonucleotides (siRNAs) targeting human KAT6A sequences (5'-GGAAUUAUUUCGUGAUAUC-3', 5'GGAGUUGAGUGUUAAAGAU-3',5'GCGCUAUACUAAUCCAAUA-3',5'GUCAGACCCUGCAGAGUUA-3') and non-targeting siRNA, used as negative control, were purchased from Dharmacon. For transfection of siRNA duplexes at 10 nM, $5x10^5$ K562 cells were seeded the day of transfection in 6 well plates with 1 mL of RPMI complete medium. For each well, 0.6 µl of 20 µM stock of siRNA and 14 µl of INTERFERin® was diluted into 200 µl medium without serum. The mixture was immediately vortexed for 10 seconds and incubated for 10 minutes at room temperature allowing

INTERFERin®/siRNA complexes to form. This complex was added into the well containing cells suspension in growth medium, and was homogenized by gently swirling the plate. Plates were incubated at 37 °C. After 5 hours, 1.5 ml of complete medium was added and cells were incubated as before. Cells were harvested at 48, 72 and 96 hours post-transfection for determination of protein and mRNA levels. Pool of scrambled siRNA transfected and non-transfected cells were used as control.

2.3.7 Electroporation

Transfection was carried out using the Neon™ Transfection System (Invitrogen) according to the instructions provided by the manufacturer. $2x10^5$ K562 cells were transfected with 1 µg of plasmid in a volume of 10 µl of appropriated resuspension buffer using a Neon tip. Electroporation was carried out at 1,350 V with four 10 ms pulse. The cells were then cultured without antibiotics.

2.3.8 Cell proliferation assay

The cytotoxicity of the I-BET and the proliferation rate of MOZ K562 knockout cells was assessed by the MTS 3-(4,5-dimethyl-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (CellTiter 96® AQueous One Solution Reagent, Promega) assay. K562 cells were seeded into a 96 well-plate at a density of 5×10^3 cells per well and incubated at 37 ^oC and 5% of CO_2 for 24 hours. Afterwards, cells were exposed to different concentrations of I-BET (1, 5, 10 µg/mL) for 48 hours (except to MOZ knockout K562 cells).

Controls containing only medium and DMSO were also prepared. After the incubation time, 20 mg/mL of MTS solution was added in each well and then cells were left incubating for an additional 3 hours. Absorbance was measured using a BioTek microplate reader at 490 nm.

2.3.9 Apoptosis assays

2.3.9.1 *Annexin-V/ PI*

K562 cells were cultured on 12-well plates at a density of 7.5 x 10^4 cells per well for 24 hours in RPMI complete medium at 37 ^oC in 5% CO_2. Cell culture medium was replaced with the medium containing I-BET 151 at concentration of 1 μg/mL, 5 μg/mL and 10 μg/mL or the same volume of the vehile DMSO or only medium (controls) and cells were incubated for 48 hours at 37 ^oC in 5% CO_2. After the incubation period, supernatants were collected and the cells were centrifuged and washed with PBS. Afterwards, cells were resuspended in 100 μL of 1x annexin-binding buffer and 2.5 μL annexin-V FITC (Thermo Scientific) and 1 μL of 100 μg/mL PI solutions were added. The negative control sample cells were left untreated and unstained, while the other two control samples were also left untreated, but the cells were stained with annexin-V and PI, respectively. Cells were incubated in the dark for 15 minutes and then 400 μL of 1x annexin-binding buffer was added. Samples were immediately analysed using a Beckman Coulter FC 500 flow cytometer. Kaluza 1.5 software was used for the data analysis. For

analysis of MOZ knockout K562 cells, only untreated cells were used.

2.3.9.2 Detection of *caspase-3/7*

7.5 x 10^4 K562 cells were treated as previously described for annexin-V/PI assay. After centrifugation, cells were washed with PBS, resuspended, and then transferred to the FACS tubes. CellEventTM Caspase-3/7 Green Detection reagent at concentration of 500 µM was added to the tubes (1 µL) and cells were incubated for 30 minutes at 37 ^{o}C protected from light. After 25 minutes of incubation, 1 µL SYTOXTM AADvanced red at a concentration of 1 mM was added and cells were incubated for a further 5 minutes at 37 ^{o}C. Samples were then analysed without washing or fixing using a Beckman Coulter FC 500 flow cytometer and data were analysed using Kaluza 1.5 software.

2.4 Biochemical techniques

2.4.1 Composition of solutions used in biochemical methods

Lysis buffer: 20 mM sodium phosphate (pH 7), 250 mM NaCl, 30 mM sodium pyrophosphate, 0.1% NP-40, 10 mM NaF, 0.1 mM Na3VO4, 1 mM PMSF, 1x Complete Protease Inhibitors EDTA- free cocktail.

RIPA lysis buffer: 50 mM Tris-HCl pH 8, 150 mM NaCl, 1% NP-40, 0.1% SDS, 1 mM PMSF, 1% sodium deoxycholate (DOC), 1x complete protease inhibitor EDTA- free cocktail.

SDS-PAGE loading buffer (4x): 62.5 mM Tris-HCl pH 6.8, 40% glycerol, 2% SDS, 14.5 mM-β-mercaptoethanol and a trace of bromophenol blue.

Tris-glycine-SDS-PAGE running buffer (10x): 250 mM Tris base, 2 M glycine, 0.037% (w/v) SDS pH 8.3.

Tris-glycine-SDS transfer buffer (standard): 48 mM Tris base, 39 mM glycine, 0.037% (w/v), SDS 20% (v/v) methanol.

Brilliant Blue® staining solution: 0.25% brilliant blue® (w/v) in 45% methanol, 10% acetic acid.

Destaining/fixing solution: 45% methanol, 10% acetic acid.

Enhanced chemiluminescence (ECL) developing solution: 1 ml Tris-HCl pH 8.5, 6 µl H2O2, 22 µl Solution A (90 mM p-Coumaric acid in DMSO), 50 µl Solution B (250 mM luminol in DMSO) to a final volume of 10 ml with ddH2O

NETN buffer: 20 mM Tris-HCl pH 8.0, 100-500 mM NaCl, 1 mM EDTA, 0.5% NP-40, 1× Complete Protease Inhibitors.

Glutathione elution buffer: Freshly prepared, 30 mM reduced glutathione, 50 mM Tris-HCl pH 9, 150 mM NaCl, 0.1% Triton X-100.

Binding buffer: 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 0.1% NP-50, 1 mM PMSF, 1× Complete Protease Inhibitors.

Blocking solution (for western blotting): 5% (w/v) milk powder in 1× PBS.

Stripping buffer (for western blotting): 62.5 mM Tris-HCl pH6.8, 2% SDS, 100 mM 2-β-mercaptoethanol.

10× HBS solution: 1.37 M NaCl, 50 mM KCl, 8 mM Na2HPO4, 55 mM D-glucose, 210 mM HEPES.

2.4.2 Preparation of whole-cell extracts

Cells were harvested by aspirating the RPMI media and scraping the cell layer off in 1 ml sterile 1xPBS. Cells were collected by centrifugation at 1100 rpm, the supernatant removed and washed twice in PBS. Cells were resuspended in cold lysis buffer (2x cell compact volume) and incubated for 30 minutes, 4 ºC with rotation. The cells were then snap frozen in liquid nitrogen, thawed in cold water and run over an eppendorf rack 10 times to ensure efficient lysis. Samples were again snap frozen in liquid nitrogen and thawed in water a further two times. If samples were viscous an optional sonication step was included (3x 20 sec on/30 sec off). Cellular debris was pelleted by centrifugation at 12,000 rpm at 4 ºC for 30 minutes. The cell-free extract (supernatant) was transferred to a fresh tube and stored at -80 ºC.

2.4.3 Protein concentration determination

Protein concentrations were determined using the Bradford assay method and Bio-rad dye concentrate according to the manufacturer's instructions. Typically reactions consisted of 2 µl of sample and 1 ml 20% dye in ddH2O. The samples were mixed thoroughly and left to stand for 20 minutes at room temperature. The absorbance was read at OD595 and the concentration calculated relative to values of diluted BSA standards of known concentrations.

2.4.4 Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Cell extracts, immunoprecipitates and purified *in vitro* translated proteins were analysed by one-dimensional polyacrylamide gel electrophoresis. SDS-PAGE gels were prepared with the appropriate percentage resolving and stacking gel (Table 2.1), depending on the molecular weight of the protein being analysed. Protein extracts were denatured by boiling for 5 minutes in 4x SDS-PAGE loading buffer before being resolved on SDS-PAGE gel of concentration 8-15% depending on the size of protein. Gels were electrophoresised using a mini-gel system (Protean II, Bio-Rad) in 1x running buffer at fixed voltage.

Table 2.1 - Solution compositions for resolving and stacking gels for SDS-PAGE. The volumes of the components stated are for the production of one 5 ml or 2 ml volume of resolving or stacking gel respectively.

	Resolving gel			Stacking gel
% Acrylamide	**8**	**12**	**15**	
ddH2O (ml)	2.3	1.6	1.1	1.4
30% acrylamide mix (ml)	1.3	2.0	2.5	0.33
1.5 M Tris pH 8.8 (ml)	1.3	1.3	1.3	-
1.0 M Tris pH 6.8 (ml)	-	-	-	0.25
10% (w/v) SDS (ml)	0.05	0.05	0.05	0.02
10% (w/v) ammonium persulphate (ml)	0.05	0.05	0.05	0.02
TEMED (ml)	0.003	0.002	0.002	0.002

2.4.5 Staining of SDS-polyacrylamide gels

Gels were stained in 0.25% Brilliant blue with gentle shaking for 5 mins, followed by two cycles of 15 minutes of incubation with destain solution. The gels were then kept overnight in distilled water with shaking to completely destain the gel.

2.4.6 Western blotting and immunodetection

Proteins separated on SDS-PAGE gels for western blotting were transferred to nitrocellulose membranes by electrophoresis in 1x transfer buffer using wet transfer apparatus (Bio-Rad). Membranes were stained in 1% Ponceau S solution and washed with

sterile water to verify transfer efficiency. Nitrocellulose membranes containing immobilised proteins were incubated in blocking buffer for 1 hour at room temperature to block non-specific antibody binding sites. Membranes were then incubated with primary antibody at 4℃ overnight. Primary antibodies were prepared in blocking buffer at dilutions indicated in Table 2.2. The nitrocellulose membrane was washed three times (3x 5 minutes) in 1x PBS containing 0.1% (v/v) Tween 20, prior to incubation with secondary antibody for 90 minutes at room temperature. Horseradish peroxidise (HRP) conjugated secondary antibodies were diluted 1:3000 in blocking buffer. Excess secondary antibody was removed by three more washes (3x 5 minutes) in 1x PBS containing 0.1% Tween 20. Protein antibody complexes were detected by ECL chemiluminescence. Membranes were covered for 1 minute with 5 ml of ECL developing solution, excess reagent was gently drained off and the membrane covered in Saran wrap. The chemiluminescent signal was detected using a luminescent image analyser (Fujifilm LAS-4000). For re-use with other primary antibodies, membranes were stripped by incubation with stripping buffer for 30 minutes at 55℃ with occasional agitation. The membranes were then thoroughly washed before being incubated in blocking solution as above and re-probed with the next primary antibody.

Table 2.2 - Origin and Dilutions of Antibodies. Antibodies utilised during the research described in this thesis were obtained as detailed.

Primary Antibody	Dilution	Supplier
Anti-MOZ (29D2)	1:1000	upstate
Anti-MOZ (E17)	1:1000	Santa Cruz (sc-5715)
Anti-FLAG	1:1000	Sigma-Aldrich (F3165)
Anti-NUR77	1:1000	Abcam (ab109180)
Anti-NR2F2	1:1000	Abcam (ab41859)
Anti-p16	1:1000	Sigma (SAB3300036)
Anti-BRD4	1:1000	Cell Signaling (E2A7X)
Anti-H3K4me3	1:1000	Abcam (ab8580)
Anti-H3K14ac	1:1000	Upstate
Anti-H3K9ac	1:1000	Abcam (ab10812)
Anti-H2A	1:1000	Abcam (ab18255)
Anti-H3K14cr	1:1000	PTM BIO (PTM-535)
Anti-GST	1:1000	Sigma-Aldrich (G7781)
Anti-β-Actin	1:1000	Sigma-Aldrich (A5441)

2.4.7 Co-immunoprecipitation

Protein extracts (750 µg) obtained from mock, co-transfected or endogenous cells were diluted to 500 µl with lysis buffer. In order to pre-clear and remove unspecific extract binding proteins, 10 µl of Protein G Agarose beads (Roche) was added to each sample and incubated for 1 hour at 4 ˚C with rotation. Samples were spun at 2,000 rpm for 2 minutes to pellet the beads and the supernatant was

then transferred to a fresh tube. 2 µg of antibody and 20 µl of Protein G beads (pre-washed 3x in lysis buffer) were added and the samples incubated overnight at 4℃ with rotation (a no antibody control was also carried out with the cotransfected sample). Immunoprecipitants were washed three times in lysis buffer and run on an SDS-PAGE gel (Section 2.5.4) before transfer and immunodetection (Section 2.5.6).

2.4.8 Small-scale bacterial expression of GST-tagged proteins

Freshly transformed colonies were inoculated into 2.5 ml of LB-Ampicillin media. Following 2 hours of growth at 37℃ with 220 rpm shaking, 1 ml was transferred into two microcentrifuge tubes. 25 µl of IPTG inducer (final concentration 0.5 mM) was added to one, while 25 µl of distilled water was added to the other as control. The remaining 0.5 ml was stored at 4℃ for future inoculation of large-scale cultures. After a further 2 hours of growth at 37℃, cells were harvested by centrifugation at 13,000 rpm for 5 minutes and the pellets were resuspended in 15 µl 4x SDS loading buffer. Samples were boiled for 5 minutes and loaded onto an SDS-PAGE gel. Induced cultures were run in conjunction with the appropriate control culture, followed by Brilliant blue staining to detect the presence of an induced band.

2.4.9 Large-scale bacterial expression of GST-tagged proteins

Bacterial colonies which showed optimal protein induction were inoculated in 50 ml LB medium (with appropriate antibiotic) and

left to grow overnight at 37℃ with 220 rpm shaking. The following day 1 L of LB (+antibiotic) was inoculated using the overnight culture and grown at 37℃ until the exponential phase (OD600 0.8-0.9). A 1 ml sample (uninduced) was transferred to a microcentrifuge tube and IPTG was added to a final concentration 0.4 mM. The cultures were incubated for a further 3-16 hours at 20-37℃ (conditions optimised for each protein) and subsequently a second 1ml sample was taken representing the induced fraction. Cells were harvested by centrifugation at 5,500 rpm for 20 minutes at 4℃ (Sovrall Evolution RC rotor SLA-1500) and the pellet was frozen until ready for purification.

2.4.10 GST-pulldown of histone extracts

Normalised levels of *in vitro* translated protein or 1.5 µg of purified GST-tagged fusion protein GST-MOZ 194-323 was incubated with approximately 1.5 µg or 735 nM of biotin conjugated histone peptide (Millipore/Peptide Protein Research) or no peptide as control. Reactions were incubated overnight with rotation at 4 °C in 750 µl Binding buffer (buffer composition section 2.5.1., p125). A slurry of 15 µl Dynabeads® (M-280 Streptavidin-Invitrogen) were prewashed twice in 1x PBS + 0.1 % Tween and added to each reaction. The incubation with Dynabeads® took place for 1 hour at 4 °C. Dynabeads® were harvested using a magnet and consequently washed (3x10 minutes) in binding buffer. Samples were then

subjected to SDS-PAGE (section 2.5.5) and results visualised by immunodetection (section 2.5.6) with an α-GST antibody (Sigma).

2.4.11 RNA Seq and data analysis

Total RNA from K562 cells transfected with MOZ siRNA and control siRNA were used for RNA-seq analysis. Library preparation was performed with TruSeq RNA Sample Preparation Kit v2 of Illumina (San Diego). The prepared libraries were validated and quantified with Agilent Bioanalyzer. All samples had RNA Integrity Number (RIN) above 9.5. Total RNA was used to prepare mRNA-Seq libraries using the TruSeq RNA Sample Prep kit following the protocol described by the manufacturer (Illumina, San Diego, CA). Libraries were constructed by the Barts and London Genome Centre and sequenced on an Illumina NextSeq-500 instrument with 150 cycle mid-output v1 reagents, according to manufacturer's instructions. Fastq files were used to map reads to the human genome (hg19, UCSC) utilizing Tophat2 with default settings. Differential gene expression was performed by comparison of biological duplicate samples in Genomics Suite using analysis of variance (ANOVA). Differential gene expression values were then filtered by log2 fold change and by unadjusted P ≤.05. The sequence reads were received from the Barts and London Genome Centre as FASTQ files containing the forward and reverse reads. TopHat [33] was used to align the sequencing reads to a reference genome. Alignments reported from TopHat were processed by the Cufflinks

software package to determine differential expression of genes and transcripts between conditions. The gene expression levels were averaged over two biological replicates and expressed as fragments-per-kilobase-of-gene-per-million-mapped reads (FPKM). Transcripts were quantified by assessing the total number of reads for the entire transcript using the program cuffdiff. The log2 ratio of fold change (log2FC) of the gene expression value between MOZ silenced samples and control was calculated according to Bullard et al. [266] study. $p<0.05$ was considered significant.

2.4.12 Bioinformatic analysis

Functional annotation analysis (Database for Annotation, Visualization and Integrated Discovery version 6.7, DAVID) was performed with input of all 76 down regulated genes. Enrichment was regarded as significant if fold enrichment was ≥ 1.5 and p-value < 0.05. The KEGG pathway analysis of the DEGs was performed using Cytoscape software (version 2.6.2) (http://www.cytoscape.org/) with the ClueGO plugin (http://www.ici.upmc.fr/cluego/cluegoDownload.shtml). FunDO analysis was performed with all differential expressed genes. The statistical test used for the enrichment was based on right-sided hypergeometric option with a Benjamini-Hochberg correction and kappa score level ≥ 0.35. Integrative Genomics Viewer [35] was used to analyse read counts.

2.5 Statistical Analysis

Data are expressed as mean ± standard error of the mean (SEM). Student's t-test was used for comparision analysis between two sample means. For multiple comparison analysis, 2-way ANOVA was used. In all cases, a p-value of less than 5% was considered statistically significant.

Chapter 3 : Identification of KAT6A-regulated genes in K562 cells using siRNA Knockdown

3.1 Background

MOZ is essential for vertebrate development and knock-out studies in mice revealed a requirement for haematopoiesis and self-renewal of hematopoietic stem cells and progenitor cells [203, 213, 267, 268]. As mentioned earlier, dysfunction of MOZ and a relationship to cancer, was first observed in acute myeloid leukaemia (AML) patients harbouring recurrent reciprocal chromosomal translocations generating MOZ fusion proteins, that were later shown to play a causative role in AML- a heterogeneous hematopoietic stem cell disorder where the myeloid cells do not enter into the differentiation process, resulting in accumulation of blast cells and absence of a fully functional granulocyte lineage [230, 231, 255, 269, 270]. Our research group showed that expression of MOZ-TIF2 in different cell lines causes broad changes in histone modifications and disturbs transcription by CBP-dependent transcription factors such as p53 and retinoic acid receptors [252, 253]. On the other hand, the transcriptional activity of Runx factors, involved in regulation of blood and bone differentiation, is enhanced by MOZ-TIF2 in reporter assays [202].

Mechanistically, it is known that MOZ acts as a cofactor for certain transcription factors, particularly those with haematopoietic or neuronal specificity, such as acute myeloid leukaemia/Runt-related transcription factor 1 (AML1/RUNX1) [188, 195] and PU.1 [203]. Thus, we decided to investigate which genes have their expression

regulated by MOZ in a leukemic cell line to fully understand its importance in cell function and disease conditions.

Initial studies focussed on a small number of genes including HOXA5 and HOXA9 as potentially requiring MOZ for their expression [217, 224, 271]. High expression of some HOX genes can be observed in acute myeloid leukemias and are essential for development and differentiation of many tissues [221, 272]. More recently transcription profiling has been performed to identify MOZ-dependent genes in selected normal mouse tissues such as memory B-cell [273], B lymphocytes [274], embryos and ES cells [224] and in human SUM-52 cells, a luminal breast cancer cell line [164]. Also, several studies have identified genes whose expression is differentially regulated by MOZ-CBP or MOZ-TIF2 expression [221, 275]. However, there was limited overlap between the gene lists generated in these studies indicating that MOZ may influence different gene networks in distinct tissues.

To address this issue, we decided to investigate MOZ function in the human erythroleukemia cell line K562, which expresses robust levels of MOZ, although interestingly shows low expression of most HOX genes. We employed a standard approach of using siRNA to perform *MOZ* knockdown followed by high-throughput RNA sequencing (RNA-seq). As the project developed, CRISPR CAS9 technology became available and we used this to generate MOZ deficient K562 cells, as will be shown in the next chapter. Here we

present evidence for MOZ-regulated genes in K562 cells that are important in haematopoiesis and leukaemia.

3.2 MOZ silencing in K562 cells using siRNA

To gain insight into the function of MOZ/KAT6A in leukemia we chose to investigate its requirement for gene expression in the human erythroleukemia cell line K562. The cells (purchased from ATCC) were originally isolated from a 53 year old female patient with chronic myelogenous leukemia in blast crisis [276]. Previous studies have demonstrated that MOZ shows robust expression in K562 cells [222] and chromatin IP data from our group showed that MOZ is enriched in K562 chromatin containing acetylated H3K14 [190]. A routine approach to study gene function is to transfect cells with small interfering RNA oligonucleotides (siRNA) that specifically target the gene in question. siRNAs are robust synthetic reagents consisting of double-stranded RNA with short single-stranded ends for short-term post-transcriptional silencing [277]. Although suppression of gene expression using siRNA is transient, this method has the advantage of been rapid, relatively easy to perform and it is highly specific [278]. Usually, synthetic RNAs used for this purpose are 19-25 bp duplexes because duplexes greater than 30 bp can induce interferon response. Once a duplex RNA is transfected into cells, it binds to the RNA induced silencing complex (RISC) and the double stranded siRNA separates into single strands. The antisense strand binds to RNA-induced silencing complex (RISC)

that cleaves the complementary target mRNA by its endonuclease enzyme component known as Argonaute. Consequently, the targeting mRNA is destroyed and the cell will have a knockdown in terms of protein expression levels [279].

Thus, to knockdown MOZ in K562 cells, a pool of four siRNAs was purchased from Dharmacon, targeting human KAT6A sequences (5'-GGAAUUAUUUCGUGAUAUC-3', 5'-GGAGUUGAG UGUUAAAGAU-3', 5'-GCGCUAUACUAAUCCAAUA-3', 5'-GUCA GACCCUGCAGAGUUA-3') and a randomised non-targeting siRNA pool as control. These siRNAs are optimised for efficacy and to avoid off target hits, thus the four siRNAs within the pool target exons 3,7,9 and 18 within the KAT6A transcript (Materials & Methods). Blast (Basic Local Alignment Search Tool - NCBI) searches using these siRNA sequences confirmed MOZ target specificity, and the lowest effective concentration of siRNA pool was used to further minimise potential off-target effects. Finally, both sense and antisense strand of the siRNAs used are chemically modificated with 2'-O-methyl (2'-OMe) nucleotides. This modification inhibits the expression of a completely-matched target gene due to an alteration in the steric conformation of the siRNA guide strand disturbing the base-pairing with target transcripts. This modification also induces a steric hindrance in duplex formation on the argonaute protein disturbing the association of the siRNA with this protein [280, 281].

Preliminary experiments performed suggested that 10nM siRNA pool was sufficient to strongly reduce MOZ protein detection

in K562 cells using RTqPCR and western blots (data not shown). Therefore, for transfection of proliferating K562 cells, $5x10^5$ cells were seeded in 6 well plates with 1 mL of RPMI complete medium, and transfected with 10nM siRNA pools using INTERFERin as described in Materials and Methods. After 5 hours, transfected cells were supplied with fresh medium and subsequently harvested at 48, 72 and 96 hours post-transfection for determination of protein and mRNA levels. Each siRNA transfection was performed in duplicate and experiments repeated 3 times.

In order to quantify MOZ transcript levels after the siRNA transfection, qRT-PCR was performed and revealed successful reduction in MOZ mRNA levels compared to control (cells transfected with non-targeting siRNA pool) in all three different time points analysed (Figure 3.1). After 48 h transfection, was observed approximately a 70% decrease in MOZ mRNA level compared to control; approximately 40% decrease after 72 h and around 70% decrease after 96 h transfection. The mRNA expression was measured relative to the expression of the endogenous reference gene GAPDH.

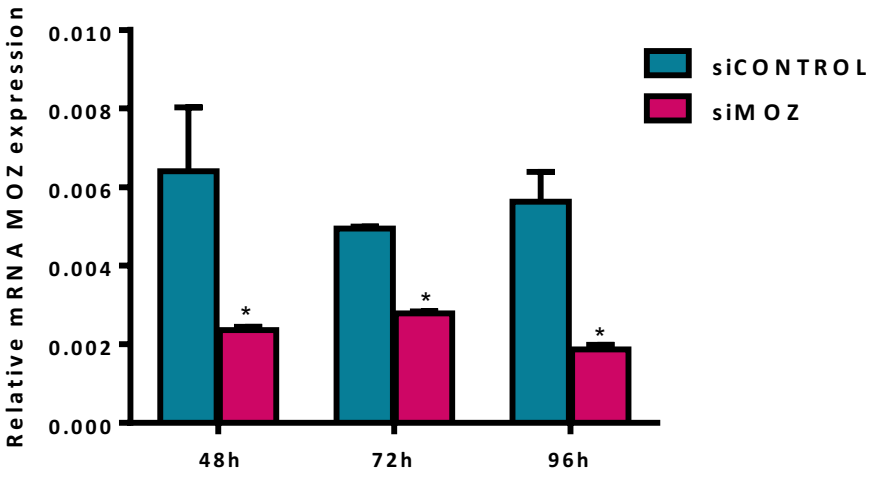


Figure 3.1 - Knockdown of KAT6A/MOZ in K562 cells. Quantitative RT-PCR validation KAT6A/MOZ silencing (pink bars) by transfecting K562 cells with a pool of KAT6A siRNAs (siMOZ). A pool of scrambled siRNAs (siSCR) was used as negative control (blue bars). Three different time points were assessed for analysis: 48, 72 and 96 hours post-transfection. Biological triplicates were performed for both MOZ siRNA and scrambled siRNA. Error bars, S.E.M. The asterisk indicates a significant difference ($P<0.05$) between control and MOZ-silenced samples.

As expected, reduction in MOZ transcript levels resulted in a strong reduction in detectable MOZ protein as confirmed by western blots (Figure 3.2). Two controls were used: non-transfected K562 cells and K562 cells transfected with non-target pool siRNA.

Measurement of MOZ protein expression was taken using densitometry and the result shows approximately 70% decrease compared to siControl after 48, 72 and 96 hours transfection. β-actin served as endogenous reference protein.

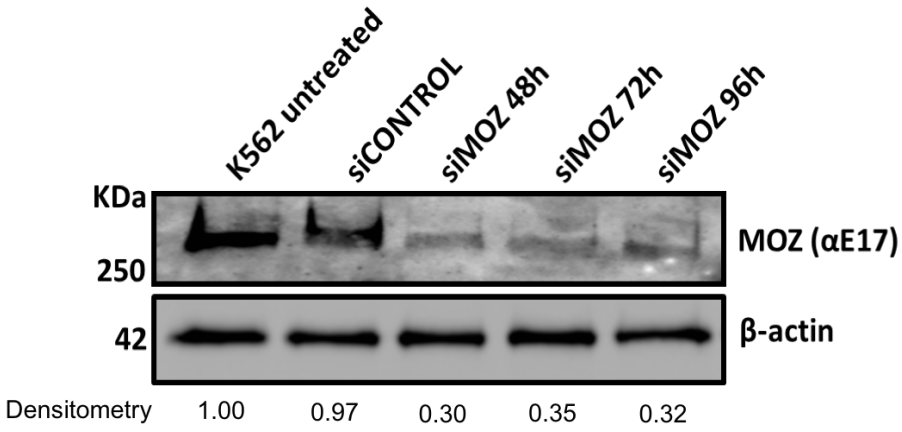


Figure 3.2 - Knockdown of MOZ protein in K562 cells. Western blot validation of MOZ-silencing cells after 48, 72 and 96 h siRNA transfection. siControl is related to cells treated with non-target pool siRNA. The total protein lysates for each siRNA treatment were equally loaded onto a protein SDS-polyacrylamide gel after measuring the concentration. The specific band for MOZ protein appeared above the 250-kDa band from the protein ladder and β-actin was used as a reference protein

3.3 Knockdown of MOZ in K562 cells reduces acetylation of H3K14 and H3K9

Having established that MOZ can be selectively depleted in transfected K562 cells using siRNA transfection, the next aim was to understand the functional consequences and to try to identify genes whose expression is regulated by this histone acetyltransferase. Our group and others have shown that MOZ acetylates H3K14 and H3K9 [195] [217] [282] [264] [190]; [283] thus to determine whether knockdown of MOZ alters the detectable levels of these histone modifications, bulk histones were extracted from K562 cells at 72h post-transfection with MOZ siRNA and siRNA control. Western blots confirmed that MOZ depletion resulted in a reduction in acetylated

H3K14 and H3K9 levels, whereas tri-methylated H3K4 level was unaffected (Figure 3.3). The incomplete reduction in H3K9ac and H3K14ac may indicate the following: a) the experiment contains untransfected cells which have unaltered MOZ expression and therefore retain these acetylation marks, or B) other histone acetyltransferases such as HBO1, MOF or TIP60 may be responsible for generating these acetylation marks across the genome. It may also be that only specific chromatin regions where MOZ is associated e.g. promoters or enhancers of genes that are regulated by MOZ are depleted in these H3 acetylation marks. Nonetheless, we can confirm that MOZ depletion leads to a reduction in global levels of H3K9 and H3K14 acetylation events in K562 cells.

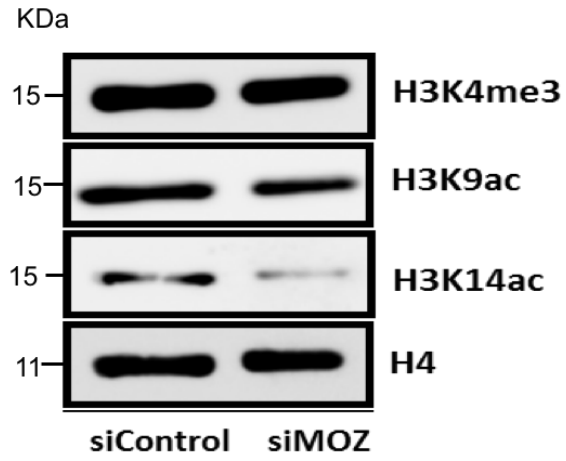


Figure 3.3 - Effect of MOZ knockdown in histone modification. K562 cells were transfected with siMOZ and control siControl for 72 hours and histone acetylation level was analyzed by Western blotting with anti-H3K4me3, anti-H3K9ac, anti-H3K14ac, antibodies. H4 was used as loading control.

Interestingly, while H3K14ac and H3K9ac are present on active genes as discussed in the Introduction, a subset of inactive

promoters are also selectively enriched for H3K14ac [284]. In consequence, since MOZ knockdown causes a decrease in H3K14 acetylation we can expect to identify genes that are differentially expressed (up or downregulated) following MOZ knockdown.

3.4 Transcription profiling of MOZ/KAT6A depleted K562 cells

To identify genes that are regulated by MOZ in K562 cells, we established transcription profiles for control siRNA treated and MOZ-depleted K562 cells using RNA-seq. Approximately $5x10^5$ K562 cells were treated with MOZ-targeting siRNA or non-targeting siRNA for 48 hours in duplicate. This time point was chosen because it presented lower level of MOZ protein as shown in Figure 3.2. Total RNA was extracted using Qiagen RNeasy kit according to manufacturers' instructions. RNA quality and quantity in each sample was assessed on nanodrop 10000 spectrophotometer (Nano Drop Technologies Inc; USA). The ratio of absorbance at 260 nm and 280 nm was used to assess the purity of RNA and a ratio of ~2.0 was accepted as good quality, once a lower ratio may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm. The presence of contaminations was also accessed by analysing the 260/230 values, which was considered acceptable values in the range of 2.0-2.2 as recommended by the manufacturer. Further RNA quality and concentration evaluation was performed using Agilent 2100 Bioanalyser "Agilent Technologies".

To access RNA integrity, RNA samples were separated on microfabricated chips and subsequently detected via laser induced fluorescence detection. An electropherogram was generated for each sample and a RNA Integrity Number (RIN) based on the ratio of the 18S to 28S ribosomal subunits was provided by the Bioanalyzer software as shown in Figure 3.4. The RIN software algorithm classifies eukaryotic total RNA, based on a numbering system from 1 to 10, with 1 being the most degraded profile and 10 being the most intact. As can be seen in Figure 3.4, all the samples had very good quality and integrity with RIN value 10.

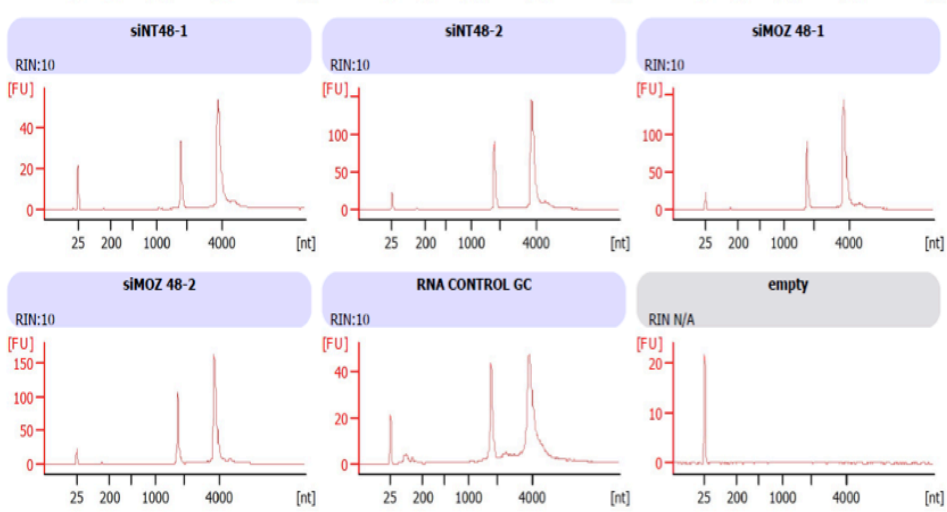


Figure 3.4 - **RNA quality analysis.** Eletropherogram showing RNA integrity from K562 cells transfected with Non-targeting siRNA (siNT) and MOZ targeting siRNA (siMOZ) for 48 hours. A marker peak is shown in 25 nucleotide region and the two ribosomal RNA peaks, the 18S and the 28S are shown around 2500 and 4000 nucleotide region respectively. RNA Integrity Number (RIN) based on the ratio of the 18S to 28S ribosomal subunits was provided by the Bioanalyzer software.

Once the RNA samples were in confirmed as high quality, we sent them to the Barts and the London Genome Centre, Queen Mary, University of London (http://www.smd.qmul.ac.uk/gc/) for sequencing. The first steps involved construction of the cDNA libraries performed with TruSeq RNA Sample Preparation Kit v2 of Illumina (San Diego) and prepared libraries were validated and quantified with Agilent Bioanalyzer. Library construction includes mRNA purification and fragmentation, first-strand cDNA synthesis, and second-strand cDNA synthesis. The double-strand cDNA was then ligated to adapters consisting of known DNA sequences and sequenced in a high-throughput manner to obtain short sequences of approximately 100 bp from both ends (pair-end sequencing). Sequencing was performed on an Illumina NextSeq-500 instrument with 150 cycles mid-output v1 reagents, according to manufacturer's instructions. Sequence reads were received from the Barts and London Genome Centre as FASTQ format, containing the forward and reverse reads to be analysed. The computational pipeline for RNA-Seq data analysis is constituted for data quality assessment, read aligns and gene expression estimation, and statistical analysis that were performed by our collaborator Dr Nigel Mongan (School of Veterinary Medicine & Sciences, University of Nottingham). Raw reads were quality and adapter-filtered using the Trim-galore wrapper for FastQC and cutadapt (http://www.bioinformatics.babraham .ac.uk/projects/trim_galore/). FASTQ files were used to map reads to the human genome (hg19, UCSC) utilizing Tophat2

(http://ccb.jhu.edu/software/tophat/index.shtml) with default settings. Alignments reported from TopHat were processed by the Cufflinks software package (http://cole-trapnell-lab.github.io/cufflinks/) to determine differential expression of genes between conditions. Gene expression levels were averaged over the two biological replicates and expressed as fragments-per-kilobase-of-gene-per-million-mapped reads (FPKM). Differential gene expression was performed by comparison of biological duplicate samples in Genomics Suite using analysis of variance (ANOVA). Differential gene expression values were then filtered by log2 fold change and by unadjusted $p \leq 0.05$. Transcripts were quantified by assessing the total number of reads for the entire transcript using the program Cuffdiff. The log2 ratio of fold change (log2FC) of the gene expression value between MOZ knockdown samples and control was calculated according to Bullard, Purdom [266] study with $p<0.05$ considered significant. Knockdown of MOZ expression was confirmed as reported by the number of reads across the MOZ genomic region encoding Exon 3 (comprising the initiation codon) using the Integrative Genomics Viewer (IGV) (Figure 3.5).

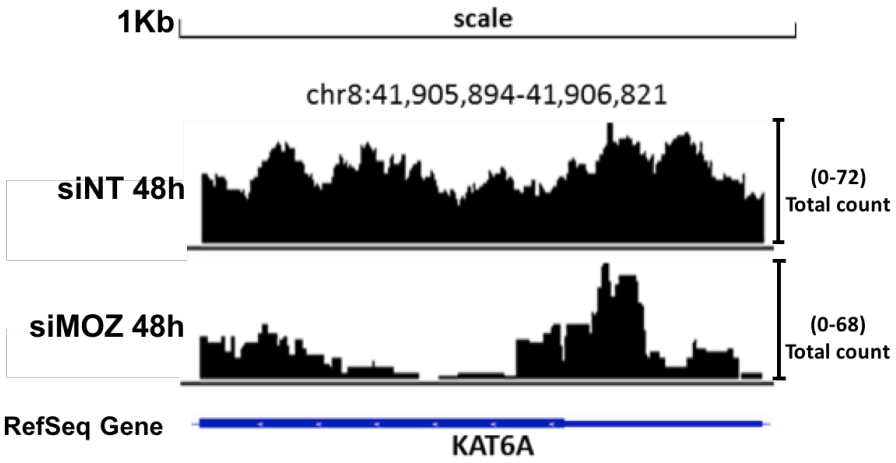


Figure 3.5 - Comparision of MOZ reads between non targeting siRNA (NT) and MOZ siRNA (siMOZ) transfected K562 cells across MOZ genomic region. Direct comparison of the read counts from MOZ-silenced and negative control samples across the MOZ genomic region encoding Exon 3 using *the Integrative Genome Viewer (IGV)*.

3.5 Detection of alternatively spliced isoforms of the MOZ/KAT6A gene

At the outset of this project, the NCBI Gene and RefSeq databases listed three isoforms of the KAT6A transcript resulting from differential splicing at the 5' of the transcript. In particular, this centred on alternative splicing events within Exon 2, which is upstream of the protein-coding region that begins within Exon 3. These were referred to as exon 2 Long (132 bp) and Exon 2 short (88 bp) due to alternative splice junctions (Figure 3.6). Under this nomenclature, KAT6A Variant 1 contains Exon 2 Long; KAT6A Variant 2 contains Exon 2 Short, and KAT6A Variant 3 lacks Exon 2 altogether. However, in 2015 this was revised in the NCBI database and reduced a single isoform containing Exon 2 long due to lack of

experimetnal evidence. However, KAT6A variants 1-3 are still represented in the UCSC Genome Browser and there is evidence for the detection of these 5' isoforms within the GTEX portal and in EST sequence databases.

To assess whether these isoforms were observed in K562 RNA Seq data, we visualised splice junctions at the 5' end of the KAT6A genes using Sashimi Plot in the Integrated Genome Viewer. This was consistent with the expression of all three splice variants in K562 cells (Figure 3.7).

