ELUCIDATING THE ROLE OF MOZ AND ITS IMPLICATIONS FOR KAT6A GLOBAL DEVELOPMENTAL DELAY SYNDROME

MR JONATHAN B WHITCHURCH, BSc.

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

SEPTEMBER 2018

Abstract

The histone acetyltransferase Monocytic Leukaemia Zinc Finger Protein (MOZ) was originally identified at the breakpoint of a chromosomal translocation (inv8) associated with acute myeloid leukaemia (AML). MOZ, encoded by the KAT6A gene, belongs to the MYST family of histone acetyltransferases (HATs) and is expressed in a range of human tissues. It is required for the transcriptional regulation of HOX genes during development and is indispensable for the self-renewal ability and differentiation of haematopoietic stem cells. In mouse studies, KAT6A inactivation results in cardiac development abnormalities and impaired B lymphocyte maturation. Furthermore, this HAT is required for dentinogenesis and the inhibition of cell senescence. Initially this protein was primarily studied in the context of cancer, leading to advances in the understanding of the structure and function of this epigenetic regulator, for example the DPF domain which confers histone substrate specificity. MOZ functions as a component of a chromatin regulatory complex containing ING5, EAF6 and the bromodomain protein BRPF1, catalysing H3K9, H3K14 and H3K23 acetylation.

Recently, pathogenic sequence variants in KAT6A, clustering in exons 17 & 18 have been implicated in global developmental delay (GDD) and intellectual disability syndromes. Most of these arise as *de novo* heterozygous mutations, resulting most commonly in premature termination codons (PTCs). Whether these variants result in nonsense-mediated mRNA decay (NMD), or the production of dominant-negative truncated MOZ proteins, is unknown.

In this study, we used CRISPR-Cas9 nickase to generate PTCs in the KAT6A gene in erythroleukemia K562 cells, to resemble those found in individuals displaying KAT6A GDD. Clones were generated in which 2 or 3 alleles of KAT6A were targeted (K562 is triploid for this gene). We observed a reduction in spliced mRNA transcript levels of KAT6A harbouring PTCs in exon 17, indicating the action of transcript degradation by NMD, as no change in pre-mRNA levels were observed.

Inhibitors of NMD were able to rescue the expression of KAT6A and MOZ target gene transcripts. In contrast, no evidence of NMD was observed in clones harbouring PTCs in KAT6A exon 18, which expressed truncated MOZ proteins. MTT assays indicated defects in all KAT6A PTC clones with regard to their rates of increase in the total number of viable, metabolically active cells over time, accompanied by changes in mRNA expression levels of selected MOZ target genes, as determined by qRT-PCR.

Analysis of publically available RNA-seq data from two individuals with this syndrome indicated significant changes in HOX gene expression. Gene Ontology terms identified from differentially expressed gene lists indicates transcriptional changes in an enrichment of genes involved in cell differentiation and Wnt signalling. Taken together, our results offer a clearer insight into the molecular mechanisms behind the KAT6A GDD syndrome, demonstrating that C-terminal regions of the MOZ protein are indeed essential for some normal molecular functions of this HAT. This paves the way for potential therapies in the future.

Publications

- Ngee Lek, Rieko Tadokoro-Cuccaro, **Jonathan B. Whitchurch**, Bismoy Mazumder, Harriet Miles, Philipp Prentice, Trevor Bunch, Karolina Zielińska, Veronika Metzler, Nigel P. Mongan, David M. Heery, Ieuan A. Hughes. (2018) ***Predicting puberty in partial androgen insensitivity syndrome: Use of clinical and functional androgen receptor indices.*** EBioMedicine, 36, DOI: 10.1016/j.ebiom.2018.09.047

- Joel Fulton, Bismoy Mazumder, **Jonathan B Whitchurch**, Cintia J Monteiro, Hilary M Collins, Chun M Chan, Maria P Clemente, Miguel Hernandez-Quiles, Elizabeth A Stewart, Winfried M Amoaku, Paula M Moran, Nigel P Mongan, Jenny L Persson, Simak Ali, David M Heery. (2017) ***Heterodimers of photoreceptor-specific nuclear receptor (PNRNR2E3) and peroxisome proliferator-activated receptor-(PPAR) are disrupted by retinal disease-associated mutations***. Cell Death & Disease, 8(3), DOI:10.1038/cddis.2017.98

- Veronika M Metzler, Simone de Brot, Robert S Robinson, Jennie N Jeyapalan, Emad Rakha, Thomas Walton, David S Gardner, Emma F Lund, **Jonathan Whitchurch**, Daisy Haigh, Jack M Lochray, Brian D Robinson, Cinzia Allegrucci, Rupert G Fray, Jenny L Persson, Niels Ødum, Regina R Miftakhova, Albert A Rizvanov, Ieuan A Hughes, Rieko Tadokoro-Cuccaro, David M Heery, Catrin S Rutland, Nigel P Mongan. (2017) ***Androgen dependent mechanisms of pro-angiogenic networks in placental and tumor development***. Placenta, DOI:10.1016/j.placenta.2017.02.018

- Emeli M. Nilsson, Kristian Bruun Laursen, **Jonathan Whitchurch**, Andrew McWilliam, Niels Ødum, Jenny L. Persson, David M. Heery, Lorraine J. Gudas, Nigel P. Mongan. (2015) ***MIR137 is an androgen regulated repressor of an extended network of transcriptional coregulators***. Oncotarget, 6(34), DOI:10.18632/oncotarget.5958

Acknowledgements

First and foremost, I would like to express my gratitude to Professor David Heery, who as my primary PhD supervisor has been a great source of knowledge and motivation. Furthermore, he has provided me with excellent opportunities to enhance my scientific and interpersonal skills, as well as the chance to partake in a range of collaborative projects, for which I am grateful. Similarly, I thank Dr Nigel Mongan for his guidance as my second supervisor and for his assistance with the experimental design and bioinformatics analyses, along with Dr Franco Falcone, whose constructive criticism and technical suggestions have been highly beneficial during my PhD journey.

A huge amount of thanks goes to Dr Hilary Collins, who has been highly supportive and assisted greatly in the technical guidance of my lab work. In addition, I am grateful to our lab technicians Barbara Rampersad & Trudi Gee, whose work has made my experiments that little bit easier and have been a source of uplifting conversation over the last 4 years.

Thanks to everyone I have had the opportunity to meet and work with in the Gene Regulation & RNA Biology laboratory over the course of my PhD - they have helped create a close and pleasant working environment, as well as being supportive and offering advice in difficult times. Bismoy Mazumder, Cintia Monteiro, Mitchell Masterson & Christopher Roberts deserve specific thanks in this regard. Also, thank you to the BBSRC and the University of Nottingham, who without their funding, I would never have been able to do this PhD.

Of course, I cannot forget Amy Smith, who has put up with my complaints and frustrations during the PhD (as well celebrating the small victories!), me coming home late, working constantly and being abroad for 3 months. Despite all this, you still said yes when I proposed! I am forever grateful - thank you.

Table of contents

ABSTRACT **1**

PUBLICATIONS **3**

ACKNOWLEDGEMENTS **4**

TABLE OF CONTENTS **5**

LIST OF FIGURES **9**

LIST OF TABLES **13**

LIST OF EQUATIONS **15**

LIST OF ABBREVIATIONS **16**

CHAPTER I - INTRODUCTION **21**

I. I. GLOBAL DEVELOPMENTAL DELAY, INTELLECTUAL DISABILITIES & *DE NOVO* MUTATIONS **21**

i. KAT6A ***22***

ii. KAT6B ***27***

I. II. CHROMATIN STRUCTURE & HISTONES **30**

I. III. REGULATION OF GENE EXPRESSION **33**

i. Transcription ***35***

ii. mRNA splicing & nucleocytoplasmic transport ***37***

iii. Protein translation ***38***

I. IV. NONSENSE-MEDIATED DECAY (NMD) **40**

I. V. POST-TRANSLATIONAL MODIFICATIONS **43**

i. Acetylation ***43***

ii. Other lysine acylations ***44***

iii. Methylation ***45***

iv. Phosphorylation ***46***

v. Ubiquitination ***47***

I. VI. HISTONE ACETYLTRANSFERASES **48**

I. VII. MOZ DOMAIN STRUCTURE & FUNCTION **51**

i. Conservation ***51***

ii. Cellular localisation ***54***

iii. DNA-binding ability ***54***

iv. Catalytic activity ***56***

v. Recognition of histone marks ***59***

vi. Protein interactions & transcriptional regulation of target genes ***62***

vii. Cell cycle regulation ***67***

I. VIII. MOZ IN DEVELOPMENT & CANCER 70
i. Embryogenesis 70
ii. Haematopoiesis & the immune system 72
iii. Dentinogenesis 73
iv. Leukaemia 74
v. Glioblastoma 77
vi. Breast cancer 78
I. IX. CRISPR-CAS9 GENOME EDITING 79
I. X. AIMS & OBJECTIVES 83

CHAPTER II - MATERIALS & METHODS 87

II. I. SOURCES OF MATERIALS & THEIR COMPOSITION 87
i. General reagents 87
ii. Bacterial reagents 87
iii. Molecular biology reagents 88
iv. Biochemical reagents 88
v. Tissue culture reagents 89
II. II. BACTERIAL PREPARATION AND CULTURE 91
i. Preparation of heat-competent Escherichia coli 91
ii. Bacterial transformations 91
II. III. MOLECULAR BIOLOGY TECHNIQUES 93
i. Annealing of oligonucleotides 93
ii. Restriction digestions 93
iii. Ethanol precipitation of DNA 94
iv. Ligation of DNA fragments 95
v. Agarose gel electrophoresis 96
vi. Purification of DNA from an agarose gel 96
vii. Polymerase chain reaction (PCR) 96
viii. Generation of mutations/deletions by recombinant PCR (site-directed mutagenesis) 98
ix. Screening of colonies via colony PCR 98
x. Small-scale purification of vectors & plasmids from bacteria (miniprep) 99
xi. Large-scale purification of vectors & plasmids from bacteria (maxiprep) 99
xii. Purification of genomic DNA from mammalian cells 100
xiii. DNA concentration quantification 100
xiv. Sequencing of DNA 100
xv. Cloning of allele PCR fragments for sequencing (pGEM-T cloning) 101

xvi. Purification of RNA from mammalian cells 102
xvii. Reverse transcription of RNA to synthesise cDNA 102
xviii. Quantitative reverse transcription PCR (qRT-PCR) 103
II. IV. MAMMALIAN CELL CULTURE 106
iii. Culturing of adherent cell lines 106
iv. Culturing of suspension cell lines 106
v. Calcium phosphate-mediated transfection of HEK293 cells 106
vi. Nucleofection of K562 cells 107
vii. Fluorescence-activated cell sorting of mammalian cells (FACS) 108
viii. MTT assays 108
ix. Cell cycle analysis via propidium iodide (PI) 109
x. Cryopreservation of mammalian cells 110
xi. Resuscitation of mammalian cells after cryopreservation 110
xii. Treatment of K562 cells with inhibitory compounds: emetine, NMDI14, MG-132 110
xiii. Fixation and nuclear staining of mammalian cells 111
II. V. BIOCHEMICAL TECHNIQUES 112
xiv. Acquisition of protein extracts from mammalian cells 112
i. Purification of histones from mammalian cells 112
xv. Protein concentration determination 112
xvi. Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) 113
xvii. Western blotting 114
II. VI. BIOINFORMATICS 115
i. Design of sgRNAs 115
II. QRT-PCR PRIMER DESIGN 115
iii. Splice site analysis 115
iv. RNA sequencing data analysis 115
v. Gene Ontology analysis 116
vi. Protein structure prediction 116

CHAPTER III - RESULTS 118

III. I STRUCTURE PREDICTION OF MOZ C-TERMINAL REGIONS 118
III. II DESIGN AND CLONING OF SGRNAS TO TARGET EXONS 17/18 127
III. III GENERATION OF KAT6A VARIANT CELL LINES 136
III. IV VALIDATION OF MUTANT CLONES 142
III. V CELL CYCLE PROGRESSION AND NUMBER OF VIABLE METABOLICALLY ACTIVE CELLS IS AFFECTED BY KAT6A PTCS 154
III. VI C-TERMINALLY TRUNCATED MOZ PROTEIN DOES NOT HAVE ABERRANT CELLULAR LOCALISATION PROPERTIES 161

III. VII TARGET GENE EXPRESSION IS ALTERED IN CELLS WITH KAT6A GDD-LIKE MUTATIONS 165
III. VIII TRUNCATED MOZ PROTEIN HAS REDUCED GENE TRANSCRIPTION REGULATION ABILITIES IN A HEK293 MOZ KO CELL LINE 174
III. IX INDIVIDUALS WITH KAT6A GDD DISPLAY ABNORMAL GENE EXPRESSION ... 179
III. X GENE ONTOLOGIES REVEAL POTENTIAL FOR DISRUPTED DEVELOPMENTAL PROCESSES AND TRANSCRIPTIONAL CONTROL 188
III. XI NONSENSE-MEDIATED DECAY AS A CONTRIBUTORY MECHANISM IN KAT6A GDD 194
CHAPTER V - DISCUSSION 205
V. I. PROTEIN STRUCTURE PREDICTION 205
V. II. CRISPR-MEDIATED GENERATION OF MODEL CELL LINES 208
V. III. VIABILITY & CELL CYCLE PROGRESSION 211
V. IV. MOZ LOCALISATION 213
V. V. MOZ TARGET GENE ANALYSIS 215
V. VI. NMD OF KAT6A MRNA TRANSCRIPTS 225
V. VII. ORIGINS OF KAT6A GDD-ASSOCIATED VARIANTS & THERAPY POTENTIAL . 227
CHAPTER VI - CONCLUDING REMARKS AND FUTURE DIRECTIONS 235
BIBLIOGRAPHY 244
APPENDIX I - KAT6A VARIANTS ASSOCIATED WITH KAT6A GDD 288
APPENDIX II - VECTORS & PLASMIDS 327
APPENDIX III - OLIGONUCLEOTIDES 328
APPENDIX IV – ANTIBODIES 338
APPENDIX V - CELL CULTURE MEDIA & SERUM 339
APPENDIX VII – DEGS SHARED BETWEEN INDIVIDUALS WITH KAT6A GDD, HEK293 MOZ KO AND K562 MOZ SIRNA CELLS 340
APPENDIX VIII - PROFESSIONAL INTERNSHIP FOR PHD STUDENTS (REFLECTION) 345

List of figures

Figure 1: KAT6A GDD-associated variants in MOZ 25
Figure 2: SBBYSS and GPS-associated variants in MORF 29
Figure 3: Overview of the association between and regulation of transcription, mRNA maturation & translation 34
Figure 4: Overview of NMD-mediated degradation of mRNA.....42
Figure 5: Domain conservation between MOZ orthologues 53
Figure 6: Arrangement of protein domains within the human MOZ primary sequence.....56
Figure 7: Post-translational modifications of the MOZ protein.....58
Figure 8: Molecular functions of MOZ protein domains.....61
Figure 9: MOZ protein interaction partners 63
Figure 10: MOZ regulates the expression of target genes 66
Figure 11: MOZ expression promotes cell cycle progression.....68
Figure 12: Biological roles of MOZ during development.....71
Figure 13: MOZ fusion proteins arising from chromosomal translocations in cancers 75
Figure 14: Shared primary amino acid identity between the paralogues MOZ & MORF 119
Figure 15: Phyre2 secondary structure & disorder prediction 124
Figure 16: Secondary structure prediction using RaptorX.....126
Figure 17: CRISPR-Cas9 and Cas9n-mediated genome editing 128
Figure 18: Selection of sgRNAs for targeting KAT6A exons 17/18 ...129
Figure 19: Map of the CRISPR-Cas9n vector pX461.....131
Figure 20: Workflow to generate KAT6A variant cell lines using CRISPR-Cas9n 132
Figure 21: Sequencing results of the KAT6A exon 17 or 18-targeting sgRNAs cloned into the pX461 vector. 133
Figure 22: sgRNA-targeted loci within KAT6A exons 17/18 & design of DNA repair template.....134
Figure 23: FACS of K562 cells transfected with CRISPR-Cas9n plasmids encoding KAT6A exon 17-targetting sgRNAs. 137

Figure 24: FACS of K562 cells transfected with CRISPR-Cas9n plasmids encoding KAT6A exon 18-targetting sgRNAs. 138
Figure 25: Assessment of genome editing in bulk FACS populations 140
Figure 26: Ser1039* mutation in the bulk FACS population 141
Figure 27: Screening of K562 single cell-derived colonies.....143
Figure 28: Workflow of T-A cloning using the pGEM-T vector 145
Figure 29: KAT6A allele screening of K562 clones E1 (17), E9 (17), E1 (18), H9 (18) & G4.....146
Figure 30: Alignment of KAT6A alleles in K562 KAT6A exon 17 mutant clones E1 and E9 148
Figure 31: Alignment of KAT6A alleles in K562 KAT6A exon 18 mutant clones E1 and H9.....149
Figure 32: MOZ protein expression in K562 KAT6A mutants.....151
Figure 33: K562 KAT6A exon 17 and 18 variants generated by CRISPR-Cas9 genome editing.....152
Figure 34: Predicted MOZ proteins expressed in our K562 KAT6A mutant cell lines.....153
Figure 35: Principle of MTT assays 155
Figure 36: Total number of viable metabolically active cells is affected in K562 cells harbouring KAT6A exon 17/18 PTC variants 157
Figure 37: Flow cytometry analysis of PI staining of K562 KAT6A mutant clones.....159
Figure 38: Cell cycle progression is affected in K562 cells harbouring KAT6A exon 17/18 PTC variants.....160
Figure 39: Regions within the MOZ protein computationally predicted to be nuclear localisation sequences.....162
Figure 40: Cellular localisation of C-terminally truncated MOZ proteins in HEK293 MOZ KO cells.....164
Figure 41: qRT-PCR amplification dissociation curves 166
Figure 42: qRT-PCR product assessment and GAPDH expression..167
Figure 43: MOZ target gene expression in K562 cells harbouring PTCs in KAT6A exons 17 or 18.....169

Figure 44: Validation of MOZ target genes identified by RNA-seq 170
Figure 45: Expression of RUNX1 target genes in K562 cells harbouring PTCs in KAT6A exons 17 or 18 172
Figure 46: HOX gene expression in K562 cells harbouring PTCs in KAT6A exons 17 or 18 173
Figure 47: Sequencing results of YFP-FLAG-MOZ Arg1024* and Arg1129* variants 175
Figure 48: FACS of MOZ KO HEK293 cells transfected with plasmids encoding YFP-FLAG-MOZ WT, Arg1024* or Arg1129* 177
Figure 49: Ability of C-terminally truncated MOZ protein to regulate target gene expression in HEK293 MOZ KO cells 178
Figure 50: DEGs in individuals with KAT6A GDD syndrome 181
Figure 51: DEGs shared between MOZ KO, KD and C-terminal PTCs 184
Figure 52: Gene Ontology terms for DEGs identified in KAT6A GDD RNA-seq data 189
Figure 53: Total KAT6A mRNA in our K562 CRISPR-Cas9-generated cell lines 195
Figure 54: qRT-PCR KAT6A allele-specific primer designs for individual alleles in KAT6A exon 17 mutant clones 196
Figure 55: qRT-PCR KAT6A allele-specific primer designs for individual alleles in KAT6A exon 18 mutant clones 197
Figure 56: Amplification using KAT6A allele-specific primers 198
Figure 57: KAT6A pre- and spliced mRNA levels expressed from variant alleles in our K562 CRISPR-Cas9-generated cell lines 200
Figure 58: Abundance of KAT6A exon 17/18 variant reads in KAT6A GDD RNA-seq data 201
Figure 59: Inhibition of NMD partially restores KAT6A mRNA levels and MOZ target gene expression 203
Figure 60: Cell cycle progression is affected in KAT6A KO cells 212
Figure 61: Nonsense KAT6A variants associated with KAT6A GDD are located at CpG sites 229

Figure 62: Abundance of 5-methylcytosine at KAT6A GDD-associated CpG sites in the normal human population 231
Figure 63: Different mechanisms of pathogenicity predicted for KAT6A GDD-associated PTC variants.....236
Figure 64: Proposed model of molecular pathogenesis for KAT6A GDD-associated PTC variants.....238

List of tables

Table 1: Reagents and composition of oligonucleotide phosphorylation mixtures 93
Table 2: Reagents used and composition of restriction digestion reactions of vectors, plasmids and PCR products 94
Table 3: Reagents and composition of ligation reactions of vectors with linear DNA fragments 95
Table 4: Reagents and composition of typical PCR reactions using Taq DNA polymerase, along with typical thermocycler settings 97
Table 5: Reagents and composition of typical PCR reactions using Phusion DNA polymerase, along with typical thermocycler settings 97
Table 6: Composition of Dpn1 restriction digestion enzyme reactions for site-directed mutagenesis 98
Table 7: The volume and concentration of vectors, plasmids, PCR products and primers required for sequencing reactions. 101
Table 8: Composition of ligation reactions for pGEM-T cloning of separate alleles from PCR products 101
Table 9: Composition of genomic DNA removal reactions from purified RNA 102
Table 10: Composition of reverse transcription reactions for cDNA synthesis 103
Table 11: Composition of master mix for qRT-PCR reactions 104
Table 12: Typical quantitative thermocycler setting for qRT-PCR experiments 104
Table 13: Composition of transfection mixtures for calcium phosphate-mediated transfection of HEK293 cells 107
Table 14: Reagents and composition of SDS-PAGE gels (resolving section) 113
Table 15: Reagents and composition of SDS-PAGE gels (stacking section) 113
Table 16: Amino acid identity shared between vertebrates at MOZ loci associated with KAT6A GDD syndrome 120

Table 17: Details of mono- and bipartite nuclear localisation sequences with the full-length WT MOZ sequence, as predicted by computational prediction tools 163
Table 18: Affected genes in both KAT6A GDD patients and our K562 CRISPR-Cas9-generated cell lines 183
Table 19: HOX gene expression in KAT6A GDD patients 186
Table 20: Differentially expressed genes of interest in KAT6A GDD patients 187
Table 21: KAT6A target genes are associated with key biological processes 190
Table 22: Differentially expressed genes associated with cell proliferation and neural processes in KAT6A GDD patients 191
Table 23: Differentially expressed FOX and Zinc Finger Proteins in KAT6A GDD patients 193

List of equations

Equation 1: Calculation of the necessary mass of linear DNA insert to be used in a ligation reaction with a vector 95
Equation 2: Calculation used to analyse the relative gene expression of qRT-PCR targets.....104

List of abbreviations

AKT	RAC-alpha serine/threonine-protein kinase
AML	Acute myeloid leukaemia
APS	Ammonium persulphate
BRPF	Bromodomain and PHD finger-containing protein
Cas9	CRISPR-associated protein 9
Cas9n	CRISPR-associated protein 9 nickase
CBP	CREB-binding protein
CDKN2A	Cyclin-dependent kinase inhibitor 2A
CenH3	Centromere-specific histone H3
CHD	Congenital heart disease
CpG	Cytosine-p-guanine
CRISPR	Clustered regularly interspaced palindromic repeats
CRISPR-Cas9	Clustered regularly interspaced palindromic repeats CRISPR-associated protein 9
crRNA	CRISPR-RNA
DGS	DiGeorge syndrome
DMP1	Dentin matrix protein 1
DPF	Double PHD finger domain
DSPP	Dentin sialophosphoprotein
EAF6	Esa-associated factor 6
EJC	Exon junction complex
Enok	Enoki mushroom
ERα	Estrogen receptor α
FACS	Fluorescence-activated cell sorting
FGF	Fibroblast growth factor
Gcn5	General control non-derepressible 5
GDD	Global developmental delay
gDNA	Genomic DNA
GNAT	Gcn5-related N-acetyltransferase
GPS	Genitopatellar syndrome
HAM	Histone acetyltransferase of the MYST family
HAT	Histone acetyltransferase

HBO1	HAT bound to ORC1
NCOA	Nuclear receptor coactivator
HDAC	Histone deacetyltransferase
HDR	Homology-directed repair pathway
hMOF	Human males-absent on the first
HOX	Homeobox protein
HRP	Horseradish peroxidase
HSCs	Haematopoietic stem cells
ING	Inhibitor of growth
KAT6A	Lysine acetyltransferase 6A
KAT6B	Lysine acetyltransferase 6B
LEUTX	Leucine-twenty homeobox
lncRNA	Long non-coding RNA
LPS	Lipopolysaccharide
M4/M5	M4 or M5 subtypes of AML according to the French-American-British (FAB) classification system
Meis1	Homeobox Protein Myeloid Ecotropic Viral Integration Site 1
mESCs	Mouse embryonic stem cells
miR-(number)	MicroRNA-(number)
miRNA	MicroRNA
MLL	Mixed-lineage leukaemia protein
MOF	Males-absent on the first
MORF	Monocytic leukaemia zinc finger protein-related factor
MOZ	Monocytic leukaemia zinc finger protein
MPO	Myeloperoxidase
MRD32	Mental retardation autosomal dominant 32
mRNA	Messenger ribonucleic acid
MYST	MOZ, Ybf2, Sas2 and Tip60 domain
MYST3	MOZ, YBF2/SAS3, SAS2 and TIP60 protein 3
NCOA2	Nuclear receptor coactivator 2
NCOA3	Nuclear receptor coactivator 3
NEMM	N-terminal region of ENOK, MOZ or MORF domain

NHEJ	Non-homologous end joining pathway
NLS	Nuclear localisation signal/sequence
NMD	Nonsense-mediated decay
p21	Cyclin-dependent kinase inhibitor 1
p300	Histone acetyltransferase p300
p53	Cellular tumour antigen p53
PAM	Protospacer adjacent motif
PCAF	p300/CBP-associated factor
PHD	Plant homeodomain
PML	Promyelocytic leukaemia
PRMT	Protein arginine methyltransferase
PSIS	Pituitary stalk interruption syndrome
PTC	Premature termination codon
PTM	Post-translational modifications
PU.1	PU-box transcription factor 1
PWWP	Proline-tryptophan-tryptophan-proline domain
RNA-seq	RNA-sequencing
RUNX	Runt-related transcription factor
RuvC	Crossover junction endodeoxyribonuclease RuvC
Sas2	Something about silencing 2
Sas3	Something about silencing 3
SBBYSS	Say-Barber-Biesecker-Young-Simpson syndrome
sgRNA	Single guide ribonucleic acid
snRNA	Small nuclear RNA
TALEN	Transcription activator-like effector nucleases
TBP	TATA-binding protein
TBX	T-box protein
TIF2	Transcriptional intermediary factor 2
TIP60	60 kDa Tat-interactive protein
trRNA	Trans-activating crRNA
TSAP	Thermosensitive alkaline phosphatase
TSS	Transcription start site
U6	RNA polymerase III promoter

VSD	Ventricular septal defect
WDR5	WD repeat-containing protein 5
WES	Whole exome sequencing
ZFN	Zinc finger nuclease

CHAPTER I

INTRODUCTION

Chapter I - Introduction

I. I. Global developmental delay, intellectual disabilities & d*e novo* mutations

Global developmental delay (GDD) occurs in 1-3% of children, with up to 25% of cases able to be explained by gene copy number variations (sometimes X-linked) (Tham *et al.*, 2015; McRae *et al.*, 2017). This is a broad term which is often used to refer to an overall delayed developmental progression, whereby an individual does not meet a developmental stage within the medically accepted timeframe. An example would be the expectation of a child to be able to sit or walk unaided. Often associated with GDD is intellectual disability (ID) (Kaufman, Ayub and Vincent, 2010). An intelligence quotient (IQ) of 70 or below is classified as intellectual disability (ID) and is thought to affect approximately 1% of children across the globe (Kaufman, Ayub and Vincent, 2010; Tham *et al.*, 2015). 60% of these individuals do not have a known cause of ID, although it has been suggested that between 25 and 50% of all cases of ID have a genetic element (Kaufman, Ayub and Vincent, 2010).

Sporadic human genetic diseases which are associated with a reduced reproductive fitness, have been demonstrated to be caused by *de novo* pathogenic gene variants (Veltman and Brunner, 2012). Increasing parental age at conception is linked to an increased likelihood of pathogenic variant occurrence (Wright *et al.*, 2015; McRae *et al.*, 2017). Additional factors such as how related the affected individual's parents are, their gender and also whether there is any pre-existing family history of developmental disorders also influences this likelihood (Wright *et al.*, 2015). In one study, 6% of affected individuals whose parents were consanguineous with respect to being first cousins or closer relations had a predicted pathogenic de novo mutation, demonstrating that increased parental relatedness could contribute to their prevalence (Wright *et al.*, 2015). Diagnostic odyssey is a term often used to describe the lengthy ordeal and time it takes for individuals with developmental delay or

intellectual disability to get a definitive diagnosis (Kim *et al.*, 2017). Whole exome sequencing (WES) greatly speeds up this process, allowing suitable rehabilitation plans to be implemented quicker, improving health of affected individuals, reducing the resources required of healthcare systems and comforting relatives (Veltman and Brunner, 2012; Johansen Taber, Dickinson and Wilson, 2014; Retterer *et al.*, 2016).

Mutations in several genes associated with epigenetic regulation of gene expression have been linked to intellectual disability syndromes, e.g. KAT6A, KAT6B and BRPF1 (Arboleda *et al.*, 2015; Gannon *et al.*, 2015; Mattioli *et al.*, 2017). Additionally, genes known to be involved in histone methylation or ubiquitination have been linked to the development of congenital heart disease when de novo mutations are present (Zaidi *et al.*, 2013). The following sections detail the KAT6A & KAT6B variants reported in the literature as a result of newer sequencing technologies, lower costs and bioinformatics. Clinical details for the associated syndromes are also provided.

i. KAT6A

The KAT6A gene encodes the histone acetyltransferase Monocytic Leukaemia Zinc Finger Protein (MOZ). This protein is expressed in a range of tissues and has been demonstrated to be crucial during embryogenesis, for stem cell maintenance, haematopoietic system development and HOX gene expression (Miller, Maves and Kimmel, 2004; Crump *et al.*, 2006; Katsumoto *et al.*, 2006; Voss *et al.*, 2009; Cho, Kim and Jang, 2016; Sheikh *et al.*, 2016). Within the last 4 years, *de novo*, heterozygous pathogenic variants in the KAT6A gene have been reported in the literature (Arboleda *et al.*, 2015; Tham *et al.*, 2015; Elenius *et al.*, 2017; Gauthier-Vasserot *et al.*, 2017; Zwaveling-Soonawala *et al.*, 2018). Individuals harbouring these variants display features of a global developmental delay syndrome, termed KAT6A GDD, otherwise known as mental retardation autosomal dominant 32 (MRD32). Clinical features include global developmental delay, dysmorphic craniofacial features, microcephaly, speech delay, cardiac

problems such as ventral septal defects (VSDs), ocular anomalies, feeding difficulties and seizures. Phenotype facial features of this syndrome include a large forehead, thin upper lip and broad nasal bridge, however many of the clinical features reported for KAT6A GDD are non-specific to this condition. The mean age of reaching developmental milestones for first social smile, sitting unaided, walking independently and ability to speak were reported to be 2.25, 12, 22 and 36 months, respectively (Wright *et al.*, 2015). This reveals a clear global developmental delay. It is thought that due to the non-specific clinical features observed in individuals harbouring pathogenic KAT6A variants, disorders associated with this gene are like to be under-diagnosed (Arboleda *et al.*, 2015). A study of 7,000 patients with various forms of neurodevelopmental disorders, reported 11 individuals in their cohort who possessed *de novo* KAT6A variants where are predicted to be pathogenic (0.157%), suggesting that this may be the frequency of KAT6A GDD cases in the broad category of neurodevelopmental disorders (McRae *et al.*, 2017).

In addition to the above features of this syndrome, several affected individuals have been found to have growth hormone deficiencies (Zwaveling-Soonawala *et al.*, 2017; Trinh *et al.*, 2018). For one patient, MRI scans revealed pituitary gland abnormalities, leading to the diagnosis of pituitary stalk interruption syndrome (PSIS) (Zwaveling-Soonawala *et al.*, 2017). A single individual with childhood apraxia of speech (CAS), a difficulty in verbally communicating, was found to be heterozygous for a KAT6A variant similar to those associated with the KAT6A GDD syndrome (Eising *et al.*, 2018). Unexpectedly, this patient was not reported to have other features of the syndrome. This does however indicate that MOZ may have a role in the development and formation of coherent speech abilities in humans. Two patients with KAT6A GDD have been reported to have recurrent intermittent neutropenia, which is the presence of abnormally few neutrophils in the blood (less than or equal to 1500 per cubic mm), leading to an increased susceptibility to infection (Manroe *et al.*, 1979; Tham *et al.*, 2015;

Gauthier-Vasserot *et al.*, 2017) (Appendix I - KAT6A variants associated with KAT6A GDD). Interestingly, mutations in the gene encoding the RUNX1 transcription factor which MOZ is known to physically interact with, have been associated with this condition (Skokowa *et al.*, 2014). This suggests impaired associations between these two proteins due to a lack of the MOZ C-terminal regions, is responsible for this clinical feature.

All the reported pathogenic KAT6A variants are *de novo* in nature, with the exception of one which has been shown to be inherited (Trinh *et al.*, 2018). Interestingly, Satoh et al. (2017) reported unaffected parents who did not harbour KAT6A variants, but had three male children who all harboured the same mutation (c.3070C>T, Arg1024*) and presented with the KAT6A GDD syndrome. This suggests that parental germline mosaicism was the origin of this pathogenic variant. The majority of pathogenic KAT6A variants are nonsense or frameshift mutations located within exons 17 (penultimate) or 18 (terminal) of the KAT6A gene, affecting the acidic glutamate/aspartate-rich region of the MOZ protein (Figure 1). Three variants do not conform to these loci: one is a splice site disruption, another is a missense mutation in the MYST domain and the third is a nonsense mutation within the first coding exon (exon 3) (Appendix I - KAT6A variants associated with KAT6A GDD). For the majority of KAT6A GDD-associated variants, a loss of 29-52% of the protein from the C-terminal end is predicted (Arboleda *et al.*, 2015). This includes the region of MOZ reported to have a transactivation function (Kitabayashi *et al.*, 2001; Miller, Maves and Kimmel, 2004). Assessment of histone acetylation patterns in patient-derived cell lines from two patients with KAT6A GDD, revealed a reduction in H3K9 acetylation in conjunction with an increase in H3K18 acetylation (Arboleda *et al.*, 2015). In addition to this, RNA-seq data from the same individuals suggested alterations in the expression of a third of the p53-dependent genes analysed, indicative of disrupted p53 signalling in this syndrome (Arboleda *et al.*, 2015).

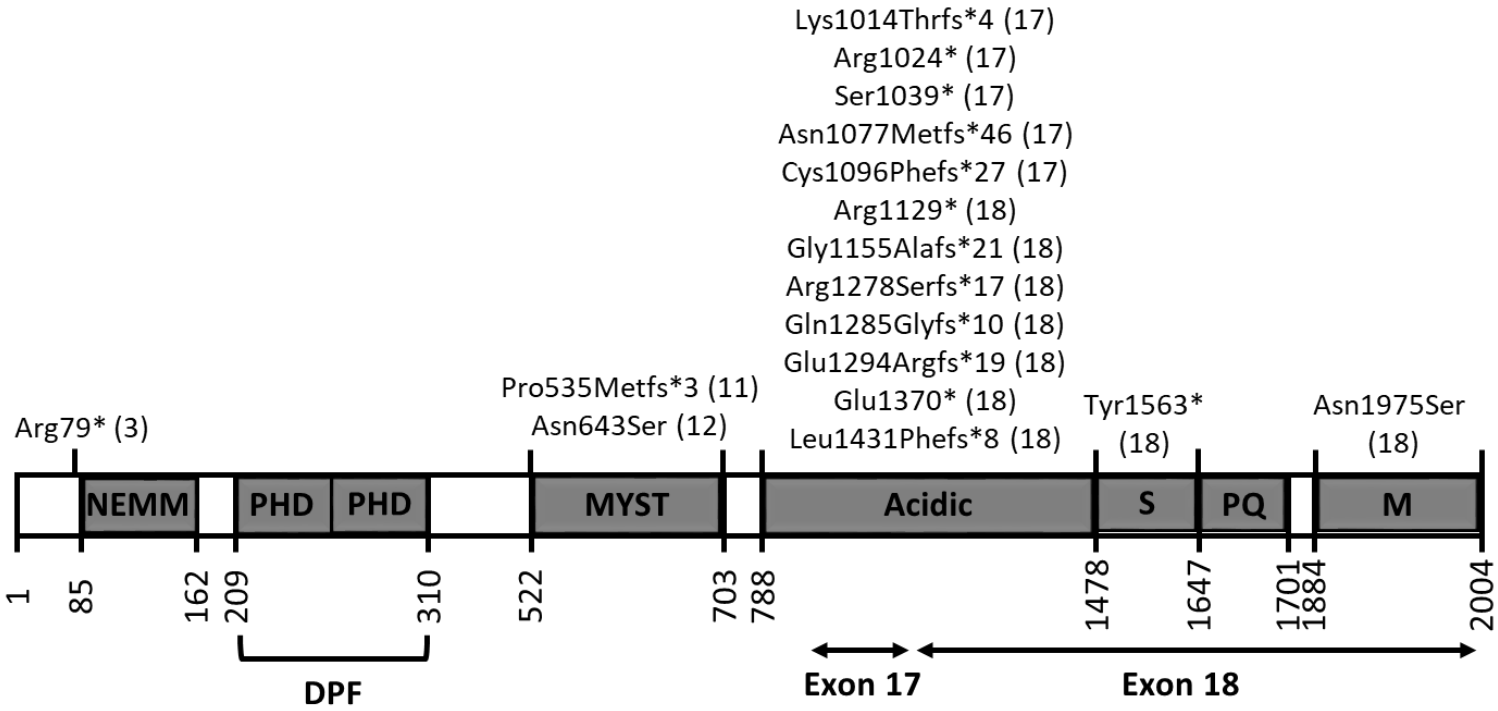


Figure 1: KAT6A GDD-associated variants in MOZ

Predicted MOZ protein mutations determined from KAT6A gene variants reported in the scientific literature, associated with the KAT6A Global Developmental Delay syndrome. Protein variants are distributed according to which domain they reside in; their corresponding exon location is noted in brackets. Note the clustering of variants within exons 17 and 18. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2 (PHD1/2). MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. Acidic residue-rich (Acidic), serine-rich (S), glutamic acid-rich (E), proline and glutamine-rich (PQ) and methionine-rich (M) regions are also indicated.

To date, the molecular mechanism(s) which link the identified genetic variants to the phenotype associated with KAT6A GDD syndrome, have not been determined. In one study, the authors discounted the role of the RNA-degradation mechanism known as nonsense-mediated mRNA decay (NMD), which is known to degrade mRNA transcripts that encode premature termination codons (PTCs, which the majority of KAT6A GDD-associated mutations are predicted to result in) (Arboleda *et al.*, 2015). NMD activity reduces the overall abundance of the targeted mRNA transcripts, thereby decreasing translation of the corresponding protein. However, there was some variability in their results and dermal fibroblasts from only two patients were examined. Since the level of NMD activity is known to differ between tissue types, cells other than fibroblasts may provide a more definitive, or differing conclusion (Zetoune *et al.*, 2008). In addition to the potential for NMD activity upon KAT6A PTC-encoding mRNA transcripts, another mechanism which could contribute to this syndrome is a loss of the amount of functional MOZ protein translated (haploinsufficiency, dose effect). This mechanism would mean a lower amount of functional MOZ protein is available to regulate target genes and acetylate its substrates. Alternatively, if truncated protein is translated from mRNAs encoding PTCs, this protein may have dominant-negative effects such as binding to the promoters of MOZ target genes, but being unable to regulate their expression due to the lack of specific protein interaction domains or correct tertiary structure, thereby blocking the binding of functional WT MOZ protein also present in the same cells.