>MOZ.Variant 1

```
     1 actgtctcca agatggcggc cgtgtcagtt tggggcatct ccgcggtccg gcccggggcc
    61 ccgggatctc ggctgtcctt cctcccggat tctttctact aatccagata cttgttgaag
   121 tgctgactag tttcttgggg aaagatgtca aggaagagtt gaaattccta gactggtgag
   181 cagagattgc ggagtacaga atagccatca cccactctgg cataagatgc acatttttct
   241 gctctggagc cgggaatgaa atattcttga gttcttacaa ctttatgacg agacccatgt
   301 gtggtgctat tgagaaattc attgggaagt tggaagacat ttcaatcaac aggttgtttt
   361 ggtttctata gtacaattgg ggtggcattc tgttttgtga aaggaggaag gacttaggcc
   421 agaaaactca tatgctatgg ttaactggtt cccagcctcc gagaatcttg ttttccatgg
   481 tgtaaaactt actcagcatc aggataaggg ataacgactc tatggatata cagaatcctt
   541 caccATGgta
```

>MOZ.Variant 2

```
     1 actgtctcca agatggcggc cgtgtcagtt tggggcatct ccgcggtccg gcccggggcc
    61 ccgggatctc ggctgtcctt cctcccggat tctttctact aatccagata cttgttgaag
   121 tgctgactag tttcttgggg aaagatgtca aggaagagtt gaaattccta gactggcata
   181 agatgcacat ttttctgctc tggagccggg aatgaaatat tcttgagttc ttacaacttt
   241 atgacgagac ccatgtgtgg tgctattgag aaattcattg ggaagttgga agacatttca
   301 atcaacaggt tgttttggtt tctatagtac aattggggtg gcattctgtt ttgtgaaagg
   361 aggaaggact taggccagaa aactcatatg ctatggttaa ctggttccca gcctccgaga
   421 atcttgtttt ccatggtgta aaacttactc agcatcagga taagggataa cgactctatg
   481 gatatacaga atccttcacc A TGg
```

>MOZ.Variant 3

```
     1 actgtctcca agatggcggc cgtgtcagtt tggggcatct ccgcggtccg gcccggggcc
    61 ccgggatctc ggctgtcctt cctcccggca taagatgcac atttttctgc tctggagccg
   121 ggaatgaaat attcttgagt tcttacaact ttatgacgag acccatgtgt ggtgctattg
   181 agaaattcat tgggaagttg gaagacattt caatcaacag gttgttttgg tttctatagt
   241 acaattgggg tggcattctg ttttgtgaaa ggaggaagga cttaggccag aaaactcata
   301 tgctatggtt aactggttcc cagcctccga gaatcttgtt ttccatggtg taaaacttac
   361 tcagcatcag gataagggat aacgactcta tggatataca gaatccttca ccATGG
```

Figure 3.6 - Sequences of MOZ/KAT6A variants showing alternative splicing at the 5' end of the transcript. Exon 1 (87 bp) is highlighted in yellow; Exon 3 is highlighted by grey shading; Exon 2 Long (132 bp) and Exon 2 Short (88bp) are shown in blue.

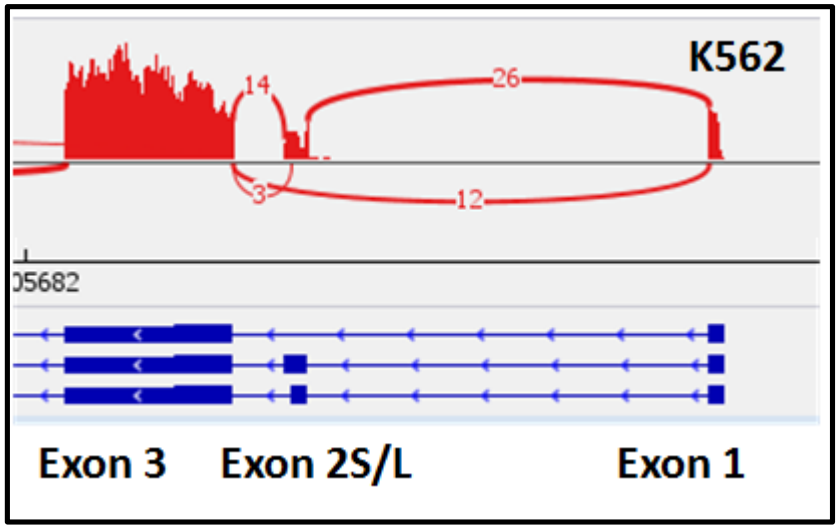


Figure 3.7 - Sashimi Plots depicting the detection of KAT6A 5' Splice variants in K562 and HEK293 RNA Seq data. Exons 1, 2S, 2L and 3 are shown. As indicated splice junctions represent ing all three variants are detected in both RNA Seq replicates for K562. However, splice events using Exon 2S (Variant 2) were below detected.

To confirm the expression of all three variants in K562 cells, we designed PCR primers to detect Variants 1-3 or variant 2 only (Materials and Methods). PCR analysis was performed using cDNA prepared from K562 cells. As shown in Figure 3.8, we successfully detected all 3 variants albeit at different levels using the non-selective primers, whereas a single band of the correct expected size was obtained using the variant 2-specific primers. Similar results were observed using cDNA from HEK293 cells. Thus, we have confirmed alternative splicing at the 5' end of the MOZ/KAT6A transcript in K562 cells, although the functional relevance of this remains to be determined.

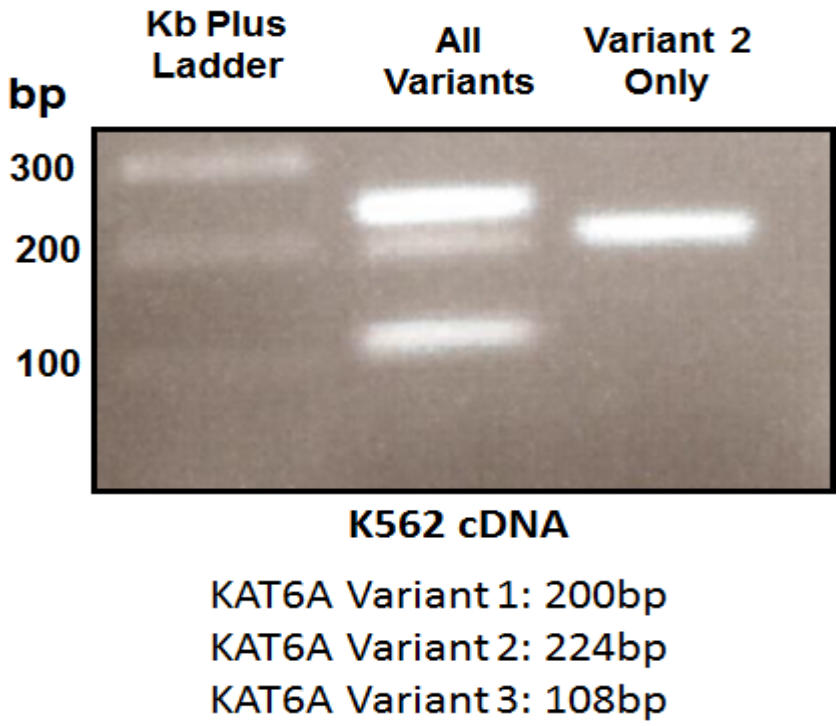


Figure 3.8 - PCR detection of KAT6A 5' splice variants in K562 cells. PCR was performed using cDNA prepared from K562 cells. the primers used were capable of amplifying all three variants (Lane 2) or were selective for Variant 2 containing the Exon 2S splice junction. The data suggest that variant 2 has the lowest expression, consistent with GSET data.

3.6 Identification of MOZ–regulated genes in K562 cells

Comparison of the transcriptomes of the K562 control and MOZ-depleted cells identified 406 genes that showed differential expression. Among these, 76 genes were deemed to be significantly down-regulated and 330 genes were up-regulated ($p<0.05$). A list of these differentially expressed genes (DEGs) is presented in Table A.2.

To visualise the DEGs, heat map was generated (assisted by my Supervisors). A heatmap was generated from replicated

experiments using the MOZ-silenced sample and control. As can be seen in Figure 3.9, MOZ knockdown had a significant effect on gene expression.

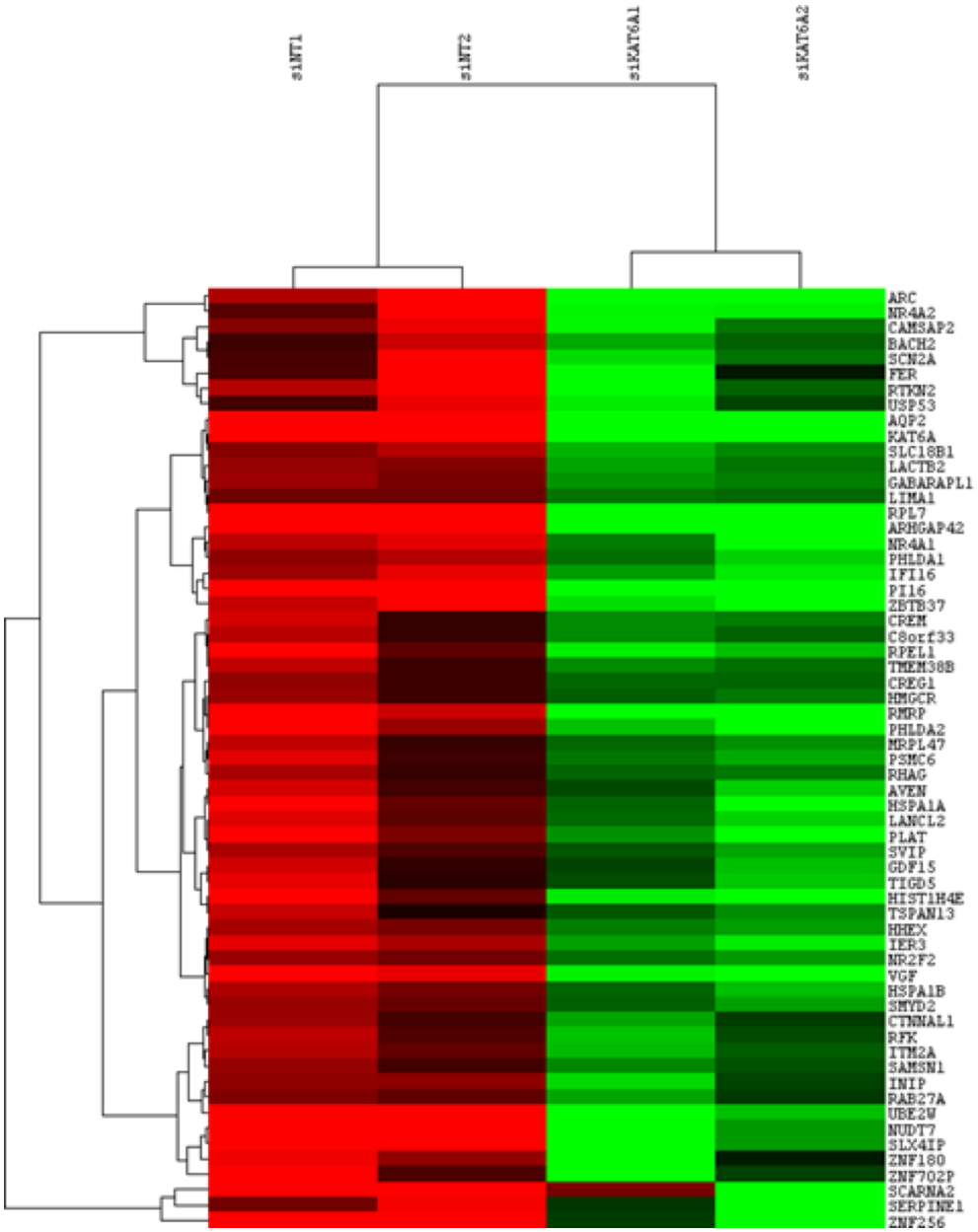


Figure 3.9 - Heatmap representing individual expression data of the genes differentially expressed in K562 MOZ silenced cells. Rows represent the 76 genes differentially expressed (downregulated) between control and MOZ silenced K562 cells ($p < x$, limma t-test), sorted by fold change. Columns represent the log2 value of the expression of each gene in each cell condition.

3.7 Validation of DEGs in K562 after KAT6A knockdown

To validate the differentially expressed genes identified by the RNA-seq data, RTqPCR was carried out as described in Materials and Methods, using the same RNA samples used for sequencing in addition to RNA from independent experiments performed under the same conditions. Initially, we focussed on 10 down-regulated genes (RPPH1, SIRT4, SERPINE1, HHEX, RAB27A, NR4A1, NR4A2, PROX1, PLAT, NUDT7) and 12 up-regulated genes (XIST, ID1, ID3, MCAM, RUNX1, LRIG3, MAP3K6, MYO1, TSPAN6, UCHL, CDKN2A, MAGED1). Several of these candidate genes chosen for their potential relevance to leukaemia. For example, NR4A1, NR4A2, HHEX and PROX1 are transcription factors that are known to be important in haematopoietic differentiation and leukaemia as discussed below. PLAT encodes the serine protease tissue plasminogen (t-PA) activator which plays a key role in blood clotting and its expression has been shown to be elevated in leukemia patients especially in AML (Wada 1993). K562 cells express and secrete tPA and its inhibitor PAI-1/2 (Oliver Blood 1989). Interestingly PLAT is located very close to KAT6A on chromosome 8p11. qRT-PCR was carried out in triplicate and normalized to GAPDH. The list of primers used for qRT-PCR validation is shown in Material and Methods.

As shown in Figure 3.10, all genes in the validation set for down-regulated genes showed a significant difference in expression compared to control across multiple experiments, confirming a strong agreement between the RNA-seq and qRT-PCR results.

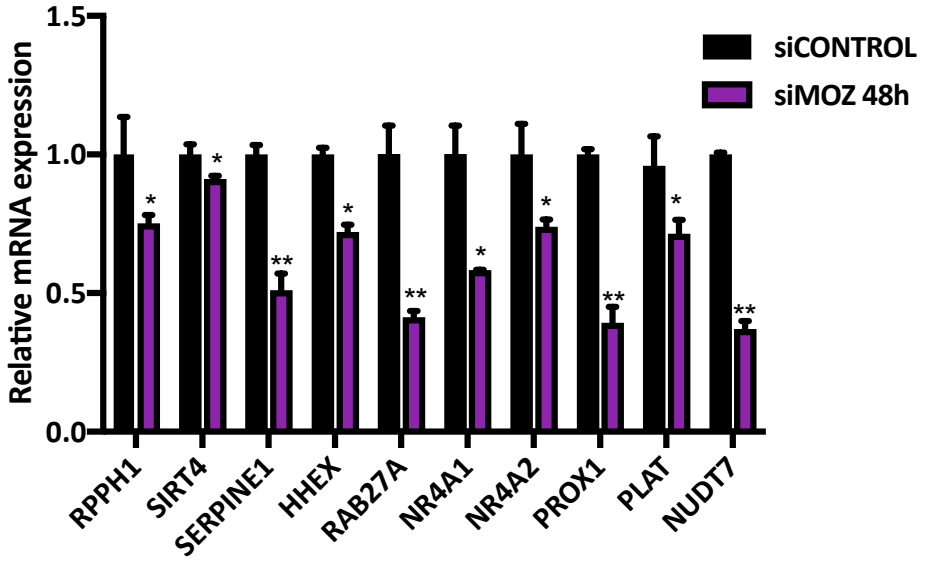


Figure 3.10 - qPCR validation of selected down-regulated genes in MOZ knockdown K562 cells. Quantitative RT-PCR showing changes in mRNA transcript levels between siControl (black bars) and siMOZ (purple bars) transfected K562 cells. Ten genes that were identified as down-regulated by RNA-seq analysis were chosen for validation. qRT-PCR data were normalized to GAPDH expression for each sample and the mean expression level in the control samples set to 1. One asterisk indicates a significant difference ($P<0.05$) and two asterisk $p<0.01$ based on student's t-test analysis between control and MOZ-silenced samples. Error bars, S.E.M.

Similarly, we chose a validation set for the DEGs that the RNA-seq indicated were upregulated upon MOZ depletion, suggesting they are directly or indirectly subject to negative regulation by MOZ. As shown in Figure 3.11 all the twelve genes analysed (XIST, ID1, ID3, MCAM, Runx1, LRIG3, MAP3K6, MYO1, TSPAN6, UCHL, CDKN2A and MAGED1) showed a significant increase in expression compared to control, as expected. Note that

for some of these genes the basal level in the control cells was very low, suggesting they are not expressed, but that MOZ depletion leads to some expression, albeit at a low level.

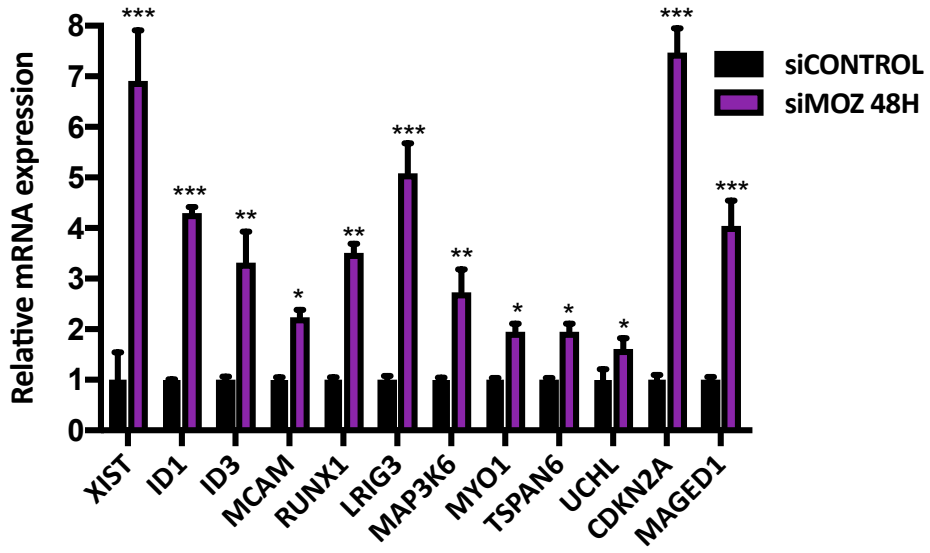


Figure 3.11 - qPCR validation data of selected up-regulated genes in MOZ knockdown K562 cells from RNA-Seq data. Quantitative RT-PCR showing changes in mRNA transcript levels between siControl (black bars) and siMOZ (purple bars) transfected K562 cells. Twelve genes that were identified as up-regulated by RNA-seq analysis were chosen for validation. qRT-PCR data were normalized to GAPDH expression for each sample and the expression level in the control samples set to 1. One asterisk indicates a significant difference ($P<0.05$), two asterisk $p<0.01$ and three asterisk $p<0.001$ based on student's t-test analysis between control and MOZ-silenced samples. Error bars, S.E.M.

We further validated 4 of the putative MOZ-dependent genes (NR4A1, NR4A2, PROX1 and HHEX*)* across the different time points to ensure that gene expression in K562 cells after 72 and 96 hours siRNA treatment (Figure 3.12). These data confirmed the decreased expression of these genes compared to controls.

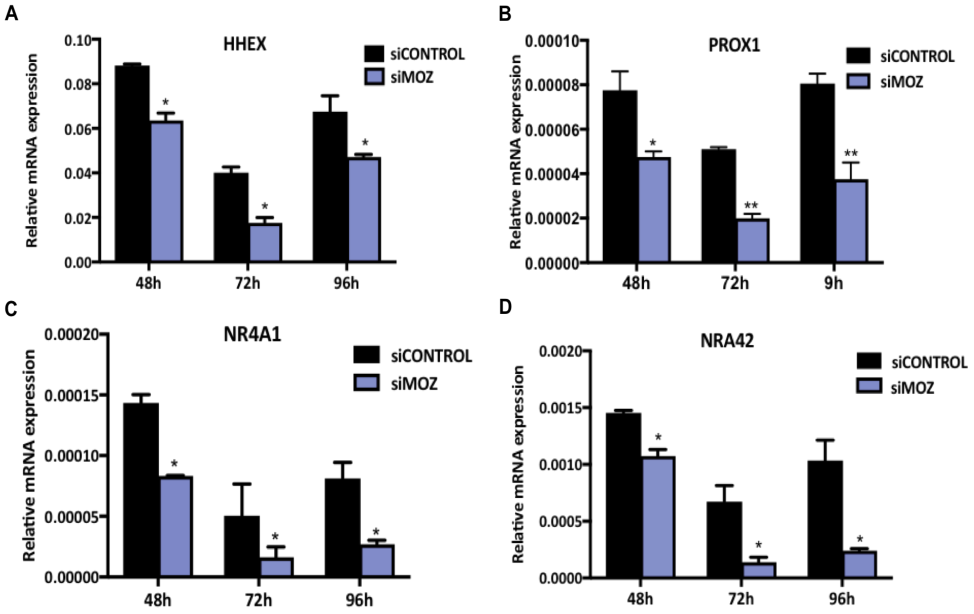


Figure 3.12 - Timecourse qPCR validation data of selected downregulated-regulated genes in MOZ knockdown K562 cells from RNA-Seq data. Quantitative RT-PCR showing changes in mRNA transcript levels of (A) HHEX (B) PROX1 (C) NR4A1 and (D) NR4A2 between siControl (black bars) and siMOZ (purple bars) transfected K562 cells. qRT-PCR data were normalized to GAPDH expression for each sample and the expression level in the control samples set to 1. Asterisk indicates a significant difference ($P<0.05$) based on student's t-test analysis between control and MOZ-silenced samples. Error bars, S.E.M.

To further validate these data, we obtained antibodies to assess the protein levels of NR2F2, HHEX, NR4A1 and CDKN2A in K562 cells treated with siMOZ or siControl after 48, 72 and 96 hours. Whole cell extracts were prepared as described in Materials and Methods. In addition, I included K562 cells that had not been subjected to transfection procedure. As shown in Figure 3.13, NR2F2, HHEX and NR4A1 protein levels were all found to be lower in MOZ-depeleted cells compared to controls, consistent with the qRT-PCR data. Moreover, expression of CDKN2A (also known as Ink4a-Arf locus), which encodes two different tumour suppressors $p16^{Ink4a}$ and $p19^{Arf}$, were found to be up-regulated in agreement with

the RNA-seq and qRT-PCR data. No changes in the level of the loading control β-actin were observed.

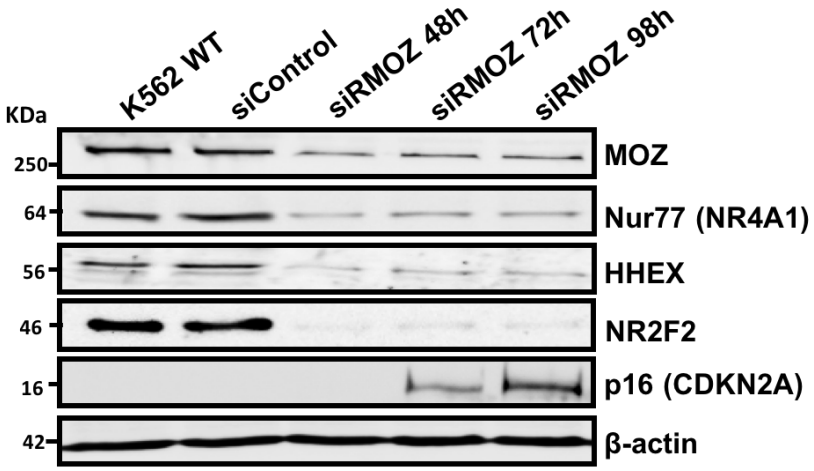


Figure 3.13 - Western blot showing expression of MOZ, NR4A1/Nur77, HHEX, NR2F2/COUP-TFII, and CDKN2A/p16 proteins in K562 cells depleted for MOZ or control after 48, 72 and 96 hours siRNA transfection. The total protein lysates for each siRNA treatment were equally loaded onto a protein SDS-polyacrylamide gel after measuring the concentration. β-actin was used as a reference protein.

3.8 Gene Ontology analyses: Functions of MOZ regulated genes

Having identified and validated a subset of genes whose expression is altered by MOZ knockdown, gene ontology analyses were performed on the list of 407 genes using a range of bioinformatics tools as described in Materials and Methods. These included: Webgestalt (http://bioinfo.vanderbilt.edu/webgestalt/option.php) Cytoscape (http://www.cytoscape.org/)

Gorilla (http://cbl-gorilla.cs.technion.ac.il/)

DAVID (https://david.ncifcrf.gov/summary.jsp)

FunDo (http://django.nubic.northwestern.edu/fundo/).

Similar results were obtained using the different platforms, regarding enrichment of genes associated with functions or pathways. For example, analysis of the 407 DEG gene set from K562 cells in Functional Disease Ontology database (FunDO), a web-based application that finds relevant diseases based on statistical analysis of the Disease Ontology (DO) annotation database (Table 3.1 and Figure 3.14) detected significantly enriched gene ontologies. The analysis shows cancer as the most significant disease term associated with MOZ knockdown in K562 cells (fold enrichment (FE) = 3,42; Bonferroni corrected p-value = 0,000071) and leukemia is one of the 22 specific cancer types listed (fold enrichment (FE) = 3,63; Bonferroni corrected p-value = 0,050). Among the potential MOZ-regulated genes associated with cancer, 85% were in the up-regulated cohort.

Table 3.1 - Functional Disease Ontology database (FunDO) identified different cancer-associated to down regulated genes under MOZ knockdown in K562 cells

DO Term	Number of Genes	Fold Enrichment	p-value	Bonferroni corrected p-value
Cancer	22	3,42	7,043E-07	0,000071
Ovarian cancer	8	9,26	0,0000026	0,00026
Embryoma	12	5,23	0,0000040	0,00040
Prostate cancer	12	3,73	0,00010	0,010
Liver cancer	8	5,52	0,00011	0,011
Breast cancer	13	3,45	0,00012	0,012
Pancreas cancer	6	7,39	0,00017	0,017
Leukemia	10	3,63	0,00049	0,050
Cervical cancer	5	7,64	0,00051	0,051
Colon cancer	9	3,63	0,00094	0,095
Neoplasm metastasis	6	4,55	0,0021	0,22
Brain tumor	6	4,38	0,0026	0,26
Melanoma	5	3,98	0,0088	0,89
Lung cancer	6	3,35	0,0095	0,96
Oral cancer	3	6,37	0,011	1,18
Nasopharyngeal cancer	2	8,49	0,023	2,33
Endometrial cancer	2	6,94	0,033	3,38
Stomach cancer	4	3,25	0,035	3,58
Squamous cell cancer	3	3,58	0,051	5,25
Esophagus cancer	2	5,09	0,058	5,93
Multiple myeloma	2	3,95	0,091	9,21

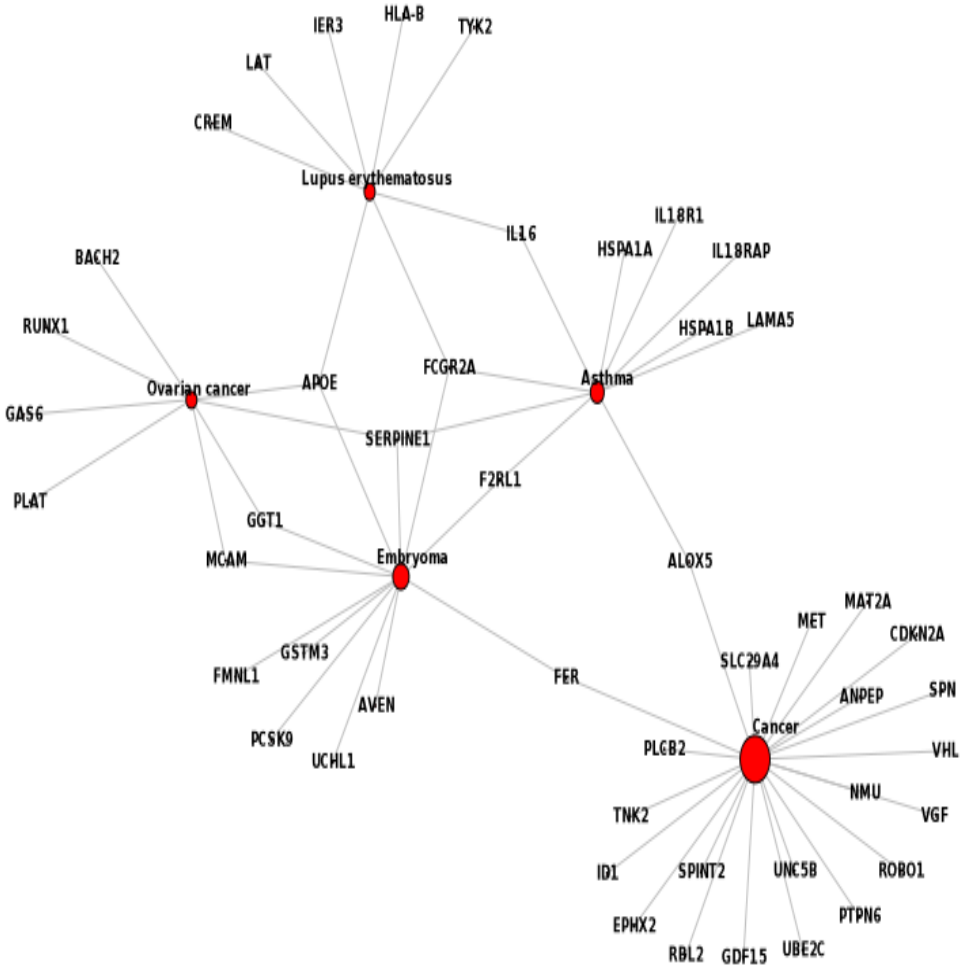


Figure 3.14 - FunDO analysis of differentially expressed genes after MOZ knockdown in K562 cells. Relevant diseases are shown based on statistical analysis of the Disease Ontology annotation database.

3.9 MOZ is required for expression of p53-regulated genes

The association of genes regulated by MOZ with cancer, as showed in the bioinformatics analysis above, may be due to the MOZ key role of MOZ and its fusion proteins in p53 activity [202, 205, 252,

258]. It is well known that p53 protein, a nuclear transcription factor, plays an important role as a tumor suppressor in response to cellular stress such as DNA damage causing cell cycle arrest or apoptosis [285-290]. Therefore, deregulation of p53 function is a major factor in tumorigenesis. Post-translational modifications of p53 are essential to guarantee its function, most specifically, p53 acetylation govern p53 transcriptional activity and selection between growth inhibitory and apoptotic gene targets [291-293].

Functional interactions of MOZ and MOZ-fusion proteins have been observed in our laboratory and other studies. For example, we showed that MOZ-TIF2 and MOZ-CBP fusion proteins can inhibit p53 mediated transcription in reporter assays, principally via the loss of CBP/p300 activity [202, 252]. Others have also reported functional interactions of MOZ with p53 to induce p21 expression and cell-cycle arrest. In response to DNA damage, levels of the p53-MOZ complex increase sufficiently to induce cell-cycle arrest [205, 258]. Furthermore, it was shown that in MOZ knockout mouse embryonic fibroblasts, cell cycle arrest in G1 and p21 expression induced by damage is significantly impaired. Rokudai, Laptenko [258] also showed that MOZ acetylates p53 at K120 and K382 and co-localises with p53 in nuclear bodies in promyelocytic leukaemia (PML), following stress. More recently, it was shown that gain-of–function p53 mutants drive expression of MOZ and MLL proteins in cancer cells [259].

Riley, Sontag [294] published an analysis article with an overview of p53-regulated genes showing a list so far of human p53-regulated genes. I compared the p53-regulated genes list with the MOZ-regulated genes in K562 cells and found 9 common genes as shown in Figure 3.15.



Figure 3.15 - Venn diagram showing the common genes regulated by MOZ in K562 cells and human p53 protein. Genes regulated by MOZ were obtained by RNA-seq analysis from MOZ knockdown K562 cells. The list of genes regulated by p53 were extracted from a review published by Riley, Sontag [294] that shows an overview of the p53-regulated genes.

Among the common genes regulated by MOZ and p53, six are involved in cell cycle regulation: Serpine-1 [295], GDF15 [296], MET [297], PLK2 [298], TP53I3 [299], and DUSP5 [300]. The other three, PERP, FDXR and IER3 play a role in apoptosis [301]. Thus, the

result showing cancer as the most significant disease term associated with MOZ knockdown in K562 cells is corroborated by its key role in p53 activity.

Breast cancer is also listed as one of significant disease associated to MOZ expression as shown in Table 3.1. It was showed that in the breast cancer cell line SUM-52, generated from a patient that had an ER-positive primary tumour in luminal B breast cancer, MOZ is one of the top genes overexpressed suggesting that MOZ functions as an oncogene in luminal breast cancer [260]. In this same study, the authors show by analysing The Cancer Genome Atlas (TCGA) data from 1000 primary breast cancers, that in 22% of cases MOZ expression is altered and it is overexpressed in 8.7% of invasive breast cancers. Also, higher level of MOZ expression leads to a statistically significant reduction in the overall survival rate. Although I used K562 cells for MOZ knockdown and RNA-seq analysis, comparison of our list of deregulated genes under MOZ knockdown to published transcription profiling data for breast cancer cell lines SUM52 and NCI-N417 mentioned above identified GDF15, PKIB, DUSP5, PHLDA1, PROX1, RPPH1, RHAG and RMRP as common down-regulated genes after MOZ knockdown. These data suggest that some genes may be regulated by MOZ in different tissue/cancer types.

Further analysis was performed with the up-regulated genes. Analysis of molecular function of up-regulated genes was performed using Cytoscape software (version 2.6.2) (http://www.cytoscape.org/)

with the ClueGO plugin (http://www.ici.upmc.fr/cluego/cluegoDownload.shtml).

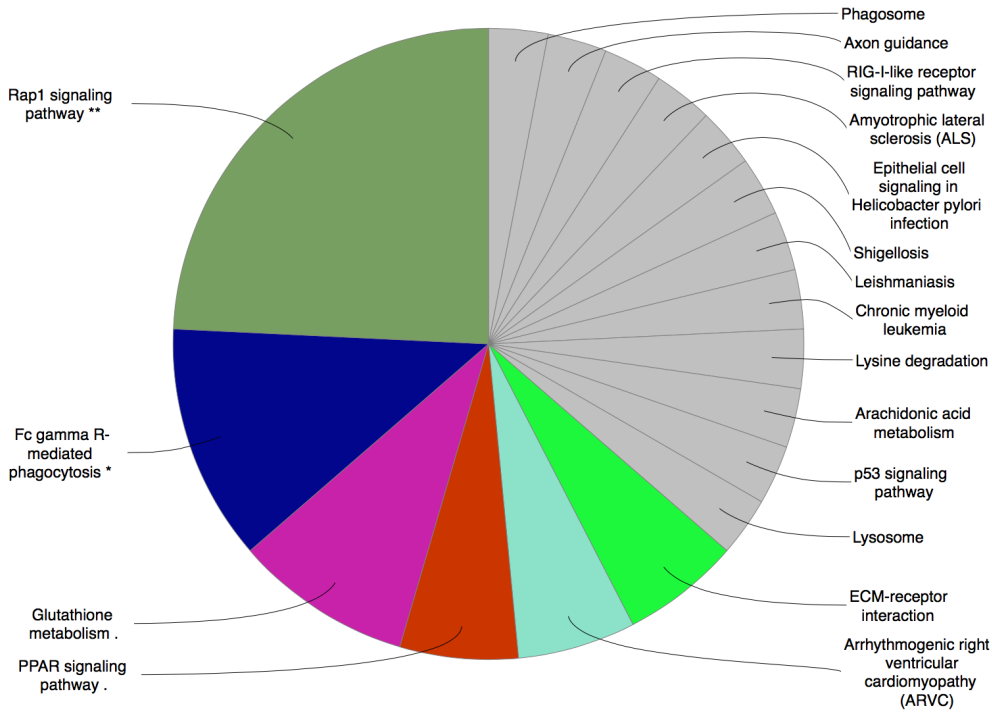


Figure 3.16 - KEGG/ ClueGO enrichment analysis of upregulated genes after MOZ knockdown in K562 cells. Relevant molecular functions are shown based on statistical analysis (Benjamini–Hochberg p value < 0.05).

RAP1

The result shows Rap1 signaling pathway as the most enriched molecular function when MOZ is knockdown in leukemic cell. Rap 1 is involved in many cellular process in mammalian cells, but more importantly, it is involved in normal cell growth,

differentiation, cell invasion and metastasis [302]. Active Rap1 activates tumor invasion and metastasis in leukaemia [303].

Glutathione metabolism

Glutathione metabolism, the major mechanism controlling cellular oxidative balance, is another enriched function under MOZ knockdown. Three genes in the glutathione S transferase mu cluster (GSTM2, GSTM3 and GSTM4) are upregulated. The glutathione S transferases are important to detoxify harmful electrophiles but they are implicated in resistance to cancer therapies [304]. It is interesting notice that Pei, Minhajuddin [305] showed that CD34+ AML cells have high expression of glutathione pathway regulatory proteins, probably as a mechanism to compensate for increased oxidative stress in leukemic cells. Consistently, CD34+ AML cells have lower levels of reduced glutathione and higher levels of oxidized glutathione compared with normal CD34+cells. Thus, the authors hypothesize that AML cells is hypersensitive to inhibition of glutathione metabolism and propose therapeutic targeting of glutathione metabolism as a potentially powerful strategy to induce selective toxicity toward a broad range of primary leukaemia cells, including malignant stem and progenitor populations.

AML1/Runx1

It is also interesting to note that MOZ-depleted K562 cells show upregulation of Runx1, a key regulator of hematopoiesis that is

the most frequently altered gene in leukaemia. Runx1/AML1 is believed to function as a tumor suppressor in myeloid neoplasms, but recently its expression has been associated to survival of AML cells [306, 307]. Moreover, Runx1/AML1 protein interaction with MOZ is required for transactivation of its target genes [202]. Taken together, these results suggest that MOZ regulates genes that play an important role in tumour development.

3.10 MOZ regulates expression of transcription factors and is associated to development and cell growth

In order to verify which genes regulated by MOZ are transcription factors, functional annotation analysis (Database for Annotation, Visualization and Integrated Discovery version 6.7, DAVID) was performed with input of all 76 down-regulated genes. Enrichment was regarded as significant if fold enrichment was ≥ 1.5 and p-value < 0.05. The result shows thirteen down-regulated transcription factors as shown in Table 3.2.

Table 3.2 - Transcriptional regulators identified by GO analysis in MOZ knockdown cells

Official Gene Symbol	Fold Change
BACH2	-0,42
CREM	-0,42
HHEX	-0,45
IFI16	-0,62
NR2F2	-0,42
NR4A1	-0,65
NR4A2	-0,79
PROX1	-0,57
TIGD5	-0,44
*ZBTB37	-0,99
**ZNF180	-0,59
***ZNF256	-1,24
****ZNF702P	-0,93

* Zbtb37 - Zinc Finger and BTB domain containing 37

** ZNF180 - Zinc Finger Protein 180

*** ZNF256 - Zinc Finger Protein 256

**** Zinc Finger Protein 702P

Gene knockout studies in mice have demonstrated that MOZ is essential for the development of hematopoietic stem cells. The phenotype characteristics of $MOZ^{-/-}$ embryos indicate failing in fetal liver hematopoiesis and reduction in hematopoietic progenitor population [203]. As mentioned earlier, dysfunction of MOZ activity, caused by reciprocal translocations generating MOZ fusion proteins, has been shown to play a causative role in acute myeloid leukaemia (AML) [235, 308]. MOZ also plays a role in neuronal development. Its

expression is up-regulated in stem cells hESC submitted to neural differentiation system [309]. Indeed, some of the transcription factors regulated by MOZ showed in Table 3.2 plays important role in hematopoietic and neural development.

3.10.1 CREM

CREM is a versatile gene that interestingly encodes a family of activators and antagonists of cAMP-inducible transcription by differential splicing [310]. The transcription factor CREM plays an important role in neuronal development [311], also is crucial for a normal program of germ cell differentiation [312]. ICER is the other transcription factor derived from an alternative promoter activation of CREM gene [313]. ICER was found to be downregulated at diagnosis of acute leukemia in a cohort of pediatric. On the other hand, the higher ICER expression was found in healthy bone marrow [314, 315]. Pigazzi, Manara [316] showed that the period of enhanced expression of ICER corresponds to a significant modification of cell proliferation and survival *in vivo*. They also provided evidence that ICER expression inhibits leukemia phenotype and controls tumor progression.

3.10.2 BACH2

The transcription factor encoded by BACH2 gene is a key regulator of the memory differentiation of T and B cells [317] and promotes B cell development by repressing the myeloid program [318]. Swaminathan, Huang [319] suggested that BACH2 is involved

in the development of haematopoietic cancers. BACH2 is located at chromosome 6q15, a region of frequent minimal deletion in individuals with B cell ALL with 6q mutations. They showed that, although BACH2 seems to either inhibit or promote cancer depending on its cellular context, in B cell acute lymphoblastic leukaemia (ALL), BACH2 functions as a tumour suppressor. Overexpression experiments indicate that BACH2 directly represses genes involved in myeloid differentiation, including CCAAT/enhancer-binding protein beta (Cebpb), aryl hydrocarbon receptor (Ahr), lymphocyte antigen 96 (Ly96), Spic, and cytokine inducible SH2-containingprotein (Cish) [320].

3.10.3 IFI16

IFI16 is a member of the HIN-200 (hematopoietic interferon-inducible nuclear antigens with 200 amino acid repeats) family of cytokines and it is associated with myeloid cell differentiation induced by cytokines and chemical agents [321]. The protein IFI16 can directly bind to the C-terminal region of p53 and enhance the transcriptional activation function of p53 [322]. Repression of IFI16 affects p53 transcriptional activity at the p21 promoter and the stability of the proteins p53 and p21. Therefore, IFI16 can affect cell cycle regulation [323]. More recently, Piccaluga, Agostinelli [324] IFI16 reduced expression is correlated with unfavorable outcome in chronic lymphocytic leukaemia. The author also suggested that IFI16

could became a part of phenotypic screening in CLL patients once it can be easily detected by immunohistochemistry or flow cytometry.

3.10.4 The nuclear receptor genes NR4A1, NR4A2 and NR2F2

NR4A1 belongs to a family of orphan receptors which expression are critical for neuronal development, organogenesis, cell fate determination and metabolic homeostasis. Its expression has been associated with a more lethal form of AML in combination with loss of Nor1/NR4A3 expression. It was further demonstrated that gene dosage is a critical factor for development of hematologic malignancies once NR4A1+/- mice develops mixed myelodysplastic/myeloproliferative (MDS/MPN) neoplasms with a more aggressive disease [325].

NR2F2 is one of the transcription factor down-regulated under MOZ knockdown as showed in Table 3.2. Furthermore, Srinivasan, Geng [326] have shown that NR2F2 is necessary for the initiation of PROX1 expression in embryonic veins by directly binding a conserved DNA domain in the regulatory region of PROX1. In addition, it was shown that a direct interaction between nuclear receptors and PROX1 is required for early maintenance of PROX1 expression in lymphatic endothelial cells during early stages of specification and differentiation. These findings support our results and indicate that MOZ regulates NR2F2 and probably NRF2 regulates PROX1 in leukaemia cells as well.

3.10.5 The homeobox genes PROX1 and HHEX

Two homeobox genes PROX1 and HHEX were also identified as regulated by MOZ. As mentioned in the introduction, homeobox genes play key role in developmental control genes regulating aspects of morphogenesis and cell differentiation in animals. Interestingly, Aranguren, Beerens [327] have shown that NR2F2 knockdown significantly decreased PROX1 expression in zebrafish. As mentioned before, NR2F2 is also down-regulated under MOZ knockdown, indicating that MOZ may control PROX1 expression via NR2F2 regulation in K562 cells.

HHEX is expressed in a range of multipotent hematopoietic progenitor cells and cell lines [328], [329]. It was found to be overexpressed in acute myeloid leukaemia and is required specifically for leukaemia growth [330]. Shields, Jackson [330] also found that loss of HHEX leads to expression of the CDKN2A, which are required for growth arrest and myeloid differentiation. HHEX binds to the CDKN2A locus and directly interacts with the Polycomb-repressive complex 2 allowing H3K27me3-mediated epigenetic repression. These results corroborate with our findings of down-regulation of HHEX and up-regulation of CDKN2A genes. Furthermore, we are showing for the first time that MOZ regulates HHEX expression and consequently CDKN2A in K562 cells.

3.11 Molecular and Biological process associated to MOZ expression

Enrichment analysis for molecular function highlighted the upregulation of H3K14ac specific deacetylases, which is of interest given the specificity of MOZ/KAT6A for H3K14 acetylation. As these genes were not part of the validation, we performed qPCR on control and siMOZ treated cells (48hrs). As shown in Figure 3.17, this confirmed a significant increase in expression of HDAC6, HDAC7 and HDAC10. Although, we did not have access to antibodies to validate this further in western blots, the result suggests that MOZ may have a direct or indirect ability to repress HDACs that remove acetyl groups from its target H3K14.

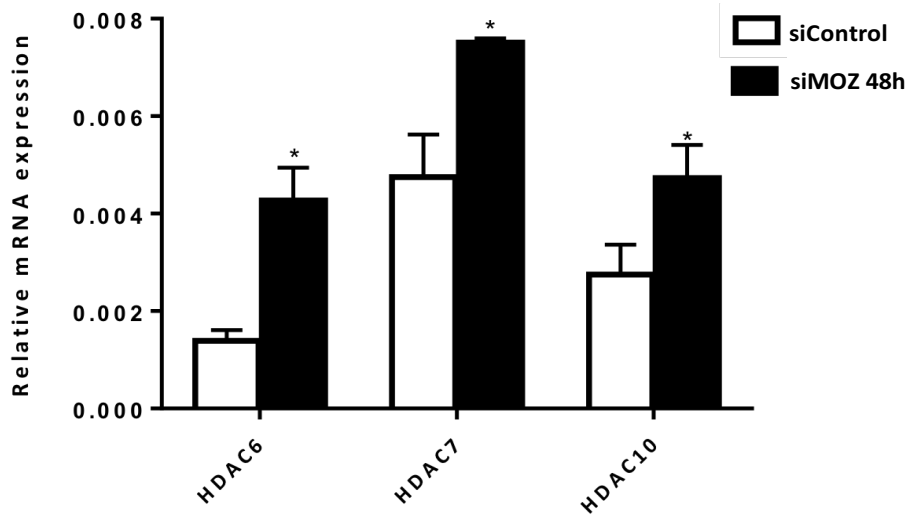


Figure 3.17 - qPCR validation data of up-regulated HDACs genes in MOZ knockdown K562 cells from RNA-Seq data. Quantitative RT-PCR showing up-regulation in HDAC6, HDAC7 and HDAC10 mRNA transcript levels between siControl (white bars) and siMOZ (black bars) transfected K562 cells after 48 hours. qRT-PCR data were normalized to GAPDH expression for each sample. The asterisk indicates significant difference ($p<0.05$) based on student's t-test analysis between control and MOZ-silenced samples. Error bars, S.E.M.

We also analysed the top biological processes associated with down-regulated genes in K562 cells using NIH DAVID Gene Ontology (GO). The result endorses the impact of down-regulated homeobox genes, showing developmental process among the top hits biological processes linked to MOZ knockdown (Table 3.3).

Table 3.3 - Top biological processes identified by GO analysis of genes whose expression Is down-regulated by MOZ in K562 cells.

GO Term	Count %	P-Value
Biological regulation	39	0.0046
Cell growth and/or maintenance	48	0.0114
Locomotion	6	0.0227
Developmental process	18	0.0474

Cell growth and/or maintenance is the second most enriched biological process associated to the knockdown genes. As mentioned above, CDKN2a is a tumor-suppressor gene involved in cell cycle control for encoding the protein p16, a cyclin-dependent kinase (Cdk) inhibitor that can block the G1 to S phase transition [331-335], and it is one of the up-regulated genes we identified under MOZ knockdown in K562 cells.

So, it is expected that MOZ knockdown cells alter the proliferation of leukemic cells. To confirm this hypothesis, flow cytometry cell cycle analyses using propidium iodide was performed in MOZ knockdown K562 cells after 48, 72 and 96 h siRNA transfection, as shown in Figure 3.18.

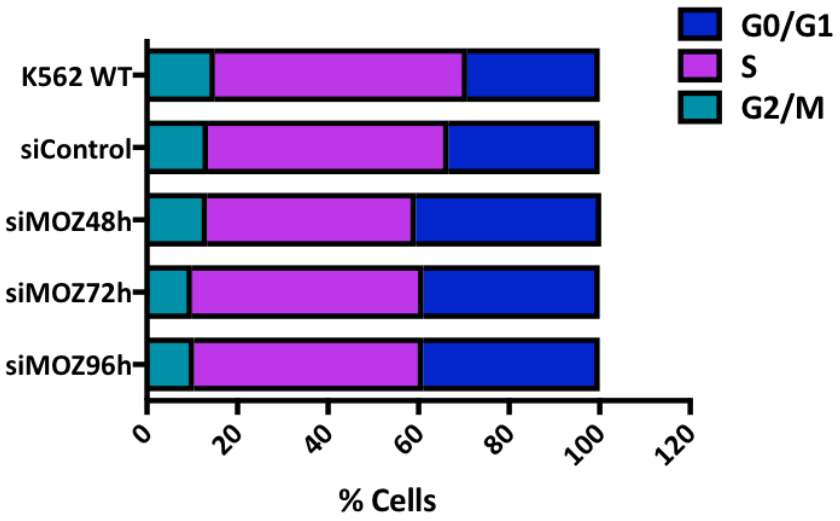


Figure 3.18 - Inhibition of cell proliferation by MOZ knockdown. Flow cytometry cell cycle analyses performed in MOZ knockdown K562 cells after 48, 72 and 96 h siRNA transfection. The percentage of cells in G1, S and G2/M (a mixed population of G2 and mitosis) phases of the cell cycle are shown. Experiments were repeated twice.

As expected, it was observed cell cycle abnormalities in the MOZ knockdown cells with 40% of the cells increase in G0/G1 stage, 10% higher compared to control cells. It confirms that MOZ plays a role in leukemic cells proliferation and can be considered a potential therapeutic target for AML. The effect on cell cycle depends on MOZ expression level, as will be presented in the next chapter about MOZ knockout in K562 cell.

Although MOZ has been found to increase premature senescence through regulation of the tumor suppressor gene p21 and p53 acetylation upon cellular stress [205, 258], other studies showed that MOZ is involved in cell senescence via regulation of CDKN2A gene expression. Perez-Campo, Costa [226] showed that

MOZ controls the proliferation of both hematopoietic and neural stem cells by modulating the transcriptional repression of p16(INK4a). Sheikh, Phipson [335] also reported that primary mouse embryonic fibroblasts isolated from Moz-deficient embryos exhibit premature senescence, which was rescued on the Ink4a-Arf$^{-/-}$ background. Nonetheless, Largeot, Perez-Campo [336] showed that the HAT activity of MOZ, preserved in the fusion protein, is essential for the CDKN2A locus transcription and consequently, for blocking senescence, a necessary step for the survival of leukemic cells.

3.12 Genes down-regulated under MOZ-silencing are involved in coenzyme A metabolic process

KEGG pathway analysis, using Cytoscape with ClueGO plugin, provided additional insight into the role of MOZ by analysing down-regulated genes from K562 cells. As showed in Figure 3.19, coenzyme A metabolic process is the major pathway enriched following MOZ knockdown. As mentioned before, acetyl Coenzyme A (acetyl-CoA) acts as a cofactor for protein acetylation. It is interesting to notice that MOZ may regulate the expression of this coenzyme that is essential to carry out protein pos-translational modification. A study performed in *Xenopus laevis* showed that at the whole-embryo level, there is a perceptible correlation between the levels of acetyl CoA and levels of acetylation of proteins including histones H3. This

suggests the level of acetyl CoA may play a role in the regulation of embryogenesis [337].

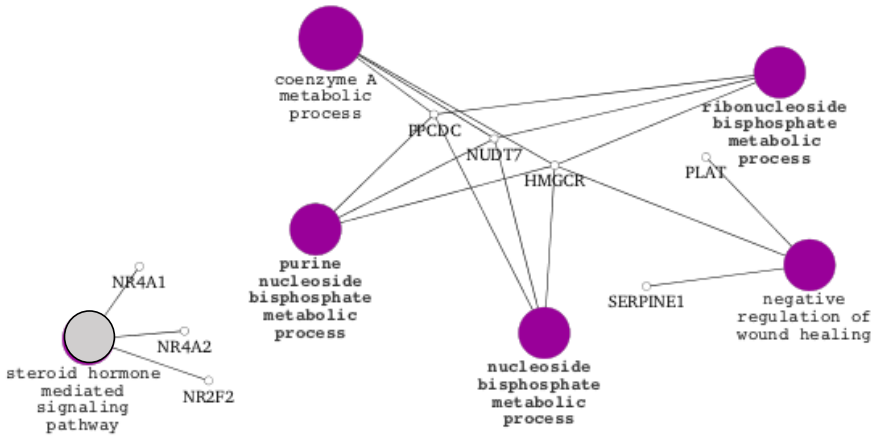


Figure 3.19 - ClueGO enrichment analysis of down-regulated genes under *MOZ* knockdown. GO/KEGG pathway functionally grouped networks with terms are indicated as nodes (Benjamini–Hochberg p value < 0.05) linked by their kappa score level (≥ 0.35), where only the label of the most significant term per group is shown. Functionally related groups partially overlap and ungrouped terms are shown in gray.

An important down-regulated gene involved in coenzyme A metabolic pathway is HMGCR, which encodes the protein 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA). This enzyme is rate-controlling in the synthesis of cholesterol and intermediates in the mevalonate pathway. Considerable numbers of studies reported deregulated or elevated activity of HMG-CoA reductase in a range of different tumors, as such hepatocellular carcinoma, lymphoma, colorectal, lung adenocarcinoma and leukemia [338-342]. Also, it has

been shown that statins, a class of drug that inhibits HMG-CoA reductase and display anti-leukemic properties can inhibit adhesion, migration, and proliferation of leukemic cells. Interestingly, studies show that p53 seems to have an important role in metabolic regulation, including mevalonate pathway [343, 344]. As it is known that MOZ increases p53 acetylation as mentioned above, it is probably that MOZ regulates HMGCR expression by regulating p53 pos-translational modification.

3.13 Summary

MOZ is a multifunctional protein that can acetylate histones and other proteins and also recognise acetyl, crotonyl and methyl marks on specific lysines in H3 and H4 [190] [191, 217, 335]. MOZ appears to perform its functions in cooperation with other chromatin regulatory proteins including BRPF1, EAF6, ING5 [196] and KDM4C/PRDM1 [345]. Current models suggest that it is recruited to specific gene promoters/ enhancers by transcription factors such as AML1/Runx1, Runx2, p53, NFkB, MAFK/NRF2, Pu.1/SPI1, c-Myc, Ets and TBX1, specific. As its role in haemaopoiesis and leukaemia is well established, we decided to identify MOZ target genes in the ertyhroleukemic cell line, K562, which expresses MOZ at robust levels. In addition, K562 is a well characterised cell line with ENCODE data available.

Previous mouse knockout and tissue-based models have implicated MOZ function in the development of a number of different tissues including haemopoeisis in foetal liver. Null mice die at birth probably because they lack oxygenated blood [213], and MOZ depletion also leads to defective development of B cells, T cells and germinal centres [273]. However, other developmental defects due to MOZ ablation are also observed include homeotic transformations resulting in additional vertebrae, lack of thymus and underdeveloped spleen, and mild craniofacial abnormalities [213]. There is also abnormality of cariac development in the aortic arch and septum,

thus loss of MOZ expression partially phenocopies DiGeorge Syndrome [264].

In human patients, recent advances in exome sequencing have identified de novo mutations in the KAT6A /MOZ gene that are associated with global developmental delay [227, 346]. These patients show a spectrum of clincal symptoms depending on the position of the mutation. This underlines the need to better understand how MOZ functions in the cell so that we can understand its role in disease mechanisms.

MOZ is known to be essential for long-term repopulation of hematopoietic stem cells [203, 213], and MOZ translocations promote continual division of leukemic cells [217]. MOZ is also altered in solid tumors and has been identified as top-ranking target amplified gene in different types of cancer [347]. MOZ is also implicated in solid tumours such as breast cancer, prostate cancer and lung cancer. Thus, in addition to its involvement as an oncogenic fusion protein in leukemia, MOZ is also clearly required for gene expression in may tissues in the adult and during development and in tumourigenesis. Thus, it is important to establish which genes it regulates in different tissues and cell types to better understand its functions. While MOZ expression is altered in many cancer types, relatively few target genes have been identiifed. At the outset of this project, MOZ target genes in leukemic cells were relatively unknown. In this chapter, we report for the first time the identification of MOZ regulated genes in K562 cells, a human erythroleukemia cell line.

To this end, *MOZ* gene was knocked down using siRNA followed by transcription profiling using RNA sequencing. Knockdown of MOZ was confirmed by both qRT-PCR and western blot, with around 70% decrease in its expression. This lead to a reduction of global H3K9 and H3K14 acetylation. Our result is not in concordance with Voss, Collin [217] data which showed that, in embryos deficient in MOZ, histone acetylation defects are relatively specific to H3K9 at Hox loci, as Hox H3K14 acetylation were not strongly affected and, in $MOZ^{-/-}$ mouse embryonic fibroblastos cells, only H3K9 presented reduction in acetylation levels [335]. However, our result is consistent with previous results published by our group which shows that MOZ DPF cooperates with the MYST domain to promote H3K9 and H3K14 acetylation, not only H3K9.

Having determined that strong knockdown of MOZ was achieved at 48hrs post siRNA treatment, we performed RNA-Seq analysis. Only 406 differentially expressed genes were identified at p value <0.05, showing significant differential expression compared to control (K562 cells transfected with non-targeting siRNA), of which 76 were down-regulated. qRT-PCR validated 10 down-regulated and 15 up-regulated genes, and western blots confirmed the altered expression of MOZ, NR4A1/Nur77, HHEX, NR2F2/COUP-TFII, and CDKN2A/p16 proteins in K562 cells depleted for MOZ or control after 48, 72 and 96 hours.

The number of DEGs identified was relatively low, and this could be for a number of reasons: 1) functional redundancy with

MORF/MST4 or other HATs; or 2) a low number of natural target genes in some tissues. Indeed, other studies that have performed transcriptome after MOZ knockdown or knockout in other cell lines or in mouse tissues also had relatively small number of DEGs, as discussed below.

Katsumoto, Aikawa [203], in order to investigate the roles of MOZ in normal hematopoiesis, generated MOZ-null mice and analysed by oligonucleotide microarray genes regulated by MOZ in E14.5 fetal liver cells. Analysis of E12.5 $MOZ^{+/+}$, $MOZ^{+/-}$, and $MOZ^{-/-}$ fetal livers identified 48 and 163 genes exhibiting more than twofold changes in expression. Among them, HoxA9 that has a key role in hematopoiesis and leukaemia pathogenesis [348, 349], had its expression level significantly decreased. Comparison amongst $MOZ^{+/+}$ and $MOZ^{+/-}$ common genes showed in this study with the genes deregulated in K562 cells under MOZ knockdown identified only RAB27B as a consensual gene. Some of the other DEGs identified in E14.5 fetal liver cells such as HOXA9 are not expressed in K562 cells.

A study performed by Turner-Ivey, Guest [260] provided evidence that MOZ functions as a novel oncogene in luminal breast cancer showing that MOZ is amplified and overexpressed in the SUM-52 breast cancer cell line and MOZ knockdown influences on proliferation and viability of this cells. Furthermore, to better understand the role of MOZ as a breast cancer oncogene, RNA-seq was performed in the SUM-52 MOZ knockdown identifying 742

genes using a 1.5 fold-change cut-off. Among them, 22 genes were identified in common with our data for MOZ knockdown in K562 cells: RMRP, GDF15, LOC441155, NFATC4, ID1, GSTM4, TSC22D3, FHL1, PLA2G6, RHAG, ROBO1, OBSCN, DHRS3, ATP2A3, VWDE, PIF1, FADS3, C11orf74, AUTS2, KCNN4, OGDHL, CHST3.

In memory B-cell, Good-Jacobson, Chen [273] showed that MOZ regulates memory cell formation by controlling memory compartment composition. Also, MOZ is required for establishing the germinal center gene expression program of these cells. RNA-seq was performed to investigate gene expression changes upon activation in B cells from $Moz^{fl/fl}Aicda^{Cre/+}$ or $Aicda^{Cre/+}$ mice. Similarly, as observed in ours result, the majority of genes differentially expressed were up-regulated, 66 in total. Just 12 genes were downregulated out of 78 DEGs. There was no overlap between the genes regulated by MOZ in this study compared to ours in K562 cells.

RNA-seq analysis was also performed in BM pre–B cells isolated from 3 WT and 3 $Moz^{+/-}$ adult mice identified 24 down-regulated genes and 92 up-regulated genes in $Moz^{+/-}$ [224], none of them was identified among our K562 DEGs. In this study, it was shown that MOZ is essential for maintaining normal transcriptional levels of MEIS1, HOXA7, and HOXA9, genes that are commonly overexpressed in leukemia. Furthermore, MOZ was required to maintain B-cell progenitor numbers, in the presence and absence of MYC overexpression. These results reinforce the importance of

performing studies in different cells and tissues to better understand MOZ function.

Gene Ontology analysis performed with MOZ targeting genes identified in K562 cells indicated an enrichment of genes associated with cancer, more specifically ovarian cancer, embryoma, prostate cancer, liver cancer, breast cancer and leukaemia, cell growth and development. KEGG pathway analysis performed with the up-regulated genes identified Rap1 signaling pathway, a cellular process involved in normal cell growth, differentiation, cell invasion and metastasis as the most enriched molecular function. Glutathione metabolism was also one of the most enriched function under MOZ knockdown.

Transcription factors play essential role in cell proliferation, differentiation and survival and their dysregulation have an important function in acute leukemia [350-353]. Among the down-regulated genes, at least 13 transcription factors were identified. Some of them are well known for playing a role in hematopoietic and neural development. For example, the transcription factor CREM plays an important role in neuronal development [311] and it is crucial for a normal program of germ cell differentiation [312]. BACH2 is essential for B cell development [318]. NR4A1 expression is critical for neuronal development, organogenesis, cell fate determination and metabolic homeostasis. Its expression has also been associated with a more lethal form of AML [325]. Two homeobox genes PROX1 and HHEX play key role in developmental control genes regulating

aspects of morphogenesis and cell differentiation in animals. HHEX is expressed in a range of multipotent hematopoietic progenitor cells and cell lines [328], [329]. It was found to be overexpressed in acute myeloid leukaemia and is required specifically for leukaemia growth [330].

The top biological processes associated with down-regulated genes in K562 cells was analysed using NIH DAVID Gene Ontology (GO). The result endorses the impact of down-regulated homeobox genes, showing developmental process as the first and cell growth and/or maintenance the second most enriched biological process associated to the knockdown genes.

Flow cytometry cell cycle analyses using propidium iodide was performed in MOZ knockdown K562 cells after 48, 72 and 96 h siRNA transfection to confirm the effect of MOZ absence in cell proliferation. The result shows cell cycle abnormalities in the MOZ knockdown cells with 40% of the cells in the G0/G1 stage, 10% higher compared to control cells. It indicated that MOZ plays a role in leukemic cells proliferation and can be considerate a potential therapeutic target for AML.

Finally, KEGG pathway analysis was performed with the down-regulated genes using Cytoscape with ClueGO plugin indicating that MOZ is also involved in coenzyme A metabolic process.

In conclusion, our data identified novel gene targets of the chromatin regulator MOZ in K562 erythroleukemic cells, many of which are implicated in haematopoietic development, cell

proliferation, have tumour suppressor or oncogenic properties when misexpressed. These findings contribute for a better understanding of MOZ function, especially in leukaemia.

Chapter 4 : CRISPR-CAS9 mediated knockout of MOZ in K562 cells

4.1 Background

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas9 system is a recently developed technology being widely used for high-throughput genome engineering. CRISPR is a component of the bacterial immune system that allows them to recognize and destroy phages [354]. In genome editing applications, Cas9 endonuclease is targeted by guide RNA sequence homology to a given locus, where it induces a double stranded break (DSB) [355]. Subsequently repair mechanisms invoked by the DSB at the specific may generate indels via non-homologous end joining (NHEJ), or alternatively the DSB can be repaired by homology-directed repair (HDR) when a repair template is provided that specifies a desired genomic modification. NHEJ has an error-prone nature that often leads to inactivating mutations. On the other hand, HDR enables precise editing with introduction of point mutation or recombination sites. It is also possible to introduce a gene of interest into a given locus by using double stranded or single stranded (oligonucleotide) repair templates [355].

CRISPR CAS9n nickase technology utilises adjacent strand-specific targeting events to induce double strand break for repair by non-homologous end joining (NHEJ). This approach is suitable for gene knockout and has been reported to greatly reduce the frequency of off target events [356, 357]. The endonuclease CAS9 generates double strand breaks through its nuclease domains RuvC

and HNH activity. In *S. pyogenes* CAS9, D10 is the critical amino acid residue for RuvC activity and H840 is critical for HNH activity [357]. CAS9 nickase is a modified version of CAS9 enzyme, which contains only one active catalytic domain resulting in a single strand break. DNA nicks, or single strand breaks, are promptly repaired by homology directed repair (HDR) using the intact complementary DNA strand as the template [358]. Therefore, it is essential to have two nickases targeting opposite strands to generate a double strand break within the target DNA. It increases target specificity because it is highly improbable that two off-target breaks will be generated within close enough proximity to constitute a double strand break [356].

In this chapter it is shown the generation of MOZ/KAT6A knockout K562 cells using CRISPR CAS9n nickase technology. Furthermore, as genome engineering using this technology causes a permanent effect that sustainably removes gene function and activity, the K562 cells generated allowed further analysis of the function of MOZ in leukaemia.

4.2 Design of synthetic guide RNA (sgRNA) for generating MOZ/KAT6A knockout using CRISPR CAS9 genome editing

By selecting adjacent sgRNA pairs targeting opposite strands it is possible to introduce InDels originating from the PAM sequences (5’ NGG). It is important to note that PAM sequence is not part of the

sgRNA, however, it needs to be in the target locus to guide the CAS9 activity [359]. In this study, the 20 nucleotides sgRNA guide sequences targeting exon 3 of KAT6A were designed using the Zhang lab target finder tool (crispr.mit.edu/nick). This tool locates possible pairs of sgRNA ranked and scored according to the efficacy of each pair, also it considers the number of predicted off target sites. For KAT6A knockout, the best sgRNA pairs targeting the ATG start codon were chosen because it is more likely to disrupt the coding sequence, thus impair MOZ gene expression. The top two pairs of sgRNA suggested were A+B and A+C (Figure 4.1). Guides A and B had individual score 68 and guide C was scored 52. The combined score for the guides A+B is 46 and for the guide A+C is 35. Computational analysis predicted zero off-target sites for these sgRNA combinations.

The sgRNAs were designed with overhang regions complementary to BbSI restriction sites for subsequent ligation to pX461 vector and an extra 5' guanine (G) to enhance U6 transcription. The sgRNA are composed by top and bottom strand oligonucleotides.

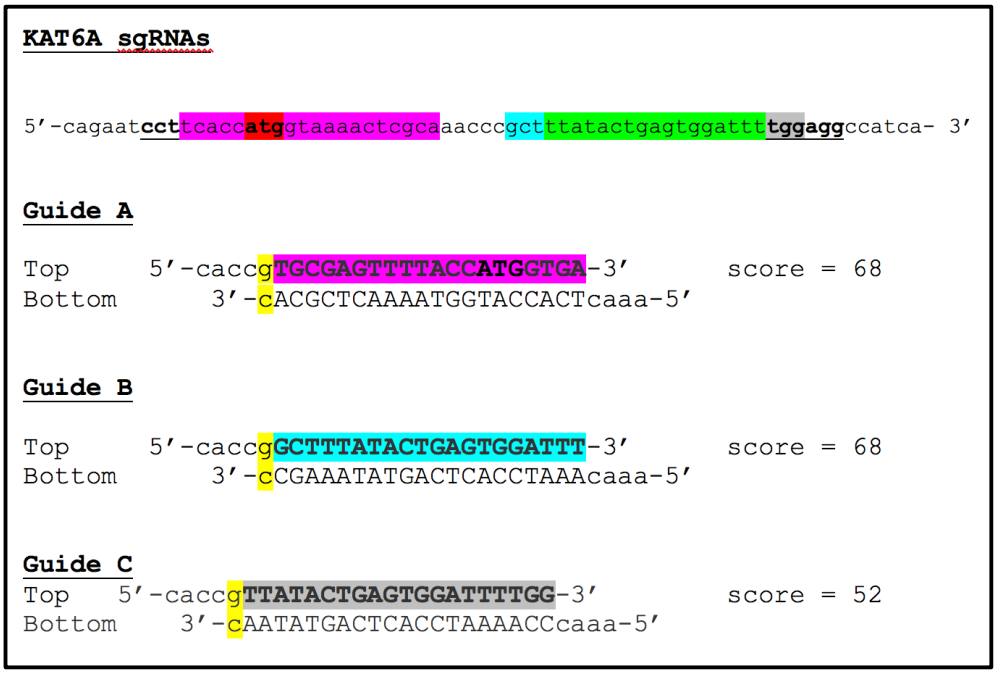


Figure 4.1 - Synthetic guide RNA (sgRNA) for KAT6A knockout. *H. sapiens* KAT6A exon 3 was run through the CRISPR design (nickase) software at crispr.mit.edu and the top two pairs of sgRNA suggested (A+B, A+C) was used for subsequent ligation into the px461 vector. Guide A appears in pink, guide B in blue and guide C in grey. The sequence shared by B and C is shown in green. ATG start codon appears in red. PAM sequence is underlined. The extra G added to the top strand to enhance U6 transcription is shown in yellow. Note that only the top part of the MOZ sequence double strand DNA is shown, however, guides B and C highlighted, in blue and grey, will bind to the opposite strand.

4.3 Generation of KAT6A targeting Cas9n vectors

Oligonucleotide pairs with BbsI compatible overhangs and containing an extra 5' G to enhance U6 transcription sequence were phosphorylated with T4 polynucleotide kinase, annealed and cloned into BbsI-digested vector.

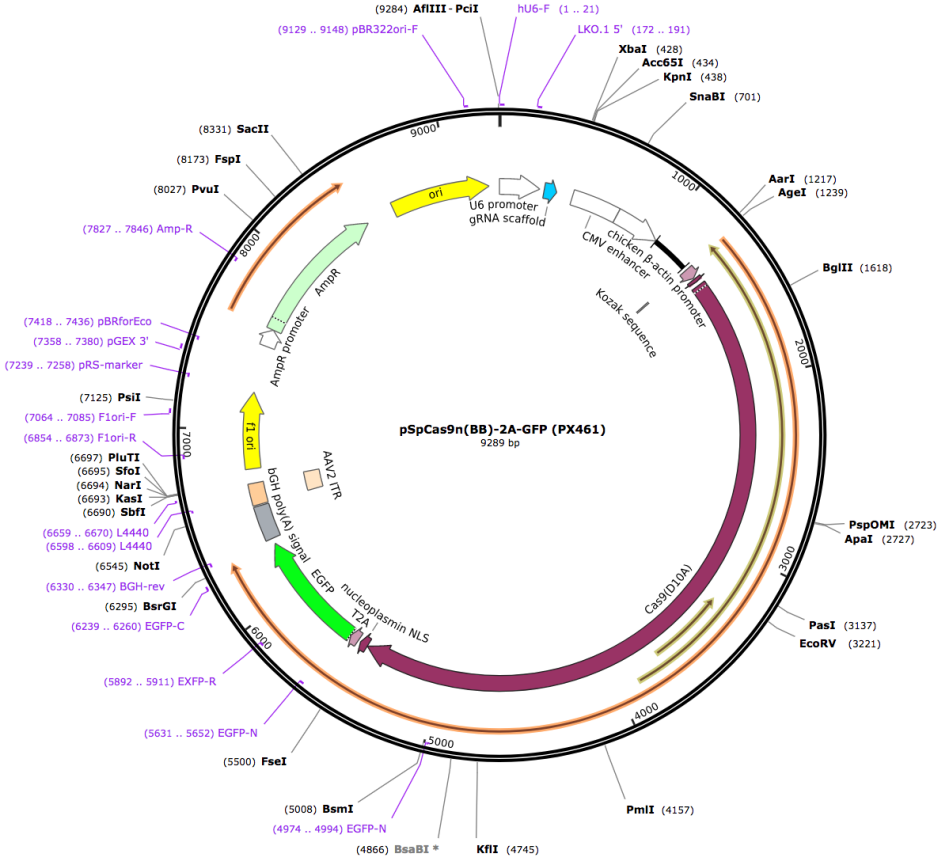


Figure 4.2 - Schematic image of pSpCas9n(BB)-2A-GFP plasmid (PX461). Zhang lab plasmid for expressing a chimeric guide RNA (gRNA) plus EGFP and human codon-optimized Cas9 with the D10A nickase mutation. Reproduced from Addgene website: https://www.addgene.org/48138/

The vector pSpCas9n(BB)-2A-GFP (PX461 from AddGene repository) used in this study encodes endonuclease Cas9, a conserved sgRNA (pSpCas9(BB)) and the sequence for U6 RNA polymerase III promoter. This vector also has a BbSI restriction site, expresses GFP to allow sorting of targeted clones and it is ampicillin resistant as shown in Figure 4.2.

In order to clone the sgRNA into the px461 vector, the plasmids were digested with BbSI restriction enzyme for 2 hours at 37ºC, then

the excised fragments were separated from possible remaining uncut vector by agarose gel and extracted to be purified using a Macherey-Nagel NucleoSpin Plasmid kit and, subsequently, ligated with the annealed phosphorylated sgRNA primers using T4 DNA ligase enzyme. The ligation mix was transformed into *E. coli* DH5α cells by heat-shock, and transformants selected on LB ampicillin agar plates. Single colonies from each transformation were inoculated in 10 mL of LB supplemented with ampicillin overnight at 37ºC with shaking (230rpm). Plasmid DNA was purified using Mini-Prep kit and validated by sequencing using a forward primer complementary to the human U6 promoter (Figure 4.3).

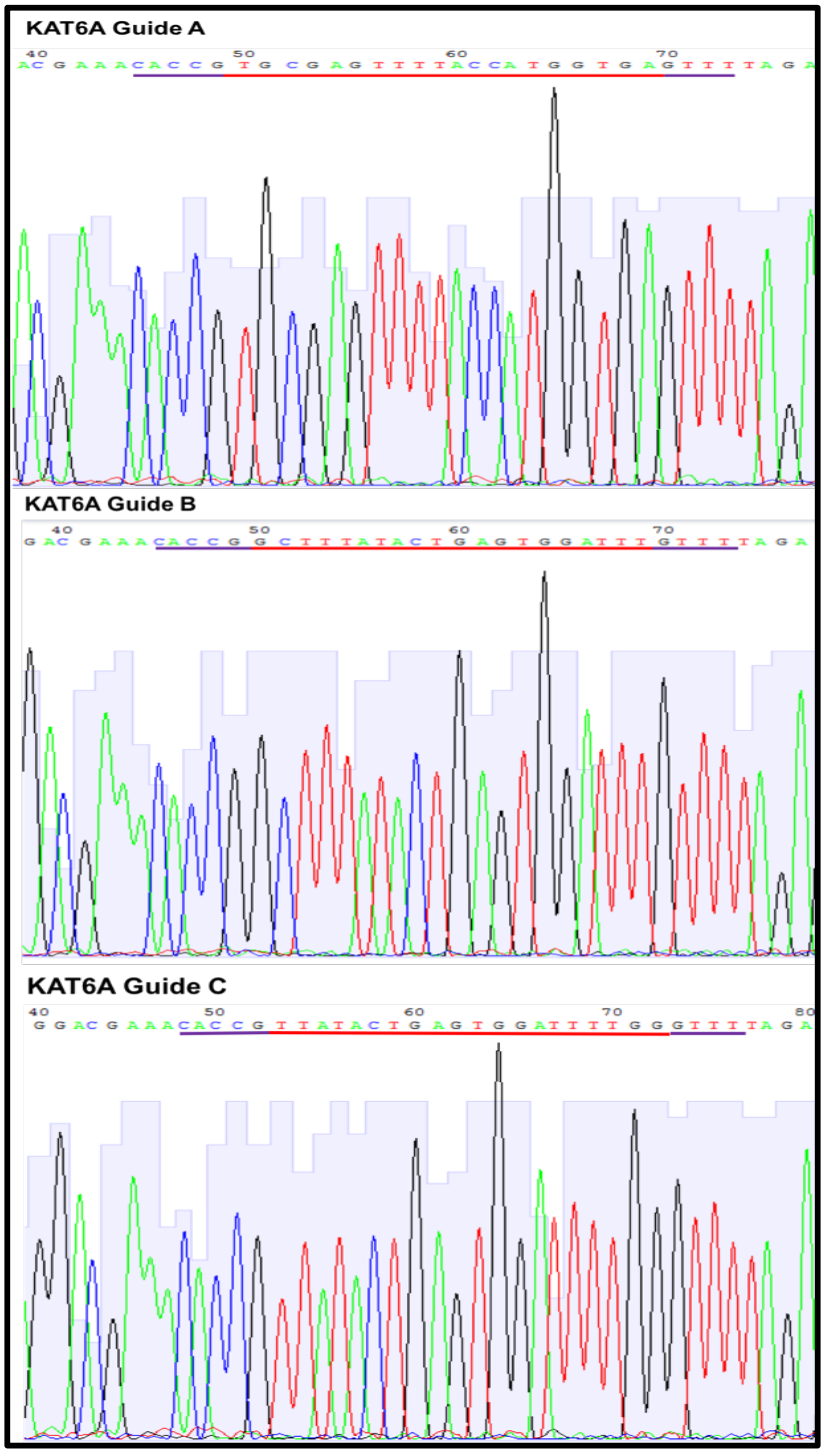


Figure 4.3 - Eletropherogram of the sgRNAs A, B and C targeting exon 3 of the KAT6A gene in K562 cells. sgRNAs A, B and C cloned into px461 vector and successfully transformed into E. coli DH5α. Guide sequences are underlined in red. Before and after each guide, sequence corresponds to BbSI restriction sites are underlined in purple

4.4 KAT6A gene editing in K562 cells using CRISPR-CAS9n

K562 cells were transfected with vectors to express A/B or A/C sgRNA pairs or A sgRNA only as control. The transfection of a single guide acted as a negative control, as only one sgRNA results in a nick with posterior repair of the DNA strand, but causes no alteration in the gene target.

Transfection was carried out using the Neon™ Transfection System (Invitrogen) according to the instructions provided by the manufacturer. Briefly, 2 x 10^5 K562 cells were transfected with 1 µg of plasmid in a volume of 10 µl using a Neon tip and electroporation. Electroporation was carried out at 1,350 V with four 10 ms pulse. The cells were then cultured without antibiotics. Electroporation transfection has the advantage of been applicable to almost all cell lines and it is relatively easy to perform. The method consists of applying high-voltage electric shock to the cell resulting in formation of pores in the cell membrane. The DNA can be introduced through the pores and ultimately, reach the nucleus. The membrane recovers its natural conformation closing the pores [360]. The electroporation efficiency depends on the plasmid quality, size and purity.

As mentioned before, the px461 vector expresses GFP to allow sorting of targeted clones. At 2 days post transfection, GFP+ cells were sorted using a Beckman Coulter MoFlo XDP cell sorter. Single GFP+ cells were seeded into 96 well plates. After culture for 3-4

weeks, three clones were chosen from isolated single K562 cells were able to form colonies and expand. One corresponds to a single guide transfection (A only), the other two correspond to the pair guide transfection A/C (ACD6 and ACF6).

To ascertain whether KAT6A was successfully targeted, genomic DNA was extracted from 5 x 10^5 cells of each clone using a Sigma GenElute Mammalian Genomic DNA Miniprep Kit (G1N350). PCR was performed with 50 ng of purified genomic DNA using the primers flanking the region 285 and 983 of KAT6A gene: 5'-ATGACGAGACCCATGTGTGG-3' and 5'-ACTGCTGGTGAAAGCCAGAG-3', which amplify the CRISPR-Cas9-targeted region on exon 3 of KAT6A. PCR was performed using Taq polymerase with annealing temperature of 55ºC for 30 cycles. PCR products were subjected to electrophoresis in 2% agarose gel at 90V for approximately 60 minutes (Figure 4.4).