Since MOZ is known to have crucial roles in the regulation of senescence, haematopoietic stem cell maintenance, correct B- and T-cell lineage specification and embryonic development, any of the mechanisms mentioned above may perturb these processes (Miller, Maves and Kimmel, 2004; Perez-Campo *et al.*, 2009; Rokudai *et al.*, 2013; Sheikh, Lee, *et al.*, 2015; Sheikh, Phipson, *et al.*, 2015; Cho, Kim and Jang, 2016; Newman *et al.*, 2016; Sheikh *et al.*, 2017). It is the one of the aims of this thesis to further elucidate the molecular mechanisms which contribute to the development of the KAT6A GDD syndrome. It may be that a single molecular mechanism is primarily responsible for the development of the KAT6A GDD syndrome, or alternatively more than one of the following proposed mechanisms contribute:

1. NMD-mediated degradation of KAT6A mRNA transcripts which encode variants that lead to the formation of in-frame PTCs.
2. MOZ haploinsufficiency due to the reduced abundance (dosage) of functional MOZ protein.
3. Translation of dominant-negative truncated MOZ polypeptides.

ii. KAT6B

Blepharophimosis-mental retardation syndromes is a group of global developmental delay syndromes also known collectively as Ohdo syndrome, with varying features. In 2006, this group was divided into more specific sub-types, of which Say-Barber-Biesecker-Young-Simpson syndrome (SBBYSS) is one (Lundsgaard *et al.*, 2017). Clinical features include a mask-like facial appearance, small mouth, small, low-set ears, broad nasal bridge, bulbous nasal tip, severe blepharophimosis and ptosis, as well as thyroid abnormalities, hypoplastic teeth and patellae and congenital heart defects. Global developmental delay and intellectual disability, as well as cleft palate and genital abnormalities are also present (Yu *et al.*, 2014; Gannon *et al.*, 2015; Li *et al.*, 2017; Niida *et al.*, 2017). Another syndrome, Genitopatellar syndrome, has many clinically overlapping features with SBBYSS, however individuals with GPS appear to present with a more severe phenotype, with additional features such as agenesis of the corpus callosum, genital abnormalities, renal cysts and absent patella (Gannon *et al.*, 2015; Okano *et al.*, 2018). There has, however been debate about whether these two syndromes are actually just different presentations of the same medical condition on a variable spectrum (Vlckova *et al.*, 2015). Additionally, there is another syndrome which has many overlapping features of SBBYSS, one known as Lin-Gettig syndrome (Lin and Gettig, 1990; Hedera and Innis, 2002; Bashir *et al.*, 2017). All three of these syndromes have recently been associated with heterozygous KAT6B pathogenic variants, suggesting that these syndromes are indeed part of the same clinical spectrum (Gannon *et al.*, 2015; Bashir *et al.*, 2017; Lonardo *et al.*, 2019).

The KAT6B gene encodes the MYST histone acetyltransferase MOZ-Related Factor (MORF) and is located on chromosome 10q22.2. MORF is a

paralogue of MOZ which also acetylates histones and is known to regulate gene transcription. All the pathogenic variants identified as being part of the KAT6B syndrome spectrum are *de novo,* with the exception of a single case of an inherited pathogenic KAT6B variant (Kim *et al.*, 2017). The majority of these KAT6B variants reported in the literature are nonsense or frameshift mutations, however an in-frame deletion, a missense mutation and a splice site variant have all been documented (Gannon *et al.*, 2015; Yilmaz *et al.*, 2015). The majority of these pathogenic variants are located within the penultimate or terminal exon of this gene, corresponding to the C-terminal regions of the MORF protein (Figure 2). This has led to the hypothesis that due to protein truncations, mutant MORF proteins lack important interaction sites, such as the one necessary to interact with the RUNX1/RUNX2 transcription factors (Campeau *et al.*, 2012; Okano *et al.*, 2018). This loss of function would therefore negatively affect transcriptional activation of RUNX-bound genes reliant upon MORF association and acetylation. However, as mentioned above in the case of MOZ, there may be other molecular mechanisms, or interplay between more than one which is/are responsible for pathogenesis.

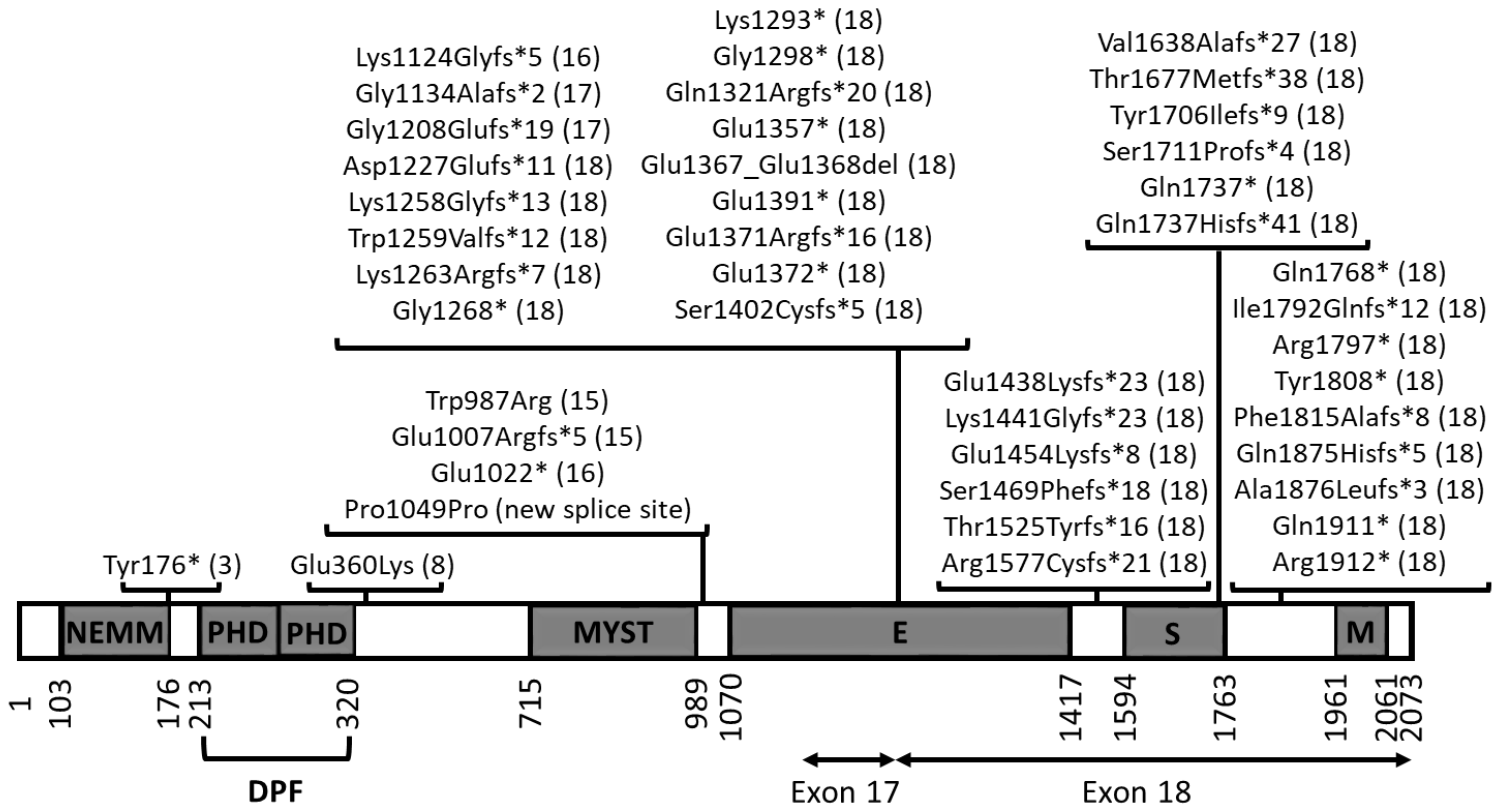


Figure 2: SBBYSS and GPS-associated variants in MORF

Predicted MORF protein mutations determined from KAT6B gene variants reported in the scientific literature, associated with the Say-Barber-Biesecker-Young-Simpson or Genitopatellar syndromes. Protein variants are distributed according to which domain they reside in; their corresponding exon location is noted in brackets. Note the clustering of variants within exons 17 and 18. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2 (PHD). MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. Glutamic acid-rich (E), serine-rich (S) and methionine-rich (M) regions are also indicated.

I. II. Chromatin structure & histones

Approximately 2 metres of linear DNA are contained within the nucleus of a diploid human cell (Kamakaka and Biggins, 2005). Compacting this amount of genetic information into such a space is no simple task, however eukaryotes have specialised proteins known as histones which function to condense the genome. A classic example of their function is in the packaging of DNA into structures we have come to recognise as chromatids and chromosomes during mitosis and meiosis (Kamakaka and Biggins, 2005). In eukaryotic cells, the accessibility of DNA is dependent on the packing density of chromatin fibres, which in turn influences its transcription. Transcriptionally active chromatin is highly decondensed and extends away from compact interphase nuclear regions; this form of chromatin is referred to as euchromatin (Eitoku *et al.*, 2008). Alternatively, heterochromatin is tightly packaged DNA encoding genes which are currently transcriptionally inactive and this conformation of DNA is found within the nuclear periphery (Eitoku *et al.*, 2008). It has been determined that half the protein mass within nuclei is actually histone proteins, indicating their importance and involvement in the nuclear architecture of eukaryotic cells and gene expression regulation (Kamakaka and Biggins, 2005). The canonical histones have been found to be expressed at a higher level during the S phase of the cell cycle, which is linked to their fundamental role in DNA compaction after replication during cell proliferation and division (Szenker, Ray-Gallet and Almouzni, 2011).

Nucleosomes are the building blocks of packaged chromatin and consist of a histone octamer with 145-147 bp of DNA wrapped around this histone core (Kamakaka and Biggins, 2005). The histone core is composed of two molecules each of histone H2A, H2B, H3 and H4. The wrapping of DNA around a histone core to form a nucleosome is not the only histone-DNA interaction involved in the compaction of genetic material; another histone protein known as histone H1, referred to as a linker histone, associates with nucleosomes at the region where the packaged DNA enters and exits the nucleosome structure, thereby acting

as an anchor for the DNA strands (Kamakaka and Biggins, 2005). This arrangement of DNA wound around nucleosomes is often termed as “beads on a string”, due to their appearance when visualised (Woodcock and Ghosh, 2010). The nucleosome (“beads”) are separated from each other by approximately 20-80 nucleotides (Woodcock and Ghosh, 2010). These “beads” then form the basis for further DNA compaction via inter-nucleosome interactions, which produces 30 nm diameter structures (Kamakaka and Biggins, 2005). The 30 nm chromatin fibres can then further be condensed to form a 100-400 nm diameter fibre and in metaphase, this compaction continues until DNA is compacted 10,000-20,000-fold (Peterson and Laniel, 2004; Woodcock and Ghosh, 2010).

Histone proteins have a globular structure consisting of a central domain that mediates interactions with other core histones, however each histone also has an N-terminal “tail” that is unstructured and consists of 20–35 amino acid residues which protrude out from the nucleosome (Song *et al.*, 2014). These N-terminal protrusions are rich in basic amino acids, such as lysine. The acetylation of lysine residues in histone N-terminal tails was first documented in 1963 and was shown to positively regulate gene transcription (Xiong *et al.*, 2016). More recently, additional acylations such as butyrylation, propionylation and crotonylation have been identified on histones by mass spectrometry (Xiong *et al.*, 2016). Efforts are now being made to establish the functional implications of these modifications. Aside from acetylation, other common post-translational modifications (PTMs) of histones include methylation (mono-, di- and tri-), phosphorylation, ubiquitination, sumoylation, ADP-ribosylation, citrullination and biotinylation (Jufvas, Strålfors and Vener, 2011).

While the majority of histone cores in nucleosomes consist of canonical components, variants of all four histones have been found to exist in nature, with some identified as having specific functions, associations with particular disease states, or even tissue-specific expression (Witt, Albig and Doenecke, 1996; Kamakaka and Biggins, 2005; Lepack *et al.*, 2016). For example, the H2A.X variant of H2A is

known to be involved in DNA repair upon exposure to ionising radiation, due to its phosphorylation (Rogakou *et al.*, 1998; Celeste *et al.*, 2002). Another H2A variant, H2A.Z is thought to be involved in the transcriptional activation of genes by preventing the binding and activity of transcriptional repressors at bound loci (Meneghini, Wu and Madhani, 2003). Like H2A.X, this variant is also involved in DNA repair signalling and is known to regulate cell progression through cell cycle checkpoints (Kamakaka and Biggins, 2005). The H3.3 variant of histone H3 is associated with regions of DNA either being actively transcribed, or poised to be and has been identified as having a higher prevalence at gene promoters (Szenker, Ray-Gallet and Almouzni, 2011). Another H3 variant called CenH3 (centromeric H3), is associated with centromeres and the ability for sister chromatid segregation (Szenker, Ray-Gallet and Almouzni, 2011). Together, histone variants have crucial roles in maintaining genome stability & regulating transcription.

I. III. Regulation of gene expression

Gene expression is a complex and intricate process, crucial for all organisms. This is demonstrated by mutations in components of the transcriptional machinery, such as transcription factors that result in disease states (Lee and Young, 2013). The process of transcription is the generation of a temporary and modifiable copy of a genomic region, which can then be utilised for a functional purpose, such as protein translation (Lewin, 2008). This allows for the processing and further regulation of genetic information and how it is expressed, for example different messenger RNA (mRNA) isoforms encoding proteins with distinct molecular functions, all from the same gene (Lewin, 2008). Therefore, transcription of DNA enhances the genetic information it encodes. Transcription itself is part of the vast regulatory network of gene and protein expression; other mechanisms of regulation include RNA processing, RNA transport, RNA localisation, mRNA degradation, mRNA translation and post-translational modifications (Lewin, 2008). A summary of the key stages from gene transcription to translation are depicted in Figure 3. Furthermore, RNAs such as microRNAs (miRNAs) and non-coding RNAs are also thought to have roles in transcription and translation (Lewin, 2008).

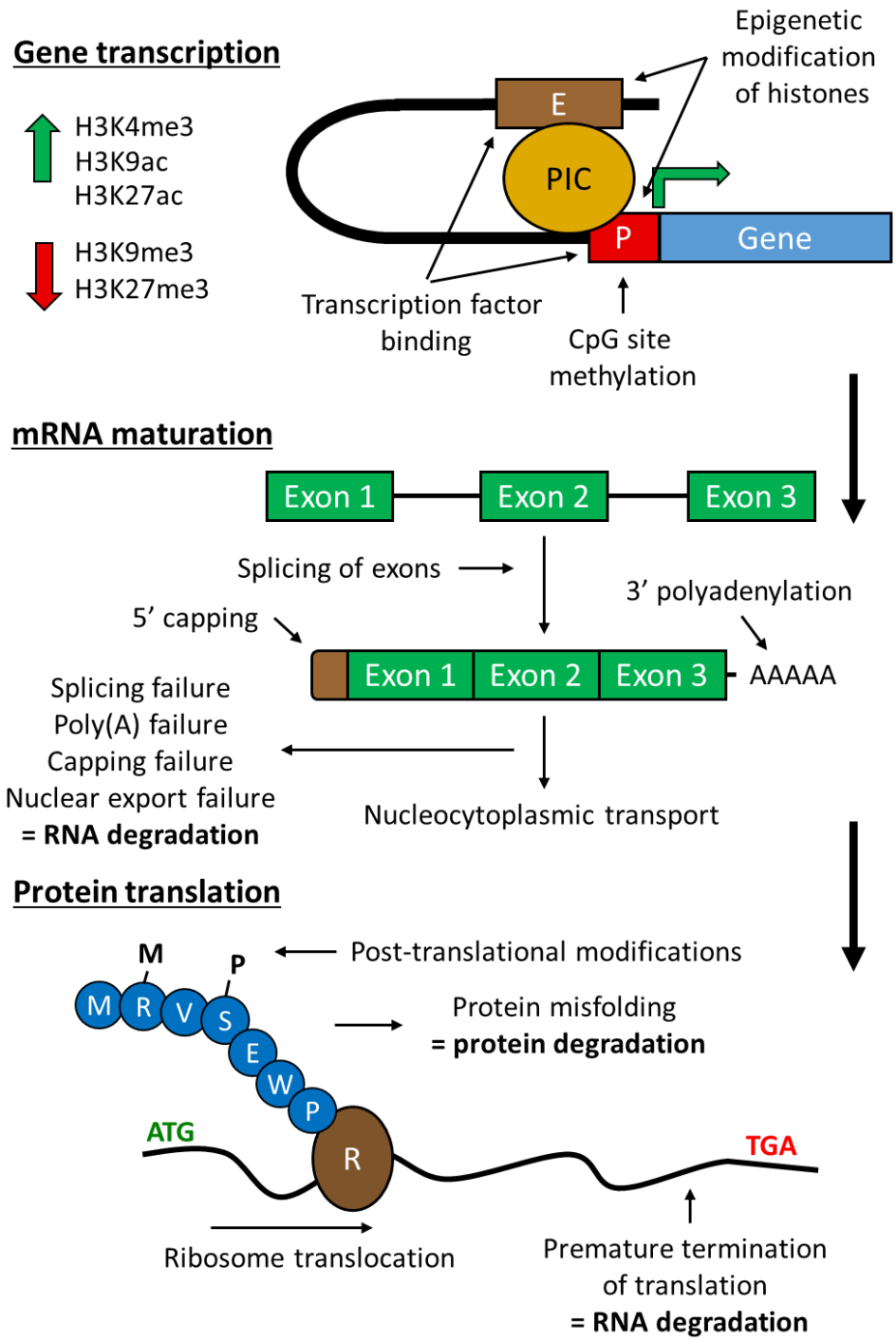


Figure 3: Overview of the association between and regulation of transcription, mRNA maturation & translation

Schematic detailing the basis and variety in regulation of gene transcription, mRNA processing and protein translation, such as histone acetylation, RNA degradation and protein post-translational modifications. Brown box "E" = enhancer, red box "P" = promoter, PIC = pre-initiation complex, black "M" & "P" = methylation / phosphorylation, ATG & TGA = start and stop codons respectively, "R" = ribosome.

i. Transcription

Promoters are located just 5' to the start site of a gene and usually possess conserved sequences such as an Initiator (Inr) region; Inr regions frequently overlap with known downstream promoter elements (DPEs) and the transcription start site (TSS) (Nechaev and Adelman, 2011). Enhancers also provide regulatory information to the transcription machinery, influencing transcription initiation. These genomic regions are often located far from the gene itself, sometimes thousands of nucleotides 5' or 3' to its regulatory target (Lewin, 2008). The juxtapositioning of enhancers and their respective promoters during transcriptional activation/repression is facilitated by protein-protein interactions of transcription factors (Lewin, 2008). The structure of proteins designated as transcription factors is important because this confers DNA-binding specificity (Frietze and Farnham, 2011). Common structures involved in protein-DNA interactions include zinc finger motifs, helix-turn-helix and helix-loop-helix domains (Frietze and Farnham, 2011). Some transcription factors are ubiquitously expressed in all tissues, whereas others are tissue- or cell-specific and hence are deemed to have specific functional roles (Lee and Young, 2013). Nuclear receptors are a type of transcription factor which can be activated by ligand-binding and thereby become activated by external stimuli (Brinkmann *et al.*, 1989; Heery *et al.*, 1997). However, the accessibility of DNA to transcription factors and transcriptional machinery, is influenced by nucleosome positioning and chromatin epigenetics (Lewin, 2008; Sainsbury, Bernecky and Cramer, 2015).