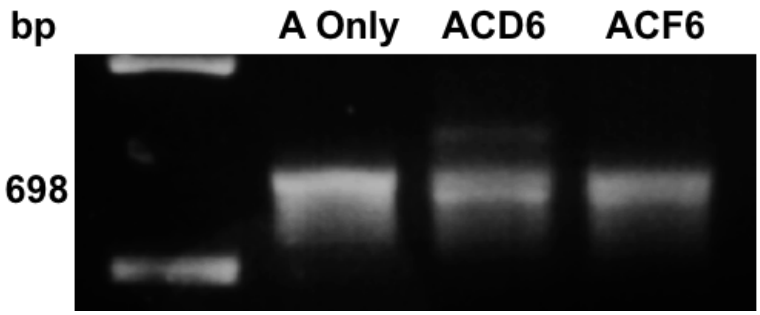


Figure 4.4 - PCR product of genomic DNA extracted from K562 cells clones tranfected with sgRNA targeting KAT6A. PCR products were run in a 2% agarose gel. A Only is the positive control transfected with only one sgRNA guide, ACD6 and ACF6 were transfected with a sgRNA pair guide in order to archieve KAT6A gene editing.

According to Naumann, Reutzel [361], there are 3 normal copies of chromosome 8 in K562 cells. As mentioned before, KAT6A is located in chromosome 8, thus it is expected to detect indels in up to three alleles. As shown in Figure 4.4, PCR product obtained from the clone ACD6 that was transfected with a pair of sgRNA guide presented three different bands size, one of them larger than the control A Only, the other one had similar size compared to control and the third one smaller than the control, suggesting deletions and insertions in the KAT6A gene. ACF6 band presented a slight difference in size compared to control, suggesting deletions in KAT6A gene.

In order to determine the exact deletions and insertions that took place in the KAT6A alleles, genomic DNA fragments from the three clones (A only, ACD6 and ACF6) were cloned into the pGEM®-T Easy vector (Promega, A1360) and transformed into *Escherichia*

coli JM109 cells (Promega, L200A), following the manufacturer's protocols. pGEM®-T Easy Vector is high-copy-number vector containing T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the α-peptide coding region of the enzyme β-galactosidase. After transformation, bacteria were plated in LB/ampicillin plates prepared with X-GAL, it allows the identification of successful cloning because the insert interrupts the coding sequence of β-galactosidase, thus colonies containing PCR products produce white colonies while blue colonies indicate unsuccessful cloning. Nine white colonies of each targeted clone were picked up and inoculated in LB/ampicillin media to grow up overnight at 37°C. Next day, 5 μL of each sample was used for colony PCR using *Taq* DNA polymerase and the same primers mentioned above. Colony PCR was the method chosen to screen for the indels because it eliminates the need to culture individual colonies and prepare plasmid DNA before analysis. The products were analysed by 2% agarose gel electrophoresis. A only clone was used as control.

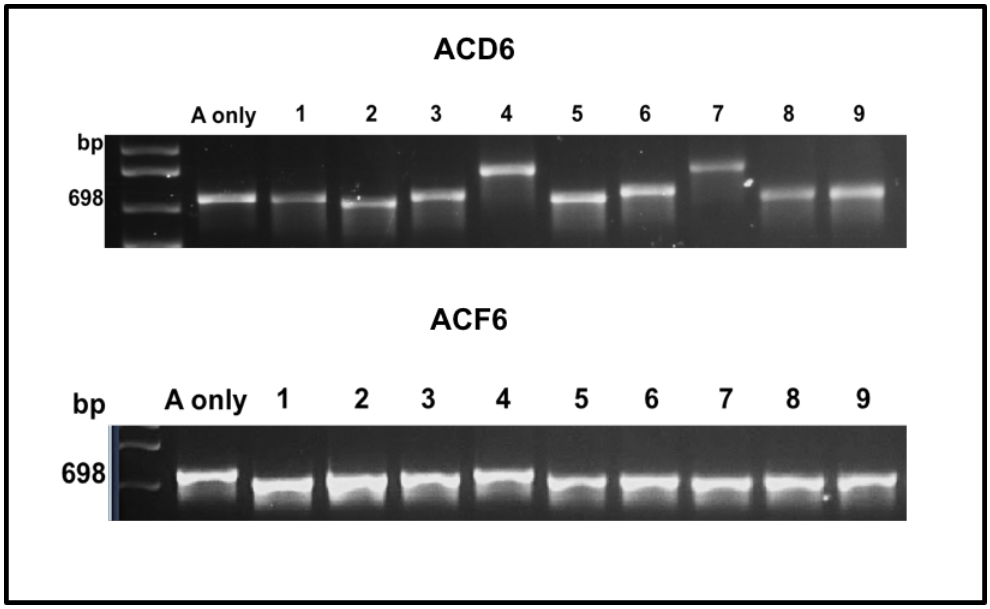


Figure 4.5 - Colony PCR of ACD6 and ACF6 targeted clones. Genomic DNA fragments of A only (control), ACD6 and ACF6 clones were cloned into the pGEM-T Easy vector and transformed into Escherichia coli JM109 cells. Nine colonies transformed with the plasmids containing ACD6 and ACF6 genomic DNA were subjected to colony PCR and the products were run in 2% agarose gel electrophoresis.

As shown in Figure 4.5, the detection of different cloned PCR fragment sizes clearly indicates genomic deletions and insertions in KAT6A gene. Insertions can be expected in ACD6 clone, as it is possible to notice bands (4 and 7) with higher mobility than the control. The band number 2 (and possibly 5,8,9) suggests a deletion, as it seems lower than the control. Clones 1, 3 and 6 appear to be similar in size to the wild type control (A only). This suggests the presence of at least 2/3 targeted alleles in K562-ACD6. Analysis of the cloned ACF6 PCR products suggests that, it is expected to detect only deletions events in KAT6A gene, since all of them are apparently lower than the wild type control.

In order to detect the specific indels events, plasmid DNA of the nine colonies showed in Figure 4.6 of each targeted clone (ACD6 and ACF6) and 2 colonies of A only were extracted using a Macherey-Nagel NucleoSpin Plasmid kit, then sequenced with the primer 5'-GTTCCCAGCCTCCGAGAATC-3' using an Applied BioSystems 3730 DNA Analyzer and PeakTrace software.

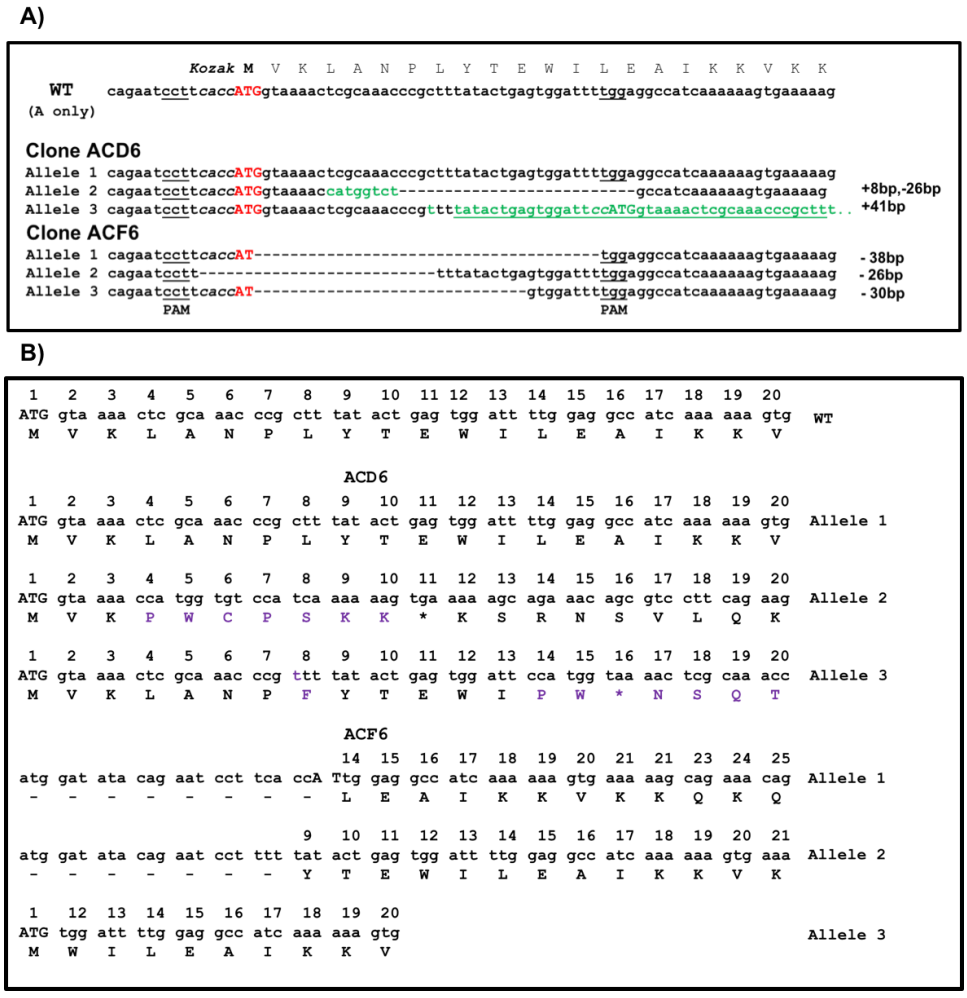


Figure 4.6 - Alignment of MOZ alleles in clones ACD6 and ACF6, relative to the WT and A only sequence. Each is shown in terms of (**A**) nucleotide sequence and (**B**) the translated protein sequence. Indels are detailed to the right of each allele and consequential alterations in protein sequences are in purple. Underlined regions represent the Cas9 PAM sequences. ATG start codon is in red, green text represents insertions, dashed deleted nucleotides.

Whereas the 'A only' control had only wild type KAT6A alleles as expected, deletions and insertions were observed in the targeted clones. ACD6 had one unadulterated allele, a second allele with a 26bp deletion and 8bp insertion with frame shift K3fs*7, and a third allele containing 48bp insertion with frame shift L8F and I13fs*2.

Clone ACF6 had no insertions, however, all the three alleles suffered deletions in the region between the PAM sequences: the first allele had a 38bp deletion which overlapped the ATG until codon L14, the second had a 26bp deletion removing before the ATG up to codon Y9 and the third had a 30bp deletion overlapping the ATG up to codon W12. The deletions of all the three alleles disrupted the ATG start codon, however, in the third allele ATG was fortuitously restored in front of W12. The second allele also lost the KOZAK consensus sequence. We confirmed that these were likely to be the only alleles present in the ACD6 ACF6 clone as each was sequenced multiple times and around 20 cloned fragments in total were sequenced for each K562 cell clone.

To further validate these results, we next used qPCR to detect MOZ transcript, and to assess the impact of these indels on expression of the MOZ/KAT6A gene. We therefore measured MOZ transcripts using qPCR primers specific to the ATG region in Exon 3 or a region of Exon 5/6 away from the targeting site (Figure 4.7).

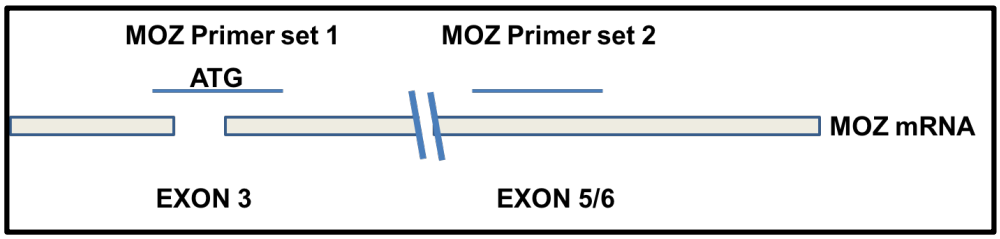
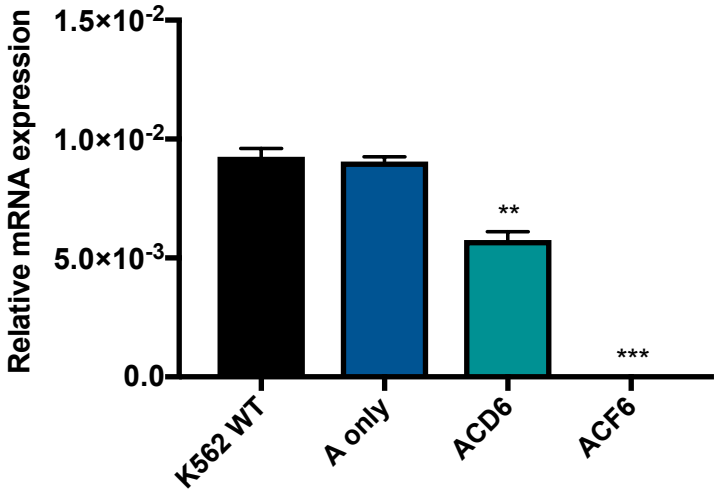


Figure 4.7 - Scheme showing the region where MOZ primers bind to MOZ transcript. MOZ primer set 1 binds to the exon 3 ATG start codon region, the targeting region used for gene editing using CRISPR CAS9n system. MOZ primer set 2 binds to a region correspondent to exon 5/6 which was not affected by CRISPR CAS9 editing.

As shown in Figure 4.8A, qRT-PCR performed using primers that span the MOZ ATG region (MOZ primer set 1) detected similar levels in untreated K562 cells and the 'A-Only' control, but failed to detect mRNAs containing this sequence in the ACF6 clone. This confirms the absence of alleles containing this sequence in ACF6. We observed a significant decrease in MOZ transcripts in ACD6 clone compared to controls, although in this case the primer can detetct allele 1 and allele 3 transcripts. When qRT-PCR was performed using a primers that detect a non-edited region (Exon 5/6) of KAT6A gene (MOZ primer set 2), MOZ transcripts were detected in ACF6, albeit at a much reduced level. In the case of the ACD6 clone, MOZ transcripts were detected at similar levels using either primer set (Figure 4.8B). This result suggests that the indels in ACF6 alleles affected the start codon appear to impact on the stability or expression of transcripts from these alleles. All gene expression values are shown relative to GAPDH expression.

A)



B)

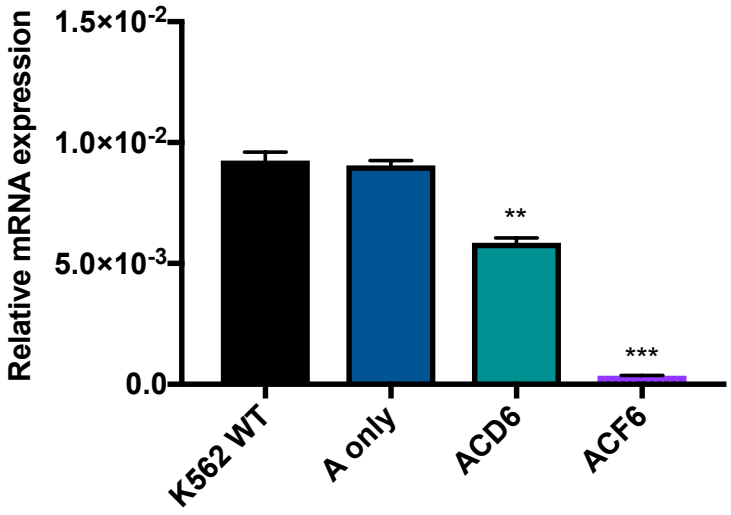


Figure 4.8 - qRT-PCR analysis of MOZ expression in edited KAT6A gene: (A) MOZ mRNA expression using primers that target the ATG start codon, **(B)** MOZ mRNA expression using primer that target a non-edited region of KAT6A gene in WT K562 cells, A only, ACD6 and ACF6 clones after KAT6A gene editing. Statistical analysis was performed via One-way ANOVA, with the p values represented as follows: $p < 0.001$ (**), $p < 0.0001$ (***). qRT-PCR data were normalized to GAPDH expression for each sample. Error bars, S.E.M. Each sample was compared to the expression of WT K562.

In order to confirm the effect of the KAT6A gene editing on MOZ protein expression, western blot was performed in K562 wild type cells, A only, ACD6 and ACF6 clones. As expected, K562 wild type and A-Only clone had similar MOZ protein expression, however, ACD6 and ACF6 had very low MOZ expression (Figure 4.9). The antibody used is a monoclonal antibody raised by our group against residues 441-460 of the MOZ protein sequence.

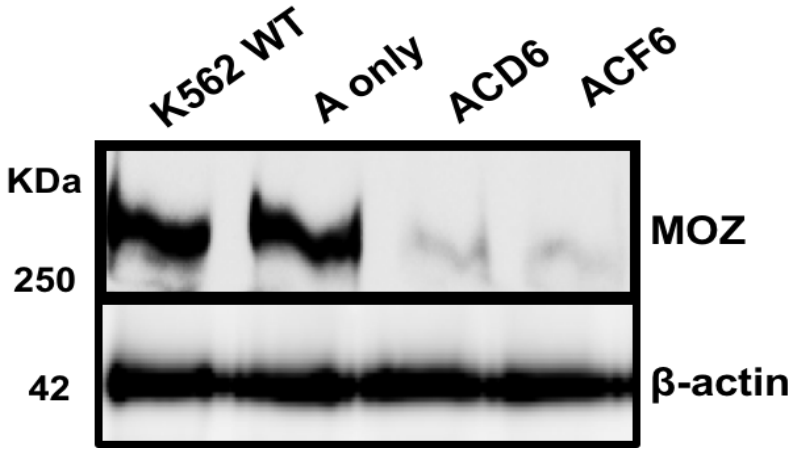


Figure 4.9 - Knockdown of MOZ protein in K562 cells using CRISPR-CAS9 system. Western blot validation of MOZ in WT K562 cells, A only, ACD6 and ACF6 clones after KAT6A gene editing. The total protein lysates for each K562 cells clone were equally loaded onto a protein SDS-polyacrylamide gel after measuring the concentration. The specific band for MOZ protein appeared above the 250-kDa band from the protein ladder and β-actin was used as a reference protein.

4.5 Validation of KAT6A genome editing events by MOZ target genes expression analysis

To further validate MOZ knockout in K562 cells by CRISPR-CAS9 gene editing, expression of MOZ gene targets in the ACD6 and ACF6 clones were assessed by qRT-PCR. As reported in the chapter 3, depletion of MOZ expression by siRNA and RNA-Seq

enabled us to detect MOZ regulated genes in K562 cells – including both up-regulated and down-regulated cohorts. We performed qRT-PCR to assess expression of some of these candidates including down-regulated genes PROX1, NR4A1, RPL7 and Serpine (Figure 4.10A) and the up-regulated genes CDKN2A, ID1 and XIST (Figure 4.10B). GAPDH was used as the housekeeping gene.

All the genes analysed presented the expected result considering the pattern of increase or decrease expression under MOZ depletion. Interestingly, the expression of some genes as NR4A1, RPL7, CDKN2A and XIST seem to be MOZ protein dose-dependent as their levels vary between ACD6 clone that still has one MOZ allele intact and ACF6 clone that has deletion in all the three MOZ alleles. On the other hand, Serpine and ID1 had the same expression level in the ACD6 and ACF6 clones.

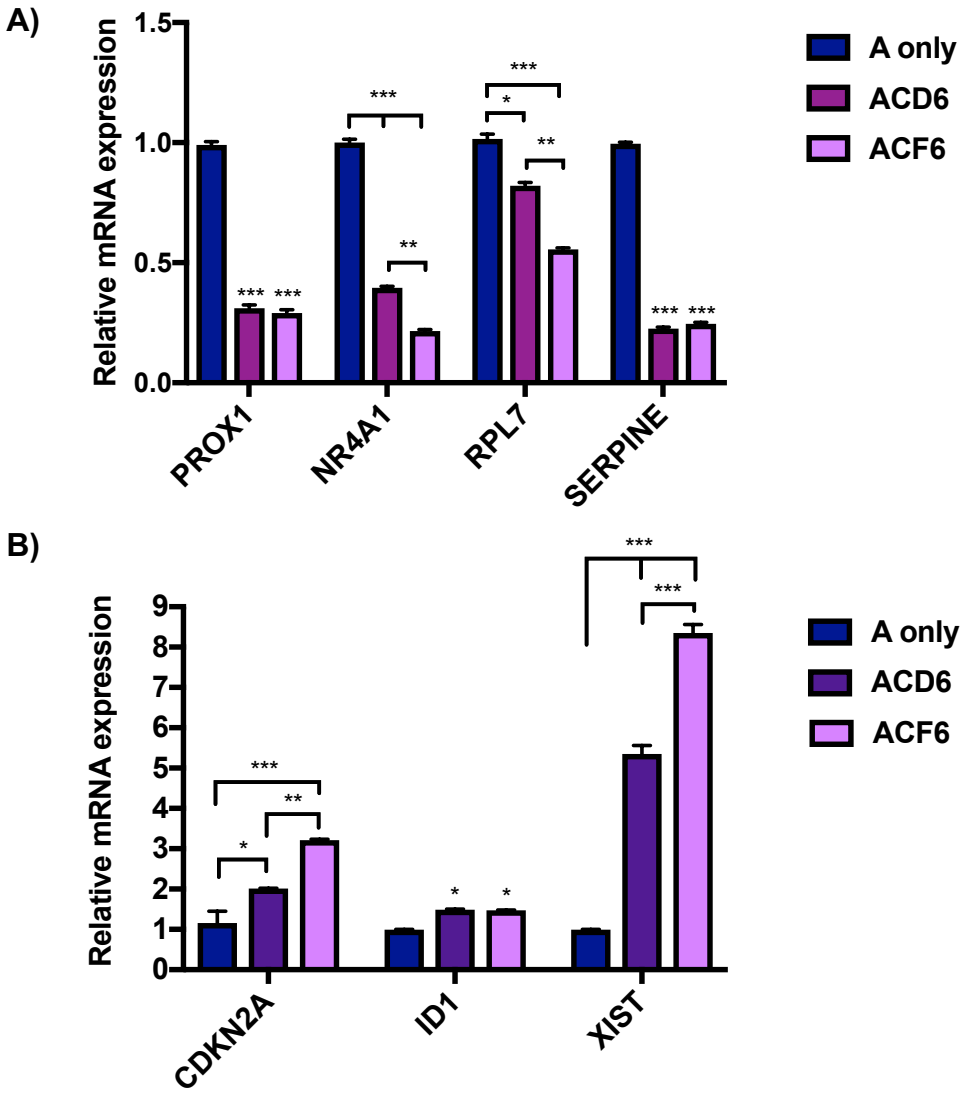


Figure 4.10 - qPCR validation of selected down and up-regulated genes in A only, ACD6 and ACF6 clones. Quantitative RT-PCR showing changes in mRNA transcript levels between the control A only and the KAT6A gene edited clones ACD6 and ACF6. **(A)** Four down-regulated and **(B)** 3 up-regulated genes were chosen for validation. qRT-PCR data were normalized to GAPDH expression for each sample and the mean expression level in the control samples set to 1. One asterisk indicates a significant difference ($P<0.05$), two asterisk $p<0.01$ and three asterisk $p<0.001$ based on student's t-test analysis between control and clone samples. Error bars, S.E.M.

In order to re-introduce exogenous MOZ expression and thus confirm the function of MOZ protein in the expression of these target genes, a rescue experiment was performed. ACF6 clone was chosen for this experiment because all its 3 MOZ alleles were edited. ACF6 was transfected with pEYFP-C1 plasmid empty vector or YFP-MOZ sequence, from our laboratory stocks [252].

A-Only and ACF6 clones transfected with GFP were used as control. Transfection was performed with 1 x 10^6 cells, 1 µg of GFP or 1µg of YFP-MOZ plasmid by electroporation using Amaxa® Cell Line Nucleofector® Kit V following the instruction. Two days after transfection, cells expressing YFP or GFP were sorted by flow cytometry. Total RNA was extracted from the sorted cells and cDNA was made to be used as template for the qPCR reaction. As shown in Figure 4.11, MOZ expression was rescued successfully and the target genes PROX1, HHEX, Serpine, SIRT4 and IDI had their expression re-established to a level similar to the control A-Only. Although PLAT and CDKN2A had statistical difference compared to control A only, the pattern of expression is the expected once PLAT is downregulated under MOZ depletion and when MOZ is rescued it is expected that PLAT increases its level, exactly what was observed in the result. The same was observed with CDKN2A, however it is up-regulated under MOZ depletion and should decrease the expression when MOZ expression is re-introduced. The result shows that CDKN2A is down-regulated when MOZ is rescued but its level is even lower than the control A only. This may possibly reflect a higher

expression of exogenous MOZ in the rescued cells compared to endogenous MOZ in the controls.

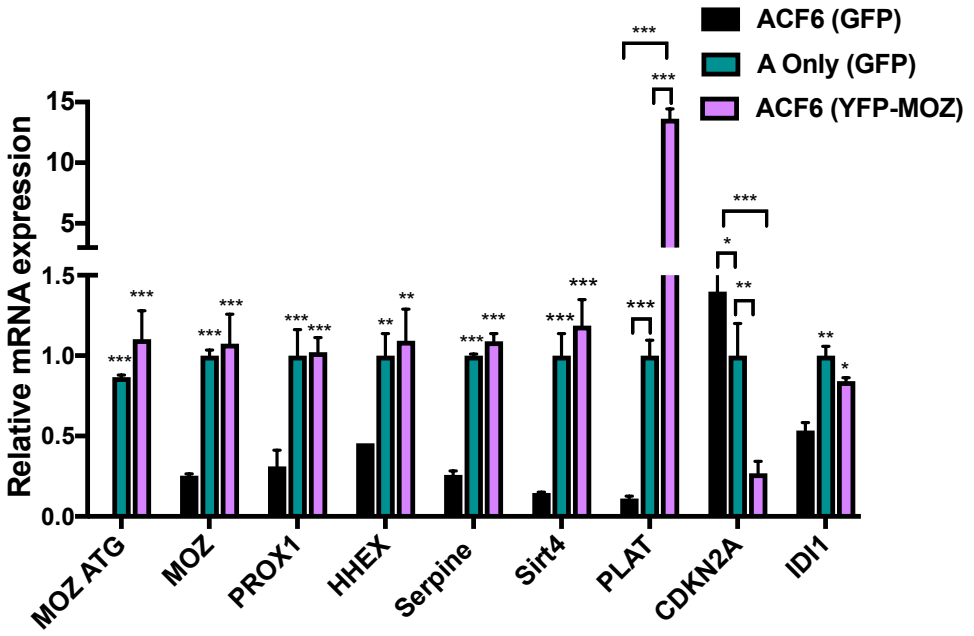


Figure 4.11 - qRT-PCR of MOZ target genes after MOZ rescue expression in K562 MOZ knockout cells. Expression of 7 MOZ target genes were assessed by qRT-PCR after MOZ expression being rescued in ACF6 clone cells. MOZ expression was also assessed using a primer that targets the ATG start codon sequence and another primer that target a sequence that was not edited in the KAT6A gene. qRT-PCR data were normalized to GAPDH expression for each sample and the mean expression level in the control samples (A-only (GFP)) set to 1. One asterisk indicates a significant difference ($P<0.05$), two asterisk $p<0.01$ and three asterisk $p<0.001$ based on Two-way ANOVA analysis between control ACF6 transfected with GFP and the other clone samples and between A only transfected with GFP and the ACF6 transfected with YFP-MOZ plasmid. Error bars, S.E.M.

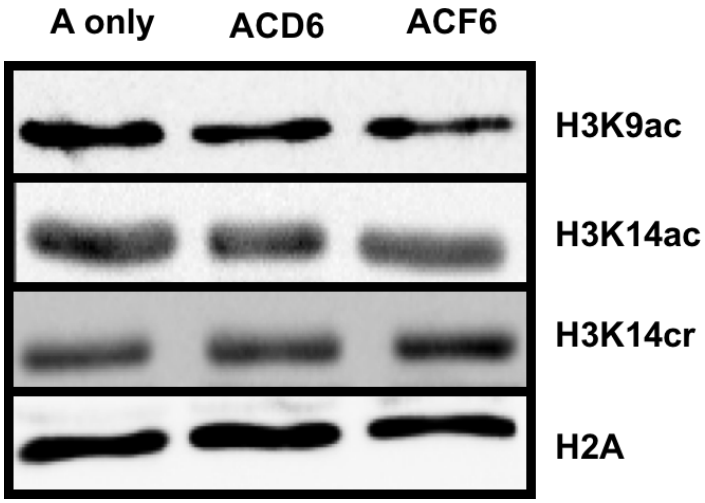
4.6 Effect of MOZ knockdown on H3K14 crotonylation levels

As discussed in chapter 3, MOZ acetylates histone H3K14 and H3K9. When MOZ was knocked down in K562 cells using siRNA, 72 hours after transfection, the levels of H3K9 and H3K14 had a visible reduction compared to control cells transfected with non-target siRNA (Figure 3.3). Here, we performed further validation of role of MOZ in H3K14 and H3K9 acetylation levels by western blot in ACD6 and ACF6 clones and compared to A only clone, used as control. As showed in Figure 4.12, both ACD6 and ACF6 had reduction in H3K9ac and H3K14ac levels, corroborating with the previous result showed in Chapter 3 following siRNA knockdown of MOZ in K562 cells. ACF6, the clone that had deletions in all the three KAT6A alleles, presented lower levels of H3K14ac and H3K9ac than ACD6 clone, which had one intact KAT6A allele. Densitometry readings showed were calculated relative to H2A using Image J software.

It was recently reported that in addition to the function of the as MOZ DPF domain in binding to acetylated H3K14, MOZ also functions as a specific reader of H3K14cr [191]. Moreover, others have shown that MOF, another MYST family acetyltransferase, has crotonyltransferase activity [362]. Thus, in addition to analysis of acetylation levels of H3K14 and H3K9, we assessed the bulk levels of H3K14 crotonylation. The western blot result shows that in ACD6

and ACF6 clones, the level of H3K14cr was not decreased, suggesting that MOZ is not essential for this histone modification. In fact, H3K14cr levels appeared to be slightly increased in in ACD6 and ACF6 clones compared to the control A only, with ACF6 clone having higher levels of H3K14cr than ACD6. This preliminary result may suggest that reduced H3K14 acetylation by MOZ allows other epigenetic regulators to impart crotonylation at H3K14. Alternatively, the expression of other crotonylases e.g. MOF may be increased upon depletion of MOZ to compensate its function. Further studies will be required to investigate this.

A)



B)

Densitometry			
	A only	**ACD6**	**ACF6**
H3K9ac	1.00	0.89	0.68
H3K14ac	1.00	0.85	0.80
H3K14cr	1.00	1.05	1.12

Figure 4.12 - Effect of KAT6A gene editing in histone modification. A) Histone acetylation and crotonylation levels of ACD6, ACF6 and A only (control) clones of K562 cells were analyzed by Western blotting with anti-H3K9ac, anti-H3K14ac and anti-H3K14cr antibodies. H2A was used as loading control. **B)** Densitometry readings were calculated, relative to H2A, using Image J software.

4.7 Impact of leukemogenic protein MOZ-TIF2 on expression of MOZ target genes

MOZ–TIF2 is a fusion protein resulting from chromosomal translocations involving KAT6A gene and the nuclear receptor co-activators TIF2 (Transcriptional Intermediary Factor) [231]. Some studies have shown that MOZ–TIF2 transforms haematopoietic progenitors and induces acute myeloid leukemia in a murine bone marrow transplant model [254, 255]. It also induces broad changes in histone acetylation destabilizing the development of haematopoietic progenitors and altering gene expression programs important for myeloid differentiation [202, 252].

To verify the effect of MOZ-TIF2 expression on some MOZ target genes, ACF6 clone was transfected with vectors for GFP or YFP-MOZ-TIF2 from our laboratory stock [202, 252]. A only and ACF6 clones transfected with GFP were used as a control. Rescue experiments were performed as described for YFP-MOZ.

The result shows that transfection was successful as MOZ expression was rescued. Once again we used two sets of primers for MOZ detection: one spanning the ATG start codon and the another target a region of the gene that is preserved in MOZ-TIF2 translocation. Interetsingly, it was observed that MOZ-TIF2 (like MOZ) can rescue the expression of all the genes analysed: MOZ, PROX1, HHEX, Serpine, Sirt4, PLAT, CDKN2A and IDI1 (Figure 4.13). PLAT was the only gene that had a statistical difference

between the A only clone control cells transfected with (GFP) and the ACF6 clone cells transfected with MOZ-TIF2 plasmid. This pattern was also observed when ACF6 clone had MOZ expression rescued, as can be seen in Figure 4.11.

Once the expression of the genes analysed after MOZ-TIF2 transfection was similar to the expression in the control A only, it is possible that the MOZ HAT activity of the fusion protein may be essential for this gene expression regulation. It is true at least for p16 gene that was shown to be repressed by MOZ-TIF2 and that HAT domain of MOZ is crucial for this repressive activity [336].

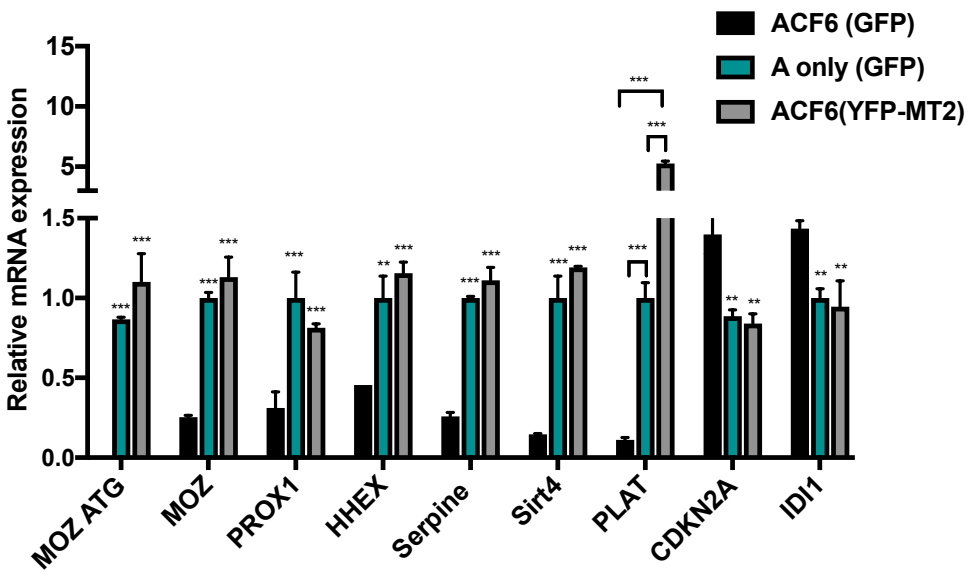


Figure 4.13 - qRT-PCR of MOZ target genes after MOZ-TIF2 transfection in K562 MOZ knockout cells. Expression of 7 MOZ target genes were assessed by qRT-PCR after MOZ-TIF2 transfection in ACF6 clone cells. MOZ expression was also assessed using a primer that targets the ATG start codon sequence and another primer that target a region of the gene that is preserved in MOZ-TIF2 translocation. qRT-PCR data were normalized to GAPDH expression for each sample and the mean expression level in the control samples (A-only (GFP)) set to 1. The control samples ACF6 (GFP) and A-only (GFP) are the same showed in Figure 4.13. Two asterisk indicates a significant ($p<0.01$) and three asterisk $p<0.001$ based on One-way ANOVA analysis between control ACF6 transfected with GFP and the other clone samples and Student t-test was used for statistical analysis between A only transfected with GFP and the ACF6 transfected with YFP-MOZ plasmid. Error bars, S.E.M.

4.8 Effect of KAT6A ablation on the K562 cells proliferation

Acute leukaemias are characterized by a rapid increase in numbers of immature blood cells. This interferes with the proliferation and/or differentiation of normal stem cells in the bone marrow

causing anemia, granulocytopenia, thrombocytopenia. The rapid leukaemic cell proliferation also causes their accumulation into the bloodstream and spread to other organs [363].

Here, we evaluated ACD6 and ACF6 proliferation and compared to A only to verify if absence of MOZ protein affects proliferation of K562 leukemic cells in culture. For this purpose, MTS assay was performed. As can be seen in Figure 4.14, ACD6 and ACF6 clones had lower proliferation levels than the A only clone. ACD6 had a reduction of approximately 50% and ACF6 70% compared to A only clone. It is also interesting notice that ACF6 grows slower than ACD6 clones.

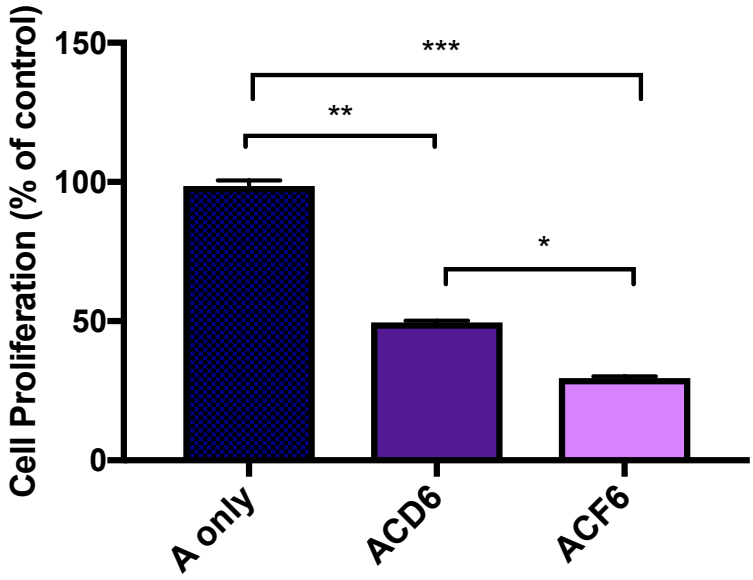


Figure 4.14 - Metabolic activity of ACD6 and ACF6 clones. Cell viability of ACD6 and ACF6 were assessed using MTS assay. A only was used as control. Statistical analysis between A only and the two other clones was performed using One-way ANOVA and between ACD6 and ACF6 was used Student t-test. One asterisk indicates a significant difference ($p<0.05$), two asterisk $p<0.01$ and three asterisk $p<0.001$. Error bars, S.E.M.

Flow cytometry cell cycle analysis using propidium iodide DNA staining was also performed on ACF6 and A only clone (control). The result indicates that around 70% of the ACF6 clone cells are in the G0-G1 cell cycle phase, around 40% more compared to A only cells. Furthermore, ACF6 had around 30% less cells in the S phase (the phase of DNA synthesis) and around 40% less cell in the G2-M phase than A only clone (Figure 4.15). The cell cycle in the G0-G1 phase observed is likely due to the derepression of the p16INK4a, encoded by CDKN2A gene, that plays an important role in cell cycle regulation by inhibiting CDK4 and CDK6 and causing blocking the progression from G1 to S phase acting as a tumor suppressor [331-335]. It is notable that the percentage of cells arrested in G1 is higher in the CRISPR edited K562 clones than observed in the siRNA-treated K562 cells (Chapter 3) in which transfection efficiency would be a variable factor. Consistent with the results presented in Chapter 3, CDKN2A expression was elevated in the ACD6 and ACF6 clones compared to control K562 cells, and can be repressed by adding back exogenous MOZ or MOZ-TIF2. Taken together these results indicate that KAT6A/MOZ is required for proliferation of K562 cells and that depletion of MOZ results in more number of cells in G1 phase.