Post-translational modification of histone N-terminal tails can facilitate the condensation or opening of chromatin for transcription (Peterson and Laniel, 2004). The human genome encodes hundreds of chromatin regulators, which "read", "write" or "erase" epigenetic marks (Yan *et al.*, 2017). As one example, histone methylation and demethylation is a constant flux and there are dynamic interactions and cooperation between the enzymes which regulate this epigenetic code (Drazic *et al.*, 2016). H4R3me2 is an activation mark which serves to

allow the recruitment of HATs like CBP or p300, facilitating the opening up of chromatin to allow enhanced gene transcription (Cheung *et al.*, 2016). On the other hand, the demethylase KDM4C removes the repressive mark H3K9me3 allowing further modifications such as acetylation, which serve as a marker for the activation of transcription (Cheung et al. 2016).

Aside from histone PTMs marking the transcriptional state of genomic loci, clustering of CpG islands within DNA sequences also annotates promoter regions (Illingworth and Bird, 2009; Sen *et al.*, 2016). CpG islands are genomic regions rich in GC base composition and promoter regions of approximately 60-70% of human genes have a high concentration of these (Illingworth and Bird, 2009). They are thought to influence chromatin regulation and accessibility via the recruitment of CpG-binding proteins (Illingworth and Bird, 2009). Methylation of these sites can prevent / alter the affinity of transcription factor binding, thereby regulating gene transcription.

Once a gene promoter has been marked as “active” by histone PTMs such as H3K27ac, H3K9ac and H3K4me3, and changes in nucleosomal positioning have occurred, the DNA is transiently available for transcription machinery interaction (Consortium, 2012). The general transcription factors TFIIB, TFIID, TFIIE, TFIIF, TFIIG and RNA PolII assemble at the gene promoter and form the pre-initiation complex (PIC) (Plaschka *et al.*, 2015). Importantly, the TATA box-binding protein (TBP), which is a component of TFIID associates with an 8 nucleotide consensus sequence (TATAWAWR, W indicates A/T, R indicates A/G nucleotides) in the TATA box region of a promoter, situated 30 nucleotides 5’ to the TSS (Basehoar, Zanton and Pugh, 2004).

There are three stages of transcription, namely initiation, elongation and termination. During initiation, the RNA Pol II complex synthesises several short abortive RNA transcripts. Once RNA Pol II is phosphorylated in its C-terminal domain however by TFIIH, the Pol II complex then transcribes beyond the promoter region into the downstream gene (elongation) (Plaschka *et al.*, 2015). Eventually, the

RNA Pol II complex reaches a terminator region and both the polymerase and newly synthesised RNA transcript are released from the DNA (termination) (Lewin, 2008; Nechaev and Adelman, 2011). Successive rounds of transcription increase the abundance of RNA transcripts.

ii. mRNA splicing & nucleocytoplasmic transport

Once transcribed, mRNA sequences are processed from pre-mRNA to mature, spliced transcripts. The processing of pre-mRNA within the nucleus involves various steps, such as polyadenylation at the 3' end, the addition of a cap at the 5' end and splicing (Lykke-Andersen and Jensen, 2015). This RNA processing is facilitated by an assortment of proteins (Dreyfuss, Kim and Kataoka, 2002). In higher eukaryotes, protein coding information is contained within gene exons, whereas intervening non-coding sequences separating the exons are called introns. To generate mature mRNA, intronic sequences are removed from the primary transcript by protein complexes referred to as spliceosomes (Le Hir *et al.*, 2000). In addition, further processing events such as 5' capping of the transcript with 7-methylguanosine and polyadenylation at the 3' end are performed (Lykke-Andersen and Jensen, 2015) (Figure 3).

Once mRNA transcripts have been transcribed, undergone splicing, capping and polyadenylation, many are exported from the nucleus via a process called nucleocytoplasmic transport (Gama-Carvalho and Carmo-Fonseca, 2001; Köhler and Hurt, 2007). This movement of mRNA is mediated by specialised proteins known as importins and exportins (Köhler and Hurt, 2007). Chan et al. (2004) showed that eIF4A3 can bind directly to TAP and REF, both of which are mRNA export factors and components of EJCs. 3'-polyadenylation of mRNAs has been found to result in more efficient nuclear export of RNA transcripts, as has the association of mRNAs with heterogeneous nuclear ribonucleoproteins (hnRNPs) (Zhao, Hyman and Moore, 1999; Gama-Carvalho and Carmo-Fonseca, 2001). In addition, the presence of 5'-cap

structures on mRNAs results in increased rates of nuclear export of the respective transcripts (Gama-Carvalho and Carmo-Fonseca, 2001).

iii. Protein translation

Just like transcription, translation is a highly regulated process and adds another level to the control of gene expression. This process is influenced by the number of ribosomes present on an mRNA transcript, the stability of the mRNA, intrinsic features of an mRNA and also the rate of translation (reviewed by Sonenberg and Hinnebusch, 2009). All of these factors come together to control the amount of a given protein which is synthesised. The protein machine which drives translation is the ribosome. This is a peptidyl transferase complex consisting of two main subunits: the 40S component that positions mRNA codons, and the 60S component which assesses these codons for anticodon matching with tRNAs and catalyses the formation of covalent peptide bonds (reviewed by Sonenberg and Hinnebusch, 2009). The 40S subunit binds to mRNA and elongation initiation factors (eIFs), resulting in a translation initiation complex. Translation initiation is regulated by two different mechanisms: RNA binding proteins and microRNAs acting in an mRNA sequence-specific way, or through modification of the eIF proteins (reviewed by Jackson, Hellen and Pestova, 2010). For example, phosphorylation of eIF2 (a GTPase) at position Ser51, results in the sequestering of eIF2B and prevents GDP→GTP replacement, rendering eIF2 inactive (reviewed by Jackson, Hellen and Pestova, 2010). The GTPase activity of eIF2 is essential for the initiation of translation, therefore reduced levels of active eIF2 decrease the overall amount of translation within cells (Jiang and Wek, 2005). Kinases known to phosphorylate eIF2 at Ser51 are reported to be activated via cell stresses such as amino acid starvation and accumulation of misfolded proteins, implying both intracellular and extracellular factors can regulate translation (reviewed by Jackson, Hellen and Pestova, 2010). Alternatively, increased phosphorylation of eIF4E by MAP-kinases Mnk1 and Mnk2 enhance cap binding complex formation, which stimulates translation of the bound

mRNA transcript (Joshi *et al.*, 1995; Scheper *et al.*, 2002). Furthermore, the poly(A) tails of mRNAs also promote translation; poly(A) tails are bound by the poly A binding protein (PABP), which stabilises mRNA and also facilitates its translation (Gorgoni and Gray, 2004).

While the majority of mRNA transcripts are thought to undergo cap-dependent translation, i.e. the presence of a 5' cap is essential for translation initiation, some mRNAs are translated independent of this via the use of internal ribosome entry sites (IRES). IRES are cis-regulatory mRNA sequences which directly allow ribosomes to bind to the translation start codon, without the need for all the components of the cap-dependent translation initiation complex (reviewed by Belsham and Sonenberg, 1996 & Sonenberg and Hinnebusch, 2009). IRES are present in viral RNAs, as well as eukaryotic RNAs, where they allow the translation of essential proteins during episodes of cellular stress or during mitosis, when cap-dependent translation is less active (reviewed by Sonenberg and Hinnebusch, 2009). Aside from protein translation, protein degradation via the 26S proteasome regulates the quantity and quality of polypeptides; misfolded or aggregating proteins that are undesired are polyubiquitinated and targeted for degradation via this method (Hershko *et al.*, 1982; Schubert *et al.*, 2000; reviewed by von Mikecz, 2006). This allows for the targeted removal of particular proteins and the recycling of amino acids for new protein synthesis.

I. IV. Nonsense-mediated decay (NMD)

Spliced mRNAs accrue exon junction complexes (EJCs) at loci 25-50 nucleotides upstream of exon-exon splice junctions (Le Hir *et al.*, 2000). The formation of EJCs is indispensable for the recognition and degradation of RNA transcripts harbouring premature termination codons (PTCs) via nonsense-mediated decay (NMD) (Chang, Imam and Wilkinson, 2007). NMD occurs in the cytoplasm and is generally considered to be a useful surveillance mechanism whereby nonsense or frameshift mutations that create PTCs within coding regions are identified and degraded, thereby preventing translation of aberrant proteins with potentially pathogenic molecular functions (Chang, Imam and Wilkinson, 2007; Lykke-Andersen and Jensen, 2015; Kurosaki and Maquat, 2016). Contrastingly, Colombo *et al.* (2017) and Lou *et al.* (2016) identified a variety of different WT human RNA transcripts which are NMD targets, such as the non-coding RNA GAS5 and the ATF4 transcription factor, as well as members of the TGFβ and BMP signalling pathways. This indicates that NMD may have a role in preventing the expression of particular isoforms, therefore helping to maintain homeostatic gene expression or influence cell fate. Similarly, NMD has a crucial role in the degradation of mRNAs transcribed from immunoglobulin and T-cell receptor genes that have undergone unproductive genomic rearrangements (a natural occurrence), producing loci which encode undesired PTCs (Weischenfeldt *et al.*, 2008). Silencing of NMD in T-cells causes an increase in these mRNA transcripts and the induction of apoptosis (Weischenfeldt *et al.*, 2008). Furthermore, NMD is indispensable for normal development, with known roles in commissural neuron axon migration via the downregulation of Robo3.2 expression and in granulocyte differentiation by downregulation of genes such as Lmnb1 after intron retention (Colak *et al.*, 2013; Wong *et al.*, 2013).

Despite its important role in regulating gene expression, NMD activity has been linked to disease and developmental conditions (Usuki *et al.*, 2006; Linde *et al.*, 2007; Martin *et al.*, 2015; Lindeboom, Supek and Lehner, 2016; Ono *et al.*, 2018). Zetoune et al. (2008) reported that

NMD intrinsically has a variable efficacy between tissue types. Additionally, it has been found that NMD exhibits variable efficacy depending on the location of PTCs both within exons and the gene as a whole: generally, NMD efficiency increases as you move 5'-3' along mRNA transcripts after the first few exons, with most efficiency in the penultimate exon; PTCs within a terminal exon are not subjected to NMD due to the lack of downstream EJCs (Lykke-Andersen and Jensen, 2015; Lindeboom, Supek and Lehner, 2016). For every exon which is susceptible, only PTCs >55 nucleotides upstream from the exon boundary are able to trigger NMD (Keeling, 2016). Furthermore, it has been reported that any exons >400 nucleotides in length have reduced NMD efficiency (Lindeboom, Supek and Lehner, 2016).

If a ribosome stalls during translation due to the presence of a PTC, prior to removing all EJCs associated with an mRNA transcript, the RNA is degraded by NMD. ATP-dependent RNA helicase up frameshift 1 (UPF1) and its paralogues UPF2 and UPF3 are essential for the initiation of NMD and recruitment of NMD protein effectors (Lykke-Andersen and Jensen, 2015; Jia *et al.*, 2017). Ribosomal stalling at a PTC and the interaction of eukaryotic release factor 3 (eRF3) with UPF1, facilitates the interaction of UPF1-UPF2-UPF3 at the next downstream EJC, stimulates SMG1-mediated phosphorylation of UPF1 and then consequential recruitment of mRNA degradation machinery (Schoenberg and Maquat, 2012; Lykke-Andersen and Jensen, 2015) (Figure 4). Endonucleolytic cleavage by SMG6, recruitment of the general decapping complex by SMG5, deadenylation by the CCR4-NOT complex and 5'-3' degradation by the XRN1 exonuclease combine to result in the complete degradation of mRNA transcripts via NMD (Lykke-Andersen and Jensen, 2015). A summary of the main effectors involved in the NMD pathway is shown in Figure 4.

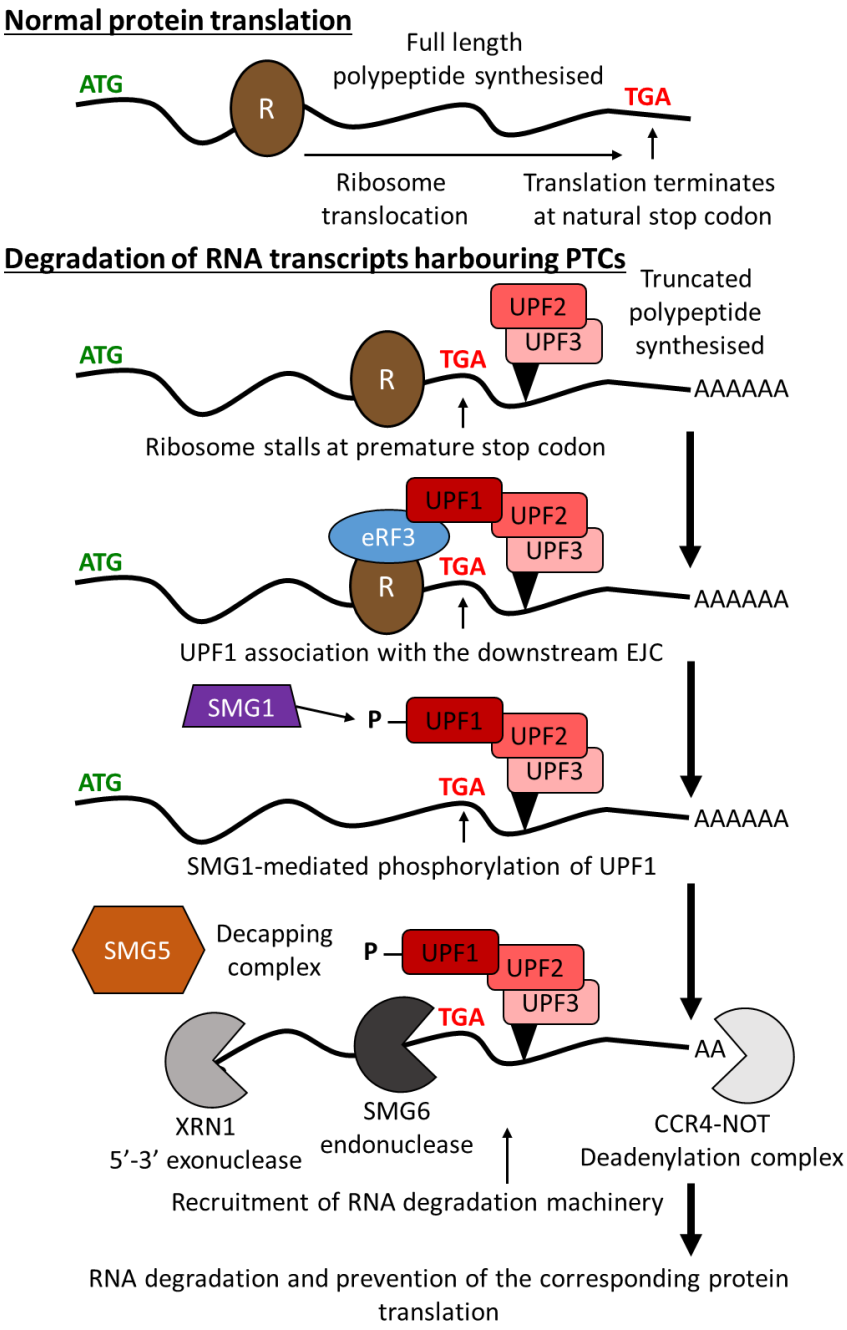


Figure 4: Overview of NMD-mediated degradation of mRNA

Schematic depicting the stalling of a ribosome at a PTC within an mRNA transcript, in contrast to normal protein translation. This is followed by recruitment and phosphorylation of UPF1 via interaction with eRF3 and SMG1. Consequentially, mRNA degradation occurs due to recruitment of RNA degradation machinery components such as the SMG5 decapping complex, CCR4-NOT deadenylation complex and endo- / exonucleases. ATG & TGA = start and stop codons respectively, black triangle represents an exon-junction complex which has been deposited at an exon-exon boundary after mRNA splicing, "R" = ribosome.

I. V. Post-translational modifications

The reversible chemical modification of proteins after translation is referred to as post-translational modification, which was first described for histones in 1964 (Allfrey, Faulkner and Mirsky, 1964). A wide variety of post-translational modifications (PTMs) exist in nature, some more prevalent than others and they are regularly associated with affecting the folding, stability, function or binding of the modified protein (Turner and O'Neill, 1995; Kouzarides, 2007; Nishi, Hashimoto and Panchenko, 2011; Dreveny *et al.*, 2014). In terms of histones, PTMs affect the recruitment of transcriptional regulators, transcription factors and epigenetic modifiers. They also serve to alter the association of nucleosomal DNA with the histone core, affecting chromatin architecture and compaction. This leads to further regulation of chromatin processes such as gene transcription, DNA replication and DNA repair. It is generally accepted that histone PTMs are dynamic in nature and often are not exclusive, with different modifications being present on histone tails together. Proteins which possess domains that can "read" these epigenetic marks are subsequently influenced in their activity and/or binding at that locus, therefore PTMs affect and are affected by a complex interplay of factors (Kouzarides, 2007). Many protein PTMs are now known to exist and with the advancement of high-sensitivity mass spectrometry, there has recently been a range of novel histone PTMs identified (Xiong *et al.*, 2016). This is an area of research which is constantly growing. Here, some of the most common or best understood PTMs are discussed.

i. Acetylation

The acetylation of histone lysine residues is one of the most thoroughly investigated covalent modifications influencing gene expression in eukaryotic cells. Acetylation is a reversible modification where lysines undergo an acetylation-deacetylation switch governed by different physiological conditions, including the metabolic state of a cell (Kouzarides, 2007). This process is extremely dynamic and is strictly

regulated by histone acetyltransferases (HATs) and histone deacetylases (HDACs). Lysine acetylation increases the hydrophobicity of the modified lysine residue, as well as its size (Verdone *et al.*, 2006). This has the effect of reducing the ability of the amino acid to form hydrogen bonds and also prevents it from interacting with negatively charged molecules. Histone acetylation neutralises the positive charge of lysine residues, thereby weakening the interactions between histones and phosphate groups of DNA and facilitating the activation of transcriptional events (Verdone *et al.*, 2006; Kouzarides, 2007). Therefore, it has been found that histone acetylation increases the accessibility of DNA to the transcriptional machinery through the reduction of inter-nucleosomal contacts that prevent the compaction and folding of higher order chromatin structures (Verdone *et al.*, 2006). A genome-wide study demonstrated that all forms of histone acetylation are positively correlated with gene expression; the repression of gene expression by deacetylation is due to the resulting compaction of the chromatin structure, as well as the binding of repressor proteins that recognise non-acetylated histone residues (Shahbazian and Grunstein, 2007; Wang *et al.*, 2008). Another important function of histone acetylation patterns is in assembly of chromatin and histone turnover. Newly synthesised histones are transiently acetylated, a factor which is important for their recognition by chaperone proteins and their correct deposition into nucleosomes (Shahbazian and Grunstein, 2007).

ii. Other lysine acylations

The method of choice for identifying new PTMs is currently mass spectrometry. In the 21st century, eight different short-chain lysine acylations have been identified via this method. Many histone PTMs are classed as short chain lysine acylations, which are similar to lysine acetylation but differ in charge, hydrophobicity or hydrocarbon chain length (Simithy *et al.*, 2017). Recently discovered lysine acylations on histones include: propionylation (Kpr), butyrylation (Kbu), 2-hydroxyisobutyrylation (Khib), succinylation (Ksucc), malonylation (Kma),

glutarylation (Kglu), crotonylation (Kcr) & β-hydroxybutyrylation (Kbhb). These eight short-chain lysine acylations have been separated into three groups, depending on their chemical properties: hydrophobic (propionyl, butyryl, crotonyl), polar (B-hydroxybutyryl, 2-hydroxyisobutyryl) or acidic (malonyl, succinyl and glutaryl).

Currently, much functional information regarding these alternative lysine acylations is lacking, however there is some evidence that suggests they affect transcription in a manner similar to that of acetylation. For example, histone crotonylation has been associated with active transcription, much like histone acetylation (Zhang *et al.*, 2016). This PTM is enriched at the promoters and enhancers of active genes in human somatic cells. Crotonylation is conserved from yeast to humans and in human histones, 28 lysine crotonylation sites have been identified (Xiong *et al.*, 2016). The epigenetic enzyme p300 has both crotonyltransferase and acetyltransferase activities, however p300-catalysed histone crotonylation is reported to be a more potent transcriptional activator than histone acetylation (Sabari *et al.*, 2015). Furthermore, other epigenetic enzymes such as MOZ and DPF2 have been demonstrated to “read” histone crotonylation marks via their PHD finger domains, although the functional consequence of this recognition has not been yet been discovered (Xiong *et al.*, 2016). Interestingly, the relative intracellular concentrations of crotonyl-CoA and acetyl-CoA determine whether a histone lysine will be crotonylated or acetylated, suggesting a link to metabolism (Sabari *et al.*, 2015, 2016).

iii. Methylation

The methylation of histones is a reversible process which does not affect the charge of the methylated protein and is found to primarily occur on the side chains of lysine and arginine residues (Cheung and Lau, 2005). Methylation is a more complex PTM compared to acetylation, purely due to the possibility of methylated residues to be mono-, di- or tri-methylated. Specifically, lysines can be mono-, di- or trimethylated, but arginines can only be mono- or di-methylated. Histone

methyltransferases (HMTs) catalyse the transfer of methyl groups from S-adenosylmethionine, whereas histone demethylases perform the reverse reaction (Bannister and Kouzarides, 2005; Greer and Shi, 2012). Histone methylation has essential roles in a whole host of nuclear processes, as such transcriptional regulation, heterochromatin formation and X-chromosome inactivation (Martin and Zhang, 2005). The methylation of histone lysine or arginine residues can have either a repressive, or activation function, depending upon the specific residue modified (Bannister and Kouzarides, 2005). There are 24 methylation sites documented in histones; H3K4, H3K36 and H3K79 are known to recruit activator proteins, whereas H3K9, H3K27 and H4K20 are associated with transcriptional repression (Bannister and Kouzarides, 2005).

iv. Phosphorylation

Serine, threonine and tyrosine residues on histone tails of all the core nucleosome components can be phosphorylated, a modification which is conserved in eukaryotes (Houben *et al.*, 2007; Rossetto, Avvakumov and Côté, 2012). Phosphorylation is the reversible addition of a negatively charged phosphate group to a protein substrate by a kinase enzyme, whereas its removal is catalysed by phosphatases (Rossetto, Avvakumov and Côté, 2012). Histone phosphorylation has been found to increase the electrostatic ion-pairing and hydrogen bonding capability of histones and has been associated with many biological processes, such as transcriptional regulation, apoptosis, cell cycle progression and DNA repair (Loomis *et al.*, 2009; Rossetto, Avvakumov and Côté, 2012). In addition, it is suggested that the phosphorylation of serine and threonine residues in histone tails facilitates chromatin condensation during mitosis and meiosis (Johansen and Johansen, 2006; Sawicka and Seiser, 2012). The phosphorylation of chromatin in cooperation with other histone marks has been shown to directly influence gene expression, providing epigenetic histone signatures which affect the subsequent chromatin activity and

recruitment of activators or repressors (Rossetto, Avvakumov and Côté, 2012).

v. Ubiquitination

The process of protein ubiquitination is reversible, performed by specialised catalytic enzymes and proteases named ubiquitin activating enzymes, ubiquitin ligases and deubiquitinating enzymes (Komander, 2009; Sowa *et al.*, 2009). The ubiquitin modification itself occurs upon lysine residues and involves the addition of a polypeptide composed of 76 amino acids. Interestingly, this ubiquitin moiety itself contains seven lysine residues and can further be ubiquitinated, thereby acting as a ubiquitin substrate. This can lead to various mono- and polyubiquitinated structures, even upon a single non-ubiquitin protein substrate. For ubiquitination to occur, ubiquitin activating enzymes (termed E1) utilise ATP to append ubiquitin to an E2 conjugating enzyme, which then acts as a substrate for the transfer of ubiquitin to the desired protein via the catalytic activity of an E3 ligase (Komander, 2009). The histones H2A and H2B have been reported to be monoubiquitinated; H2AK119 monoubiquitination is associated with transcriptional repression, while H2BK120 monoubiquitination is associated with transcriptional activation (Zhang, 2003; Minsky *et al.*, 2008). Originally, ubiquitin protein tags were thought to primarily function as marks for rapid protein degradation via the proteasome, however now they have become associated with additional biological processes such as transcriptional control, DNA repair and endocytosis (Kouzarides, 2007; Komander, 2009).

I. VI. Histone acetyltransferases

The first histone acetyltransferases were identified in 1996 and included histone acetyltransferase 1 (Hat1), CREB-binding protein (CBP), p300/CBP-associated factor (PCAF) & general control non-derepressible 5 (Gcn5) (Kleff *et al.*, 1995; Bannister and Kouzarides, 1996; Borrow *et al.*, 1996; Brownell *et al.*, 1996). There are three major groups of mammalian histone acetyltransferases: the GCN5/PCAF family, p300/CBP family and MYST family (Yang, 2004). It has been found that many histone acetyltransferases function catalytically *in vivo* as part of large multiprotein complexes (Struhl, 1998; Grant *et al.*, 1999). The HAT domain of histone acetyltransferases consists of three main functional regions: a central region that has the ability to bind to acetyl-CoA and facilitate its catalysis to acetyl-amino acid + CoA, and N- and C-terminal regions flanking this, which are involved in histone binding (Yan *et al.*, 2000; Marmorstein and Zhou, 2014). The flanking regions vary between HATs; this has been suggested to confer the different binding specificities found amongst the HAT families (Yan *et al.*, 2000; Marmorstein and Zhou, 2014).

Functionally, histone acetyltransferase activity is generally associated with increased levels of histone acetylation and transcription of genes at acetylated loci (Struhl, 1998; Legube and Trouche, 2003). They are well known to facilitate HOX gene transcription and have roles in stem cell or differentiation transcriptional programmes (Voss and Thomas, 2009; Paggetti *et al.*, 2010; Mallo and Alonso, 2013; Seifert *et al.*, 2015; Cho, Kim and Jang, 2016). Indeed, they are essential for correct organismal development, with many of them being embryonic lethal (Howe, 2001; Kueh *et al.*, 2011; Gao *et al.*, 2014). Additionally, many have been associated with acetylation to signal for DNA repair, or the regulation of the cell cycle (Ogryzko *et al.*, 1996; Ikura *et al.*, 2000; Howe, 2001; Rokudai *et al.*, 2009; Shiota *et al.*, 2010; Selvi *et al.*, 2015; Huang, Abmayr and Workman, 2016; Meyer *et al.*, 2016).

In the MYST family, there are five human HATs. These are: HAT bound to ORC1 (HBO1), Monocytic leukaemia zinc finger protein (MOZ),

MOZ related factor (MORF), human males-absent on the first (hMOF) and Tat Interacting Protein 60kDa (TIP60). In yeast, there are three MYST family HATs and the first MYST HAT protein, Sas3 was identified in yeast in 1996 (Reifsnyder *et al.*, 1996). Plants also possess MYST family HATs, namely HAM1 and HAM2 who share 98% amino acid similarity (Peng *et al.*, 2018). The name MYST is derived from this family's 4 founding members: MOZ, Ybf2, Sas2, and Tip60. All MYST HATs, with the exception of Esa1, share a conserved MYST domain comprising of both an acetyl-CoA-binding motif and a C2HC-type zinc finger motif ($CXXCX_{12}HXXXC$) (Yang, 2015). This zinc finger is necessary for their HAT activity and interestingly, the MYST domain's catalytic activity is reliant upon autoacetylation of a lysine residue within itself (Takechi and Nakayama, 1999; Sun *et al.*, 2011; Yang *et al.*, 2012).

MYST HATs bind to scaffold proteins which allow active HAT complex assembly. One such example is BRPF1, a protein conserved from humans to *Caenorhabditis elegans* (Yang and Ullah, 2007). There are three paralogues, BRPF1/2/3, all of which share a PZP module and bromodomain, but vary in their HAT associations (Yan *et al.*, 2016; Lloyd and Glass, 2018). BRPF1 binds to the MYST HATs MOZ and MORF via its N-terminal motif; its C-terminal allows binding to Inhibitor of Growth 5 (ING5) and the human orthologue of yeast Esa1-Associated Factor 6 (EAF6) (Ullah *et al.*, 2008). This demonstrates that BRPF1 promotes complex formation by acting as a scaffold protein (Ullah *et al.*, 2008). Conversely, BRPF2 and BRPF3 prefer to form complexes with HBO1 instead of MOZ or MORF, despite their high sequence similarity to BRPF1 (Mishima *et al.*, 2011; Feng *et al.*, 2016).

For some MYST HATs, there appears to be a degree of functional redundancy when it comes to the acetylation of certain histone lysine residues. For example, MOZ is known to acetylate H3K9 and H3K14 in vitro, however other HATs appear to be able to compensate for loss of this HAT, as global acetylation levels of H3K9 and H3K14 do not decrease significantly in its absence (Ullah *et al.*, 2008; Voss *et al.*, 2009; Sheikh, Lee, *et al.*, 2015). This is consistent with the limited impact

observed when Sas3 is lost in yeast, although loss of Gcn5 along with disrupted HAT activity of Sas3 results in decreased acetylation of these two loci (Howe, 2001). Feller et al. (2015) reported that depletion of any of the histone acetyltransferases in *Drosophila*, with the exception of Enok, has a minor reduction in H3 or H4 acetylation levels of less than 10%. Contrastingly, depletion of Enok caused a reduction of approximately 35%, which correlates with H3K23ac (the epigenetic mark facilitated by Enok) being a highly abundant acetyl mark in flies (Feller *et al.*, 2015). This clearly demonstrates that for at least some HATs, there is a degree of functional overlap.

I. VII. MOZ domain structure & function

The human KAT6A gene is located on chromosome 8p11.2 and encodes the HAT known as MOZ. It consists of 2004 amino acids in 18 exons, of which the first two are non-protein coding and has a molecular weight of ~225 kDa.

i. Conservation

MOZ co-purifies with ING5, EAF6 and BRPF1, BRPF2 or BRPF3 in the same complex, resembling the transcription-activating yeast NuA3 complex (Ullah *et al.*, 2008). The MOZ orthologue Enok was identified in *Drosophila* and has been shown to form a complex with orthologues of BRPF1 (Br140), ING5 (Ing5) and EAF6 (Eaf6) (Scott, Lee and Luo, 2001). This indicates that MOZ/MOZ orthologue complexes are highly conserved in eukaryotes. Indeed, orthologues of MOZ form HAT complexes which consist of between 4 and 7 unmodified or modified histone reader domains, the number of which varies between yeast, flies and humans (Huang, Abmayr and Workman, 2016). The domains in question include PHD fingers, PWWP domains or Bromodomains and function to regulate the activity of the respective orthologue upon particular substrates.

Within vertebrates, orthologues of MOZ are well conserved at the level of their primary amino acid sequence. The N-terminal regions of MOZ which consist of the MYST, DPF and NEMM domains share a high sequence homology with *Drosophila* Enok, Zebrafish MOZ and Rat MOZ, its orthologues in these organisms (Yang, 2015) (Figure 5). In contrast, the acidic region and serine/methionine-rich domains are absent from Zebrafish and *Drosophila* (Figure 5). This suggests that the MOZ C-terminus varies between mammals and other vertebrates. The amino acid sequence of rat MOZ has 89% overall homology to that of human MOZ, with each characterised domain having homology of >75% (Ohta *et al.*, 2005). The most well studied domains and those responsible for MOZ's catalytic activity have the highest sequence homology between orthologues. For example, the NEMM, DPF and MYST domains have

89%, 98% and 100% homology in rats and humans (Ohta *et al.*, 2005) (Figure 5). The C-terminal regions of MOZ proteins are more divergent in their primary sequence in other organisms, compared to their N-terminal domains. Surprisingly, the orthologues of MOZ (HAM1/HAM2) in plants lack these C-terminal regions and the DPF domain completely, with a chromodomain replacing the NEMM domain (Figure 5); their MYST domains have ~70% amino acid similarity to human MOZ, revealing they are more divergent as a whole compared to vertebrates (Yang, 2004). However, this high homology within the MYST domain clearly suggests a requirement for specific residues to maintain its structure and function.

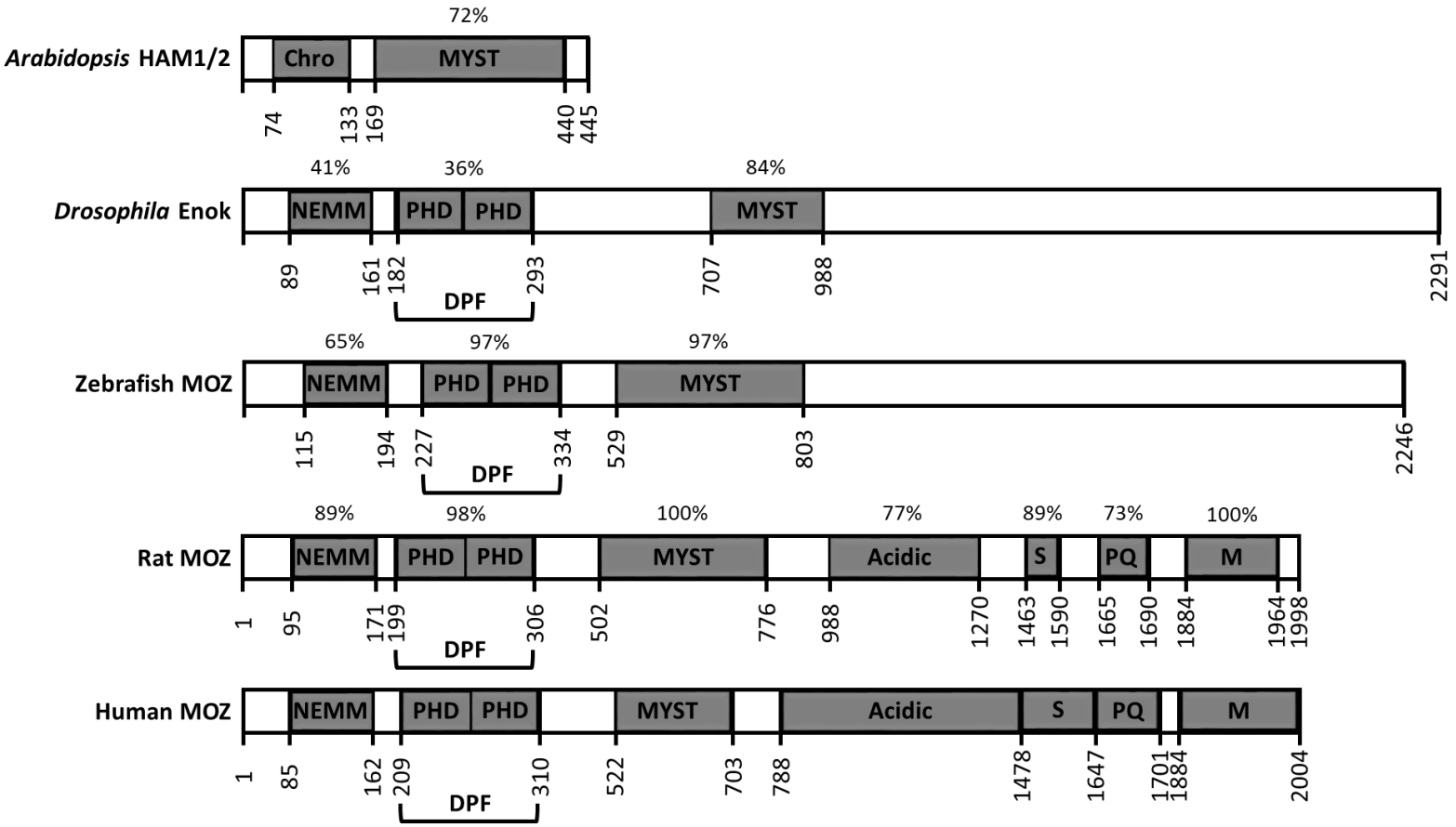


Figure 5: Domain conservation between MOZ orthologues

Schematic of MOZ orthologues showing the percentage of domain amino acid similarity, relative to human MOZ. Data acquired from UniProt. Chro = Chromodomain. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2 (PHD1/2). MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. Acidic residue-rich (Acidic), serine-rich (S), proline and glutamine-rich (PQ) and methionine-rich (M) regions are also indicated.

ii. Cellular localisation

The import of proteins to the nucleus is tightly controlled, with the class of proteins known as importins being responsible for amino acid sequence recognition and the subsequent trafficking of cargo into the nucleus. These amino acid sequences recognised by importins are referred to as nuclear localisation sequences (NLS) and can be classed as either mono-, or bipartite. Monopartite sequences are either a single cluster of at least four consecutive basic amino acids (R/K), or three which incorporate a fourth residue that can be any amino acid. Alternatively, a bipartite NLS is two basic amino acids, followed by what is referred to as a 10-12 amino acid linker, then at least 3 basic amino acids in the next five consecutive residues. Examples of monopartite NLS include that of the SV40 large T antigen (PKKKRKV), as well as c-Myc (PAAKRVKLD), whereas Nucleoplasmin harbours a bipartite NLS (KRPAATKKAGQAKKKK).

MOZ is a nuclear protein, which matches its predominant function as a transcriptional regulator and histone acetyltransferase. To date, no experimentally defined mono-, or bipartite NLS within MOZ have been identified. Localisation experiments have reported that this HAT localises within the nucleus in a characteristic manner as dot-like “speckles” (Kitabayashi *et al.*, 2001; Collins *et al.*, 2006). It has been demonstrated that a region within the N-terminal NEMM domain is sufficient for MOZ nuclear localisation, as deletion of amino acids 52 - 166 destroys the ability of MOZ to be predominantly nuclear, instead dispersing throughout both the nucleus and cytoplasm (Kitabayashi *et al.*, 2001). Despite this, it has also been shown that MOZ lacking the N-terminal 310 amino acids is still able to localise in the nucleus, suggesting that this region within the NEMM domain is not the only amino acid sequence to possess nuclear targeting properties (Kitabayashi *et al.*, 2001).

iii. DNA-binding ability

Holbert et al. (2007) reported that MOZ binds to DNA via its HAT domain, specifically making contacts in the helix-turn-helix and zinc finger

motifs. Helix-turn-helix motifs are DNA-binding motifs which are conserved through evolution and have been found to be present in both prokaryotes (repressor proteins) and eukaryotes (transcription factors) (Holbert *et al.*, 2007). Calculation of the electrostatic potential of the zinc finger and helix-turn-helix motifs in the MOZ HAT domain, suggests that the surface of these regions are positive, giving credibility to the notion that these regions form contacts with negatively charged DNA (Holbert *et al.*, 2007). The MOZ HAT domain has a high affinity for cruciform DNA, compared to that of Z-form DNA, single-stranded DNA and β-form double-stranded DNA. This has been suggested to be due to the characteristic of cruciform DNA to be more flexible than other forms, allowing it to come into contact with both the zinc finger and helix-turn-helix DNA-binding motifs (Holbert *et al.*, 2007). Mutation of particular residues located within the zinc finger (535-561) or helix-turn-helix motifs (710-735) of the MOZ HAT domain reduced the ability of this domain to bind to cruciform DNA, thereby confirming their importance in facilitating interactions between this HAT and DNA. Analysis of the DNA binding abilities of the HAT domains of Esa1, as well as GCN5-type family members revealed that the DNA binding characteristics of the MOZ HAT domain are not conserved between histone acetyltransferases (Holbert *et al.*, 2007). It cannot however be ruled out that members of the MYST family may possess other DNA-binding abilities, although possibly with different preferences and affinities to MOZ. The binding of MOZ to a specific DNA sequence has not been reported.