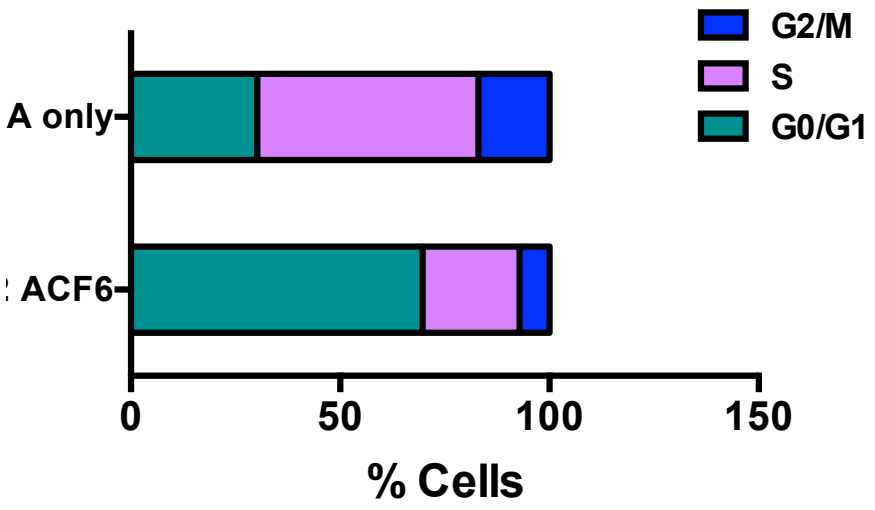


Figure 4.15 - Cell cycle analysis of ACF6 clone. Propidium iodide (PI) staining and flow cytometric analysis performed in ACF6 and the control A only. Percentage of cells are shown in the cell cycle phases: G0/G1, S and G2/M.

Senescence is defined as an arrested state in which the cells remain viable but not stimulated to divide. It is characterized by molecular and morphological cell changes as an irreversible cell cycle arrest, increase of number and size of lysosome and increase of cell granulation [364, 365]. Increase of senescence-associated β-galactosidase activity (SA-β-gal) is also observed in senescent cells and is the most used biomarker, along with p16 expression, for senescent and aging cells [366]. There is some controversy in the literature regarding the induction of senescence with opposing findings for the role of MOZ in this process [205, 226, 258, 335] [268, 336, 367].

To assess if we were able to detect evidence for sensescence in the MOZ depeleted cells we measured SA-β-gal assay was

performed on ACD6 and ACF6 clones and compared to A only control. However, we were unable to detect any significant change in the number of senescent cells in control K562 or the MOZ-depleted cell lines using this assay, suggesting that cell cycle arrest and decrease in cell proliferation caused by MOZ ablation is not associated with senescence, at least under the conditions used (Figure 4.16). It would be interesting having a positive control, as a culture of normal human fibroblasts induced to senesce by oxidative stress or ionizing radiation.

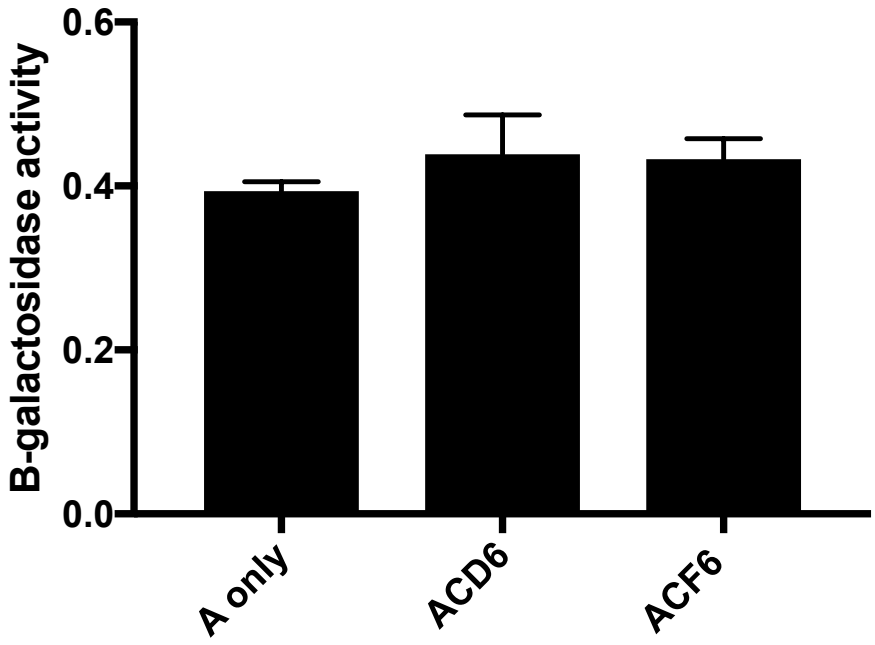


Figure 4.16 - β-galactosidase assay to verify senescence in ACD6 and ACF6 clones. Quantitative analysis of ß-galactosidase activity of ACD6 and ACF6 clones. One-way ANOVA was used to verify statistical difference between A only and the clones ACD6 and ACF6. Error bars, S.E.M.

Cell apoptosis response is critical for suppressing tumorigenesis and cell proliferation in complex eukaryotes [368, 369]. As senescence did not appear to be enhanced in ACD6 and ACF6 clone cells, we next performed apoptosis analysis using annexin-V / PI assay by flow cytometry. This method allows the identification of viable, apoptotic and dead cells. Live cells are negative for annexin-V and PI, early apoptotic cells are positive for annexin-V, cells in late stages of apoptosis are positive for annexin-V and PI and necrotic cells are positive only for PI. Three controls were used: K562 cells unstained, treated with annexin-V FITC only and treated with PI only.

In Figure 4.17, the result is shown considering the proportions (%) of different cell populations in each stage compared to controls. Basal levels of apoptosis and cells autofluorescence (negative control) were taken into consideration for the analysis. The result indicates that only 5% of the cells of ACD6 and ACF6 clones were in late apoptosis stage, 95% were live cells. The control A only clone presented 99% of live cells.

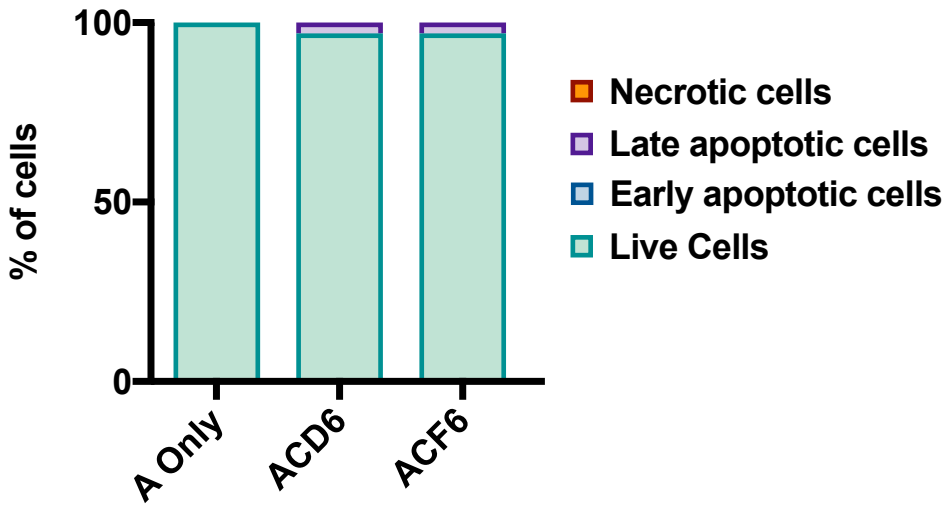


Figure 4.17 - Apoptosis assay using annexin-V / PI assessed for the ACD6 and ACF6 clones. ACD6 and ACF6 apoptosis analysis was performed by annexin-V / PI assay using flow cytometry. Live cells are negative for annexin-V and PI, early apoptotic cells are positive for annexin-V, cells in late stages of apoptosis are positive for annexin-V and PI and necrotic cells are positive only for PI.

In order to confirm the above result, caspase activity was assessed using CellEventTM caspase-3/7 green detection reagent combined with SYTOXTM AADvanced red, with this combination is possible discriminate activated caspase 3/7 from necrotic cells using flow cytometry. Viable cells were negative for CellEventTM Caspase 3/7 and SYTOXTM red, apoptotic cells were positive for CellEventTM Caspase 3/7 and necrotic cells were positive for SYTOXTM red and for CellEventTM Caspase 3/7. Unstained K562 cells and controls for CellEvent Caspase 3/7 and Sytox were used.

As shown in Figure 4.19, 100% of the control A only clone cells were viable and the MOZ edited gene clones ACD6 and ACF6 also presented 100% of viable cells, indicating that MOZ ablation does not lead cells to apoptosis.

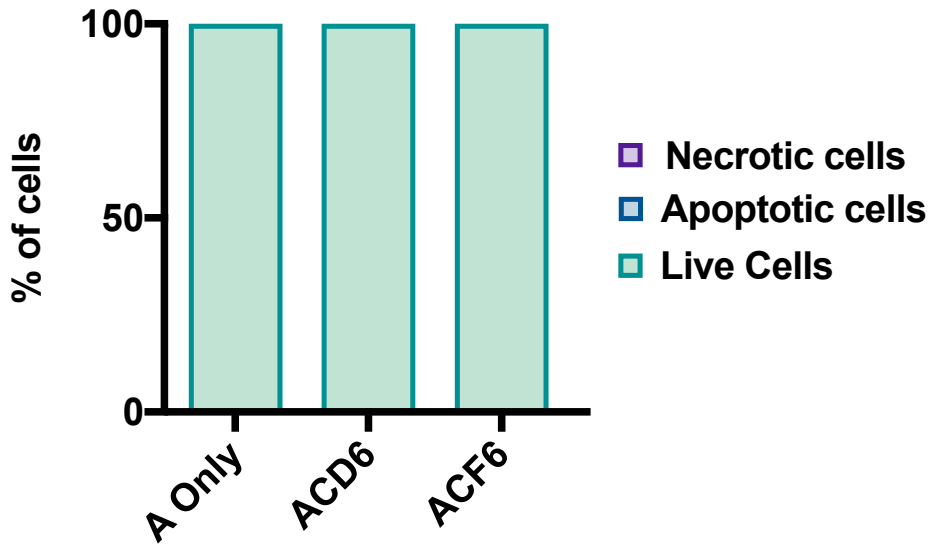


Figure 4.18 - Caspase activity assessed for the ACD6 and ACF6 clones. Caspase activity was assessed using CellEventTM caspase-3/7 green detection reagent combined with SYTOXTM AADvanced red, with this combination is possible discriminate activated caspase 3/7 from necrotic cells using flow cytometry. Viable cells are negative for CellEventTM Caspase 3/7 and SYTOXTM red, apoptotic cells are positive for CellEventTM Caspase 3/7 and necrotic cells are positive for SYTOXTM red and for CellEventTM Caspase 3/7.

4.9 Effect of MOZ ablation on K562 cells differentiation

K562 cells are multipotent and previous studies have shown that they can undergo differentiation upon exposure to different chemical agents such as hemin or 5-azacytidine [370-373]. This results in the synthesis of hemoglobins [374]. Recent studies have revealed a role for the corepressor NCOR in erythroid differentiation and haem syntheisis [375]. It was noted that cell pellets from ACF6 cells had a red colour in comparison to the control cells, similar to

that induced by exposure to hemin (Figure 4.20 shown for comparison with [375]). While there was insufficient time to confirm if the red colour indicates increased haemoglobin synthesis and thus differentiation of K562 cells, if this can be confirmed it would be consistent with the hypothesis that MOZ is necessary to maintain the undifferentiated state of haematopoietic progenitors [267]. Future plans will include the comparison of globin gene expression and haemoglobin levels in the control and ACF6 K562 cells.

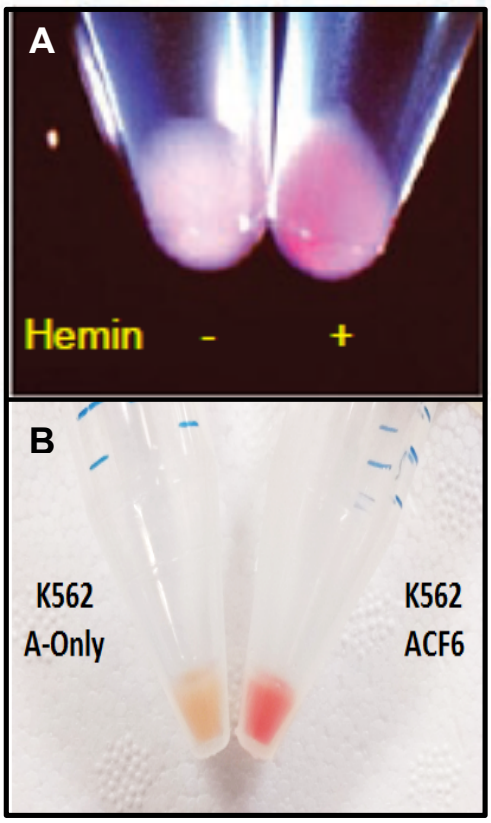


Figure 4.19 - Difference in K562 cell pellet colour between WT and treated cells. A) K562 cells were induced for erythroid differentiation with hemin and the pellet presented red color while the control (cells non-treated with hemin) maintained the white color [375]. B) Similar red color is observed in MOZ knockout K562 cells (ACF6) clone pellet while A-only clone control also preserved the white colored pellet.

4.10 Summary

In order to generate MOZ deficient K562 cells CRISPR CAS9 nickase technology was employed. This approach has been reported to greatly reduce the frequency of off target events [356, 357]. K562 cells were then transfected with vectors to express A/B or A/C sgRNA pairs or A sgRNA only as control. Three clones from isolated single K562 cells were able to form colonies and expand: One corresponds to a single guide transfection (A only) and the other two correspond to the pair guide transfection A/C (ACD6 and ACF6).

Genotyping analysis confirmed that all alleles in the A-Only control were wild type whereas indels were detected in clones ACD6 and ACF6. It was confirmed the absence of the deleted region spanning the ATG in MOZ transcripts in ACF6 and the strongly reduced/ absent expression of MOZ in both targeted cell lines. Interestingly, the indels were located between the two PAM sequences in the guide pairs showing this approach can be used to generate substantial indels, whereas full nuclease CAS9 can often generate shorter indels.

Although ACD6 had one unedited allele, the protein expression level of MOZ in western blots was very low. However, the effects on gene expression were more acute in ACF6 which has all three alleles edited, albeit potentially with one in frame deletion of the first 11 amino acids. Given the size of MOZ (approximately 250 kDa) and its migration in SDS PAGE gels, a short deletion would not

be possible to detect, thus we cannot be sure if the truncated protein is expressed. Interestingly, a recent paper showed that the N-terminal sequence of MOZ and MOZ-TIF2 is essential for the recruitment of PRMT1 and KDM4C cofactors which methylate H3K4 and demethylate H3K9 respectively [345]. Thus even if this allele is expressed at low level it may be functionally compromised.

In Chapter 3, knockdown of MOZ in K562 was coupled with RNA-Seq detection to detect MOZ regulated genes. We therefore validated the misregulation of these genes following MOZ ablation in K562 cells after CRISPR-CAS9 gene editing. Consistent with the previous data, expression of NR4A1, RPL7, Serpine were strongly reduced in ACD6 and ACF6, whereas expression of CDKN2A, ID1 and XIST were increased compared to the unedited control. Interestingly, the level expression of some genes as NR4A1, RPL7, CDKN2A, XIST and PLAT seems to be MOZ level expression dependent as their levels vary between ACD6 clone that still has one MOZ allele intact and ACF6 clone that has deletion in all the three MOZ alleles. On the other hand, Serpine and ID1 showed the similar expression level in the ACD6 and ACF6 clones.

We then performed rescue experiments in which YFP-tagged MOZ (which functions similarly to untagged MOZ [252] was expressed in the ACF6 and ACD6 lines. This resulted in a robust rescue of expression of MOZ-dependent genes to a level greater or equal to the controls.

In previous work from our group, it was shown that the MOZ

protein can acetylate H3K9 and H3K14 *in vitro* in the context of core histones or histone tail peptides. This involves not only the HAT domain located between amino acids 510-810, but also the DPF domain which can bind the H3 tail dependent on its modification status. It was shown in HAT assays that the rate of acetylation of H3 is strongly enhanced (8-10 fold) by the presence of the DPF due to its ability to bind the H3 tail and induce alpha helical folding of the tail to promote H3K9 and H3K14 acetylation. Consistent with this, the MOZ-ablated cell lines of K562 showed an overall reduction in cellular levels of H3K9ac and H3K14ac. ACF6 had lower levels of H3K14ac and H3K9ac than ACD6 clone. This result is probably due to the intact MOZ allele that ACD6 still preserves causing higher expression of MOZ and consequently higher acetylation activity in ACD6 clone compared to ACF6. It is likely that other HATs including MYST HATs may also acetylate these sites, so that some compensation is possible to permit essential gene regulation and other processes to continue. However, this global reduction may contribute to the proliferation defects and possibly hamper the role of MOZ in maintaining pluripotent state of K562.

Histone lysine crotonylation was recently identified as a non-acetyl acylation that is widespread in mammalian genomes and marks active chromatin and is enriched at promoter and enhancer region [137]. However, histone crotonylation functions and distinction from acetylation are still largely unknown. The MOZ DPF domain was shown to be a specific reader of H3K14cr. Xiong, Panchenko [191]

observed 4-fold enhancement in MOZ DPF domain binding of H3H14cr over H3K14ac, and both acyl groups dock in the same pocket on the DPF surface. Moreover, it has been shown that another MYST family protein (MOF) can catalyse crotonylation of H3 [362], opening up the possiblility that other MYST proteins may have similar activities. Therefore we also assessed the effect of KAT6A ablation on H3K14 crotonylation levels.

Our results showed that H3K14cr levels were not reduced but in contrast were marginally higher in the ACD6 and ACF6 clones compared to the control A only. Also, the levels of crotonylation in ACF6 were higher than in ACD6. This result suggests that the H3K14cr levels may increased under MOZ knockout to compensate the reduced levels of acetylated H3K14. It may be done by DPF2 protein, for example, that is also a specific reader of H3K14cr and binds preferentially to H3K14cr over H3K14ac [191]. Alternatively, there may be competition between acetyltransferases and crotonylases to modify this site. While we can conclude from these results that MOZ does not appear to crotonylate H3K14, it is possible that crotonylation of lysines at other locations in histones or other proteins may be possible. This could be established by proteomic analysis or by using pan-crotonyl antibodies to compare MOZ^{+} and MOZ^{-} cell lines.

We assessed the ability of MOZ-TIF2 fusion protein to rescue normal expression of MOZ target genes. MOZ-TIF2 conserves the histone acetyltransferase MYST domain [231], but was shown to acts

as a dominant inhibitor of the transcriptional activities of CBP-dependent activators [252]. The capacity of MOZ-TIF2 to immortalize primitive myeloid progenitors is dependent of the C2HC nucleosome recognition motif of MOZ, however, MOZ HAT activity is not essential suggesting protein interaction or chromatin interaction may be more critical [255]. However other studies have shown that the N-terminal region of MOZ-TIF2 also contributes to leukaemia development by recruitment of KDM4c and PRMT1 or by association with BRPF1 to regulate HOX gene expression [367]. The HOX expression regulation is MOZ-dependent histone acetylation [376].

Our rescue experiments have shown that both MOZ and MOZ-TIF2 can rescue expression of MOZ-dependent genes including PROX1, HHEX, Serpine, Sirt4 and PLAT. This suggests that the C-terminal domain of MOZ is not essential to regulate the expression of these genes in the K562 cell line. Our group previously showed that MOZ and MOZ-TIF2 both activate AML1 regulated reporter genes, but have opposing effects on p53 and RAR/RXR reporters, due to the detrimental effect of MOZ-TIF2 on CBP/p300 levels [252, 253].

The enhanced CDKN2A observed in in ACF6 compared to wild type K562 was repressed by exogenous MOZ-TIF2, although we did not observe the same extent of repression as for rescue by wild type MOZ. This may be due to a slightly lower level expression of MOZ-TIF2 in the rescue experiment as may be indicated by the qPCR data for exon 3 containing transcripts. However, consistent

with other studies the results indicate that both MOZ and MOZ-TIF2 can function in the repression of CDNK2A. This is likely to be dependent on the activity of the HAT domain of MOZ, which is crucial for the repression of p16 [336].

MOZ has being reported to play an important role in cell proliferation. Mice lacking MOZ HAT activity fail to maintain a normal number of hematopoietic precursors and exhibit defect in B-cell development [214] [226]. Partial inhibition of MOZ is capable to reduce proliferation of MEIS1 and HOX-driven lymphoma and leukemia cells [215].

Consistent with this, MTS assay reported here show that MOZ deficient K562 cells show a proliferative reduction of approximately 50% in the ACD6 clone and 30% in ACF6 compared to A only clone. It indicates not only that MOZ plays a role in K562 cell proliferation but also that the gene dosage needs to be taken in consideration as ACD6 clone, which preserves one MOZ allele, presented more proliferative capacity than ACF6 that has all the 3 alleles edited. Cell cycle analysis was also performed in ACF6 cells by flow cytometry using propidium iodide DNA staining and the result shows that around 70% of the cells were arrested in the G0-G1, the period in cell cycle in which cells stand in an resting or non-cycling state. Compared to A only control, around 40% more cells were arrested in G0-G1 phase. Considering the S phase, the part of the cell cycle in which DNA is replicated, ACF6 had around 30% less cells compared to control. Furthermore, ACF6 had around 40% less cell in the G2-M

phase than A only clone.

Although it was demonstrated that around 70% of the ACF6 clone cells is arrested in the G0-G1 phase and ACF6 and ACD6 clones presented lower cell proliferation rate compared to A only, senescence has not been found in the ACF6 and ACD6 clones. It is important to notice that was no positive control in this experiment, for this reason, the result is not conclusive.

As discussed in chapter 3, there is a conflicting report in the literature about the role of MOZ in the senescence. While Perez-Campo, Costa [226] suggest that MOZ is important to suppress senescence, Rokudai, Aikawa [205] suggest that cells are not able to senesce and undertake apoptosis under MOZ absence.

Apoptosis analysis was also performed in the ACF6, ACD6 and A only (control) clone cells by annexin-V/PI and caspase 3/7 activity methods using flow cytometry. The results did not indicate a role for KAT6A/MOZ in the apoptosis process in these cells.

In summary, the findings show successful generation of MOZ knockout K562 cells using CRISPR-CAS9 system, confirming that MOZ regulates the expression of some genes identified by siRNA/ RNA–Seq, as discussed in the chapter 3. Our results also confirm the role of MOZ in maintaining levels of H3K14ac and H3K9ac and an apparent inverse relationship the levels of H3K14 crotonylation, which will require further analysis. Furthermore, the results show that MOZ is important for cell growth and the gene dosage influence on the cell proliferation rate. MOZ ablation in K562 cells causes G1

arrest concomitant with CDKN2A/p16INK4A expression arrest but, does not appear to induce senescence apoptosis.

The MOZ/KAT6A deficient cells generated in this study, and the proof of principle that re-expression of MOZ can rescue endogenous gene expression provides an excellent cell model to study the function of different domains and activities of the MOZ protein in repression and activation of target genes. For example, colleagues are using these cell lines to study the effect of mutations disabling specific MOZ functions (e.g. BRPF1 interactions) on MOZ activity. The CRISPR approach is also being used to introduce mutations that mimic the clinical variants in Exon 17 and Exon 18 associated with global developmental delay.

Preliminary evidence suggests an increase in haemoglobin consistent with the hypothesis that MOZ helps maintain the undifferentiated state of these multipotent cells. For this reason, our next step will use these approaches to interrogate the role of MOZ in K562 cells differentiation, and the molecular complexes involved. Taken together, the results confirm MOZ as a promising therapeutic target for leukaemia and for the exploration of epigenetic regulation of cell differentiation and gene expression.

Chapter 5 : IBET-151 Inhibits Expression of MOZ-Dependent Genes: Evidence for MOZ-BRD4 complexes

5.1 Background

As discussed in previous chapters, histone modifications play a key role in determining the conformation and function of chromatin, and as a consequence, are involved in regulation of different cellular responses such as transcriptional regulation, DNA repair and replication, depending on the type and the site of the covalent mark [68, 377-379]. The ability of chromatin regulators to recognize combinatorial heterotypic post-translational modifications in their histone substrates is essential to their function. These include bromodomains, chromodomains, tudor domains and PHD fingers [97]. Because aberrant chromatin modifications are associated with disease, the proteins that recognise chromatin modifications are considered promising therapeutic targets in diseases such as cancer.

The bromodomain is a 110 residue module present in more than 50 nuclear proteins that functions as a reader of acetylated lysines in histones and other proteins. One family of proteins known as the BET proteins (bromodomain and extra-terminal domain) includes the BRD2,3,4 and BRDT proteins. These proteins are associated with actively expressed genes and are likely to link acetylated histones at enhancers with the transcriptional machinery [380, 381].

Small molecule inhibitors that prevent the BET bromodomain interaction with acetyl-lysine have been developed to treat malignancies due its ability to affect the transcription of a limited group of genes, thus resulting in specific inhibition of disease

promoting genes [382]. IBET-151 (GSK1210151A) is one of BET inhibitors that have been tested in preclinical cancer models for MLL-fusion leukaemia [383]. The target of this inhibitor is the interaction of bromodomain proteins with acetylated lysine, including acetylated H3K14, known as active mark and is associated with ongoing transcription (Wang *et al*. 2008; Heintzman and Ren, 2009) [384] showed that this inhibitor has efficacy against human and murine MLL-fusion leukaemic cell lines through the induction of early cell cycle arrest and apoptosis. I-BET 151 also demonstrated therapeutic value *in vivo*, providing survival benefit in two distinct mouse models of murine MLL-AF9 and human MLL-AF4 leukaemia.

Because MOZ and BRD2-4 can recognise H3K14ac, we decided to investigate whether MOZ-dependent genes are sensitive to iBET151, and whether there is any evidence for cooperation of MOZ and BRD4 in gene regulation. We also assessed whether iBET inhibitors had any effect on the recognition of H3K14ac by the MOZ DPF domain. Our results show preliminary evidence that MOZ forms a complex with BRD4 protein using co-immunoprecipitation experiments. Interestingly we found that expression of MOZ–dependent genes in K562 cells was inhibited by I-BET151 treatment. Moreover, we detected iBET151 dose-dependent inhibition of K562 cell proliferation. In the previous chapters, it was suggested that ablation of MOZ activity could be more explored as potential treatment for leukaemia once it inhibits leukemic cell proliferation and promotes differentiation. As MOZ inhibitors are still not available as

therapeutic agents, here we suggest that I-BET151 could be used to inhibit MOZ activity through inhibiton of MOZ/BRD4 complex.

5.2 I-BET 151 treatment affects expression of MOZ target genes

It was reported by Dawson, Prinjha [383] and others that I-BET151 is a potent and selective inhibitor of specific interaction between the BET bromodomain and acetylated lysine residues on histones. *In vitro* experiments reported that BRD4 binds stongly to histone H3 tails peptides acetylated at H3K9 and H3K14 [174].

Our group has also shown that acetylated H3K14 can also bind strongly to the MOZ PHD domain [190] indicating that MOZ and BRD4 share an ability to recognise acetylated H3K14. Thus, although the MYST DPF domain is structurally distinct from bromodomains, both domains form a pocket that can accommodate acetyl lysines. Therefore, the possibility remined that there might be some molecular similarlity in the modes of interaction. For this reason, we decided to investigate whether I-BET151 might have an inhibitory effect on the interaction of MOZ DPF domain with H3 peptides containing H3K14ac and, consequently, affect the expression of some MOZ target genes.

In order to assess the ideal concentration of I-BET151 to treat the cells maintaining their viability, K562 cells were treated with 1 μM, 5 μM and 10 μM of I-BET151 and with the same volume of the vehicle (DMSO) for 24 hours. Cytotoxicity was evaluated by MTS

assay as described in the Materials and Methods. As shown in Figure 5.1, 1 μM of I-BET151 was not observed to be toxic for the cells, at least in terms of proliferation and metabolic activity. However, at both 5 μM and and 10 μM concentrations of I-BET a decrease in the cell viability of 60% and 80% respectively was observed. K562 cells did not display any sensitivity towards the vehicle DMSO.

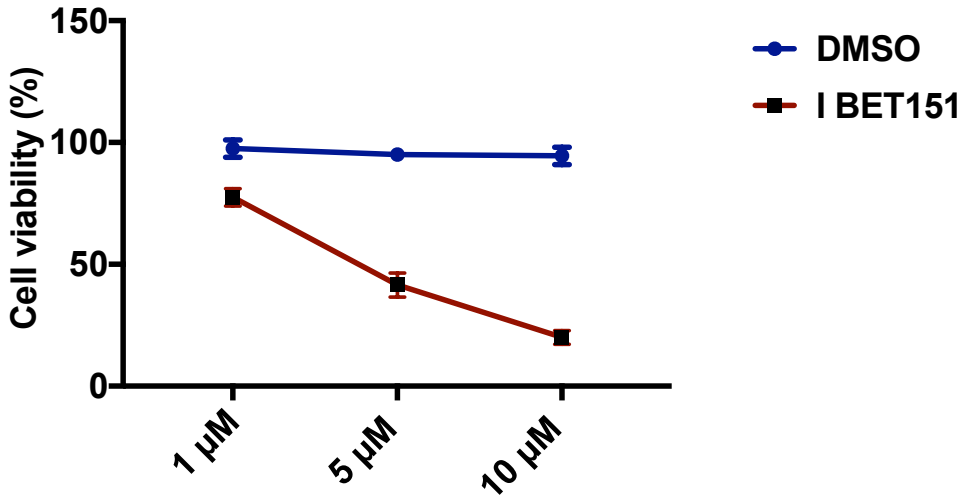


Figure 5.1 - Cytotoxicity of I-BET151 in K562 cells and quantification by MTS assay. K562 cells were treated with 1 μM, 5 μM and 10 μM of I-BET151 and same volume of the vehicle (DMSO) was used as control. Dots and lines are shown as the mean ± S.E.M of cell activity value. Differences between every two means are significant ($p<0.05$).

We next assessed the the effects of I-BET151 treatment on gene expression in K562 cells, focusing on genes that were identified in the transcriptome profiling (Chapter 3) as MOZ-dependent (PROX1, NR4A1, NR4A2, NUDT7, MLL2) or genes that were de-repressed after MOZ ablation (XIST, ID1, ID3, RUNX1 and CDKN2A)

Transcript levels were determined by qRT-PCR after treatment of K562 cells with with 1 μM of I-BET151 for 24h (i.e. a subcytotoxic concentration). All the genes analyzed showed statistically significant alteration in the expression following treatment with I-BET151 in comparision to controls (cells treated with DMSO), with the exceptions of MLL2 and CDKN2A (Figure 5.2).

This result indicates that I-BET151 treatment of K562 cells inhibits the expression of genes whose expression is dependent on MOZ. This could indicate that both MOZ and BRD4 are involved in the expression of these genes, and therefore regulatory complexes of MOZ and BRD4 may exist. Alternatively it remained possible that I-BET151 blocks the interaction of both MOZ and BRD4 to acetylated histones. In addition, I-BET151 was found to induce a moderate upregulation of several genes that appear to be strongly repressed by MOZ, e.g. ID1 and ID3 (Fig 5.2). Inhibitor of differentiation proteins ID1 and ID3 are negative regulators of E protein transcription factors and their disruption in haematopoietic progentitors leads to severe disturbance of ertythropoeisis [385]. Early reports showed that ID1 and ID3 have only low level expression in K562 and other myeloid leukaemia cells. However, these genes are upregulated upon differentiation on human bone marrow cells to B-lymphocytes [386].

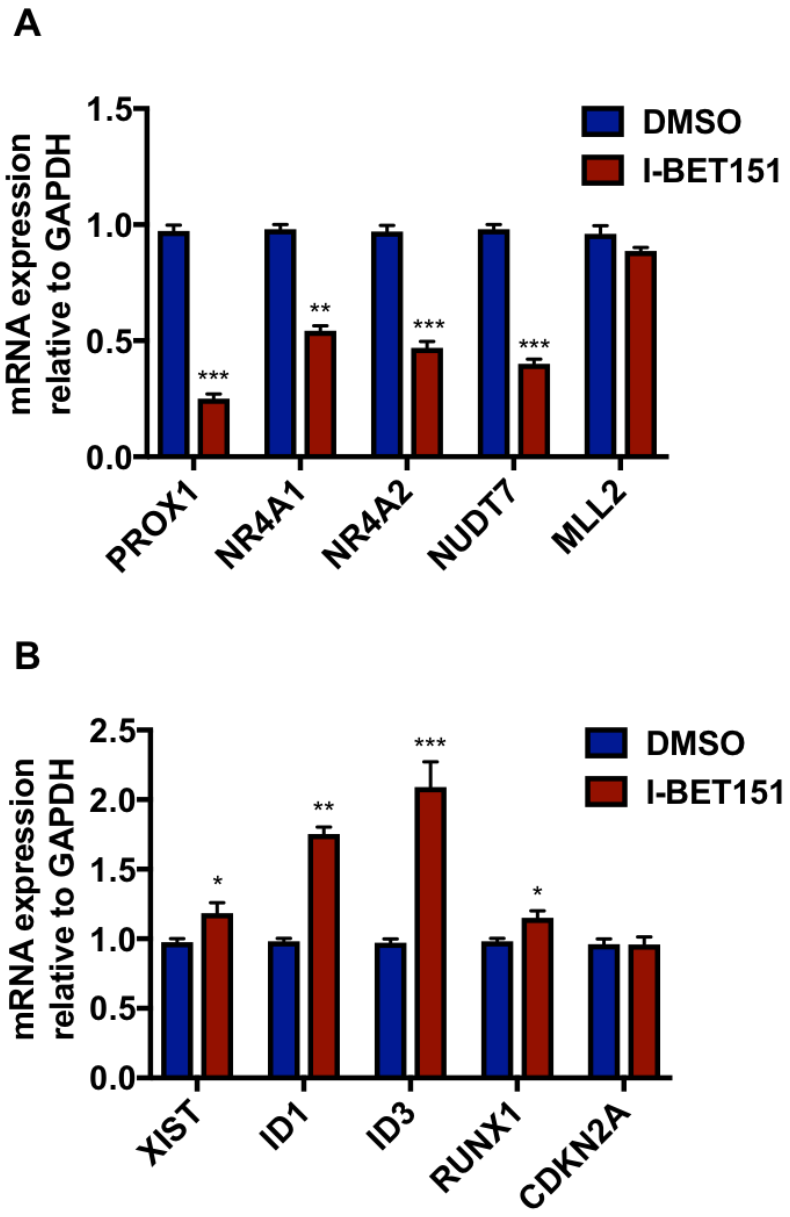


Figure 5.2 - qRT-PCR analysis of MOZ target genes expression after treatment with I-BET151. K562 cells were treated with 1 μM of I-BET151 and same volume of the vehicle (DMSO), used as control. The mRNA expression of some MOZ target genes were assessed by qPCR. Statistical analysis was performed via Student's t-test, with the p values represented as follows: $p < 0.001$ (**), $p < 0.0001$ (***). qRT-PCR data were normalized to GAPDH expression for each sample. Error bars, S.E.M

5.3 IBET-151 does not inhibit the binding of MOZ DPF to acetylated H3K14

In order to determine if I-BET151 altered the expression of MOZ target genes by inhibiting the binding of MOZ DPF domain to H3K14ac, *in vitro* histone binding assays were performed as described in Material and Methods. Briefly, 15 µg of purified GST-fusion constructs of MOZ DPF (194 –323) protein was incubated with increasing concentrations of the I-BET151 (10^{-6}, 10^{-5} and 10^{-4} M) or the same volume of the vehicle DMSO (control) and 1.5 µg of biotin-conjugated histone peptides H3 (1-21) or H3K14ac or no peptide as control, overnight with rotation at 4°C in binding buffer. After capture of complexes on Dynabeads and washing of non-bound materials, samples were subjected to SDS-PAGE and results visualised by immunodetection using α-GST antibody.

As expected, in the absence of peptide no GST protein was retained on the beads. Consistent with previous data from the group [190] an interaction of MOZ DPF with unmodified H3 peptide (1-22) was observed (lane 3), which was stimulated by acetylated H3K14 (compare lanes 3 and 7). While DMSO had no effect on the binding (data not shown) treatment of the binding mixes with I-BET151 had an unexpected effect on the formation of complexes with the unmodified H3 peptide. In contrast, I-BET151 did not inhibit interaction of MOZ DPF with H3K14ac peptide at 1-10 µM, although reduced binding was observed at 100 µM IBET151. Given that the

control was also affected, we cannot conclude that I-BET151 has a direct effect on the ability of MOZ DPF to associate with H3K14ac in vitro, at least in these assays. Therefore, the explanation for the effects of I-BET151 on MOZ target genes may be more likely to involve coregulation by BRD4 and MOZ, potentially in complex.

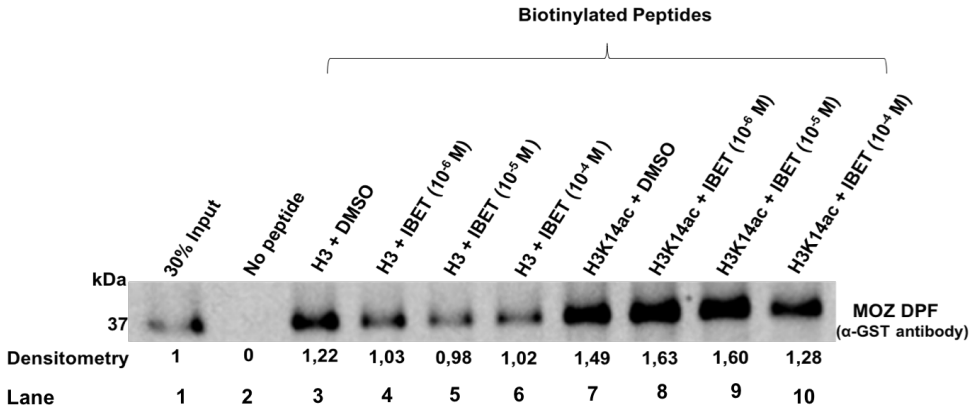


Figure 5.3 - *In vitro* binding assay of GST-MOZ DPF and H3K14ac. Western blot analysis of the binding assay from recombinant GST-MOZ DPF fusion protein with biotinylated histone peptides of H3 (1 – 21) and H3K14ac incubated with I-BET151 inhibitor or DMSO (control) equal to the volume of the highest concentration of inhibitor. Densitometry readings were calculated relative to H3 using Image J.

5.4 Detection of MOZ / BRD4 complexes in K562 cells

As I-BET151 displays potent binding affinity to BRD2, BRD3 and BRD4 [383] and BRD4 interacts with chromatin-associated acetylated histones to regulate target gene transcription [387], we decided to investigate if the effect of I-BET151 on MOZ target genes,

as showed above, is due to an existence of cellular complexes of MOZ and BRD4.

Thus, we initially confirmed expression of BRD4 in K562 cells via western blot (Figure 5.4A). This showed we could readily detect a band at approximately 200 KDa corresponding to the expected size for full length BRD4 isoform A (Uniprot).

In order to assess MOZ-BRD4 interaction, total endogenous protein was extracted from K562 cells and co-immunoprecipitation was performed using BRD4 antibody to pulldown the protein and MOZ antibody for protein detection using western blot. Input lane verify the expression of 10% of the amount of MOZ protein used for Co-IP assay in the cell lysates. The result showed in Figure 5.4B revealed that IP of BRD4 allows co-precipitation of MOZ proteins, confirming the existence of endogenous complexes containing BRD4 and MOZ. In contrast no MOZ was found in the IP controls.

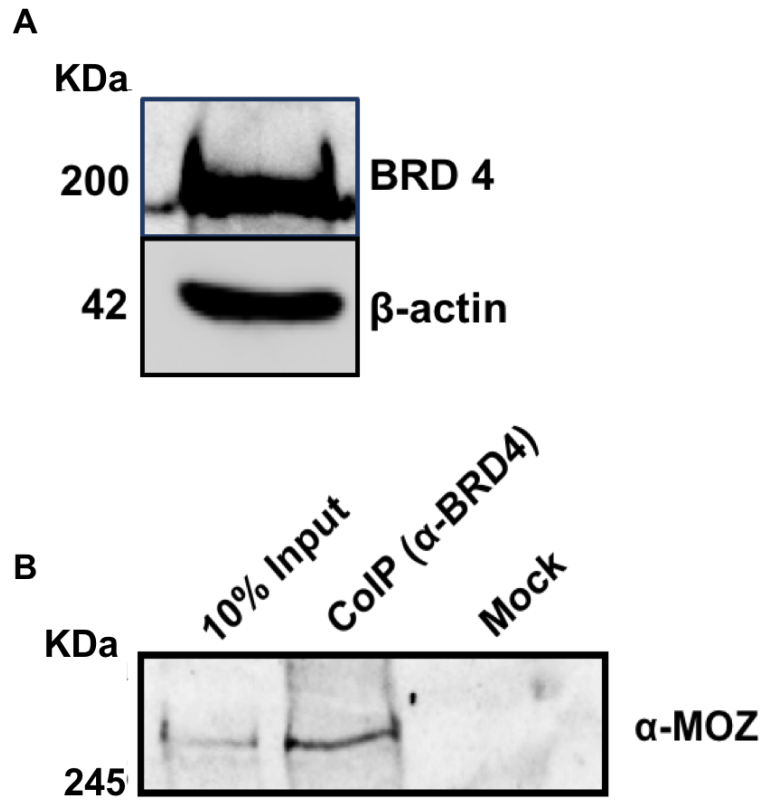


Figure 5.4 - Expression of co-precipitation of MOZ/BRD4 complexes in the K562 cells (A) Western blot displaying expression of BRD4 in extracts from K562 cells. (B) Co-immunoprecipitation of endogenous MOZ and BRD4 proteins from K562 extract. Input (10%) shows expression levels of MOZ using α-MOZ (29D2), Immunoprecipitation with α-BRD4 (as indicated) or IgG (mock) was performed. β-actin was used as loading control.