Both human and mouse MOZ appear to contain an extended AT-hook motif at the same position (although vary by a single amino acid). For human MOZ, this motif is located at residues 316-358, sandwiched between the DPF and MYST domains, with the sequence “GRKLLQKKAAQIKRRYTNPIGRPKNRLKKQNTVSKGPFSKVRT” (Filarsky *et al.*, 2015) (Figure 6). An AT-hook motif is a DNA-binding motif consisting of a GRP tripeptide sequence which is flanked by basic residues. Most eukaryotic proteins found to harbour an AT-hook motif are associated with chromatin regulatory functions such as histone

modification or chromatin remodelling (Filarsky *et al.*, 2015). AT-hook motifs have been suggested to act as anchors for enzymes which modify chromatin, binding them to AT-rich DNA sequences. AT-hook-like motifs (extended AT-hook motifs) sequences in which there are basic amino acids 12-15 residues away from the core GRP sequence on either side, making this motif approximately 3 times larger than that of the canonical AT-hook motif (Filarsky *et al.*, 2015). These extended motifs are known to have a stronger binding affinity for RNA, compared to DNA, however for MOZ this motif has not been investigated in detail (Filarsky *et al.*, 2015).

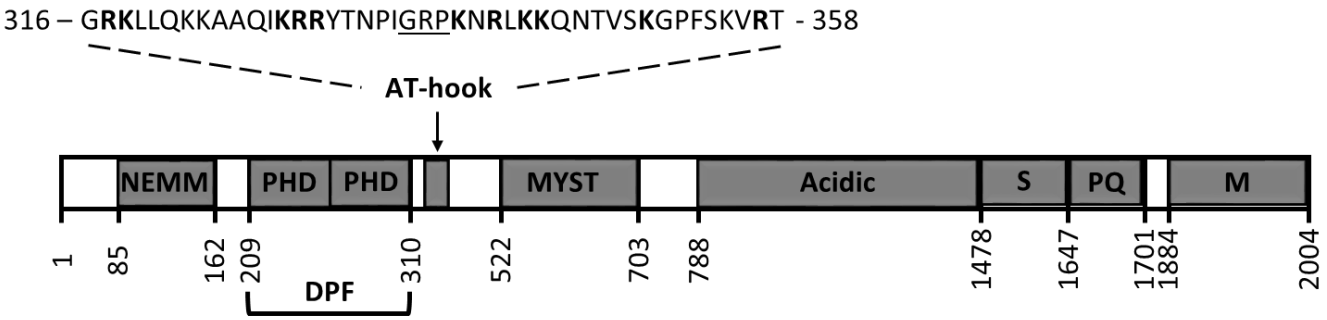


Figure 6: Arrangement of protein domains within the human MOZ primary sequence

Linear diagram of protein domains within human MOZ, relative to their amino acid number. Data acquired from UniProt. Note that the central GRP AT-hook tripeptide is underlined and surrounding basic amino acids are in bold. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2 (PHD1/2). MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. Acidic residue-rich (Acidic), serine-rich (S), proline and glutamine-rich (PQ) and methionine-rich (M) regions are also indicated.

iv. Catalytic activity

It has been demonstrated *in vitro* and determined *in vivo* that MOZ can catalyse the acetylation of histone H3K9 and H3K14, with some reports of H2A, H2B and H4 acetyltransferase activity also *in vitro* (Champagne, Pelletier and Yang, 2001; Holbert *et al.*, 2007; Voss *et al.*, 2009; Qiu *et al.*, 2012; Dreveny *et al.*, 2014). Holbert et al. (2007) demonstrated that the MOZ MYST domain can acetylate peptides

resembling 19 amino acids of the N-terminal tails of histones H3 and H4 *in vitro*. They also provided data indicative of the ability of the HAT domain of MOZ to bind to assembled nucleosomes, concluding that the DNA binding and acetylation activities of the MOZ HAT domain can occur simultaneously. It has been suggested that the binding of MOZ to DNA may enhance its acetyltransferase activity (Holbert *et al.*, 2007). Autoacetylation of MOZ is known to enhance its acetyltransferase activity, however little is known about the functional or structural consequences of other detected post-translational modifications on this HAT (Figure 7).

Shogren-Knaak et al. (2006) proposed that H4K16 acetylation helps to disrupt chromatin compaction, resulting in an alteration from heterochromatin which is not transcribed to euchromatin with transcriptionally active genes. MOZ is known to facilitate the acetylation of H4K16, which in conjunction with the finding that this HAT can bind to and acetylate nucleosomes just as well as free histones, led Holbert et al. (2007) to suggest that the HAT activity of MOZ upon H4K16 may result in chromatin decondensation from a condensed, transcriptionally inactive state and as a consequence, facilitate an increase in gene transcription.

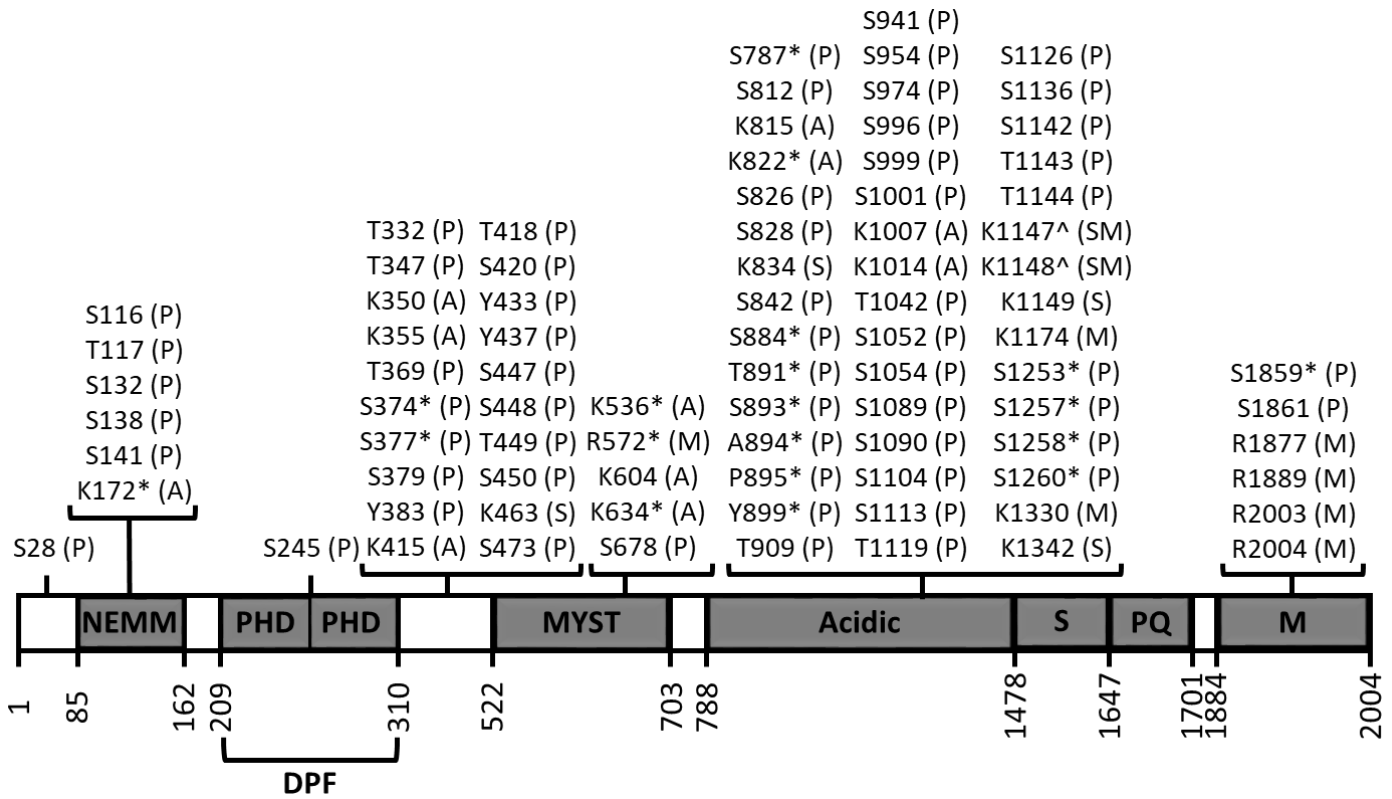


Figure 7: Post-translational modifications of the MOZ protein

Reported post-translational modifications of the human MOZ protein. * = expected for the human MOZ protein due to shared amino acid identity with Rat or Mouse MOZ with this PTM. A = acetylation, P = phosphorylation, M = methylation, S = Sumoylation. Data acquired from Peri et al. 2004; Hornbeck et al. 2015; Hendriks et al. 2017. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2 (PHD1/2). MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. Acidic residue-rich (Acidic), serine-rich (S), glutamic acid-rich (E), proline and glutamine-rich (PQ) and methionine-rich (M) regions are also indicated.

v. Recognition of histone marks

The double PHD finger domain (DPF) of MOZ is known to bind H3K14ac as well as H3R2 (unmodified), which results in recruitment of this HAT to gene promoters and facilitates their subsequent transcription (Qiu *et al.*, 2012; Dreveny *et al.*, 2014). MOZ itself can acetylate H3K14, thereby its recruitment to regions of the genome already harbouring H3K14ac marks by its own DPF, is suggestive of a mechanism which allows this epigenetic mark to propagate at particular genetic loci, perhaps to elicit a strong transcriptional response (Dreveny *et al.*, 2014). Contrastingly, H3K4me3 marks prevent the binding of the MOZ DPF domain to the N-terminal tail of histone H3 (Dreveny *et al.*, 2014).

Xiong et al. (2016) reported that the DPF domain of MOZ can bind H3K14cr, H3K14bu and H3K14pr, although the highest affinity for these acylations appears to be H3K14cr; the affinity of the MOZ DPF was found to be four times stronger for H3K14cr than that of H3K14ac (Xiong *et al.*, 2016). The reader pocket of the MOZ DPF domain is highly size-selective, with the planar alkene H3K14 crotonyl moiety fitting nicely into the binding pocket due to its flat and slim shape. This is in contrast to the H3K14bu group, which has a more bulky shape and as a result encounters steric clashes in the binding pocket, necessitating the flipping of the butyryl group in orientation to facilitate binding (Xiong *et al.*, 2016). H3K14pr can fit more easily than its butyryl counterpart, due to it having a side chain one carbon atom shorter. As a result, the DPF of MOZ has a two-fold higher binding preference for H3K14cr than H3K14bu or H3K14pr, with an enthalpy change 2.5-fold more for the crotonyl group and reduced steric repulsion (Xiong *et al.*, 2016). Of interest, crotonylation of H3K23 in conjunction with H3K14cr increases the affinity of the MOZ DPF to bind crotonylated peptides 1.6-fold, compared to that of H3K14cr alone (Xiong *et al.*, 2016). This suggests a mild additive effect of H3K23cr. However it is unknown whether the MOZ DPF can bind to H3K23cr itself, or perhaps the presence of H3K14cr and H3K23cr together results in a conformational change in the substrate, increasing its binding affinity.

MOZ target genes like HOXA5 and HOXA9 have higher levels of H3K14cr at their promoters than non-MOZ target genes such as HOXB1 and HOXD12 (Xiong *et al.*, 2016). MOZ mutants S210M and D282A are both unable to bind H3K14cr due to mutation of the binding pocket; these mutants were found to lack enrichment at the promoters of MOZ target genes, whereas wild-type MOZ was clearly enriched at these loci (Xiong *et al.*, 2016). These findings demonstrate that the DPF domain of MOZ is important for the ability of this HAT to recognise and bind H3K14cr marks, and this enables its localisation to specific gene promoters. Presumably, the localisation of MOZ to its target gene promoters, facilitated by enhanced levels of H3K14cr helps drive their transcription, however this is currently unknown. A summary of the known functions of each MOZ protein domain is depicted in Figure 8.

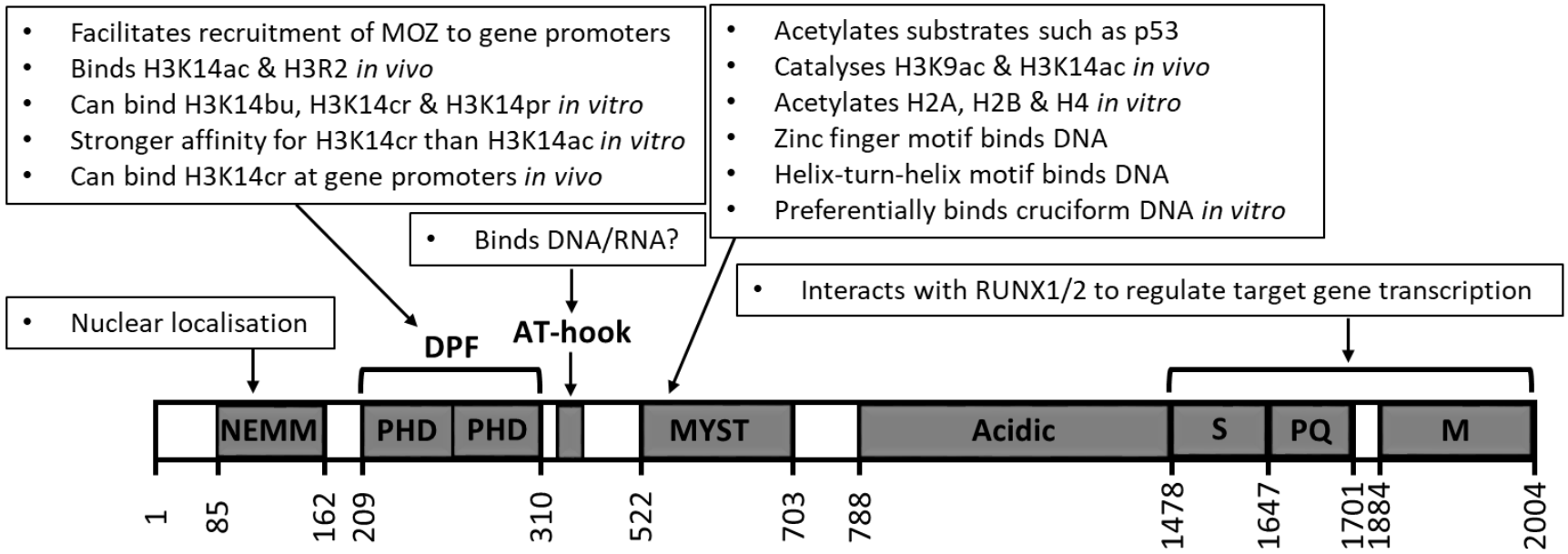


Figure 8: Molecular functions of MOZ protein domains

A summary of the known molecular functions of each domain present in the human MOZ protein, taken from published articles. Note that there is very limited knowledge regarding the function of the C-terminal domains and the AT-hook motif has not been experimentally verified. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2 (PHD1/2). MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. Acidic residue-rich (Acidic), serine-rich (S), proline and glutamine-rich (PQ) and methionine-rich (M) regions are also indicated.

vi. Protein interactions & transcriptional regulation of target genes

It has been reported that c-Jun, a basic region leucine zipper (bZIP) protein involved in cell proliferation and differentiation can interact with rat MOZ (Ohta *et al.*, 2005) (Figure 9). More specifically, it was found that the DNA-binding domain of c-Jun was required for this interaction, whereas the activation domain was disposable in this respect. Furthermore, similar analysis revealed an interaction between both rat MOZ and C/EBPα and C/EBPβ (Ohta *et al.*, 2005) (Figure 9). C/EBPα & β are CCAAT/enhancer binding proteins that also belong to the bZIP family of proteins. These findings imply that MOZ may be able to be recruited to gene targets by its association with bZIP proteins. MOZ can interact and cooperate with other epigenetic enzymes such as the histone methyltransferase MLL, to co-ordinate gene expression (Paggetti *et al.*, 2010). In addition, this HAT can bind to several epigenetic "readers" like BRPF1 and WDR5, which function to facilitate complex formation (Ullah *et al.*, 2008; Paggetti *et al.*, 2010) (Figure 9). These interactions and subsequent MOZ catalytic activity leads to alterations in gene transcription and they have been reported to be necessary for cell differentiation (Yang, 2015).

Pelletier et al. (2002) reported that the SM domain of MOZ can be immunoprecipitated with RUNX2, indicating an interaction between these two proteins (Figure 9). RUNX2 is a paralogue of RUNX1 and RUNX3, all of which are crucial transcription factors in body morphogenesis and differentiation (Inoue, Shiga and Ito, 2008; Chuang, Ito and Ito, 2013; Voon, Hor and Ito, 2015). RUNX1 is also thought to interact with MOZ due to sequence homology with RUNX2. It has been reported that the SM domain of MOZ is able to increase RUNX2 transcriptional activation, thereby the MOZ-RUNX interactions are thought to promote transcription of RUNX-bound genes (Pelletier *et al.*, 2002; Bristow and Shore, 2003). Furthermore, Heair et al. (2015) identified MOZ as an acetyltransferase of RUNX2, suggesting MOZ regulates the activity of this transcription factor. A summary of the known protein interactions with MOZ are depicted in Figure 9.

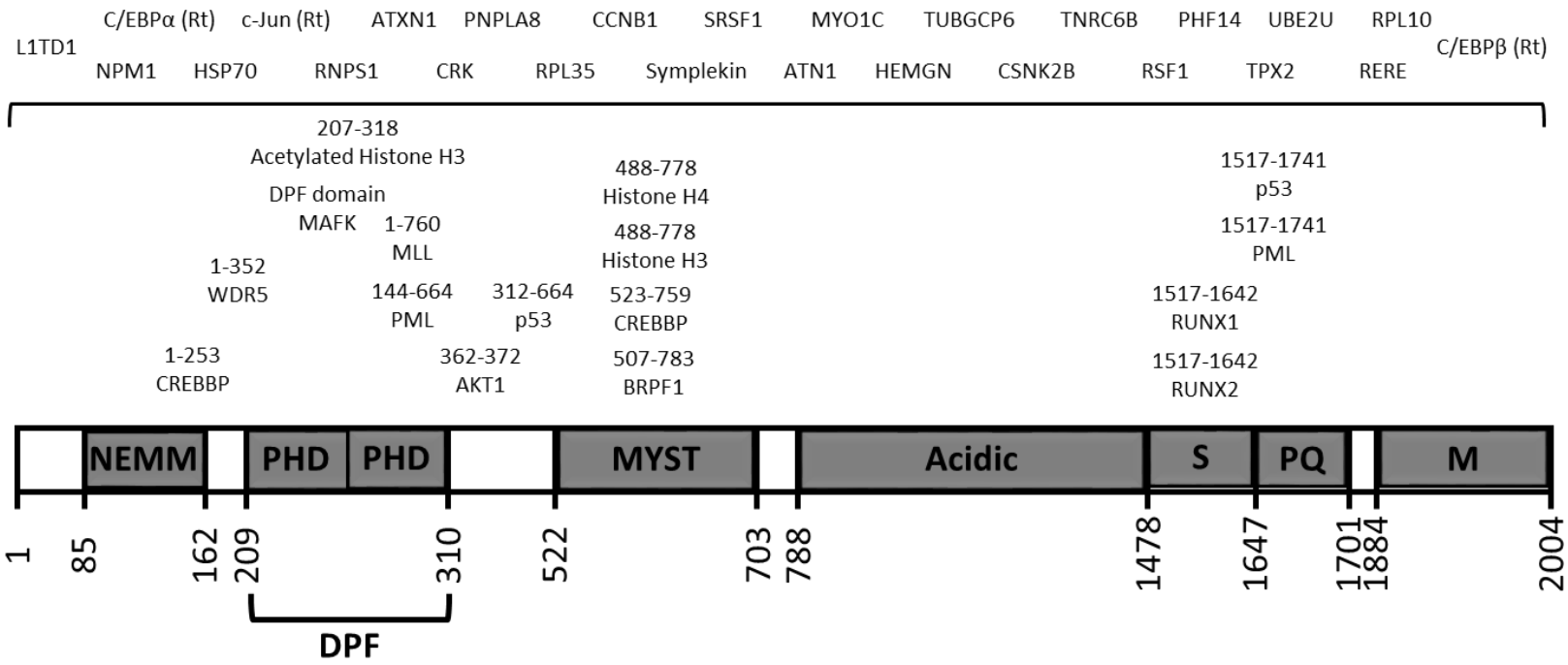


Figure 9: MOZ protein interaction partners

Schematic of proteins which are reported to interact with MOZ. Data acquired from UniProt, BioGRID and the respective publications. Some interactions were determined by proximity techniques, therefore may not be direct, rather both proteins are present in the same complex. All interactions are for the human MOZ protein, except where Rt is mentioned (Rat MOZ). Amino acid residues of MOZ which interact are shown if known. Some proteins have multiple binding regions and are therefore shown more than once. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2 (PHD1/2). MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. Acidic residue-rich (Acidic), serine-rich (S), glutamic acid-rich (E), proline and glutamine-rich (PQ) and methionine-rich (M) regions are also indicated.