To try to determine which region of MOZ is responsible for potential interactions with BRD4, we used a series of epitope-tagged MOZ constructs available in our laboratory. Due to ease of transfection , we utilized HEK293 cells for these experiments. Thus, we performed transient transfections of HEK293 cells to express various YFP-FLAG-tagged constructs including MOZ full length, MOZTIF2 (retains the NEMM, PHD zinc finger and the HAT (MYST) domains of MOZ [255]), MOZ 1-810 and MOZ 1-779 (also contain NEMM, PHD and HAT domains), MOZ 570-810 (MOZ MYST domain

only) as shown in Figure 5.5A. Cells transfected with empty pcDNA3.0 FLAG vector were used as control. The expression of full length BRD4 (at approx. 200 kDa) in HEK293 extracts was first confirmed by western blot (Figure 5.5B). Cells were harvested and lysed to release the soluble proteins 48 hours posttransfection. Co-immunoprecipitations were performed using anti-FLAG antibodies to capture MOZ proteins and any associated BRD4 protein was detected using anti –BRD4 antibodyt.

While the quality of the western blot result was not optimal, nonetheless, the preliminary results are reported here. No coIP of BRD4 in the control empty vector although better controls for future attempts might include unrelated FLAG-tagged proteins rather than FLAG vector. In any case, we did observe signals corresponding to co-precipitated BRD4 proteins in lanes containing full-length MOZ and MOZ-TIF2 which would suggest that the N-terminal portion of MOZ 1-1118 retained in MOZ-TIF2 is sufficient for this interaction. Very weak bands were observed for FLAG-MOZ 1-810 and 1-779, although these would need to be reproduced. Unfortunately there was insuffcieint time to complete replications of these data to try to improve the signals, but that will be the focus of future experimetns in the group. On balance, taken together with the effects of I-BET151 on MOZ target genes, we believe these coIP data provide supporting evidence for the existence of interactions of MOZ and BRD4 in cells, either via direct or indirect association in chromatin complexes. In

future work, direct interaction of MOZ and BRD4 might also be assessed and mapped using yeast-2-hybrid and similar methods.

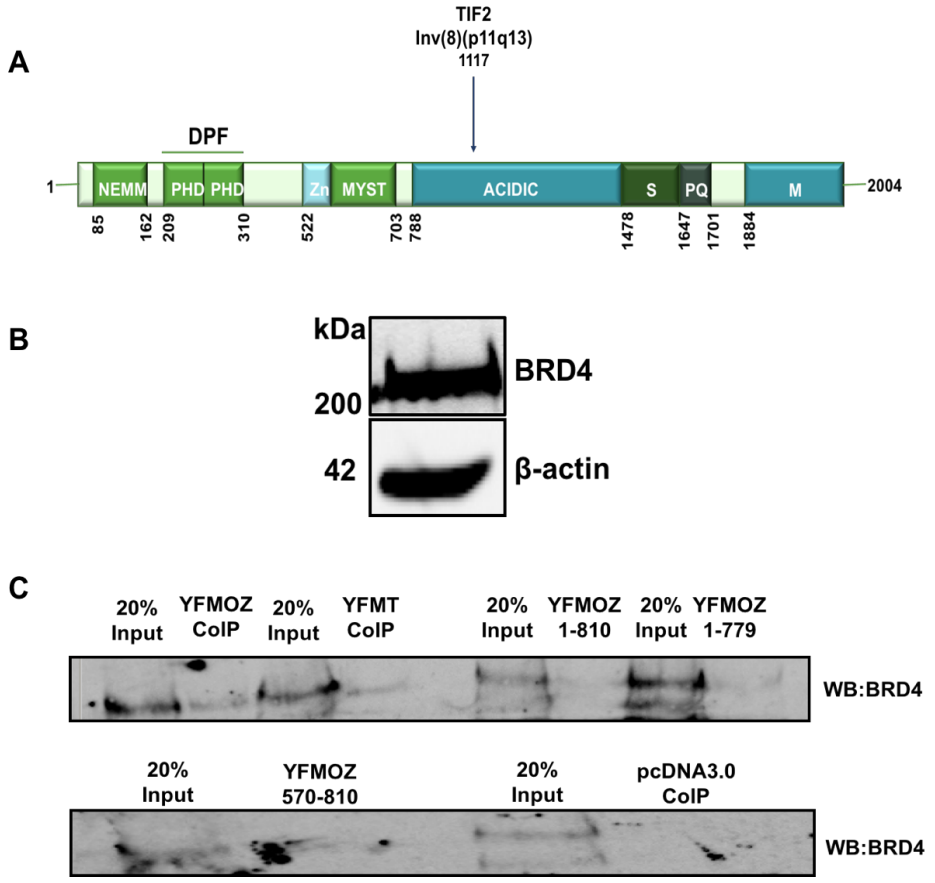


Figure 5.5 - Attempted co-immunoprecipitation of YFP-FLAG-MOZ constructs with BRD4 proteins in the HEK293 cells. (A) Modular structure of MOZ. The number at the domain ends represent the amino-acid position and MOZ-TIF2 breakpoint is indicated with arrow. (B) Western blot displaying expression of BRD4 in extracts from HEK293 cells. (C) Co-immunoprecipitation of full length MOZ and N-terminal MOZ constructions and endogenous BRD4 proteins from HEK293 extract. Input (20%) shows expression levels of BRD4 using α-BRD4, immunoprecipitation was performed with α-FLAG.

5.5 I-BET151 inhibits K562 cells proliferation and causes cell cycle arrest

Aberrant proliferation, inhibition of apoptosis and differentiation arrest of hematopoietic progenitor cells are the main

characteristics that leads to acute myeloid leukemia (AML) [388]. Once it was shown that I-BET151 modulates MOZ target genes we wanted to verify if I-BET151 also inhibits K562 cell proliferation as observed in MOZ knockout cells (Chapter 4).

To this end, $5.0x10^3$ K562 cells were treated with 1 μM of I-BET151 for 24, 48 and 72h and the proliferation levels were evaluated by MTS assay and compared to K562 cells treated with the same volume of the vehicle (DMSO). The result shows that the concentration of I-BET151 used for the experiment reduced the cell proliferation of around 20% in all the time-points analysed compared to control (Figure 5.6). The decrease noticed in the K562 cells proliferation treated with I-BET151 was lower compared to the MOZ knockout cells that presented more than 50% reduction compared to control (Chapter 4).

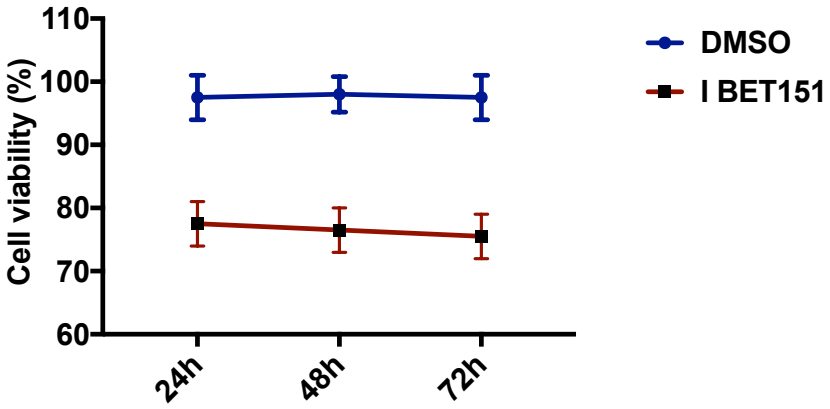


Figure 5.6 - MTS proliferation assay performed upon K562 cells treated with I-BET151. K562 cells were treated with 1 μM of I-BET151 for 24, 48 and 72h and same volume of the vehicle (DMSO) was used as control. Dots and lines are shown as the mean ± S.E.M of cell activity value. Differences between every two means are significant ($p<0.05$).

Flow cytometry cell cycle analysis using propidium iodide DNA staining was also performed on K562 cells treated with 1 μM of I-BET151 for 24 h and compared to the K562 cells treated with the same volume of DMSO (control). The result indicates that around 50% of the cells treated with I-BET151 are in the G0-G1 cell cycle phase, around 20% more compared to the cells treated with DMSO. Considering the S phase, treatment with I-BET151 caused reduction of around 20% in this phase compared to control. No difference was noticed in the G2-M phase between cells treated with I-BET151 and DMSO (Figure 5.6). Interestingly, as shown in Chapter 4, MOZ knockout K562 cells presented considerable difference in cell proliferation and cell cycle arrest possible due to the derepression of the p16INK4a, encoded by CDKN2A gene, however I-BET151 treatment does not alter this gene expression.

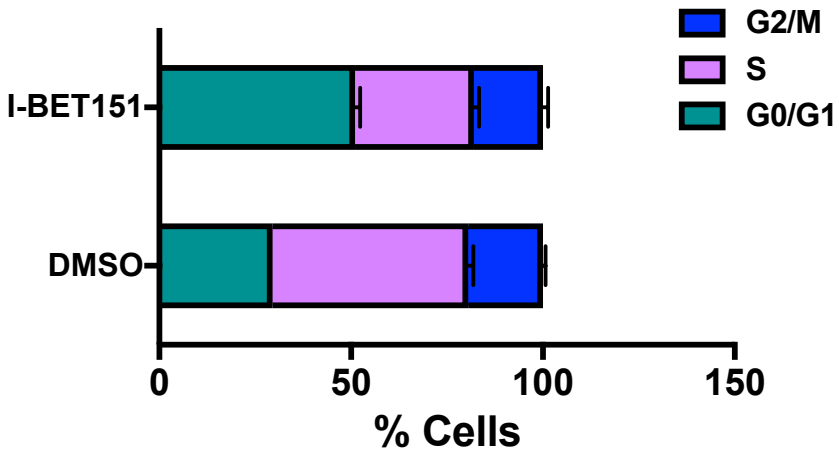


Figure 5.7 - Cell cycle analysis of K562 cells treated with I-BET151. Propidium iodide (PI) staining and flow cytometric analysis was performed in K562 cells treated with 1 μM of I-BET151 and with the vehicle DMSO. Percentage of cells are shown in the cell cycle phases: G0/G1, S and G2/M.

5.6 K562 cells undego apoptosis upon treatment with I-BET151

In order to verify if the diminished cell proliferation is due to apoptosis response to the I-BET151 treatment, apoptosis analysis using annexin-V / PI assay by flow cytometry was performed to identify viable, apoptotic and dead cells. As explained in Chapter 4, live cells are negative for annexin-V and PI, early apoptotic cells are positive for annexin-V, cells in late stages of apoptosis are positive for annexin-V and PI and necrotic cells are positive only for PI. Three controls were used: K562 cells unstained, treated with annexin-V FITC only and treated with PI only. Basal levels of apoptosis and cells autofluorescence (negative control) were taken into consideration for the analysis.

K562 cells were treated with 1 μM of I-BET151 for 24 h and compared to the K562 cells treated with the same volume of DMSO. The result shown in Figure 5.8 indicates that 98% of the cells treated with 1 μM of I-BET151 were live cells, only 2% were in late apoptosis. The control cells treated with DMSO presented 99% of live cells. This result is very similar to observed in the MOZ knockout cells (result showed in Chapter 4).

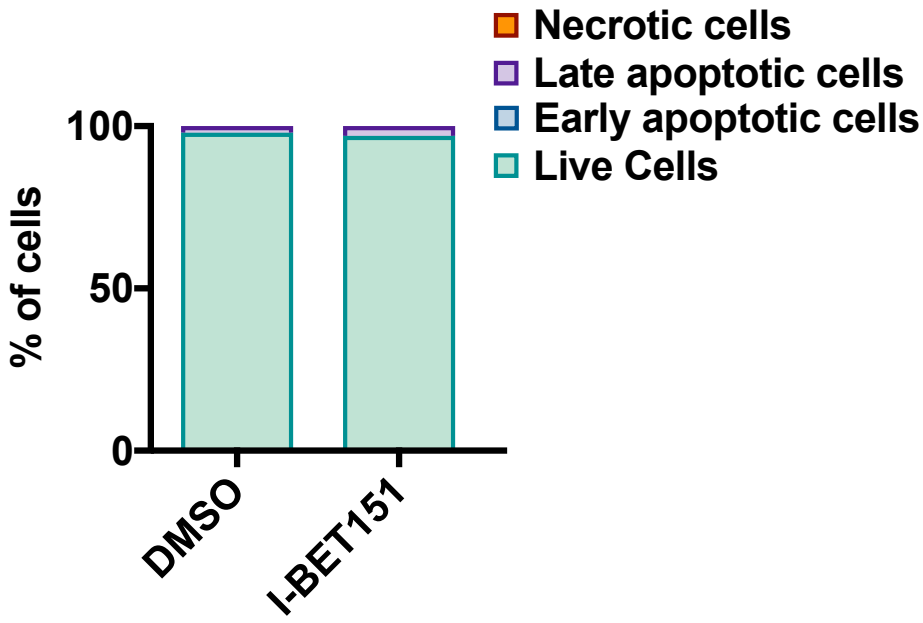


Figure 5.8 - Apoptosis assay using annexin-V / PI assessed for the K562 cells treated with I-BET151. Analysis of apoptosis in K562 cells treated with 1 μM of I-BET151 for 24 h and the same volume of the vehicle DMSO was performed by annexin-V / PI assay using flow cytometry. Live cells are negative for annexin-V and PI, early apoptotic cells are positive for annexin-V, cells in late stages of apoptosis are positive for annexin-V and PI and necrotic cells are positive only for PI.

In order to confirm that I-BET151 treatment did not lead K562 cells to apoptosis, cells treated with 1 μM of I-BET151 for 24 h and the same volume of the vehicle DMSO had the caspase 3/7 activity assessed by flow cytometry using CellEventTM caspase-3/7 green detection reagent combined with SYTOXTM AADvanced red. Viable cells were negative for CellEventTM Caspase 3/7 and SYTOXTM red, apoptotic cells were positive for CellEventTM Caspase 3/7 and necrotic cells were positive for SYTOXTM red and for CellEventTM Caspase 3/7. Unstained K562 cells and controls for CellEvent Caspase 3/7 and Sytox were used.

As shown in Figure 5.9, 100% of the control (K562 cells treated with DMSO) was viable and the K562 cells treated with I-BET also presented 100% of viable cells, indicating that K562 cells treated with 1 μM of I-BET151 do not undergo apoptosis.

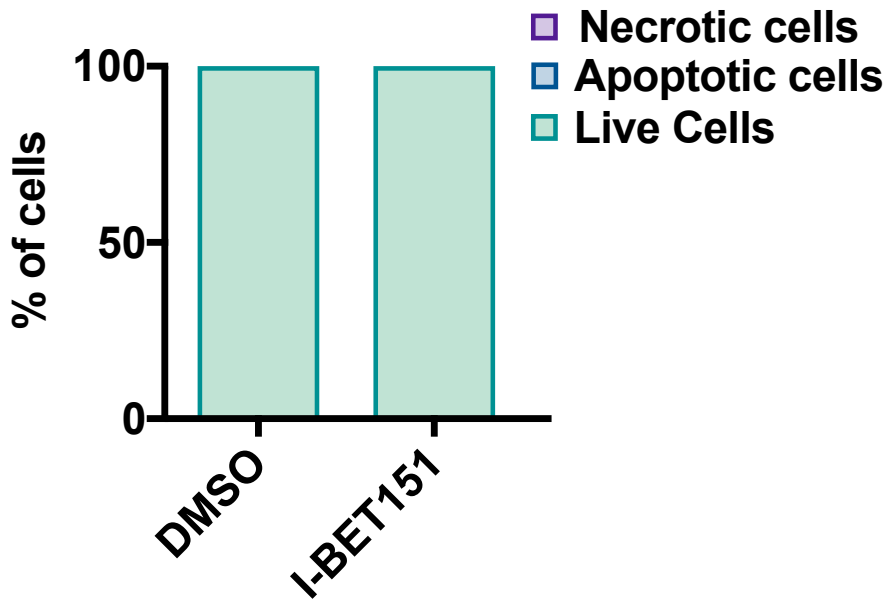


Figure 5.9 - Caspase activity assessed for K562 cells treated with I-BET151. Caspase 3/7 activity was assessed in K562 cells treated with 1 μM of I-BET151 for 24 h and the same volume of the vehicle DMSO (control) using CellEventTM caspase-3/7 green detection reagent combined with SYTOXTM AADvanced red, with this combination is possible discriminate activated caspase 3/7 from necrotic cells using flow cytometry. Viable cells are negative for CellEventTM Caspase 3/7 and SYTOXTM red, apoptotic cells are positive for CellEventTM Caspase 3/7 and necrotic cells are positive for SYTOXTM red and for CellEventTM Caspase 3/7

5.7 Summary

It is well known that MOZ interacts with other proteins and functions as a catalytic subunit in many nuclear multisubunit complexes. MOZ was first isolated as part of the AML1-CBFβ transcription factor complex, by which MOZ stimulates AML1-mediated transcripition [195]. MOZ was shown to be the catalytic subunit of a tetrameric ING5 (inhibitor of growth 5) complex, which specifically acetylates nucleosomal histone H3K14 [192]. Other subunits of the complexes are EAF6 (homolog of Esa1-associatedfactor 6) and BRPF1/2/3 (bromodomain-PHD finger protein 1, 2 or 3), which interacts with MOZ through the MYST domain, and bridging the interaction between MOZ and ING5 proteins and the fourth partner Esa1-Associated Factor 6 (hEAF6) [196, 197]. MOZ and ING5 complex play an important role in the *in vivo* regulation of T-box DNA binding Transcription Factor 1 gene (*Tbx1*), responsible for heart and aortic arch development, in the 22q11 deletion [217].

Several groups have demonstrated that MOZ interacts with many transcription factors, particularly with haematopoietic specificity, such as acute myeloid leukaemia/Runt-related transcription factor 1 (AML1/RUNX1) [195, 201, 202] and PU.1 [203]. In addition, MOZ appears to enhance the transcriptional activity of p53, MAFK, NFκb and others [206]. The interaction of MOZ with the p65 subunit of NF-kappaB enhances expression of NF-kappaB-

dependent promoters and may play an important role in the pathogenesis of certain acute myeloid leukemia [207]. Regarding the MOZ-p53 interaction, MOZ is an acetyltransferase of p53 at K120 and K382 and colocalizes with p53 in promyelocytic leukemia (PML) nuclear bodies under cellular stress. MOZ-PML-p53 interaction enhances MOZ-mediated acetylation of p53, and this ternary complex enhances p53-dependent p21 expression and cell arrest [205, 208].

In this Chapter, we are providing preliminary evidence for a hitherto unknown evidence for the existence of a link between MOZ and BRD4 proteins. Bromodomain protein 4 (BRD4) is a chromatin-binding protein implicated in transcriptional regulation in multiple cancers [389-391]. It also plays a role activating transcription by phosphorylating RNA polymerase II (Pol II) and by functioning as a scaffold for other transcriptional regulators [387, 392]. Furthermore, BRD4 was reported to have histone acetyltransferase (HAT) activity and to acetylate H3K122, which is critical for nucleosome stability, that helps to initiate gene activation [393].

We first examined the effects of the bromodomain inhibitor I-BET151 on the expression of MOZ-regulated genes. An optimal subcytoxic concentration of I-BET151 [1 μM] was determined by MTS assay.

Under these conditions the expression of some MOZ target genes that are up and down-regulated under MOZ ablation (PROX1, NR4A1, NR4A2, NUDT7, MLL2, XIST, ID1, ID3, RUNX1 and

CDKN2A) were assessed in K562 cells by qRT-PCR. Except for MLL2 and CDKN2A, these genes showed altered expression after I-BET151 treatment, which paralleled the effects of MOZ ablation on their expression. This suggested to us a possible coregulation by MOZ and BRD4 proteins of a cohort of genes in K562 cells.

To explain the effect of I-BET151 treatment observed on MOZ target gene expression, we hypothesized that MOZ could form a complex with BRD4 and the deregulation observed on the gene expression could be due to an inhibition on this complex. In order to test this hypothesis, the expression of BRD4 was first confirmed in K562 cells via western blot. Following the confirmation of BRD4 protein expression, total endogenous protein was extracted from K562 cells to perform co-immunoprecipitation. The result confirmed the existence of endogenous complexes of BRD4 and MOZ.

Because the interaction analysis was performed by co-immunoprecipitation, it is not possible to conclude whether MOZ and BRD4 protein has a direct interaction. In fact it is possible that they interact indirectly through transcription factors e.g. it was shown that both MOZ [205] and BRD4 [394] bind directly to p53. Alternatively the association of both proteins with histones and DNA may facilitate their corecruitment to gene regulatory regions such as promoters and enhancers.

In addition to coIP of native MOZ-BRD4 protein complex in K562 cells, potential interactions epitope tagged MOZ proteins with BRD4 was also observed in transiently transfected HEK293 cells.

Indeed interactions were observed with both YFP-FLAG-MOZ and YFP-FLAG-MOZ-TIF2, suggesting the N-terminal portion of MOZ is sufficient for BRD4 association, whatever the mechanism. Attempts at coIP with shorter MOZ constructs were not convincingly conclusive and need to be repeated.

In the Chapter 4, it was showed that MOZ ablation causes inhibition on K562 cells proliferation, which is desirable once aberrant proliferation is one of the characteristics that leads to acute myeloid leukemia. For this reason, we wanted to investigate if I-BET151 also can reduce proliferation. K562 cell proliferation was evaluated by MTS assay performed after K562 cells treatment using 1 μM of I-BET151 for 24, 48 and 72h. The result showed that I-BET151 reduced the cell proliferation of around 20% compared to control. Although the result was positive, it was lower than noticed in K562 MOZ knockout cells that had more than 50% reduction compared to the control (Chapter 4).

Cell cycle was also performed to investigate if I-BET151 arrests K562 cells to G0-G1 phase as observed in K562 MOZ knockout cells. To this end, flow cytometry cell cycle analysis using propidium iodide DNA staining was performed on K562 cells treated with 1 μM of I-BET151 and compared to the K562 cells treated with the same volume of DMSO (control). Again, the result was positive, indicating that around 50% of the cells treated with I-BET151 were arrested in the inactive G0-G1 cell cycle phase, around 20% more compared to the cells treated with DMSO. Considering the S phase,

treatment with I-BET151 caused reduction of around 20% cells in this phase compared to control. In regard to the G2-M phase, no difference was noticed between cells treated with I-BET151 and DMSO.

Interestingly, although I-BET151 treatment did not affect CDKN2A expression in K562 cells, it had an effect on cell proliferation and cell cycle. Carretta, Brouwers-Vos [395] using 1 μM of I-BET151 (the same concentration that we used to treat K562 cells) to treat CB $CD34^+$ transduced MLL-AF9 cells co-cultured with MS5 cells found that genes that negatively control cell cycle such as CDKN1A and CDKN1B were up-regulated. Some other genes enriched for GO terms such as R-SMAD binding, G2/M transition of mitotic cell cycle and negative regulation of cell proliferation were also up-regulated. This may explain why we observed a decrease in proliferation and cell cycle in K562 cells treated with I-BET151 even when the expression of CDKN2A appeared affected at the level of transcription.

Apoptosis analysis was also performed in the K562 cells treated with I-BET151 by annexin-V/PI and caspase 3/7 activity methods using flow cytometry. The results did not indicate a role for KAT6A/MOZ in the apoptosis process in these cells. This result is in accordance with Carretta, Brouwers-Vos [395] that also treated K562 cells with 1 μM of I-BET151 for 72 h and did not detected Annexin V positive cells.

In normal progenitor cells, MOZ appears to have a key role in maintaining the non-differentiated and stem-like properties. In acute myeloid leukaemias, MOZ has a direct role in paediatric AMLs containing recurrent translocations that express MOZ fusion proteins. In other leukaemias and also in c-MYC driven B cell lymphomas, wild type MOZ expression/overexpression may help to establish a leukemic stem cell niche [215, 249, 396]. The data presented here shows that ablation of MOZ has a detrimental effect on the proliferation of leukemic cells in culture, thus we speculate that blocking MOZ activity may be beneficial for leukaemia treatment because it decreases the proliferation potential and 'stemness' of leukemic cells. Inhibition of the acetyltransferase activity of MOZ has only partial effect in animal models of MOZ-TIF2 leukemias [255] whereas interference with MOZ domains involved in assembly of multiprotein complexes may be a more fruitful approach [345]. Several approaches are now beginning to identify inhibitors of MOZ via small molecule screens e.g. https://www.monash.edu/__data/assets/pdf_file/0008/543149/Potent-inhibitors-of-MOZ.pdf

Our preliminary results suggest that I-BET151 could also be used to inhibits MOZ-regulated gene expression pathways through the inhibition of MOZ-BRD4 complex which also affects the proliferation of erythroleukemic cells. In the longer term, the development of precision drugs that enable dual targeting of both MOZ and BRD4 may be beneficial for patients and help prevent the

emergence of resistance commonly observed with single target therapies.

Chapter 6 : Concluding remarks

MOZ is a histone acetyltransferase considered a key regulator of haematopoietic stem cell development. Chromosomal abnormalities affecting this gene are associated with acute myeloid leukaemia and other cancers. Several groups have demonstrated that MOZ acts as a co-activator for selected DNA binding transcription factors, particularly with haematopoietic specificity, such RUNX1 [195, 201, 202]. Knockout of the MOZ gene in mice is embryonic lethal and is associated with reduced expression of genes such as HOXA9 and defective haematopoiesis [213]. HOX genes are highly expressed in acute myeloid leukaemias and are essential for development and differentiation of many tissues [221, 272]. Although transcription profiling has been performed to identify MOZ-dependent genes in selected tissues [274] [224] [164], MOZ target genes in leukemic cells were relatively unknown at the outset of this study [190].

In this study, we reported for the first time the identification of MOZ target genes in K562, a human erythroleukemia cell line. MOZ was knockdown using siRNA and transcription profiling was performed using RNA sequencing. We identified 406 genes regulated by MOZ, most of them are implicated in haematopoietic development and cell proliferation. Among of the overall genes identified, 76 were down-regulated. Comparison amongst MOZ target genes found in our study and other transcription profiling studies performed to identify MOZ-dependent genes in selected normal mouse tissues such as B lymphocytes [274], embryos and ES cells

[224] and in human SUM-52 cells, a luminal breast cancer cell line [164] showed limited overlap between the genes indicating that MOZ may perform different functions in distinct tissues.

Gene Ontology analysis performed with MOZ targeting genes found in our study indicated an enrichment of genes associated with cancer, more specifically ovarian cancer, embryoma, prostate cancer, liver cancer, breast cancer and leukaemia, also cell growth and development. KEGG pathway analysis performed with the up-regulated genes identified Rap1 signaling pathway, a cellular process involved in normal cell growth, differentiation, cell invasion and metastasis as the most enriched molecular function. Around of 13 transcription factors were identified among the down-regulated genes, some of them plays a role in hematopoietic and neural development.

The role MOZ plays in acetylation of H3K14 is controversial once it was shown that in embryos deficient in MOZ, histone acetylation defects are relatively specific to H3K9 at HOX loci, as HOX H3K14 acetylation were not strongly affected and, in $MOZ^{-/-}$ mouse embryonic fibroblasts cells, only H3K9 presented reduction in acetylation levels [335]. However, previous results published by our group showed that MOZ DPF cooperates with the MYST domain to promote H3K9 and H3K14 acetylation, not only H3K9 [190]. Our result showed that MOZ knockdown leads to a reduction of global H3K9 and H3K14 acetylation, in concordance with results from our group.

We also generated a permanent MOZ knockout K562 cells using CRISPR CAS9 nickase technology, as shown in chapter 4. The expression of some MOZ target genes identified by siRNA/ RNA–Seq was assessed in these cells confirming the results mentioned above. Rescue experiments in which YFP-tagged MOZ (which functions similarly to untagged MOZ [252]) was expressed in the MOZ knockout cells resulting in a rescue of expression of MOZ-dependent genes to a level similar to observed in wild type K562 cells (control). YFP-tagged MOZTIF2 was also expressed in MOZ knockout cells to assess its effect on MOZ target genes. MOZ-TIF2 conserves the histone acetyltransferase MYST domain [231]. Our results show that MOZ-TIF2 can also rescue expression of MOZ-dependent genes suggesting that the C-terminal domain of MOZ is not essential to regulate the expression of the genes analyzed in the K562 cell line.

Levels of acetylated H3K14 and H3K9 were also decreased in these cells, in conformation with the results found in knockdown cells using MOZ siRNA. Moreover, we found an increase in the levels of histone lysine crotonylation. This histone modification was recently identified as a non-acetyl acylation that is widespread in mammalian genomes and marks active chromatin and is enriched at promoter and enhancer region [137]. MOZ DPF domain was shown to be a specific reader of H3K14cr Xiong, Panchenko [191].

MOZ has being reported to play an important role in cell proliferation. Mice lacking MOZ HAT activity failure to maintain a

normal number of hematopoietic precursors and exhibit defect in B-cell development [214] [226]. Partial inhibition of MOZ is capable to reduce proliferation of MEIS1 and HOX-driven lymphoma and leukemia cells [215]. Indeed, we also showed that MOZ knockout decrease proliferation of K562 cells increasing the number of cells arrested in the G0-G1 phase of cell cycle with severe decrease of cells in the S-phase. However, apoptosis was not observed.

Finally, we also observed that I-BET151 treatment mimics the effect of MOZ knockout in K562 cells in terms of gene expression profile and cell proliferation. IBET-151 have been tested in preclinical cancer models for MLL-fusion leukaemia [383]. This compound inhibits the interaction of bromodomain proteins, including BRD4, with acetylated lysine [79, 397]. We found preliminary evidence that MOZ forms a complex with BRD4 protein using co-immunoprecipitation experiments. We suggested that I-BET151 could be used to inhibit MOZ-regulated gene expression pathways through the inhibition of MOZ-BRD4 complex which also affects the proliferation of erythroleukemic cells.

Together our findings provided better understanding of MOZ function, especially in leukaemia.

Appendix

Table A.1. Oligonucleotides Sequences using for qRT-PCR

Oligo Name	Forward	Reverse
MOZ	5'-TGTCAGTTTGGGGCATCTC-3'	5'-CAGCACTTCAACAAGTATCTGGA-3'
RPPH1	5'-CCGGAGCTTGGAACAGACT-3'	5'-GTAGTCTGAATTGGGTTATGAGGTC-3'
SIRT4	5'-GGGGTGTGAAGTGTCCGTAG-3'	5'-AATGGAGGCTTTCGAGCAC-3'
SERPINE 1	5'-GCTTCCCAGAAATTTGGGTTA-3'	5'-TTAAAAGTCAAGGAAGATGAAGAGGT-3'
HHEX	5'-CGGACGGTGAACGACTACA-3'	5'-AGAAGGGGCTCCAGAGTAGAG-3'
RAB27A	5'-AGTGATCTGGAGGACCAGAGAG-3'	5'-TCCATATTTCTCTGCGAGTGC-3'
NR4A1	5'-CCACTGCCTCCTTCAACC-3'	5'-GGGCATCTATCTAGCCTCACAG3'
NR4A2	5'-ATGAAGAGAGACGCGGAGAA-3'	5'-AAAAGCAATGGGGAGTCCA-3'
PROX1	5'-GAGCCTCCGTGGAACTCA-3'	5'-TGGGCACAGCTCAAGAATC-3'
PLAT	5'-AGCTGTGGGGAGCTCAGA-3'	5'-TGCTTCACAGCGTCCCTTA-3'
NUDT7	5'-CGTCCGGTCAGAGAAGCTA-3'	5'-GGGTCACGCTTACCTCCAG-3'
ID1	5'-CCAGAACCGCAAGGTGAG-3'	5'- GGTCCCTGATGTAGTCGATGA-3'
ID3	5'-CATCTCCAACGACAAAAGGAG-3'	5'- CTTCCGGCAGGAGAGGTT-3'
MCAM	5'-GGGTACCCATTCCTCAAGT-3'	5'-CAGTCTGGGACGACTGAATG-3'
RUNX1	5'-CTCCCTGAACCACTCCACTG-3'	5'-TGGGGATGGTTGGATCTG-3'
LRIG3	5'-ACAGCTGCACAGCTCAGAAC-3'	5'-CGCAAAAATGATGGTGTTTCT-3'
MAP3K6	5'-GGCTCTGGATGCCTTCTACA-3'	5'-ACAGGGAGGGCTGTACCA-3'
MYO1	5'-AGACTGTGCGCAACAACAA-3'	5'-CGGCTGAACCTGGATCTCAA-3'
TSAPAN6	5'-TTGGAGTTGCTTGCTTCCA-3'	5'-TTGGGTTACACTATCTCATACTGGTT-3'
UCHL1	5'-CCTGAAGACAGAGCAAAATGC-3'	5'-TGTCATCTACCCGACATTGG-3'
CDKN2A	5'-GTGGACCTGGCTGAGGAG-3'	5'-CTTTCAATCGGGGATGTCTG-3'
MAGED1	5'-CCGTCCTGAGCTGCTACTG-3'	5'-AAGCAAGGCGCTGTCTTCTA-3'
HDAC6	5'-AGTTCACCTTCGACCAGGAC-3'	5'-GCCAGAACCTACCCTCAG-3'
HDAC7	5'-CTCCAGCAGCAGCACCCTCAG-3'	5'-CCAGAGGAAGCAGCACAGT-3'
HDAC10	5'-TGGGAAGCTCCTGTACCTCTT-3'	5'-GGCTGGAGTGGCTGCTATAC-3'
XIST	5'-GGCTCCAAATTAATCATATCTTTCC-3'	5'-AACAACAAGCCTATTCTTCTTCTGAGG-3'
GAPDH	5'-AGGTGAAGGTCGGAGTCAAC-3'	5'-GATGACAAGCTTCCCGTTCT-3'

Table A.2. MOZ target genes in K562 cells

gene_id	log2(fold_change)	p_value
ADAMTS14	0.773013	5.00E-05
APP	4.28711	5.00E-05
GABRE	1.44527	5.00E-05
H1F0	inf	5.00E-05
KAT6A	-1.21831	5.00E-05
MEG3	1.13968	5.00E-05
PLS3	4.61502	5.00E-05
RPL7	-3.31079	5.00E-05
SPN	0.773458	5.00E-05
VGF	-0.951394	5.00E-05
XIST	5.42724	5.00E-05
DES	0.877425	0.0001
ATHL1	0.918009	0.00015
AMOT	4.85822	0.0003
PI4KAP1	0.913285	0.0004
TUBB4A	0.685156	0.0004
RMRP	-1.08188	0.0005
ZNF692	0.655229	0.0007
FHL1	3.58309	0.00075
DOK3	0.834335	0.0008
TUBGCP6	0.735277	0.0008
UNC5B	0.859284	0.0008
HIST1H4C	-1.15285	0.0009
ABCA2	0.82109	0.001
FDXR	0.670608	0.001
HSPA1A	-0.560214	0.00105
GOLGA8B	0.914044	0.0011
MARCKS	2.3237	0.0011
HDAC7	0.659922	0.0012
HERC2P2	0.750378	0.0012
ACAP3	0.648709	0.00135
LINC01000	0.673608	0.0014
MAN2A2	0.586673	0.0015
MAPK8IP3	0.659824	0.0015
FAM193B	0.638127	0.00165
KLHL17	0.778247	0.0017
MYO1F	0.663414	0.0019
SERPINB6	0.627873	0.0019
CAPRIN2	0.965209	0.00235
NLRP1	0.570934	0.0026
NR4A1	-0.651286	0.0027
SH3PXD2A	1.02557	0.00295
DHTKD1	0.58471	0.003
PFN2	1.05026	0.0033

AIG1	2.35053	0.00375
ID1	0.613459	0.00375
LOC728392	0.867815	0.00375
RSRP1	0.617582	0.00375
PLCB2	1.14709	0.00385
UBE2C	2.60927	0.00385
CLASRP	0.534918	0.0041
NRBP2	0.655982	0.00415
SPPL2B	0.570565	0.00415
LTBP4	0.517774	0.00425
CHST3	0.534267	0.00435
DUSP5	-0.712745	0.00445
HELZ2	0.796963	0.0045
RBL2	0.592401	0.0045
LINC01311	0.814107	0.00455
LPIN3	1.04738	0.00465
TNNI3	0.579958	0.00465
TMEM173	0.693336	0.0048
APOC1	0.521757	0.005
ROBO1	3.86904	0.00545
NR2C1	0.710519	0.00555
KIFC2	0.630439	0.0056
MCAM	0.515828	0.00575
SCN2A	-0.533534	0.0058
ZBED9	0.821545	0.0058
MAGED1	4.48956	0.00585
OPLAH	0.579106	0.00585
LRIG3	2.61937	0.00595
LENG8	0.496914	0.0061
LOC643201	0.698754	0.00615
FADS1	0.499343	0.00625
EFCAB4A	0.838946	0.00635
FAM65A	0.797186	0.0064
FCGR2A	0.627161	0.00655
IL18RAP	0.84531	0.00665
PLA2G6	0.773614	0.0067
PPAP2B	0.693397	0.0067
MICALL2	0.543934	0.00675
RAB34	3.49631	0.00695
CILP	1.51877	0.007
MAN2C1	0.541182	0.00745
CORO2A	0.547195	0.00755
KIFC3	0.50447	0.0076
SEC31B	1.01703	0.0076
GIGYF1	0.485463	0.0079
TCIRG1	0.513615	0.00805
WDR91	0.549618	0.0081
CACNA1I	0.922539	0.00825

SLX4IP	-1.59361	0.0083
WDR90	1.32059	0.0084
PNISR	0.47523	0.00845
PHLDA1	-0.503918	0.00865
DHRS3	0.53499	0.00875
GABBR1	0.655203	0.0089
HHEX	-0.449468	0.00925
MYH3	1.49416	0.0093
ZNF648	0.578286	0.00935
AUTS2	0.970727	0.0094
NR4A2	-0.790596	0.0094
LDB1	0.450728	0.00955
WASH1	0.975752	0.0096
NEAT1	0.500751	0.0099
VHL	0.442352	0.0099
NXN	1.63989	0.01005
CCDC103	0.93037	0.01015
UBE2W	-0.902243	0.01015
PRORSD1P	1.42364	0.01025
CAMSAP2	-0.578562	0.0103
TNFRSF14	1.33664	0.01035
SLC46A3	0.758066	0.01055
CCR7	-1.02198	0.0106
CDKN2A	4.28047	0.0106
CCNL2	0.544616	0.0107
SLC18B1	-0.498931	0.01085
HLA-B	0.966044	0.011
LPPR3	0.66187	0.011
PHLDB1	0.539249	0.01125
HERC2P9	0.713553	0.01145
IER3	-0.624993	0.0116
SLC29A4	0.73183	0.0116
TUBAL3	0.778615	0.0116
CDK5RAP3	0.576939	0.01165
ERVK13-1	0.817297	0.0118
ID3	0.595995	0.012
NCALD	0.793807	0.01205
PI16	-1.64045	0.01205
PPP1R12C	0.457843	0.01205
ATP2A3	0.502584	0.01225
LAMB2	0.484647	0.0123
HSPA1B	-0.458295	0.01245
GDF15	-0.408461	0.01255
NR2F2	-0.418571	0.01255
INO80B-WBP1	0.936412	0.0126
MAT2A	0.440316	0.0128
CPNE7	0.503867	0.0129
FSTL1	2.37956	0.0129