MLL is a methyltransferase which catalyses H3K4 methylation, a histone mark that consequentially antagonises PRMT6-mediated H3R2 methylation; MLL activity upon H3K4 is associated with the inhibition of transcriptional repression (Guccione *et al.*, 2007). Leukaemia cells in which MLL has undergone chromosomal rearrangements, display abnormal expression of HOXA5, HOXA7 & HOXA9 and several studies have shown that this methyltransferase directly regulates HOX genes (Ernst *et al.*, 2004; Kohlmann *et al.*, 2005; Milne *et al.*, 2005). MOZ is also known to regulate the expression of many HOX genes, including those mentioned above and in Zebrafish was found to influence the expression of the Hox clusters 1-4, e.g. Hoxb1a, Hoxa2b, Hoxb3a and Hoxb4a (Miller, Maves and Kimmel, 2004; Voss *et al.*, 2009; Paggetti *et al.*, 2010). Paggetti *et al.* (2010), demonstrated that MOZ and MLL can colocalize and interact with each other, cooperating to regulate the expression of HOX genes. Furthermore, they reported that MOZ can interact with WDR5, a protein that binds to the MLL-associated epigenetic mark H3K4me3. Therefore, MOZ is essential for the correct regulation of HOX gene transcription through interactions with other epigenetic “readers” and “writers” to increase acetylation at target loci.

MLL is a member of the Trithorax (TrxG) group of proteins, which are thought to promote the active transcription of genes, whereas the Polycomb (PcG) group of proteins repress transcription (reviewed by Voigt, Tee and Reinberg, 2013). The most-studied PcG proteins are PRC1 & PRC2. Both groups were originally identified as regulators of Hox gene expression (reviewed by Schuettengruber *et al.*, 2007). The activity of these two groups antagonises each other, creating a molecular competition between the activating and repressing epigenetic marks within gene promoters (Bernstein *et al.*, 2006). H3K4me3 (MLL-mediated, transcription activating) and H3K27me3 (PRC2-mediated, transcription repressing) marks have been identified in the same promoter regions, suggesting these promoters are repressed, but poised to be quickly activated (Bernstein *et al.*, 2006; Mikkelsen *et al.*, 2007). Promoters in such a state are deemed to be bivalent. As MOZ can

interact with MLL and recognise H3K4me3 marks via its interaction with WDR5, it has been suggested that MOZ activates the expression of genes that have been previously, or are simultaneously methylated by MLL (Paggetti *et al.*, 2010) (Figure 9). This may thereby propagate active chromatin marks and push the bivalent state towards active gene transcription. Interestingly however, Kang *et al.* (2017) reported that the *Drosophila* orthologue of MOZ, Enok could be co-purified with PRC1. This indicates that MOZ may not only interact with positive mediators of transcription activation such as MLL, but also proteins that act to repress transcription.

Aside from the transcriptional regulation of Hox genes, MOZ has also been demonstrated to bind or localise to promoter regions of other genes and influence their expression. Examples of such genes include the macrophage inflammatory protein 1α (MIP-1α) gene which encodes a proinflammatory cytokine, the dentin sialophosphoprotein (Dspp) gene which facilitates the formation of dentin in teeth, and the T-Box 1 (TBX1) transcription factor that is essential for normal heart development (Bristow and Shore, 2003; Voss *et al.*, 2012; Heair *et al.*, 2015). For Dspp, it has been demonstrated that enhanced transcription of this target gene can be mediated by the interaction of MOZ with the transcription factor RUNX2, clearly indicating that protein-protein interactions are important for the ability of MOZ to regulate gene expression (Heair *et al.*, 2015). Currently, there are only a small number of genes documented in the scientific literature which are known to be directly regulated in their expression by MOZ. This HAT is crucial for normal embryogenesis and is embryonic lethal, therefore it stands to reason that there are many more MOZ target genes yet to be identified (Miller, Maves and Kimmel, 2004; Voss *et al.*, 2009; Vanyai, Thomas and Voss, 2015). Indeed, our lab has recently identified a large number of genes which appear to be regulated by MOZ using RNA-seq, qRT-PCR validation, CRISPR-Cas9 gene KO, MOZ overexpression and siRNA KD (Monteiro 2018, PhD thesis - unpublished); mRNA expression of several of these genes is shown in Figure 10A, along with protein expression in Figure 10B. Of

great interest are the HHEX and PROX1 transcription factors, as well as the tumour suppressor CDKN2A, as this indicates MOZ regulates the expression of crucial transcriptional regulators that are important for development and lineage specification (Risebro *et al.*, 2009; Lavado *et al.*, 2010; Paz *et al.*, 2010; Goodings *et al.*, 2015).

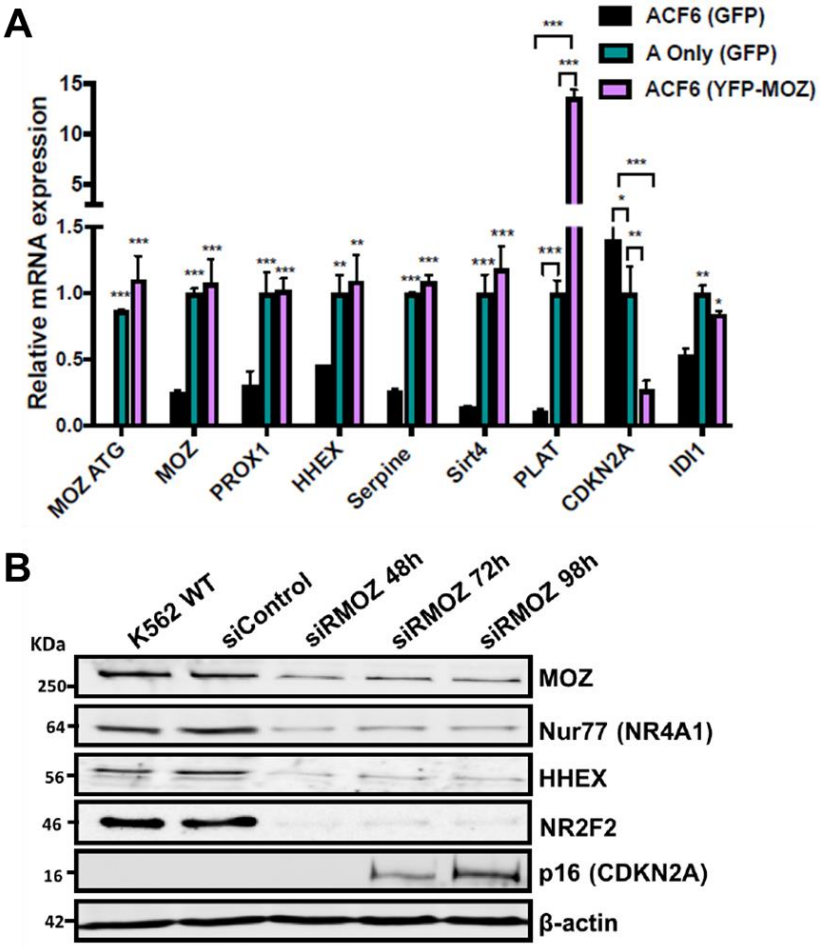


Figure 10: MOZ regulates the expression of target genes

qRT-PCR analysis of mRNA expression **(A)** and western blotting of proteins **(B)** known to be MOZ target genes. MOZ expression is shown for validation of KO or KD. mRNA expression was normalised to GAPDH, β-actin was used as a protein loading control. A only = K562 CRISPR-Cas9 control cells which are WT for the KAT6A gene, ACF6 = KAT6A KO K562 cells, GFP = transfected with a control GFP expression plasmid, YFP-MOZ = transfected with a full length WT MOZ expression plasmid N-terminally tagged with YFP to rescue MOZ target gene expression. siRMOZ = siRNA-mediated knockdown of MOZ expression in WT K562 cells for the number of hours indicated. Error bars represent standard error of the mean. Statistical analysis performed using two-way ANOVA. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$. n = 3 for all experiments here. Taken from Monteiro (2018), PhD thesis – unpublished.

vii. Cell cycle regulation

The *Drosophila* orthologue of MOZ, Enok, in the presence of its active complex has been demonstrated to promote the progression of neuroblasts in flies from G_1 to S phase of the cell cycle during proliferation (Scott, Lee and Luo, 2001). In addition, inactivating the HAT domain of the MOZ orthologue in yeast, Sas3, causes cells to accumulate at the G_2/M phase (Howe, 2001). This clearly shows that the MYST family of HATs is important for cell proliferation and cell cycle progression. The acetylation of different substrates by MOZ can have either a positive or negative influence upon the arrest of cellular growth. For example, MOZ acetylates and stabilises p53 at positions K120 and K382 via its interaction with p53 and PML, which then acts to facilitate the expression of p21 and induces senescence via the acetylated p53 (Rokudai *et al.*, 2013). Alternatively, MOZ inhibits cellular senescence and promotes stem cell maintenance by indirectly repressing the INK4A-ARF (CDKN2A) locus, which encodes a cyclin-dependent kinase known to promote senescence and inhibit proliferation (Perez-Campo *et al.*, 2014; Sheikh, Phipson, *et al.*, 2015). For this inhibition, MOZ itself actually activates the transcription of genes which encode potent inhibitors of CDKN2A transcription, by acetylating H3K9 at their promoter regions (Sheikh, Phipson, *et al.*, 2015). We have previously shown that knockout of MOZ in K562 cells profoundly reduces the rate of increase in the total number of viable, metabolically active cells (Figure 11A) and siRNA KD of MOZ reduces the proportion of the cell population in S phase (Monteiro 2018, PhD thesis - unpublished) (Figure 11B). Consequentially, an increase in the population of cells in G_0/G_1 was also observed, indicative of the requirement for MOZ in promoting proliferation (Figure 11B).

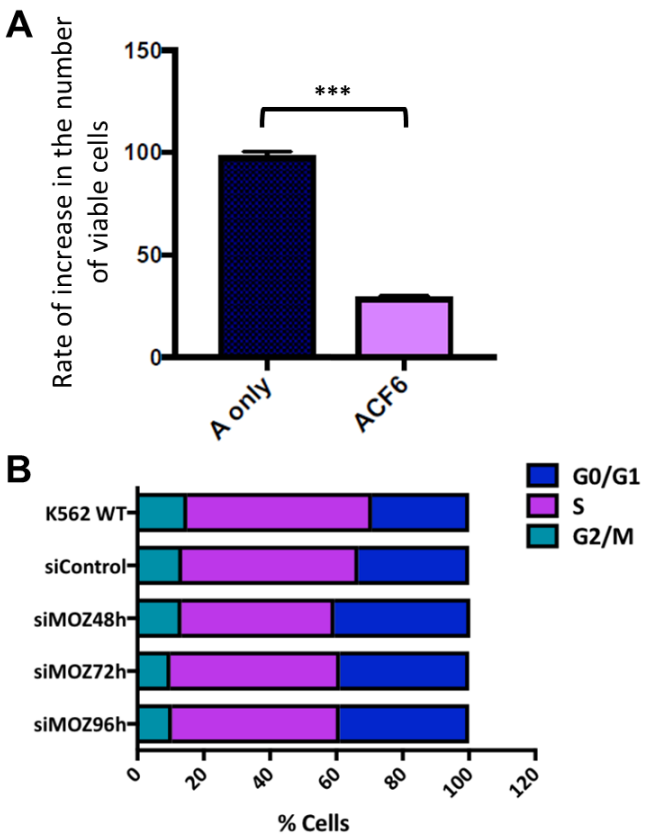


Figure 11: MOZ expression promotes cell cycle progression

Rate of increase in the number of viable, metabolically active cells, as measured by MTT assay for WT and MOZ KO K562 cells over a 24 hour period **(A)**. Cell cycle analysis by propidium iodide staining of WT K562 or siRNA-mediated MOZ KD cells **(B)**. A only = K562 control CRISPR-Cas9 cells, ACF6 = K562 MOZ KO cells. siMOZ = siRNA-mediated MOZ KD in WT K562 cells for the number of hours indicated. Note the reduced rate of increase in viable, metabolically active cells upon MOZ KO. Also note the reduction in the proportion of the cell population in S phase upon MOZ siRNA KD, as well as the increased proportion of those in G_0/G_1. Error bars represent standard error of the mean. *** = $p<0.001$. n = 3 for all experiments here. Taken from Monteiro (2018), PhD thesis – unpublished.

Senescence is the irreversible entry of a cell into the G_0 stage of the cell cycle, which can occur in response to DNA damage or when a cell can no longer replicate its genome due to telomere shortening, known as the Hayflick limit (Hayflick, 1965). Quiescence on the other hand, is the reversible entry of a cell into G_0, where it resides until cell cycle activation by an extrinsic signal. Unfortunately, the term senescence is occasionally used interchangeably with quiescence to mean that a cell population has stopped (or reduced) proliferating, without defining whether that is permanently or temporarily. Several publications have demonstrated that a lack of MOZ expression specifically affects senescence in their cell populations, although to date there have been no studies investigating MOZ in quiescence (Rokudai *et al.*, 2013; Perez-Campo *et al.*, 2014; Sheikh, Phipson, *et al.*, 2015). Therefore, it is unknown whether this HAT has any involvement in the induction, maintenance or exit from quiescence.

I. VIII. MOZ in development & cancer

i. Embryogenesis

MOZ-deficient mice are embryonic lethal and MOZ is highly expressed in embryonic stem cells (ESCs) (Cho, Kim and Jang, 2016). The expression of MOZ mRNA increases in mouse ESCs (mESCs) during differentiation, indicative of a functional role in this process (Cho, Kim and Jang, 2016). Knockdown of MOZ results in a reduction in the expression of ectoderm-related genes such as MAP2 and PAX6 during differentiation, however the expression of some endoderm-related genes like GATA4 increases (Cho, Kim and Jang, 2016). This indicates MOZ is involved in germ layer formation during embryogenesis and influences cell lineage specification.

Histone acetyltransferases are also well known to be important for correct development of the heart and cardiovascular system. For example, both Kat2a (Gcn5) and Kat2b (Pcaf) can acetylate the transcription factor Tbx5 in zebrafish at lysine 339, which is necessary for its nuclear retention (Ghosh *et al.*, 2018). Inhibition of the acetylation of this residue has a negative impact upon cardiac development, resulting in a morphologically deformed heart with reduced function (Ghosh *et al.*, 2018). In contrast to this, MOZ has not been implicated in the acetylation of the TBX5 protein, but instead is reported to be essential for the normal activation of transcription of both Tbx1 & Tbx5 (Voss *et al.*, 2012; Vanyai, Thomas and Voss, 2015). Loss of Moz expression depletes Tbx1 mRNA to 50-60% of the wild type level, while Tbx5 is reduced by 75% (Voss *et al.*, 2012). MOZ catalytic activity is required for H3K9ac deposition and the subsequent expression of these two transcription factors (Voss *et al.*, 2012). In the absence of mesodermal Moz expression, several genes encoding transcription factors important for heart development are downregulated, such as Pitx2, Gata4 & Frzb (Vanyai, Thomas and Voss, 2015). A summary of the known roles of MOZ during embryogenesis are depicted in Figure 12.

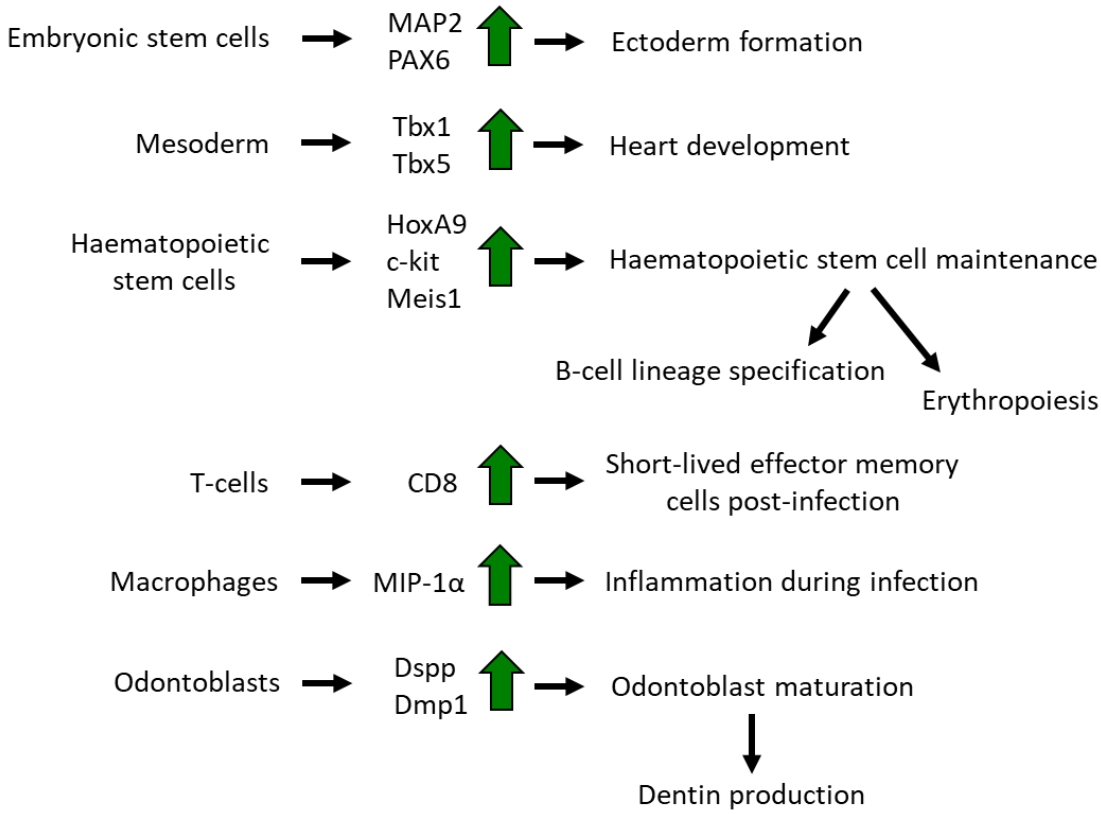


Figure 12: Biological roles of MOZ during development

A summary of the biological processes associated with MOZ expression during development. Data was curated from the scientific literature. Tissue/cell type is depicted on the left, genes upregulated upon MOZ expression/activity are represented next to green arrows, and biological outcomes are on the right. It is clear that MOZ is important for normal immune system activity, with involvement in B-cell specification, memory T-cell formation and inflammation. This HAT is also crucial for embryonic development (MOZ is embryonic lethal), where it is known to regulate genes responsible for gastrulation (specifically ectoderm formation) and heart development.

ii. Haematopoiesis & the immune system

At embryonic day 15 (E15), MOZ-deficient mice have a characteristically small liver (Katsumoto *et al.*, 2006). This is consistent with their significantly reduced production of mature erythrocytes. In MOZ-deficient foetal livers, haematopoietic stem cells (HSCs) and CD19+ B-lineage cells are vastly reduced when compared to their WT counterparts, however in contrast, granulocyte and monocyte populations are increased (Katsumoto *et al.*, 2006). Furthermore, the myeloid progenitor KLSF (c-kit+, Lin-, Sca-1-, FcRy+) population is higher in MOZ-deficient foetal livers, indicating that MOZ has an important role in the development and differentiation of lymphoid and myeloid cells (Katsumoto *et al.*, 2006). When MOZ-deficient foetal liver cells are transplanted into irradiated mice, the recipients fail to develop a reconstituted hematopoietic system. This shows that MOZ is necessary for the self-renewal capability of HSCs, which is backed up by the reduced expression of HoxA9, c-kit and c-Mpl in MOZ-deficient mice, genes which are known to be required for HSC self-renewal ability (Lawrence *et al.*, 1997, 2005; Katsumoto *et al.*, 2006) (Figure 12).