PLK2	0.537843	0.013
APOE	0.423474	0.01305
MROH6	0.886437	0.01345
FAM118A	0.467608	0.0137
GPSM3	0.644538	0.0137
CHTF18	0.481517	0.01375
EPHX2	0.488484	0.01405
FLJ22184	2.49993	0.0142
NOL4L	0.56563	0.0142
TMEM143	0.778139	0.01435
MAP3K6	0.620534	0.0146
ARSA	0.647103	0.0148
GSTM3	0.913022	0.0151
METTL7A	0.462063	0.01525
PI4KAP2	0.487781	0.01525
ARHGAP42	-1.48116	0.01545
GOLGA8A	0.529749	0.01575
FAM110A	0.736568	0.01585
ZDHHC8	0.431798	0.0159
ZNF248	1.08395	0.0159
PAPLN	0.899765	0.01595
PRSS30P	1.56076	0.016
ITM2A	-0.440697	0.0161
TAF1C	0.458875	0.01615
DLK1	0.413779	0.0162
TRPM4	0.587942	0.0162
MID1	2.11808	0.0163
RNASE1	0.534897	0.0163
RIMS3	0.663796	0.01645
GPM6B	3.91821	0.0165
DDX11	0.463952	0.01655
LOC100996291	1.22971	0.0166
SLC36A4	1.31779	0.01665
LOC729737	0.600932	0.01675
OGDHL	1.58553	0.01675
AQP2	-1.88316	0.01685
CLEC4M	1.60262	0.01705
FBLN1	2.80383	0.01705
HDHD3	0.533292	0.0171
OBSCN	0.476921	0.01715
SPINT2	1.9243	0.01725
LOC100131564	0.669681	0.01755
HDAC6	0.443714	0.0177
ITGA2B	1.30642	0.0177
HDAC10	0.57014	0.01805
CPT1B	0.992249	0.01815
HMGA2	1.60511	0.0183
PTPN6	3.48294	0.01845

ELMO3	1.04182	0.01875
HRC	0.529873	0.01885
GAS6	1.18557	0.0189
CTBP2	2.03564	0.01895
VANGL2	1.50292	0.01895
RASSF5	0.625658	0.019
CMBL	0.409579	0.0191
YJEFN3	1.27297	0.01935
PLCD1	0.489106	0.0194
HES4	1.20244	0.0196
SIGIRR	0.669958	0.0196
MEG9	2.04835	0.01965
UNC13D	0.404587	0.01975
ZNF251	0.519545	0.01985
CROCCP2	0.484574	0.01995
PROX1	-0.57283	0.01995
TRABD	0.446976	0.0203
EPHB6	-0.986103	0.02035
TNKS1BP1	0.438115	0.02055
SLC22A31	2.40853	0.0206
ZBTB37	-0.992902	0.02075
VWDE	0.455223	0.0209
C21orf58	0.612635	0.02125
GGT1	0.937245	0.02125
PHLDA2	-0.744546	0.02125
SH3BP2	0.953895	0.02145
TMEM38B	-0.403254	0.0216
LOC642846	0.908344	0.0219
PIDD1	0.530675	0.022
MET	1.03567	0.02205
WDR92	0.50396	0.02205
ZNF702P	-0.929103	0.02205
CPZ	0.466503	0.02215
MZF1	0.586138	0.02225
RFK	-0.426176	0.02225
PLEKHG4	0.464324	0.02235
FASTK	0.468456	0.0228
IFI16	-0.618964	0.0228
GSTM4	0.651406	0.023
LAT	1.116	0.0231
PCSK9	0.44885	0.02355
NUDT7	-1.34142	0.0236
ADCY6	0.847949	0.02365
PTPRU	0.995663	0.02365
FAM219B	0.49999	0.02395
PTPRVP	0.653047	0.0243
RILP	0.517083	0.0243
IL18R1	0.729847	0.02435

SGSM2	0.467739	0.0244
TMEM259	0.373496	0.02495
TSPAN6	5.64711	0.025
AGRN	0.393651	0.02515
ARC	-1.7222	0.02535
TPT1-AS1	1.23929	0.0254
FCHO1	0.444884	0.0255
CITED1	0.930426	0.02575
NCR1	1.19315	0.02585
PERP	1.52195	0.02625
CDH2	1.93296	0.0265
F2RL1	1.85621	0.02665
APOL1	0.431695	0.0268
HOXC10	1.14683	0.0268
WDR27	0.626226	0.02705
C1orf54	1.86951	0.0271
MOCS2	-0.4303	0.0271
PBXIP1	0.55383	0.02715
CUL9	0.532726	0.0272
GRIN2C	1.43549	0.0272
AP1G2	0.574781	0.02725
ATP8B3	0.949494	0.02725
SCARNA2	-1.10216	0.02735
C8orf33	-0.377424	0.0274
HP1BP3	0.373996	0.02745
OSBPL7	0.542809	0.02765
RUNX1	0.475822	0.0279
CREM	-0.425805	0.02795
ELFN1-AS1	1.0037	0.02865
MAPK15	1.06176	0.02875
NBPF6	1.68335	0.029
CKAP4	0.825365	0.0291
CYP20A1	0.427889	0.0294
ZNF180	-0.59168	0.02945
ANPEP	1.39632	0.02985
STC2	0.786128	0.0299
GSDMB	0.933847	0.02995
TNK2	0.551306	0.03015
ZNF785	0.67507	0.03045
LPCAT4	0.586102	0.03085
MRPL47	-0.392961	0.03095
NEFM	3.45918	0.03105
BAHCC1	0.493603	0.03115
ABCA7	0.418174	0.0312
KAT2A	0.390955	0.03195
TRERF1	0.733882	0.03225
UCHL1	4.6444	0.03245
PKIB	-0.466262	0.0329

GSTM2	1.20186	0.033
ACVRL1	0.967645	0.03305
IQGAP3	0.420716	0.03305
Sep-08	0.355302	0.0331
LOC441155	0.553539	0.0332
MAPK12	0.460804	0.0333
KIAA1147	0.787697	0.03345
SZT2	0.453145	0.03355
RHAG	-0.355359	0.0336
CRABP2	2.19499	0.03375
SEMA6C	1.01292	0.03385
FAM160B2	0.435497	0.0341
LOC100499484-C9ORF174	0.708607	0.03415
RNF215	0.616375	0.03415
GLI4	0.554249	0.0342
LANCL2	-0.499981	0.0342
DDX26B	0.484354	0.03445
NPTXR	0.614099	0.03455
ZNF839	0.54214	0.03455
FNBP4	0.384369	0.03495
GABARAPL1	-0.438294	0.035
RHPN1	0.798182	0.0351
SERPINE1	-0.558938	0.0351
NMU	0.348446	0.03515
AGAP6	0.542703	0.03535
SMYD2	-0.404749	0.0355
TARBP1	0.436405	0.0356
FADS3	0.428118	0.03575
CD22	0.538472	0.03605
DYNC1H1	-0.351151	0.03605
FER	-0.533505	0.0361
SVIP	-0.395863	0.03615
PNKP	0.424669	0.0363
TNFAIP2	0.472485	0.03635
RTKN2	-0.837701	0.03655
TRMT10B	0.631769	0.0367
NPIPB5	0.650009	0.03685
ALOX5	2.07882	0.0369
RECQL4	0.35752	0.0374
PPCDC	-0.630483	0.03745
PICK1	0.554593	0.0375
DDX12P	0.576816	0.03815
INIP	-0.441889	0.03815
EXOC3L4	0.754645	0.03825
ABHD3	0.483867	0.0383
CCDC150	0.617411	0.03835
IKBKE	1.73599	0.0384

SAMSN1	-0.34386	0.0384
AMH	0.617322	0.03845
FMNL1	0.529909	0.0386
LST1	2.43381	0.0386
FKBP11	0.926263	0.03875
TYK2	0.365942	0.0393
BACH2	-0.42373	0.0394
NFATC4	0.552553	0.0394
IL17RC	0.72256	0.0396
SEMA4B	0.469862	0.0396
TP53I3	0.585384	0.0396
GUSBP11	0.607687	0.03975
ZBTB40	0.355846	0.0398
ZDHHC2	1.14343	0.0401
TMEM175	0.496519	0.04025
TSPAN4	0.591505	0.04025
RAB18	0.369565	0.04045
PLAT	-0.69436	0.04065
ZNF767P	0.550429	0.04075
KCNN4	0.624447	0.04085
CFH	0.721127	0.0411
USP53	-0.480717	0.04125
RAB27A	-0.350577	0.0413
PSMC6	-0.460617	0.04155
AP1S3	0.784255	0.0417
HECW2	-0.406382	0.0417
MAMDC4	0.819829	0.04175
KMT2D	-0.348883	0.0418
ZNF256	-1.24357	0.04185
HMGCR	-0.342121	0.0424
DDX58	-1.07657	0.04245
IL16	0.502764	0.04265
PLD2	0.621214	0.04265
HOXB6	1.49005	0.0428
PPARD	0.492473	0.04305
TCN2	0.630595	0.04305
PIF1	0.574588	0.04315
LAMA5	0.369765	0.0432
BTNL9	0.458896	0.04345
EFHC1	0.494595	0.04345
POLR2J3	1.46803	0.04365
KLK1	0.805869	0.0437
TSC22D3	0.601618	0.04395
RPPH1	-0.585442	0.04415
ALDH3A2	1.40278	0.0442
ARHGAP33	0.553909	0.0443
SLC6A15	4.80803	0.04455
CPT1A	0.461196	0.0446

PTGDS	2.25946	0.04485
RPEL1	-0.689945	0.0451
SLC25A28	0.51873	0.0458
LIME1	0.804021	0.0459
PFKFB4	0.490899	0.0461
LRRC4B	0.508247	0.0463
TFR2	0.356323	0.04635
ALPK1	1.00888	0.0464
TIGD5	-0.439596	0.04645
PLOD2	1.24986	0.04695
APOBR	0.724889	0.0472
PGM1	5.78593	0.0473
CREG1	-0.328703	0.0474
NRROS	0.487322	0.0474
SIRT4	-1.64782	0.04745
AVEN	-0.441279	0.04755
MIR17HG	0.536684	0.04755
TSPAN13	-0.369671	0.04785
SMARCD3	0.949985	0.048
MR1	0.361074	0.04805
ASB8	0.387136	0.0481
HIST1H4E	-1.04587	0.04845
LINC01021	1.56914	0.049
CTNNAL1	-0.353571	0.04915
HPS5	-0.409658	0.04935
ASAP3	0.4051	0.04945
LACTB2	-0.441415	0.0496
LIMA1	-0.345703	0.04985

References

1. Sainsbury, S., C. Bernecky, and P. Cramer, *Structural basis of transcription initiation by RNA polymerase II.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 129-43.
2. Nechaev, S. and K. Adelman, *Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation.* Biochim Biophys Acta, 2011. **1809**(1): p. 34-45.
3. Plaschka, C., et al., *Architecture of the RNA polymerase II-Mediator core initiation complex.* Nature, 2015. **518**(7539): p. 376-80.
4. He, Y., et al., *Structural visualization of key steps in human transcription initiation.* Nature, 2013. **495**(7442): p. 481-6.
5. Basehoar, A.D., S.J. Zanton, and B.F. Pugh, *Identification and distinct regulation of yeast TATA box-containing genes.* Cell, 2004. **116**(5): p. 699-709.
6. Cavallini, B., et al., *A yeast activity can substitute for the HeLa cell TATA box factor.* Nature, 1988. **334**(6177): p. 77-80.
7. Horikoshi, M., et al., *Cloning and structure of a yeast gene encoding a general transcription initiation factor TFIID that binds to the TATA box.* Nature, 1989. **341**(6240): p. 299-303.
8. Hahn, S., et al., *Isolation of the gene encoding the yeast TATA binding protein TFIID: a gene identical to the SPT15 suppressor of Ty element insertions.* Cell, 1989. **58**(6): p. 1173-81.
9. Kuras, L., et al., *TAF-Containing and TAF-independent forms of transcriptionally active TBP in vivo.* Science, 2000. **288**(5469): p. 1244-8.
10. Rhee, H.S. and B.F. Pugh, *Genome-wide structure and organization of eukaryotic pre-initiation complexes.* Nature, 2012. **483**(7389): p. 295-301.
11. Blair, R.H., J.A. Goodrich, and J.F. Kugel, *Single-molecule fluorescence resonance energy transfer shows uniformity in TATA binding protein-induced DNA bending and heterogeneity in bending kinetics.* Biochemistry, 2012. **51**(38): p. 7444-55.
12. Conaway, R.C. and J.W. Conaway, *General initiation factors for RNA polymerase II.* Annu Rev Biochem, 1993. **62**: p. 161-90.
13. Imbalzano, A.N., K.S. Zaret, and R.E. Kingston, *Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA.* J Biol Chem, 1994. **269**(11): p. 8280-6.
14. Kugel, J.F. and J.A. Goodrich, *Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIIE, TFIIH, and ATP on negatively supercoiled DNA.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9232-7.
15. Guo, J. and D.H. Price, *RNA polymerase II transcription elongation control.* Chem Rev, 2013. **113**(11): p. 8583-603.
16. Fuda, N.J., M.B. Ardehali, and J.T. Lis, *Defining mechanisms that regulate RNA polymerase II transcription in vivo.* Nature, 2009. **461**(7261): p. 186-92.

17. Canella, D., et al., *Defining the RNA polymerase III transcriptome: Genome-wide localization of the RNA polymerase III transcription machinery in human cells.* Genome Res, 2010. **20**(6): p. 710-21.
18. Willis, I.M., *RNA polymerase III. Genes, factors and transcriptional specificity.* Eur J Biochem, 1993. **212**(1): p. 1-11.
19. Herr, A.J., et al., *RNA polymerase IV directs silencing of endogenous DNA.* Science, 2005. **308**(5718): p. 118-20.
20. Wierzbicki, A.T., et al., *RNA polymerase V transcription guides ARGONAUTE4 to chromatin.* Nat Genet, 2009. **41**(5): p. 630-4.
21. Robert, F., et al., *A human RNA polymerase II-containing complex associated with factors necessary for spliceosome assembly.* J Biol Chem, 2002. **277**(11): p. 9302-6.
22. Kornberg, R.D., *The molecular basis of eucaryotic transcription.* Cell Death Differ, 2007. **14**(12): p. 1989-97.
23. Hahn, S., *Structure and mechanism of the RNA polymerase II transcription machinery.* Nat Struct Mol Biol, 2004. **11**(5): p. 394-403.
24. Langlais, D., et al., *The Stat3/GR interaction code: predictive value of direct/indirect DNA recruitment for transcription outcome.* Mol Cell, 2012. **47**(1): p. 38-49.
25. Gilmore, T.D., *Introduction to NF-kappaB: players, pathways, perspectives.* Oncogene, 2006. **25**(51): p. 6680-4.
26. Saklatvala, J., *The p38 MAP kinase pathway as a therapeutic target in inflammatory disease.* Curr Opin Pharmacol, 2004. **4**(4): p. 372-7.
27. Frietze, S. and P.J. Farnham, *Transcription factor effector domains.* Subcell Biochem, 2011. **52**: p. 261-77.
28. Adcock, I.M. and G. Caramori, *Cross-talk between pro-inflammatory transcription factors and glucocorticoids.* Immunol Cell Biol, 2001. **79**(4): p. 376-84.
29. Barnes, P.J., *Reduced histone deacetylase in COPD: clinical implications.* Chest, 2006. **129**(1): p. 151-5.
30. Perkins, N.D., *Post-translational modifications regulating the activity and function of the nuclear factor kappa B pathway.* Oncogene, 2006. **25**(51): p. 6717-30.
31. Ito, K., K.F. Chung, and I.M. Adcock, *Update on glucocorticoid action and resistance.* J Allergy Clin Immunol, 2006. **117**(3): p. 522-43.
32. Illingworth, R.S. and A.P. Bird, *CpG islands--'a rough guide'.* FEBS Lett, 2009. **583**(11): p. 1713-20.
33. Thomson, J.P., et al., *CpG islands influence chromatin structure via the CpG-binding protein Cfp1.* Nature, 2010. **464**(7291): p. 1082-6.
34. Deweese, J.E., M.A. Osheroff, and N. Osheroff, *DNA Topology and Topoisomerases: Teaching a "Knotty" Subject.* Biochem Mol Biol Educ, 2008. **37**(1): p. 2-10.
35. Peterson, C.L. and M.A. Laniel, *Histones and histone modifications.* Curr Biol, 2004. **14**(14): p. R546-51.

36. Luger, K., *Structure and dynamic behavior of nucleosomes.* Curr Opin Genet Dev, 2003. **13**(2): p. 127-35.
37. Marzluff, W.F., *Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts.* Curr Opin Cell Biol, 2005. **17**(3): p. 274-80.
38. Song, F., et al., *Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units.* Science, 2014. **344**(6182): p. 376-80.
39. Li, B., M. Carey, and J.L. Workman, *The role of chromatin during transcription.* Cell, 2007. **128**(4): p. 707-19.
40. Hizume, K., et al., *Removal of histone tails from nucleosome dissects the physical mechanisms of salt-induced aggregation, linker histone H1-induced compaction, and 30-nm fiber formation of the nucleosome array.* Ultramicroscopy, 2009. **109**(8): p. 868-73.
41. Morales, V. and H. Richard-Foy, *Role of Histone N-Terminal Tails and Their Acetylation in Nucleosome Dynamics.* Molecular and Cellular Biology, 2000. **20**(19): p. 7230-7237.
42. Biswas, M., et al., *Role of histone tails in structural stability of the nucleosome.* PLoS Comput Biol, 2011. **7**(12): p. e1002279.
43. Iwasaki, W., et al., *Contribution of histone N-terminal tails to the structure and stability of nucleosomes.* FEBS Open Bio, 2013. **3**: p. 363-9.
44. Woodcock, C.L. and R.P. Ghosh, *Chromatin higher-order structure and dynamics.* Cold Spring Harb Perspect Biol, 2010. **2**(5): p. a000596.
45. Routh, A., S. Sandin, and D. Rhodes, *Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure.* Proc Natl Acad Sci U S A, 2008. **105**(26): p. 8872-7.
46. Tremethick, D.J., *Higher-order structures of chromatin: the elusive 30 nm fiber.* Cell, 2007. **128**(4): p. 651-4.
47. Finch, J.T. and A. Klug, *Solenoidal model for superstructure in chromatin.* Proc Natl Acad Sci U S A, 1976. **73**(6): p. 1897-901.
48. Langmore, J.P. and J.R. Paulson, *Low angle x-ray diffraction studies of chromatin structure in vivo and in isolated nuclei and metaphase chromosomes.* J Cell Biol, 1983. **96**(4): p. 1120-31.
49. Widom, J. and A. Klug, *Structure of the 300A chromatin filament: X-ray diffraction from oriented samples.* Cell, 1985. **43**(1): p. 207-13.
50. Gerchman, S.E. and V. Ramakrishnan, *Chromatin higher-order structure studied by neutron scattering and scanning transmission electron microscopy.* Proc Natl Acad Sci U S A, 1987. **84**(22): p. 7802-6.
51. Ghirlando, R. and G. Felsenfeld, *Hydrodynamic studies on defined heterochromatin fragments support a 30-nm fiber having six nucleosomes per turn.* J Mol Biol, 2008. **376**(5): p. 1417-25.

52. Woodcock, C.L., *Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length.* J Cell Biol, 1994. **125**(1): p. 11-9.
53. Williams, S.P., et al., *Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length.* Biophys J, 1986. **49**(1): p. 233-48.
54. Jansen, A. and K.J. Verstrepen, *Nucleosome positioning in Saccharomyces cerevisiae.* Microbiol Mol Biol Rev, 2011. **75**(2): p. 301-20.
55. Talbert, P.B. and S. Henikoff, *Histone variants on the move: substrates for chromatin dynamics.* Nat Rev Mol Cell Biol, 2017. **18**(2): p. 115-126.
56. Waterborg, J.H. and A.J. Robertson, *Common features of analogous replacement histone H3 genes in animals and plants.* J Mol Evol, 1996. **43**(3): p. 194-206.
57. Talbert, P.B. and S. Henikoff, *Histone variants--ancient wrap artists of the epigenome.* Nat Rev Mol Cell Biol, 2010. **11**(4): p. 264-75.
58. Verdone, L., et al., *Histone acetylation in gene regulation.* Brief Funct Genomic Proteomic, 2006. **5**(3): p. 209-21.
59. Steiner, F.A. and S. Henikoff, *Diversity in the organization of centromeric chromatin.* Curr Opin Genet Dev, 2015. **31**: p. 28-35.
60. McKinley, K.L. and I.M. Cheeseman, *The molecular basis for centromere identity and function.* Nat Rev Mol Cell Biol, 2016. **17**(1): p. 16-29.
61. Adam, M., et al., *H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions.* Mol Cell Biol, 2001. **21**(18): p. 6270-9.
62. Soboleva, T.A., et al., *A unique H2A histone variant occupies the transcriptional start site of active genes.* Nat Struct Mol Biol, 2011. **19**(1): p. 25-30.
63. Ratnakumar, K., et al., *ATRX-mediated chromatin association of histone variant macroH2A1 regulates alpha-globin expression.* Genes Dev, 2012. **26**(5): p. 433-8.
64. Yelagandula, R., et al., *The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis.* Cell, 2014. **158**(1): p. 98-109.
65. Eitoku, M., et al., *Histone chaperones: 30 years from isolation to elucidation of the mechanisms of nucleosome assembly and disassembly.* Cell Mol Life Sci, 2008. **65**(3): p. 414-44.
66. Bartova, E., et al., *Differentiation of human embryonic stem cells induces condensation of chromosome territories and formation of heterochromatin protein 1 foci.* Differentiation, 2008. **76**(1): p. 24-32.
67. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.
68. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.

69. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.
70. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41-5.
71. Turner, B.M., *Cellular memory and the histone code.* Cell, 2002. **111**(3): p. 285-91.
72. Bonaldi, T. and R. Noberini, *Recent advances in mass spectrometry analysis of histone post-translational modifications: potential clinical impact of the PAT-H-MS approach.* Expert Rev Proteomics, 2016. **13**(3): p. 245-50.
73. Onder, O., et al., *Progress in epigenetic histone modification analysis by mass spectrometry for clinical investigations.* Expert Rev Proteomics, 2015. **12**(5): p. 499-517.
74. Janssen, K.A., S. Sidoli, and B.A. Garcia, *Recent Achievements in Characterizing the Histone Code and Approaches to Integrating Epigenomics and Systems Biology.* Methods Enzymol, 2017. **586**: p. 359-378.
75. Lee, J.S., E. Smith, and A. Shilatifard, *The language of histone crosstalk.* Cell, 2010. **142**(5): p. 682-5.
76. Allfrey, V.G., R. Faulkner, and 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 of the United States of America, 1964. **51**(5): p. 786-794.
77. Petty, E. and L. Pillus, *Balancing chromatin remodeling and histone modifications in transcription.* Trends Genet, 2013. **29**(11): p. 621-9.
78. Parthun, M.R., *Hat1: the emerging cellular roles of a type B histone acetyltransferase.* Oncogene, 2007. **26**(37): p. 5319-28.
79. Wang, Z., et al., *Combinatorial patterns of histone acetylations and methylations in the human genome.* Nat Genet, 2008. **40**(7): p. 897-903.
80. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.
81. Das, C., et al., *CBP/p300-mediated acetylation of histone H3 on lysine 56.* Nature, 2009. **459**(7243): p. 113-7.
82. Berndsen, C.E., et al., *Catalytic mechanism of a MYST family histone acetyltransferase.* Biochemistry, 2007. **46**(3): p. 623-9.
83. Li, X., et al., *Nucleus-Translocated ACSS2 Promotes Gene Transcription for Lysosomal Biogenesis and Autophagy.* Mol Cell, 2017. **66**(5): p. 684-697 e9.
84. Shahbazian, M.D. and M. Grunstein, *Functions of site-specific histone acetylation and deacetylation.* Annu Rev Biochem, 2007. **76**: p. 75-100.
85. Marks, P., et al., *Histone deacetylases and cancer: causes and therapies.* Nat Rev Cancer, 2001. **1**(3): p. 194-202.

86. Richon, V.M. and J.P. O'Brien, *Histone deacetylase inhibitors: a new class of potential therapeutic agents for cancer treatment.* Clin Cancer Res, 2002. **8**(3): p. 662-4.
87. Bernstein, B.E., J.K. Tong, and S.L. Schreiber, *Genomewide studies of histone deacetylase function in yeast.* Proc Natl Acad Sci U S A, 2000. **97**(25): p. 13708-13.
88. Greer, E.L. and Y. Shi, *Histone methylation: a dynamic mark in health, disease and inheritance.* Nat Rev Genet, 2012. **13**(5): p. 343-57.
89. Wei, H., et al., *Protein arginine methylation of non-histone proteins and its role in diseases.* Cell Cycle, 2014. **13**(1): p. 32-41.
90. Bannister, A.J. and T. Kouzarides, *Reversing histone methylation.* Nature, 2005. **436**(7054): p. 1103-6.
91. Araki, Y. and T. Mimura, *The Histone Modification Code in the Pathogenesis of Autoimmune Diseases.* Mediators Inflamm, 2017. **2017**: p. 2608605.
92. Bedford, M.T. and S. Richard, *Arginine methylation an emerging regulator of protein function.* Mol Cell, 2005. **18**(3): p. 263-72.
93. Martin, C. and Y. Zhang, *The diverse functions of histone lysine methylation.* Nat Rev Mol Cell Biol, 2005. **6**(11): p. 838-49.
94. Qian, C. and M.M. Zhou, *SET domain protein lysine methyltransferases: Structure, specificity and catalysis.* Cell Mol Life Sci, 2006. **63**(23): p. 2755-63.
95. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.
96. Komander, D., *The emerging complexity of protein ubiquitination.* Biochem Soc Trans, 2009. **37**(Pt 5): p. 937-53.
97. Taverna, S.D., et al., *How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers.* Nat Struct Mol Biol, 2007. **14**(11): p. 1025-1040.
98. Cheung, P., C.D. Allis, and P. Sassone-Corsi, *Signaling to chromatin through histone modifications.* Cell, 2000. **103**(2): p. 263-71.
99. Cruickshank, M.N., P. Besant, and D. Ulgiati, *The impact of histone post-translational modifications on developmental gene regulation.* Amino Acids, 2010. **39**(5): p. 1087-105.
100. Houben, A., et al., *Phosphorylation of histone H3 in plants--a dynamic affair.* Biochim Biophys Acta, 2007. **1769**(5-6): p. 308-15.
101. Johansen, K.M. and J. Johansen, *Regulation of chromatin structure by histone H3S10 phosphorylation.* Chromosome Res, 2006. **14**(4): p. 393-404.
102. Loomis, R.J., et al., *Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation.* Mol Cell, 2009. **33**(4): p. 450-61.
103. Rogakou, E.P., et al., *DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139.* J Biol Chem, 1998. **273**(10): p. 5858-68.

104. Rossetto, D., N. Avvakumov, and J. Cote, *Histone phosphorylation: a chromatin modification involved in diverse nuclear events.* Epigenetics, 2012. **7**(10): p. 1098-108.
105. Cao, J. and Q. Yan, *Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer.* Front Oncol, 2012. **2**: p. 26.
106. Sowa, M.E., et al., *Defining the human deubiquitinating enzyme interaction landscape.* Cell, 2009. **138**(2): p. 389-403.
107. Zhang, Y., *Transcriptional regulation by histone ubiquitination and deubiquitination.* Genes Dev, 2003. **17**(22): p. 2733-40.
108. Minsky, N., et al., *Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells.* Nat Cell Biol, 2008. **10**(4): p. 483-8.
109. Meas, R. and P. Mao, *Histone ubiquitylation and its roles in transcription and DNA damage response.* DNA Repair (Amst), 2015. **36**: p. 36-42.
110. Smeenk, G. and N. Mailand, *Writers, Readers, and Erasers of Histone Ubiquitylation in DNA Double-Strand Break Repair.* Front Genet, 2016. **7**: p. 122.
111. Uckelmann, M. and T.K. Sixma, *Histone ubiquitination in the DNA damage response.* DNA Repair (Amst), 2017. **56**: p. 92-101.
112. Northam, M.R. and K.M. Trujillo, *Histone H2B mono-ubiquitylation maintains genomic integrity at stalled replication forks.* Nucleic Acids Res, 2016. **44**(19): p. 9245-9255.
113. Vann, K.R. and T.G. Kutateladze, *Histone H3 Dual Ubiquitylation Mediates Maintenance DNA Methylation.* Mol Cell, 2017. **68**(2): p. 261-262.
114. Seeler, J.S. and A. Dejean, *Nuclear and unclear functions of SUMO.* Nat Rev Mol Cell Biol, 2003. **4**(9): p. 690-9.
115. Shiio, Y. and R.N. Eisenman, *Histone sumoylation is associated with transcriptional repression.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13225-30.
116. Meulmeester, E. and F. Melchior, *Cell biology: SUMO.* Nature, 2008. **452**(7188): p. 709-11.
117. Nathan, D., et al., *Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications.* Genes Dev, 2006. **20**(8): p. 966-76.
118. Dhall, A., et al., *Sumoylated human histone H4 prevents chromatin compaction by inhibiting long-range internucleosomal interactions.* J Biol Chem, 2014. **289**(49): p. 33827-37.
119. Hottiger, M.O., et al., *Toward a unified nomenclature for mammalian ADP-ribosyltransferases.* Trends Biochem Sci, 2010. **35**(4): p. 208-19.
120. Vyas, S., et al., *Family-wide analysis of poly(ADP-ribose) polymerase activity.* Nat Commun, 2014. **5**: p. 4426.

121. Gibson, B.A. and W.L. Kraus, *New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs.* Nat Rev Mol Cell Biol, 2012. **13**(7): p. 411-24.
122. Caldecott, K.W., *Single-strand break repair and genetic disease.* Nat Rev Genet, 2008. **9**(8): p. 619-31.
123. Okano, S., et al., *Spatial and temporal cellular responses to single-strand breaks in human cells.* Mol Cell Biol, 2003. **23**(11): p. 3974-81.
124. Kanno, S., et al., *A novel human AP endonuclease with conserved zinc-finger-like motifs involved in DNA strand break responses.* EMBO J, 2007. **26**(8): p. 2094-103.
125. Hassa, P.O., et al., *Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going?* Microbiol Mol Biol Rev, 2006. **70**(3): p. 789-829.
126. Cohen-Armon, M., et al., *DNA-independent PARP-1 activation by phosphorylated ERK2 increases Elk1 activity: a link to histone acetylation.* Mol Cell, 2007. **25**(2): p. 297-308.
127. Hymes, J. and B. Wolf, *Human biotinidase isn't just for recycling biotin.* J Nutr, 1999. **129**(2S Suppl): p. 485s-489s.
128. Hymes, J., K. Fleischhauer, and B. Wolf, *Biotinylation of histones by human serum biotinidase: assessment of biotinyl-transferase activity in sera from normal individuals and children with biotinidase deficiency.* Biochem Mol Med, 1995. **56**(1): p. 76-83.
129. Wolf, B. and G.S. Heard, *Biotinidase deficiency.* Adv Pediatr, 1991. **38**: p. 1-21.
130. Zempleni, J. and D.M. Mock, *BIOTIN | Physiology A2 - Caballero, Benjamin*, in *Encyclopedia of Food Sciences and Nutrition (Second Edition)*. 2003, Academic Press: Oxford. p. 516-523.
131. Camporeale, G., et al., *K12-biotinylated histone H4 marks heterochromatin in human lymphoblastoma cells.* J Nutr Biochem, 2007. **18**(11): p. 760-8.
132. Filenko, N.A., et al., *The role of histone H4 biotinylation in the structure of nucleosomes.* PLoS One, 2011. **6**(1): p. e16299.
133. Kobza, K., et al., *K4, K9 and K18 in human histone H3 are targets for biotinylation by biotinidase.* FEBS J, 2005. **272**(16): p. 4249-59.
134. Ramachandran, G.N. and A.K. Mitra, *An explanation for the rare occurrence of cis peptide units in proteins and polypeptides.* J Mol Biol, 1976. **107**(1): p. 85-92.
135. Nelson, C.J., H. Santos-Rosa, and T. Kouzarides, *Proline isomerization of histone H3 regulates lysine methylation and gene expression.* Cell, 2006. **126**(5): p. 905-16.
136. Shaw, P.E., *Peptidyl-prolyl cis/trans isomerases and transcription: is there a twist in the tail?* EMBO Rep, 2007. **8**(1): p. 40-5.
137. Tan, M., et al., *Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification.* Cell, 2011. **146**(6): p. 1016-28.

138. Sabari, B.R., et al., *Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation.* Mol Cell, 2015. **58**(2): p. 203-15.
139. Flynn, E.M., et al., *A Subset of Human Bromodomains Recognizes Butyryllysine and Crotonyllysine Histone Peptide Modifications.* Structure, 2015. **23**(10): p. 1801-14.
140. Bao, X., et al., *Identification of 'erasers' for lysine crotonylated histone marks using a chemical proteomics approach.* Elife, 2014. **3**.
141. Feldman, J.L., J. Baeza, and J.M. Denu, *Activation of the protein deacetylase SIRT6 by long-chain fatty acids and widespread deacylation by mammalian sirtuins.* J Biol Chem, 2013. **288**(43): p. 31350-6.
142. Montellier, E., et al., *Histone crotonylation specifically marks the haploid male germ cell gene expression program: post-meiotic male-specific gene expression.* Bioessays, 2012. **34**(3): p. 187-93.
143. Sabari, B.R., et al., *Metabolic regulation of gene expression through histone acylations.* Nature reviews. Molecular cell biology, 2017. **18**(2): p. 90-101.
144. Lee, K.K. and J.L. Workman, *Histone acetyltransferase complexes: one size doesn't fit all.* Nat Rev Mol Cell Biol, 2007. **8**(4): p. 284-95.
145. Roth, S.Y., J.M. Denu, and C.D. Allis, *Histone acetyltransferases.* Annu Rev Biochem, 2001. **70**: p. 81-120.
146. Allis, C.D., et al., *Proteolytic processing of histone H3 in chromatin: a physiologically regulated event in Tetrahymena micronuclei.* Cell, 1980. **20**(1): p. 55-64.
147. Vossaert, L., et al., *Identification of histone H3 clipping activity in human embryonic stem cells.* Stem Cell Res, 2014. **13**(1): p. 123-34.
148. Duarte, L.F., et al., *Histone H3.3 and its proteolytically processed form drive a cellular senescence programme.* Nat Commun, 2014. **5**: p. 5210.
149. Howe, C.G. and M.V. Gamble, *Enzymatic cleavage of histone H3: a new consideration when measuring histone modifications in human samples.* Clin Epigenetics, 2015. **7**: p. 7.
150. Tvardovskiy, A., et al., *Top-down and Middle-down Protein Analysis Reveals that Intact and Clipped Human Histones Differ in Post-translational Modification Patterns.* Mol Cell Proteomics, 2015. **14**(12): p. 3142-53.
151. Santos-Rosa, H., et al., *Histone H3 tail clipping regulates gene expression.* Nat Struct Mol Biol, 2009. **16**(1): p. 17-22.
152. Clapier, C.R. and B.R. Cairns, *The biology of chromatin remodeling complexes.* Annu Rev Biochem, 2009. **78**: p. 273-304.
153. Narlikar, G.J., R. Sundaramoorthy, and T. Owen-Hughes, *Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes.* Cell, 2013. **154**(3): p. 490-503.

154. Lorch, Y., B. Maier-Davis, and R.D. Kornberg, *Mechanism of chromatin remodeling.* Proc Natl Acad Sci U S A, 2010. **107**(8): p. 3458-62.
155. Saha, A., J. Wittmeyer, and B.R. Cairns, *Chromatin remodelling: the industrial revolution of DNA around histones.* Nat Rev Mol Cell Biol, 2006. **7**(6): p. 437-47.
156. Close, P., et al., *The emerging role of lysine acetylation of non-nuclear proteins.* Cell Mol Life Sci, 2010. **67**(8): p. 1255-64.
157. Yang, X.J., *The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases.* Nucleic Acids Res, 2004. **32**(3): p. 959-76.
158. Anamika, K., et al., *Lessons from genome-wide studies: an integrated definition of the coactivator function of histone acetyl transferases.* Epigenetics Chromatin, 2010. **3**(1): p. 18.
159. Lafon, A., et al., *MYST opportunities for growth control: yeast genes illuminate human cancer gene functions.* Oncogene, 2007. **26**(37): p. 5373-84.
160. Utley, R.T. and J. Cote, *The MYST family of histone acetyltransferases.* Curr Top Microbiol Immunol, 2003. **274**: p. 203-36.
161. Yan, Y., et al., *Crystal structure of yeast Esa1 suggests a unified mechanism for catalysis and substrate binding by histone acetyltransferases.* Mol Cell, 2000. **6**(5): p. 1195-205.
162. Yuan, H., et al., *MYST protein acetyltransferase activity requires active site lysine autoacetylation.* EMBO J, 2012. **31**(1): p. 58-70.
163. Rea, S., G. Xouri, and A. Akhtar, *Males absent on the first (MOF): from flies to humans.* Oncogene, 2007. **26**(37): p. 5385-94.
164. Turner-Ivey, B., et al., *KAT6A, a Chromatin Modifier from the 8p11-p12 Amplicon is a Candidate Oncogene in Luminal Breast Cancer.* Neoplasia. **16**(8): p. 644-655.
165. Wu, J., et al., *Bisubstrate Inhibitors of the MYST HATs Esa1 and Tip60.* Bioorg Med Chem, 2009. **17**(3): p. 1381-6.
166. Chrivia, J.C., et al., *Phosphorylated CREB binds specifically to the nuclear protein CBP.* Nature, 1993. **365**(6449): p. 855-9.
167. Eckner, R., et al., *Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor.* Genes Dev, 1994. **8**(8): p. 869-84.
168. Yuan, L.W. and A. Giordano, *Acetyltransferase machinery conserved in p300/CBP-family proteins.* Oncogene, 2002. **21**(14): p. 2253-60.
169. Thompson, P.R., et al., *Transcriptional coactivator protein p300. Kinetic characterization of its histone acetyltransferase activity.* J Biol Chem, 2001. **276**(36): p. 33721-9.
170. Chan, H.M. and N.B. La Thangue, *p300/CBP proteins: HATs for transcriptional bridges and scaffolds.* J Cell Sci, 2001. **114**(Pt 13): p. 2363-73.

171. Goodman, R.H. and S. Smolik, *CBP/p300 in cell growth, transformation, and development.* Genes Dev, 2000. **14**(13): p. 1553-77.
172. Grossman, S.R., et al., *p300/MDM2 complexes participate in MDM2-mediated p53 degradation.* Mol Cell, 1998. **2**(4): p. 405-15.
173. Ryan, C.M., et al., *Functional interaction of CREB binding protein (CBP) with nuclear transport proteins and modulation by HDAC inhibitors.* Cell Cycle, 2006. **5**(18): p. 2146-52.
174. Liu, X., et al., *The structural basis of protein acetylation by the p300/CBP transcriptional coactivator.* Nature, 2008. **451**(7180): p. 846-50.
175. Brownell, J.E., et al., *Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation.* Cell, 1996. **84**(6): p. 843-51.
176. Horiuchi, J., et al., *ADA1, a novel component of the ADA/GCN5 complex, has broader effects than GCN5, ADA2, or ADA3.* Mol Cell Biol, 1997. **17**(6): p. 3220-8.
177. Xie, L., et al., *The roles of bacterial GCN5-related N-acetyltransferases.* Crit Rev Eukaryot Gene Expr, 2014. **24**(1): p. 77-87.
178. Blanco, J.C., et al., *The histone acetylase PCAF is a nuclear receptor coactivator.* Genes Dev, 1998. **12**(11): p. 1638-51.
179. Xu, W., D.G. Edmondson, and S.Y. Roth, *Mammalian GCN5 and P/CAF acetyltransferases have homologous amino-terminal domains important for recognition of nucleosomal substrates.* Mol Cell Biol, 1998. **18**(10): p. 5659-69.
180. Guelman, S., et al., *Host cell factor and an uncharacterized SANT domain protein are stable components of ATAC, a novel dAda2A/dGcn5-containing histone acetyltransferase complex in Drosophila.* Mol Cell Biol, 2006. **26**(3): p. 871-82.
181. Vetting, M.W., et al., *Structure and functions of the GNAT superfamily of acetyltransferases.* Arch Biochem Biophys, 2005. **433**(1): p. 212-26.
182. Dyda, F., D.C. Klein, and A.B. Hickman, *GCN5-related N-acetyltransferases: a structural overview.* Annu Rev Biophys Biomol Struct, 2000. **29**: p. 81-103.
183. Neuwald, A.F. and D. Landsman, *GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein.* Trends Biochem Sci, 1997. **22**(5): p. 154-5.
184. Santos-Rosa, H., et al., *Mechanisms of P/CAF auto-acetylation.* Nucleic Acids Res, 2003. **31**(15): p. 4285-92.
185. Borrow, J., et al., *The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein.* Nat Genet, 1996. **14**(1): p. 33-41.

186. Chaffanet, M., et al., *MOZ is fused to p300 in an acute monocytic leukemia with t(8;22).* Genes Chromosomes Cancer, 2000. **28**(2): p. 138-44.
187. Holbert, M.A., et al., *The human monocytic leukemia zinc finger histone acetyltransferase domain contains DNA-binding activity implicated in chromatin targeting.* J Biol Chem, 2007. **282**(50): p. 36603-13.
188. Yoshida, H. and I. Kitabayashi, *Chromatin regulation by AML1 complex.* Int J Hematol, 2008. **87**(1): p. 19-24.
189. Sanchez, R. and M.M. Zhou, *The PHD finger: a versatile epigenome reader.* Trends Biochem Sci, 2011. **36**(7): p. 364-72.
190. Dreveny, I., et al., *The double PHD finger domain of MOZ/MYST3 induces alpha-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification.* Nucleic Acids Res, 2014. **42**(2): p. 822-35.
191. Xiong, X., et al., *Selective recognition of histone crotonylation by double PHD fingers of MOZ and DPF2.* Nat Chem Biol, 2016. **12**(12): p. 1111-1118.
192. Doyon, Y., et al., *ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation.* Mol Cell, 2006. **21**(1): p. 51-64.
193. Smith, E.R., et al., *A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16.* Mol Cell Biol, 2005. **25**(21): p. 9175-88.
194. Doyon, Y., et al., *Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans.* Mol Cell Biol, 2004. **24**(5): p. 1884-96.
195. Kitabayashi, I., et al., *Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ-CBP fusion protein.* Embo j, 2001. **20**(24): p. 7184-96.
196. Ullah, M., et al., *Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes.* Mol Cell Biol, 2008. **28**(22): p. 6828-43.
197. Laue, K., et al., *The multidomain protein Brpf1 binds histones and is required for Hox gene expression and segmental identity.* Development, 2008. **135**(11): p. 1935-1946.
198. Cho, H.I., M.S. Kim, and Y.K. Jang, *The BRPF2/BRD1-MOZ complex is involved in retinoic acid-induced differentiation of embryonic stem cells.* Exp Cell Res, 2016. **346**(1): p. 30-9.
199. Yan, K., et al., *The Chromatin Regulator BRPF3 Preferentially Activates the HBO1 Acetyltransferase but Is Dispensable for Mouse Development and Survival.* J Biol Chem, 2016. **291**(6): p. 2647-63.
200. Voss, Anne K., et al., *MOZ Regulates the Tbx1 Locus, and Moz Mutation Partially Phenocopies DiGeorge Syndrome.* Developmental Cell, 2012. **23**(3): p. 652-663.

201. Yoshida, H. and I. Kitabayashi, *Chromatin regulation by AML1 complex.* International Journal of Hematology, 2008. **87**(1): p. 19-24.
202. Collins, H.M., et al., *MOZ-TIF2 alters cofactor recruitment and histone modification at the RARbeta2 promoter: differential effects of MOZ fusion proteins on CBP- and MOZ-dependent activators.* J Biol Chem, 2006. **281**(25): p. 17124-33.
203. Katsumoto, T., et al., *MOZ is essential for maintenance of hematopoietic stem cells.* Genes Dev, 2006. **20**(10): p. 1321-30.
204. Pelletier, N., et al., *MOZ and MORF histone acetyltransferases interact with the Runt-domain transcription factor Runx2.* Oncogene, 2002. **21**(17): p. 2729-40.
205. Rokudai, S., et al., *Monocytic leukemia zinc finger (MOZ) interacts with p53 to induce p21 expression and cell-cycle arrest.* J Biol Chem, 2009. **284**(1): p. 237-44.
206. Ohta, K., et al., *Histone acetyltransferase MOZ acts as a co-activator of Nrf2-MafK and induces tumour marker gene expression during hepatocarcinogenesis.* Biochem J, 2007. **402**(3): p. 559-66.
207. Chan, E.M., et al., *MOZ and MOZ-CBP cooperate with NF-kappaB to activate transcription from NF-kappaB-dependent promoters.* Exp Hematol, 2007. **35**(12): p. 1782-92.
208. Rokudai, S., et al., *MOZ increases p53 acetylation and premature senescence through its complex formation with PML.* Proc Natl Acad Sci U S A, 2013. **110**(10): p. 3895-900.
209. Kastan, M.B., et al., *Participation of p53 protein in the cellular response to DNA damage.* Cancer Res, 1991. **51**(23 Pt 1): p. 6304-11.
210. Kuerbitz, S.J., et al., *Wild-type p53 is a cell cycle checkpoint determinant following irradiation.* Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7491-5.
211. Clarke, A.R., et al., *Thymocyte apoptosis induced by p53-dependent and independent pathways.* Nature, 1993. **362**(6423): p. 849-52.
212. Katsumoto, T., N. Yoshida, and I. Kitabayashi, *Roles of the histone acetyltransferase monocytic leukemia zinc finger protein in normal and malignant hematopoiesis.* Cancer Sci, 2008. **99**(8): p. 1523-7.
213. Thomas, T., et al., *Monocytic leukemia zinc finger protein is essential for the development of long-term reconstituting hematopoietic stem cells.* Genes Dev, 2006. **20**(9): p. 1175-86.
214. Perez-Campo, F.M., et al., *The histone acetyl transferase activity of monocytic leukemia zinc finger is critical for the proliferation of hematopoietic precursors.* Blood, 2009. **113**(20): p. 4866-74.
215. Sheikh, B.N., et al., *MOZ regulates B-cell progenitors and, consequently, Moz haploinsufficiency dramatically retards MYC-induced lymphoma development.* Blood, 2015. **125**(12): p. 1910-21.

216. Crump, J.G., et al., *Moz-dependent Hox expression controls segment-specific fate maps of skeletal precursors in the face.* Development, 2006. **133**(14): p. 2661-9.
217. Voss, A.K., et al., *Moz and retinoic acid coordinately regulate H3K9 acetylation, Hox gene expression, and segment identity.* Dev Cell, 2009. **17**(5): p. 674-86.
218. the Ab they used was not specific
219. Largeot, A., et al., *Symplekin, a polyadenylation factor, prevents MOZ and MLL activity on HOXA9 in hematopoietic cells.* Biochim Biophys Acta, 2013. **1833**(12): p. 3054-3063.
220. Ayton, P.M. and M.L. Cleary, *Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9.* Genes Dev, 2003. **17**(18): p. 2298-307.
221. Camos, M., et al., *Gene expression profiling of acute myeloid leukemia with translocation t(8;16)(p11;p13) and MYST3-CREBBP rearrangement reveals a distinctive signature with a specific pattern of HOX gene expression.* Cancer Res, 2006. **66**(14): p. 6947-54.
222. Paggetti, J., et al., *Crosstalk between leukemia-associated proteins MOZ and MLL regulates HOX gene expression in human cord blood CD34+ cells.* Oncogene, 2010. **29**(36): p. 5019-31.
223. Cao, R., Y. Tsukada, and Y. Zhang, *Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing.* Mol Cell, 2005. **20**(6): p. 845-54.
224. Sheikh, B.N., et al., *MOZ and BMI1 play opposing roles during Hox gene activation in ES cells and in body segment identity specification in vivo.* Proc Natl Acad Sci U S A, 2015. **112**(17): p. 5437-42.
225. Sheikh, B.N., et al., *MOZ and BMI1 act synergistically to maintain hematopoietic stem cells.* Exp Hematol, 2017. **47**: p. 83-97 e8.
226. Perez-Campo, F.M., et al., *MOZ-mediated repression of p16(INK) (4) (a) is critical for the self-renewal of neural and hematopoietic stem cells.* Stem Cells, 2014. **32**(6): p. 1591-601.
227. Tham, E., et al., *Dominant mutations in KAT6A cause intellectual disability with recognizable syndromic features.* Am J Hum Genet, 2015. **96**(3): p. 507-13.
228. Vanyai, H.K., T. Thomas, and A.K. Voss, *Mesodermal expression of Moz is necessary for cardiac septum development.* Dev Biol, 2015. **403**(1): p. 22-9.
229. Heair, H.M., et al., *MicroRNA 665 Regulates Dentinogenesis through MicroRNA-Mediated Silencing and Epigenetic Mechanisms.* Mol Cell Biol, 2015. **35**(18): p. 3116-30.
230. Kitabayashi, I., et al., *Fusion of MOZ and p300 histone acetyltransferases in acute monocytic leukemia with a t(8;22)(p11;q13) chromosome translocation.* Leukemia, 2001. **15**(1): p. 89-94.

231. Carapeti, M., et al., *A novel fusion between MOZ and the nuclear receptor coactivator TIF2 in acute myeloid leukemia.* Blood, 1998. **91**(9): p. 3127-33.
232. Esteyries, S., et al., *NCOA3, a new fusion partner for MOZ/MYST3 in M5 acute myeloid leukemia.* Leukemia, 2008. **22**(3): p. 663-5.
233. Chinen, Y., et al., *The leucine twenty homeobox (LEUTX) gene, which lacks a histone acetyltransferase domain, is fused to KAT6A in therapy-related acute myeloid leukemia with t(8;19)(p11;q13).* Genes Chromosomes Cancer, 2014. **53**(4): p. 299-308.
234. Deguchi, K., et al., *MOZ-TIF2-induced acute myeloid leukemia requires the MOZ nucleosome binding motif and TIF2-mediated recruitment of CBP.* Cancer Cell. **3**(3): p. 259-271.
235. Shipley, J.L. and J.N. Butera, *Acute myelogenous leukemia.* Exp Hematol, 2009. **37**(6): p. 649-58.
236. Sill, H., et al., *Therapy-related myeloid neoplasms: pathobiology and clinical characteristics.* Br J Pharmacol, 2011. **162**(4): p. 792-805.
237. Dohner, H., et al., *Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet.* Blood, 2010. **115**(3): p. 453-74.
238. Patel, J.P., et al., *Prognostic relevance of integrated genetic profiling in acute myeloid leukemia.* N Engl J Med, 2012. **366**(12): p. 1079-89.
239. Cancer Genome Atlas Research, N., et al., *Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia.* N Engl J Med, 2013. **368**(22): p. 2059-74.
240. Gilliland, D.G. and J.D. Griffin, *The roles of FLT3 in hematopoiesis and leukemia.* Blood, 2002. **100**(5): p. 1532-42.
241. Takahashi, S., *Current findings for recurring mutations in acute myeloid leukemia.* J Hematol Oncol, 2011. **4**: p. 36.
242. Kihara, R., et al., *Comprehensive analysis of genetic alterations and their prognostic impacts in adult acute myeloid leukemia patients.* Leukemia, 2014. **28**(8): p. 1586-95.
243. Cook, A.M., et al., *Role of altered growth factor receptor-mediated JAK2 signaling in growth and maintenance of human acute myeloid leukemia stem cells.* Blood, 2014. **123**(18): p. 2826-37.
244. Ghoshal Gupta, S., H. Baumann, and M. Wetzler, *Epigenetic regulation of signal transducer and activator of transcription 3 in acute myeloid leukemia.* Leuk Res, 2008. **32**(7): p. 1005-14.
245. Yamada, O. and K. Kawauchi, *The role of the JAK-STAT pathway and related signal cascades in telomerase activation during the development of hematologic malignancies.* JAKSTAT, 2013. **2**(4): p. e25256.
246. Schuringa, J.J., et al., *Constitutive Stat3, Tyr705, and Ser727 phosphorylation in acute myeloid leukemia cells caused by the autocrine secretion of interleukin-6.* Blood, 2000. **95**(12): p. 3765-70.

247. Spiekermann, K., et al., *Overexpression and constitutive activation of FLT3 induces STAT5 activation in primary acute myeloid leukemia blast cells.* Clin Cancer Res, 2003. **9**(6): p. 2140-50.
248. Steensma, D.P., et al., *JAK2 V617F is a rare finding in de novo acute myeloid leukemia, but STAT3 activation is common and remains unexplained.* Leukemia, 2006. **20**(6): p. 971-8.
249. Koduru, P.R., et al., *Cytogenetic and Cytogenomic Microarray Characterization of Chromothripsis in Chromosome 8 Affecting MOZ/NCOA2 (TIF2), FGFR1, RUNX1T1, and RUNX1 in a Pediatric Acute Myeloid Leukemia.* J Pediatr Hematol Oncol, 2017. **39**(4): p. e227-e232.
250. Yang, X.J. and M. Ullah, *MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells.* Oncogene, 2007. **26**(37): p. 5408-19.
251. Avvakumov, N. and J. Cote, *The MYST family of histone acetyltransferases and their intimate links to cancer.* Oncogene, 2007. **26**(37): p. 5395-407.
252. Kindle, K.B., et al., *MOZ-TIF2 inhibits transcription by nuclear receptors and p53 by impairment of CBP function.* Mol Cell Biol, 2005. **25**(3): p. 988-1002.
253. Kindle, K.B., H.M. Collins, and D.M. Heery, *MOZ-TIF2-mediated destruction of CBP/p300 is blocked by calpain inhibitor 2.* Leukemia, 2010. **24**(7): p. 1359-1361.
254. Huntly, B.J., et al., *MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors.* Cancer Cell, 2004. **6**(6): p. 587-96.
255. Deguchi, K., et al., *MOZ-TIF2-induced acute myeloid leukemia requires the MOZ nucleosome binding motif and TIF2-mediated recruitment of CBP.* Cancer Cell, 2003. **3**(3): p. 259-71.
256. Aikawa, Y., et al., *PU.1-mediated upregulation of CSF1R is crucial for leukemia stem cell potential induced by MOZ-TIF2.* Nat Med, 2010. **16**(5): p. 580-585.
257. Ye, M., et al., *Hematopoietic Differentiation Is Required for Initiation of Acute Myeloid Leukemia.* Cell Stem Cell, 2015.
258. Rokudai, S., et al., *MOZ increases p53 acetylation and premature senescence through its complex formation with PML.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(10): p. 3895-3900.
259. Zhu, J., et al., *Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth.* Nature, 2015. **525**(7568): p. 206-11.
260. Turner-Ivey, B., et al., *KAT6A, a Chromatin Modifier from the 8p11-p12 Amplicon is a Candidate Oncogene in Luminal Breast Cancer.* Neoplasia (New York, N.Y.), 2014. **16**(8): p. 644-655.
261. Yu, L., et al., *Identification of MYST3 as a novel epigenetic activator of ERalpha frequently amplified in breast cancer.* Oncogene, 2017. **36**(20): p. 2910-2918.

262. Singh, P.B. and S.D. Georgatos, *HP1: facts, open questions, and speculation.* J Struct Biol, 2002. **140**(1-3): p. 10-6.
263. Lindsay, E.A. and A. Baldini, *Recovery from arterial growth delay reduces penetrance of cardiovascular defects in mice deleted for the DiGeorge syndrome region.* Hum Mol Genet, 2001. **10**(9): p. 997-1002.
264. Voss, A.K., et al., *MOZ regulates the Tbx1 locus, and Moz mutation partially phenocopies DiGeorge syndrome.* Dev Cell, 2012. **23**(3): p. 652-63.
265. Hanahan, D., *Studies on transformation of Escherichia coli with plasmids.* J Mol Biol, 1983. **166**(4): p. 557-80.
266. Bullard, J., et al., *Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments.* BMC Bioinformatics, 2010. **11**(1): p. 1-13.
267. Sheikh, B.N., et al., *MOZ (KAT6A) is essential for the maintenance of classically defined adult hematopoietic stem cells.* Blood, 2016.
268. Sheikh, B.N., et al., *MOZ and BMI1 act synergistically to maintain hematopoietic stem cells.* Exp Hematol, 2017. **47**: p. 83-97.e8.
269. Rozman, M., et al., *Type I MOZ/CBP (MYST3/CREBBP) is the most common chimeric transcript in acute myeloid leukemia with t(8;16)(p11;p13) translocation.* Genes Chromosomes Cancer, 2004. **40**(2): p. 140-5.
270. Crowley, J.A., et al., *Detection of MOZ-CBP fusion in acute myeloid leukemia with 8;16 translocation.* Leukemia, 2005. **19**(12): p. 2344-5.
271. Miller, C.T., L. Maves, and C.B. Kimmel, *moz regulates Hox expression and pharyngeal segmental identity in zebrafish.* Development, 2004. **131**(10): p. 2443-61.
272. Jung, N., et al., *An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis.* Nat Commun, 2015. **6**.
273. Good-Jacobson, K.L., et al., *Regulation of germinal center responses and B-cell memory by the chromatin modifier MOZ.* Proceedings of the National Academy of Sciences, 2014. **111**(26): p. 9585-9590.
274. Sheikh, B.N., et al., *MOZ regulates B-cell progenitors and, consequently, Moz haploinsufficiency dramatically retards MYC-induced lymphoma development.* Blood, 2015. **125**(12): p. 1910-1921.
275. Murati, A., et al., *Genome profiling of acute myelomonocytic leukemia: alteration of the MYB locus in MYST3-linked cases.* Leukemia, 2009. **23**(1): p. 85-94.
276. Klein, E., et al., *Properties of the K562 cell line, derived from a patient with chronic myeloid leukemia.* Int J Cancer, 1976. **18**(4): p. 421-31.
277. Carthew, R.W. and E.J. Sontheimer, *Origins and Mechanisms of miRNAs and siRNAs.* Cell, 2009. **136**(4): p. 642-655.

278. Mocellin, S. and M. Provenzano, *RNA interference: learning gene knock-down from cell physiology.* Journal of Translational Medicine, 2004. **2**(1): p. 39.
279. Siomi, H. and M.C. Siomi, *On the road to reading the RNA-interference code.* Nature, 2009. **457**(7228): p. 396-404.
280. Jackson, A.L., et al., *Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing.* RNA, 2006. **12**(7): p. 1197-1205.
281. Iribe, H., et al., *Chemical Modification of the siRNA Seed Region Suppresses Off-Target Effects by Steric Hindrance to Base-Pairing with Targets.* ACS Omega, 2017. **2**(5): p. 2055-2064.
282. Qiu, Y., et al., *Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ reveals a regulatory mechanism for HOXA9 transcription.* Genes Dev, 2012. **26**(12): p. 1376-91.
283. Meyer, B., et al., *Histone H3 Lysine 9 Acetylation Obstructs ATM Activation and Promotes Ionizing Radiation Sensitivity in Normal Stem Cells.* Stem Cell Reports. **7**(6): p. 1013-1022.
284. Karmodiya, K., et al., *H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells.* BMC Genomics, 2012. **13**(1): p. 424.
285. Harris, S.L. and A.J. Levine, *The p53 pathway: positive and negative feedback loops.* Oncogene, 2005. **24**(17): p. 2899-908.
286. Finlay, C.A., P.W. Hinds, and A.J. Levine, *The p53 proto-oncogene can act as a suppressor of transformation.* Cell, 1989. **57**(7): p. 1083-93.
287. Baker, S.J., et al., *Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas.* Science, 1989. **244**(4901): p. 217-21.
288. Nigro, J.M., et al., *Mutations in the p53 gene occur in diverse human tumour types.* Nature, 1989. **342**(6250): p. 705-708.
289. Eliyahu, D., et al., *Wild-type p53 can inhibit oncogene-mediated focus formation.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(22): p. 8763-8767.
290. Banin, S., et al., *Enhanced phosphorylation of p53 by ATM in response to DNA damage.* Science, 1998. **281**(5383): p. 1674-7.
291. Reed, S.M. and D.E. Quelle, *p53 Acetylation: Regulation and Consequences.* Cancers, 2015. **7**(1): p. 30-69.
292. Tang, Y., et al., *Acetylation is indispensable for p53 activation.* Cell, 2008. **133**(4): p. 612-26.
293. Barlev, N.A., et al., *Acetylation of p53 activates transcription through recruitment of coactivators/histone acetyltransferases.* Mol Cell, 2001. **8**(6): p. 1243-54.
294. Riley, T., et al., *Transcriptional control of human p53-regulated genes.* Nat Rev Mol Cell Biol, 2008. **9**(5): p. 402-412.

295. Giacoia, E.G., et al., *PAI-1 leads to G1-phase cell-cycle progression through cyclin D3/cdk4/6 upregulation.* Mol Cancer Res, 2014. **12**(3): p. 322-34.
296. Wang, S., et al., *Growth differentiation factor 15 promotes blood vessel growth by stimulating cell cycle progression in repair of critical-sized calvarial defect.* Sci Rep, 2017. **7**(1): p. 9027.
297. Dai, L., et al., *Targeting HGF/c-MET induces cell cycle arrest, DNA damage, and apoptosis for primary effusion lymphoma.* Blood, 2015. **126**(26): p. 2821-31.
298. Cizmecioglu, O., et al., *Plk2 regulated centriole duplication is dependent on its localization to the centrosome and a functional polo-box domain.* Cell Cycle, 2008. **7**(22): p. 3548-3555.
299. Lee, J.H., et al., *The p53-inducible gene 3 (PIG3) contributes to early cellular response to DNA damage.* Oncogene, 2010. **29**(10): p. 1431-1450.
300. Ueda, K., H. Arakawa, and Y. Nakamura, *Dual-specificity phosphatase 5 (DUSP5) as a direct transcriptional target of tumor suppressor p53.* Oncogene, 2003. **22**(36): p. 5586-5591.
301. Attardi, L.D., et al., *PERP, an apoptosis-associated target of p53, is a novel member of the PMP-22/gas3 family.* Genes Dev, 2000. **14**(6): p. 704-18.
302. Zhang, Y.-L., et al., *Roles of Rap1 signaling in tumor cell migration and invasion.* Cancer Biology & Medicine, 2017. **14**(1): p. 90-99.
303. Infante, E., S.J. Heasman, and A.J. Ridley, *Statins inhibit T-acute lymphoblastic leukemia cell adhesion and migration through Rap1b.* J Leukoc Biol, 2011. **89**(4): p. 577-86.
304. Lu, S.C., *Glutathione synthesis.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2013. **1830**(5): p. 3143-3153.
305. Pei, S., et al., *Targeting aberrant glutathione metabolism to eradicate human acute myelogenous leukemia cells.* J Biol Chem, 2013. **288**(47): p. 33542-58.
306. Goyama, S., et al., *Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells.* The Journal of Clinical Investigation, 2013. **123**(11): p. 4979-4979.
307. Ben-Ami, O., et al., *Addiction of t(8;21) and inv(16) acute myeloid leukemia to native RUNX1.* Cell Rep, 2013. **4**(6): p. 1131-43.
308. Deguchi, K., et al., *MOZ-TIF2-induced acute myeloid leukemia requires the MOZ nucleosome binding motif and TIF2-mediated recruitment of CBP.* Cancer Cell, 2003. **3**(3): p. 259-271.
309. Weng, M.K., et al., *Extensive Transcriptional Regulation of Chromatin Modifiers during Human Neurodevelopment.* PLoS ONE, 2012. **7**(5): p. e36708.
310. Foulkes, N.S. and P. Sassone-Corsi, *More is better: activators and repressors from the same gene.* Cell, 1992. **68**(3): p. 411-4.
311. Aguado, F., et al., *The CREB/CREM transcription factors negatively regulate early synaptogenesis and spontaneous network activity.* J Neurosci, 2009. **29**(2): p. 328-33.

312. Sassone-Corsi, P., *CREM: a master-switch governing male germ cells differentiation and apoptosis.* Semin Cell Dev Biol, 1998. **9**(4): p. 475-82.
313. Molina, C.A., et al., *Inducibility and negative autoregulation of CREM: an alternative promoter directs the expression of ICER, an early response repressor.* Cell, 1993. **75**(5): p. 875-86.
314. Pigazzi, M., et al., *cAMP response element binding protein (CREB) overexpression CREB has been described as critical for leukemia progression.* Haematologica, 2007. **92**(10): p. 1435-7.
315. Cheng, J.C., et al., *Potential role of CREB as a prognostic marker in acute myeloid leukemia.* Future Oncol, 2007. **3**(4): p. 475-80.
316. Pigazzi, M., et al., *ICER expression inhibits leukemia phenotype and controls tumor progression.* Leukemia, 2008. **22**(12): p. 2217-2225.
317. Shinnakasu, R., et al., *Regulated selection of germinal-center cells into the memory B cell compartment.* Nat Immunol, 2016. **17**(7): p. 861-869.
318. Itoh-Nakadai, A., et al., *The transcription repressors Bach2 and Bach1 promote B cell development by repressing the myeloid program.* Nat Immunol, 2014. **15**(12): p. 1171-1180.
319. Swaminathan, S., et al., *BACH2 mediates negative selection and p53-dependent tumor suppression at the pre-B cell receptor checkpoint.* Nature medicine, 2013. **19**(8): p. 1014-1022.
320. Itoh-Nakadai, A., et al., *The transcription repressors Bach2 and Bach1 promote B cell development by repressing the myeloid program.* Nat Immunol, 2014. **15**(12): p. 1171-80.
321. Dawson, M.J. and J.A. Trapani, *IFI 16 gene encodes a nuclear protein whose expression is induced by interferons in human myeloid leukaemia cell lines.* J Cell Biochem, 1995. **57**(1): p. 39-51.
322. Johnstone, R.W., et al., *Functional interaction between p53 and the interferon-inducible nucleoprotein IFI 16.* Oncogene, 2000. **19**(52): p. 6033-42.
323. Kwak, J.C., et al., *IFI16 as a negative regulator in the regulation of p53 and p21(Waf1).* J Biol Chem, 2003. **278**(42): p. 40899-904.
324. Piccaluga, P.P., et al., *IFI16 reduced expression is correlated with unfavorable outcome in chronic lymphocytic leukemia.* Apmis, 2017. **125**(6): p. 511-522.
325. Mullican, S.E., et al., *Abrogation of nuclear receptors Nr4a3 and Nr4a1 leads to development of acute myeloid leukemia.* Nat Med, 2007. **13**(6): p. 730-5.
326. Srinivasan, R.S., et al., *The nuclear hormone receptor Coup-TFII is required for the initiation and early maintenance of Prox1 expression in lymphatic endothelial cells.* Genes Dev, 2010. **24**(7): p. 696-707.
327. Aranguren, X.L., et al., *Transcription factor COUP-TFII is indispensable for venous and lymphatic development in zebrafish*

and Xenopus laevis. Biochemical and Biophysical Research Communications, 2011. **410**(1): p. 121-126.
328. Bedford, F.K., et al., *HEX: a novel homeobox gene expressed during haematopoiesis and conserved between mouse and human.* Nucleic Acids Res, 1993. **21**(5): p. 1245-9.
329. Crompton, M.R., et al., *Identification of a novel vertebrate homeobox gene expressed in haematopoietic cells.* Nucleic Acids Research, 1992. **20**(21): p. 5661-5667.
330. Shields, B.J., et al., *Acute myeloid leukemia requires Hhex to enable PRC2-mediated epigenetic repression of Cdkn2a.* Genes & Development, 2016. **30**(1): p. 78-91.
331. Ai, L., et al., *The p16 (CDKN2a//INK4a) Tumor-Suppressor Gene in Head and Neck Squamous Cell Carcinoma: A Promoter Methylation and Protein Expression Study in 100 Cases.* Mod Pathol, 2003. **16**(9): p. 944-950.
332. Collado, M., M.A. Blasco, and M. Serrano, *Cellular senescence in cancer and aging.* Cell, 2007. **130**(2): p. 223-33.
333. Agarwal, P., et al., *Tumor suppressor gene p16/INK4A/CDKN2A-dependent regulation into and out of the cell cycle in a spontaneous canine model of breast cancer.* J Cell Biochem, 2013. **114**(6): p. 1355-63.
334. Lukas, J., et al., *Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16.* Nature, 1995. **375**(6531): p. 503-6.
335. Sheikh, B.N., et al., *MOZ (MYST3, KAT6A) inhibits senescence via the INK4A-ARF pathway.* Oncogene, 2015. **34**(47): p. 5807-20.
336. Largeot, A., et al., *Expression of the MOZ-TIF2 oncoprotein in mice represses senescence.* Experimental Hematology, 2016. **44**(4): p. 231-237.e4.
337. Tsuchiya, Y., et al., *Changes in acetyl CoA levels during the early embryonic development of Xenopus laevis.* PLoS One, 2014. **9**(5): p. e97693.
338. Larsson, O., *HMG-CoA reductase inhibitors: role in normal and malignant cells.* Crit Rev Oncol Hematol, 1996. **22**(3): p. 197-212.
339. Kawata, S., et al., *Increase in the active form of 3-hydroxy-3-methylglutaryl coenzyme A reductase in human hepatocellular carcinoma: possible mechanism for alteration of cholesterol biosynthesis.* Cancer Res, 1990. **50**(11): p. 3270-3.
340. Harwood, H.J., Jr., et al., *In vivo regulation of human leukocyte 3-hydroxy-3-methylglutaryl coenzyme A reductase: increased enzyme protein concentration and catalytic efficiency in human leukemia and lymphoma.* J Lipid Res, 1991. **32**(8): p. 1237-52.
341. Vitols, S., et al., *Multilevel regulation of low-density lipoprotein receptor and 3-hydroxy-3-methylglutaryl coenzyme A reductase gene expression in normal and leukemic cells.* Blood, 1994. **84**(8): p. 2689-98.
342. Caruso, M.G., et al., *Enhanced 3-hydroxy-3-methyl-glutaryl coenzyme A reductase activity in human colorectal cancer not*

expressing low density lipoprotein receptor. Anticancer Res, 1999. **19**(1a): p. 451-4.
343. Puzio-Kuter, A.M., *The Role of p53 in Metabolic Regulation.* Genes & Cancer, 2011. **2**(4): p. 385-391.
344. Laezza, C., et al., *p53 regulates the mevalonate pathway in human glioblastoma multiforme.* Cell Death Dis, 2015. **6**: p. e1909.
345. Cheung, N., et al., *Targeting Aberrant Epigenetic Networks Mediated by PRMT1 and KDM4C in Acute Myeloid Leukemia.* Cancer Cell, 2016. **29**(1): p. 32-48.
346. Arboleda, V.A., et al., *De novo nonsense mutations in KAT6A, a lysine acetyl-transferase gene, cause a syndrome including microcephaly and global developmental delay.* Am J Hum Genet, 2015. **96**(3): p. 498-506.
347. Zack, T.I., et al., *Pan-cancer patterns of somatic copy number alteration.* Nat Genet, 2013. **45**(10): p. 1134-40.
348. Lawrence, H.J., et al., *Mice bearing a targeted interruption of the homeobox gene HOXA9 have defects in myeloid, erythroid, and lymphoid hematopoiesis.* Blood, 1997. **89**(6): p. 1922-30.
349. So, C.W., et al., *Leukemic transformation of hematopoietic progenitors by MLL-GAS7 in the absence of Hoxa7 or Hoxa9.* Blood, 2004. **103**(8): p. 3192-9.
350. Cammenga, J., et al., *RUNX1 DNA-binding mutants, associated with minimally differentiated acute myelogenous leukemia, disrupt myeloid differentiation.* Cancer Res, 2007. **67**(2): p. 537-45.
351. Milella, M., et al., *MEK blockade converts AML differentiating response to retinoids into extensive apoptosis.* Blood, 2007. **109**(5): p. 2121-9.
352. Harhaj, N.S., S.C. Sun, and E.W. Harhaj, *Activation of NF-kappa B by the human T cell leukemia virus type I Tax oncoprotein is associated with ubiquitin-dependent relocalization of I kappa B kinase.* J Biol Chem, 2007. **282**(6): p. 4185-92.
353. Berman, J.N. and A.T. Look, *Targeting transcription factors in acute leukemia in children.* Curr Drug Targets, 2007. **8**(6): p. 727-37.
354. Horvath, P. and R. Barrangou, *CRISPR/Cas, the immune system of bacteria and archaea.* Science, 2010. **327**(5962): p. 167-70.
355. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.
356. Shen, B., et al., *Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects.* Nat Methods, 2014. **11**(4): p. 399-402.
357. Chiang, T.W., et al., *CRISPR-Cas9(D10A) nickase-based genotypic and phenotypic screening to enhance genome editing.* Sci Rep, 2016. **6**: p. 24356.
358. Hsu, P.D., E.S. Lander, and F. Zhang, *Development and applications of CRISPR-Cas9 for genome engineering.* Cell, 2014. **157**(6): p. 1262-78.

359. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nat Protoc, 2013. **8**(11): p. 2281-2308.
360. Potter, H. and R. Heller, *Transfection by electroporation.* Curr Protoc Mol Biol, 2010. **Chapter 9**: p. Unit9 3.
361. Naumann, S., et al., *Complete karyotype characterization of the K562 cell line by combined application of G-banding, multiplex-fluorescence in situ hybridization, fluorescence in situ hybridization, and comparative genomic hybridization.* Leuk Res, 2001. **25**(4): p. 313-22.
362. Liu, X., et al., *MOF as an evolutionarily conserved histone crotonyltransferase and transcriptional activation by histone acetyltransferase-deficient and crotonyltransferase-competent CBP/p300.* Cell Discov, 2017. **3**: p. 17016.
363. Clarkson, B., et al., *Studies of cellular proliferation in human leukemia. I. Estimation of growth rates of leukemic and normal hematopoietic cells in two adults with acute leukemia given single injections of tritiated thymidine.* J Clin Invest, 1967. **46**(4): p. 506-29.
364. Schmitt, C.A., et al., *A senescence program controlled by p53 and p16INK4a contributes to the outcome of cancer therapy.* Cell, 2002. **109**(3): p. 335-46.
365. Caino, M.C., J. Meshki, and M.G. Kazanietz, *Hallmarks for senescence in carcinogenesis: novel signaling players.* Apoptosis, 2009. **14**(4): p. 392-408.
366. Shay, J.W. and I.B. Roninson, *Hallmarks of senescence in carcinogenesis and cancer therapy.* Oncogene, 2004. **23**(16): p. 2919-33.
367. You, L., et al., *Deficiency of the chromatin regulator BRPF1 causes abnormal brain development.* J Biol Chem, 2015. **290**(11): p. 7114-29.
368. Sager, R., *Senescence as a mode of tumor suppression.* Environmental Health Perspectives, 1991. **93**: p. 59-62.
369. Campisi, J., *Cellular senescence as a tumor-suppressor mechanism.* Trends Cell Biol, 2001. **11**(11): p. S27-31.
370. Rutherford, T.R., J.B. Clegg, and D.J. Weatherall, *K562 human leukaemic cells synthesise embryonic haemoglobin in response to haemin.* Nature, 1979. **280**(5718): p. 164-5.
371. Housset, M., M.T. Daniel, and L. Degos, *Small doses of ARA-C in the treatment of acute myeloid leukaemia: differentiation of myeloid leukaemia cells?* Br J Haematol, 1982. **51**(1): p. 125-9.
372. Watanabe, T., et al., *Effects of 1-beta-D-arabinofuranosylcytosine and phorbol ester on differentiation of human K562 erythroleukemia cells.* Mol Pharmacol, 1985. **27**(6): p. 683-8.
373. del Senno, L., et al., *Human leukemic K562 cells: differential effects of 5-azacytidine on DNA methylation of epsilon-, gamma-globin and 7SL RNA genes.* Boll Soc Ital Biol Sper, 1984. **60**(8): p. 1613-9.

374. Fibach, E., et al., *Hemin-induced acceleration of hemoglobin production in immature cultured erythroid cells: preferential enhancement of fetal hemoglobin.* Blood, 1995. **85**(10): p. 2967-74.
375. Zhang, D., E. Cho, and J. Wong, *A critical role for the co-repressor N-CoR in erythroid differentiation and heme synthesis.* Cell Res, 2007. **17**(9): p. 804-14.
376. Shima, H., et al., *Bromodomain-PHD finger protein 1 is critical for leukemogenesis associated with MOZ-TIF2 fusion.* Int J Hematol, 2014. **99**(1): p. 21-31.
377. Campos, E.I. and D. Reinberg, *Histones: annotating chromatin.* Annu Rev Genet, 2009. **43**: p. 559-99.
378. Yi, C. and C. He, *DNA repair by reversal of DNA damage.* Cold Spring Harb Perspect Biol, 2013. **5**(1): p. a012575.
379. Eker, A.P., et al., *DNA repair in mammalian cells: Direct DNA damage reversal: elegant solutions for nasty problems.* Cell Mol Life Sci, 2009. **66**(6): p. 968-80.
380. Taniguchi, Y., *The Bromodomain and Extra-Terminal Domain (BET) Family: Functional Anatomy of BET Paralogous Proteins.* Int J Mol Sci, 2016. **17**(11).
381. Roberts, T.C., et al., *BRD3 and BRD4 BET Bromodomain Proteins Differentially Regulate Skeletal Myogenesis.* Sci Rep, 2017. **7**(1): p. 6153.
382. Filippakopoulos, P. and S. Knapp, *Targeting bromodomains: epigenetic readers of lysine acetylation.* Nat Rev Drug Discov, 2014. **13**(5): p. 337-56.
383. Dawson, M.A., et al., *Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia.* Nature, 2011. **478**(7370): p. 529-33.
384. Dawson, M.A., et al., *Recurrent mutations, including NPM1c, activate a BRD4-dependent core transcriptional program in acute myeloid leukemia.* Leukemia, 2014. **28**(2): p. 311-20.
385. Zhao, Q., et al., *Combined Id1 and Id3 Deletion Leads to Severe Erythropoietic Disturbances.* PLoS One, 2016. **11**(4): p. e0154480.
386. Kersten, C., et al., *BMP-6 inhibits human bone marrow B lymphopoiesis--upregulation of Id1 and Id3.* Exp Hematol, 2006. **34**(1): p. 72-81.
387. Itzen, F., et al., *Brd4 activates P-TEFb for RNA polymerase II CTD phosphorylation.* Nucleic Acids Res, 2014. **42**(12): p. 7577-90.
388. Gong, J.N., et al., *The role, mechanism and potentially therapeutic application of microRNA-29 family in acute myeloid leukemia.* Cell Death Differ, 2014. **21**(1): p. 100-12.
389. Bandopadhayay, P., et al., *BET bromodomain inhibition of MYC-amplified medulloblastoma.* Clin Cancer Res, 2014. **20**(4): p. 912-25.

390. Venkataraman, S., et al., *Inhibition of BRD4 attenuates tumor cell self-renewal and suppresses stem cell signaling in MYC driven medulloblastoma.* Oncotarget, 2014. **5**(9): p. 2355-71.
391. Pastori, C., et al., *BET bromodomain proteins are required for glioblastoma cell proliferation.* Epigenetics, 2014. **9**(4): p. 611-20.
392. Zhang, W., et al., *Bromodomain-containing protein 4 (BRD4) regulates RNA polymerase II serine 2 phosphorylation in human CD4+ T cells.* J Biol Chem, 2012. **287**(51): p. 43137-55.
393. Devaiah, B.N., et al., *BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin.* Nat Struct Mol Biol, 2016. **23**(6): p. 540-8.
394. Wu, S.Y., et al., *Phospho switch triggers Brd4 chromatin binding and activator recruitment for gene-specific targeting.* Mol Cell, 2013. **49**(5): p. 843-57.
395. Carretta, M., et al., *BRD3/4 inhibition and FLT3-ligand deprivation target pathways that are essential for the survival of human MLL-AF9+ leukemic cells.* PLoS One, 2017. **12**(12): p. e0189102.
396. Andrade, F.G., et al., *Identification of the MYST3-CREBBP fusion gene in infants with acute myeloid leukemia and hemophagocytosis.* Rev Bras Hematol Hemoter, 2016. **38**(4): p. 291-297.
397. Heintzman, N.D. and B. Ren, *Finding distal regulatory elements in the human genome.* Curr Opin Genet Dev, 2009. **19**(6): p. 541-9.