Sheikh et al. (2015) reported that mice haploinsufficient for Moz expression have smaller spleens than their WT counterparts, suggesting impaired B-cell development or maturation. While the overall bone marrow cellularity is unaffected in these mice, the number of pre-B-cells is reduced by 1/3 and immature B cells by ½ (Sheikh, Lee, *et al.*, 2015). Pre-B-cells isolated from Moz haploinsufficient mice display a 50% reduction in the mRNA expression of Hoxa9, as well as downregulated Meis1 expression; the expression of these genes is associated with a high self-renewal ability and inhibition of differentiation, suggesting Moz is responsible for the expression of genes promoting an undifferentiated B-cell fate (Lawrence *et al.*, 1997; Ariki *et al.*, 2014; Sheikh, Lee, *et al.*, 2015). Indeed, mice lacking Hoxa9 have been reported previously to have reduced numbers of B-cell progenitors from the pro-B-cell stage (Lawrence *et al.*, 1997). Similarly, loss of Meis1 in the hematopoietic

compartment leads to a reduction in all B-cell progenitors (Ariki *et al.*, 2014).

Research has implicated MOZ in the differentiation and activation of T cell populations, as well as in development of the thymus (Newman *et al.*, 2016). KAT6A-deficient CD8+ T cells downregulate surface CD8 co-receptor expression during clonal expansion, which correlates with a reduced number of Cd8α mRNA transcripts and reduced acetylation of H3K9 at the Cd8 locus (Newman *et al.*, 2016). While conditional Kat6a knockout mice which lack Moz expression in lymphocytes do not have any significant difference in the numbers of cells in the peripheral T cell compartments or the thymus, the population of P14 cells (a T cell subtype) is markedly reduced 60 days post-viral infection (Newman *et al.*, 2016). Differentiation of the P14 cells revealed that Kat6a KO reduced their ability to differentiate into short-lived effector memory cells (Newman *et al.*, 2016). This is indicative of a crucial role for MOZ in the adaptive immune system (Figure 12).

Assessment of Kat6a expression in murine macrophages revealed that expression of this HAT is upregulated at both the mRNA, and protein levels during macrophage activation with lipopolysaccharide (LPS) (Shukla *et al.*, 2018). This indicates that the stimulation of macrophages due to bacterial infection results in increased MOZ expression. The promoter of the macrophage inflammatory protein 1α (MIP-1α) gene, encoding a proinflammatory cytokine, harbours two RUNX binding sites; it has been shown that MOZ and RUNX1 act synergistically in a dose-dependent manner at this promoter to facilitate MIP-1α expression (Wolpe and Cerami, 1989; Bristow and Shore, 2003). It has therefore been proposed that MOZ is involved in both inflammation and immune response to infection (Figure 12).

iii. Dentinogenesis

The miRNA miR-665 is known to repress early and late odontoblast marker genes and proteases involved in dentin maturation. It has also been demonstrated that MOZ is a target of this miRNA;

overexpression of miR-665 prevents MOZ-induced chromatin remodelling and H3K9ac deposition, which in turn represses gene transcription (Heair *et al.*, 2015). Expression of miR-665 reduces MOZ recruitment to the Dspp and Dmp1 promoters, two genes involved in the formation of dentin (Heair *et al.*, 2015). The interaction of RUNX2 with MOZ has been shown to facilitate the transcription of genes required for the differentiation of cells into mature odontoblasts (Pelletier *et al.*, 2002; Heair *et al.*, 2015) (Figure 12). These findings have led to the suggestion that MOZ is involved in dentinogenesis via odontoblast maturation.

iv. Leukaemia

A subset of patients with acute myeloid leukaemia (AML) have been discovered to have chromosomal translocations of the KAT6A gene with CREBBP (CBP), EP300 (p300), or NCOA2 (TIF2) (Liang *et al.*, 1998; Chaffanet *et al.*, 2000; Rozman *et al.*, 2004). Additional cases whereby KAT6A is translocated with LEUTX, ASXL2 or NCOA3 (SRC3) have also been reported (Esteyries *et al.*, 2008; Katoh, 2013; Chinen *et al.*, 2014) (Figure 13A). These translocations result in the expression of an N-terminal MOZ fusion protein, whereby the C-terminal domains of MOZ are replaced with C-terminal domains of the corresponding fusion partner (see example of MOZ-TIF2 in Figure 13B). AML caused by KAT6A chromosomal translocations are particularly aggressive; the median survival post-diagnosis for these patients ranges from 2 – 10 months, demonstrating that the deregulation of MOZ has catastrophic effects (Kindle *et al.*, 2005; Collins *et al.*, 2006). The HAT activity of MOZ is known to influence HOX gene expression and the deregulation of the expression of HOX genes has been noted to be correlated with acute leukemia (Camós *et al.*, 2006; Voss *et al.*, 2009; Cheung *et al.*, 2016). Epigenetic modifiers such as MOZ are coordinated in their recruitment to specific loci, therefore changes in their expression levels or even the functional domains they possess (fusion proteins) can establish oncogenic transcriptional programs for cancer development (Cheung *et al.*, 2016).

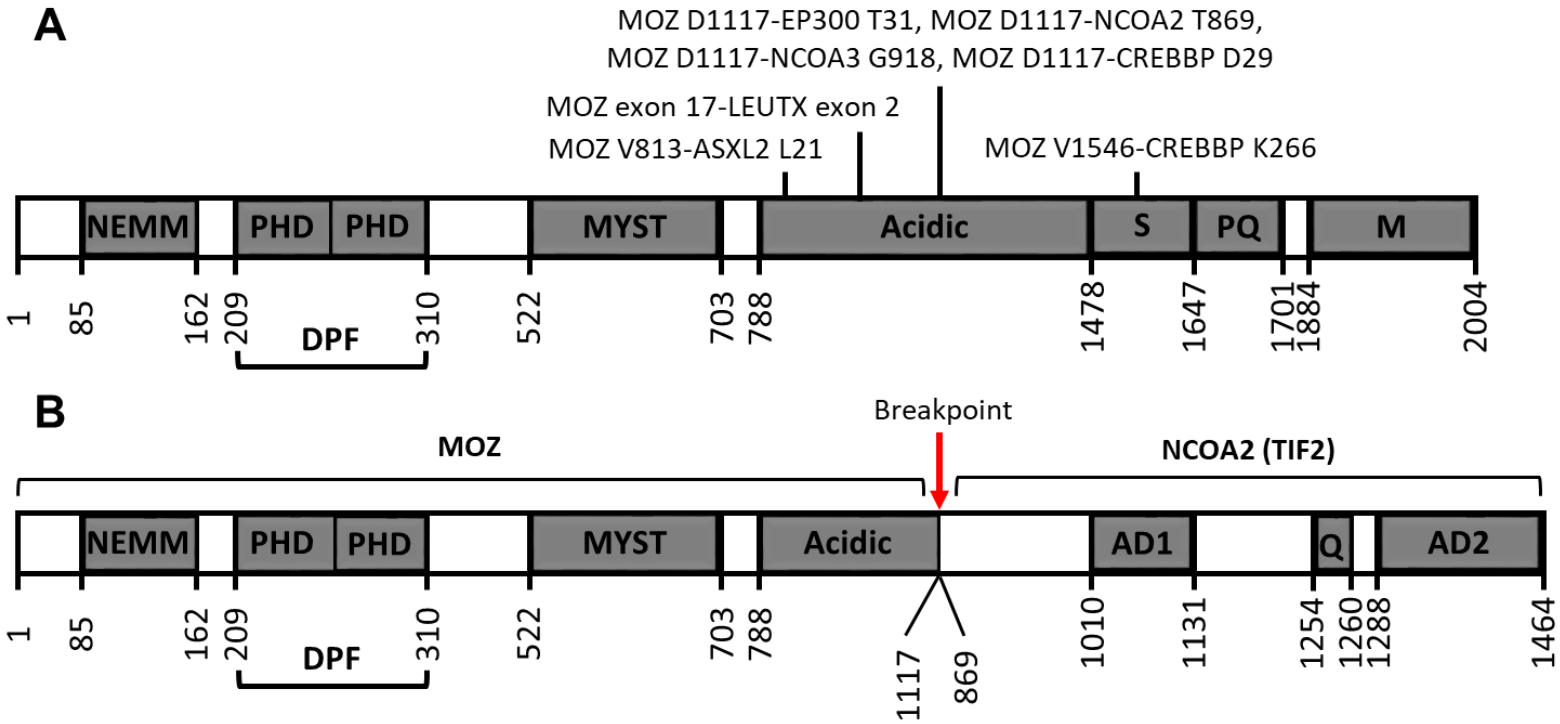


Figure 13: MOZ fusion proteins arising from chromosomal translocations in cancers

(A) Fusion proteins involving MOZ in cancer resulting from chromosomal translocations, showing the MOZ protein breakpoints relative to domains. The boundaries of fused MOZ N-terminal polypeptide sequences to C-terminal fusion partners are detailed. **(B)** Representative example of the MOZ-TIF2 fusion protein in AML showing the corresponding retained protein domains. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2 (PHD1/2). MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. AD1/2 = Activation domain 1 or 2, respectively. Acidic residue-rich (Acidic), serine-rich (S), glutamic acid-rich (E), glutamine-rich (Q), proline and glutamine-rich (PQ) and methionine-rich (M) regions are also indicated.

The (8;16)(p11;p13) translocation, which translationally generates the MOZ-CBP fusion protein, results in the activation of genes involved in cell cycle control and the regulation of transcription, thought to be due to the combined histone acetyltransferase activities of both MOZ and CBP in the same protein (Coenen *et al.*, 2013). The expression of this fusion protein is associated with a high rate of mortality (Gervais *et al.*, 2008). The MOZ-CBP fusion protein inhibits RUNX1-dependent activation of the myeloperoxidase (MPO) promoter, but can activate NF-kB-dependent transcription (Kitabayashi *et al.*, 2001; Chan *et al.*, 2007). Inhibition of RUNX1-mediated transcription and myeloid cell differentiation by the MOZ-CBP fusion protein have been suggested as the primary pathogenic mechanisms behind leukemogenesis in this AML type (Rozman *et al.*, 2004; Camós *et al.*, 2006; Gervais *et al.*, 2008).

p300 (a paralogue of CBP) shares 63% amino acid sequence homology with CBP, has similar cellular functions and is somewhat similar in its structure; this has led to the suggestion that both MOZ-CBP and MOZ-p300 fusion proteins have a similar pathophysiology (Lundblad *et al.*, 1995; Chan and La Thangue, 2001). Cases of AML arising from MYST3-CREBBP fusion genes have a close transcriptional signature to that of AML cases with MLL chromosomal rearrangements (Camós *et al.*, 2006; Serravalle *et al.*, 2010; Coenen *et al.*, 2013). The similarities include an upregulation of the HOXA9, HOXA10 and MEIS1 genes, while other members of the homeobox family are downregulated (Serravalle *et al.*, 2010).

Expression of the MOZ-TIF2 fusion protein is thought to confer self-renewal activity to committed progenitor cells, as transplanting haematopoietic stem cells or myeloid progenitors that have been transfected with a MOZ-TIF2 retroviral vector into mice, induces the development of AML (Deguchi *et al.*, 2003). Indeed, the MOZ zinc finger within the MYST domain is indispensable for MOZ-TIF2-induced leukaemic transformation (Deguchi *et al.*, 2003). The MOZ-TIF2 fusion protein represses retinoic acid receptor-dependent and p53-dependent transcription (Kindle *et al.*, 2005; Collins *et al.*, 2006). Cheung et al.

(2016) reported that the interaction between MOZ-TIF2 and the H4R3 methyltransferase PRMT1 (mediated by the N-terminal 79 amino acids of MOZ), is necessary for leukemic transformation. It has been revealed that HOXA9 and other downstream targets of MOZ-TIF2 have high levels of H4R3me2 and recruitment of PRMT1 (Cheung *et al.*, 2016). This suggests that MOZ-TIF2 interactions with other epigenetic regulators are a key mechanism of leukaemic development.

v. Glioblastoma

Lv et al. (2017) reported high KAT6A mRNA and MOZ protein levels in a glioblastoma dataset. High KAT6A expression was strongly correlated to glioblastoma cancer progression and a worse prognosis of approximately 10 months, compared to 13.5 months for patients with low KAT6A expression. MOZ overexpression increases AKT activation, via the increased expression of the kinase PIK3CA and subsequent phosphorylation of AKT; this drives cell proliferation and promotes cell survival (Lv *et al.*, 2017). Therefore, in glioblastoma, KAT6A upregulates PIK3CA expression, which in turn activates the PI3K/AKT signalling pathway, promoting tumourigenesis. In addition, increased H3K23ac in a model of glioblastoma was correlated with increased KAT6A expression, indicating that KAT6A acetylation of H3K23 is essential for the expression of PIK3CA and glioblastoma tumourigenesis (Lv *et al.*, 2017). The tripartite motif-containing 24 (TRIM24) protein is known to bind H3K23ac and has been previously shown to upregulate the mRNA expression level of PIK3CA in glioblastomas and activate the PI3K/AKT signalling pathway (Zhang *et al.*, 2015). Lv et al. (2017), demonstrated that in glioblastoma, TRIM24 and H3K23ac are both present at the PIK3CA promoter, however knockdown of MOZ inhibited TRIM24 binding and reduced H3K23ac marks at this locus. This indicates that in KAT6A-high expressing glioblastomas, MOZ facilitates the association of TRIM24 with H3K23ac, thereby enhancing activation of the PI3K/AKT signalling pathway and promoting tumourigenesis.

vi. Breast cancer

The chromosome amplicon consisting of loci 8p11-p12 contains approximately 55 genes and is present in 12-15% of breast cancers (Turner-Ivey *et al.*, 2014). It is associated with poor prognosis in primary breast tumours and a lower 5-year metastasis-free survival rate (Turner-Ivey *et al.*, 2014). The KAT6A gene is situated within this amplicon. Turner-Ivey et al. (2014) found that KAT6A was a consistently top hit in their shRNA screen in the SUM-52 breast cancer cell line for genes which are amplified and overexpressed. This suggests that this gene functions as an oncogene in luminal breast cancers containing the 8p11-p12 amplicon. Data from 1,000 primary breast cancers revealed that in 22% of cases, the expression of KAT6A is altered and it is overexpressed in 8.7% of invasive breast cancers at the mRNA level; expression of MOZ at the protein level is increased in 15% of cases (Turner-Ivey *et al.*, 2014; Yu *et al.*, 2016). This increased KAT6A expression is associated with a statistically significant reduction in the overall survival rate (Turner-Ivey *et al.*, 2014). Experimentally, KAT6A knockdown mammospheres are smaller than their control counterparts and the 3D structures normally formed by mammospheres after two weeks in culture, fail to materialise, showing a requirement of KAT6A for normal mammosphere morphology. Additionally, there is a more than a 10-fold decrease in SUM-52 clonogenic capacity after KAT6A knockdown, which has been associated with increased differentiation (Turner-Ivey *et al.*, 2014). Yu et al. (2016) reported that MOZ has HAT activity at the ERα promoter in breast cancer and that the ERα promoter displays H3K9ac enrichment, indicating that MOZ may bind to this region and increase the expression of ERα via the acetylation of surrounding histone tails. RNA sequencing revealed several other transcription factors such as TBX3, GATA3 and MYC as MOZ target genes in breast cancer cells, suggesting that MOZ promotes breast tumorigenesis via the alteration of luminal mammary differentiation and self-renewal capacity (Turner-Ivey *et al.*, 2014).

I. IX. CRISPR-Cas9 genome editing

It is thought that most bacteria and virtually all archaea have developed an adaptive immunity via some form of programmable endonuclease system (Barrangou *et al.*, 2007). The Clustered Regularly Interspaced Palindromic Repeats (CRISPR) CRISPR-associated enzyme 9 (-Cas9) system is an RNA-guided "immune system" identified in prokaryotes (Barrangou *et al.*, 2007; Koonin, Makarova and Zhang, 2017). Upon infection, a Cas protein complex inserts a piece of foreign invading DNA into a CRISPR array; this segment of DNA is now known as a spacer (Barrangou *et al.*, 2007). The spacer is expressed as precursor CRISPR RNA (pre-cr), which is processed to form a mature CRISPR RNA (crRNA) 20-24 nucleotides in length. These crRNAs form complexes with a scaffold RNA, known as a tracrRNA, which function to "load" the crRNA onto the Cas9 enzyme (Ran *et al.*, 2013). This tracrRNA-crRNA complex (referred to as a single guide RNA or sgRNA in engineered Cas endonucleases), binds to a Cas9 endonuclease enzyme and serves to direct its catalytic specificity to target DNA/RNA sequences, such as foreign DNA from an invading virus (Barrangou *et al.*, 2007; Moore, 2015).

Commonly referred to as just CRISPR, this genome editing tool was originally discovered in 2007 in the bacterial species *Streptococcus pyogenes* (Barrangou *et al.*, 2007). However, it was not repurposed from its prokaryotic origins into a useful and programmable endonuclease for use in scientific research until several years later (Mali *et al.*, 2013; Ran *et al.*, 2013). This genome editing system has now successfully been utilised in a range of cell lines from different organisms, including that of humans to effectively create specific genetic variants or knockout desired genes (Mali *et al.*, 2013; Wang *et al.*, 2013; Gundry *et al.*, 2016; Yoon *et al.*, 2018). There are two major forms of the CRISPR system, termed "classes" which use different forms of Cas effector enzymes (Koonin, Makarova and Zhang, 2017). Separately, there are also different "types" of CRISPR, depending upon their mechanism of action and specific endonuclease and these types exist within both classes (Makarova *et al.*,

2015). Each of the class systems shares the Cas1 and Cas2 protein complex, which functions to produce the spacer, however the processing of pre-cr RNA into mature crRNA is where they begin to diverge. Class 1 systems have a complex involving several Cas proteins with the crRNA, which then acts together upon its target(s). Alternatively, the class 2 systems employ a single Cas protein bound to crRNA and this acts as the effector (Makarova *et al.*, 2015). It is the class II, type II CRISPR-Cas9 system which has been adapted from *S. pyogenes* into the ever more commonly used genome editing technique (Mali *et al.*, 2013).

Importantly, the sequences able to be cleaved by Cas endonucleases are limited by the need for a protospacer adjacent motif (PAM) immediately 3' to the target sequence (Mali *et al.*, 2013). In the case of Cas9, this motif is a 3 bp sequence of either 5'-NGG-3' or 5'-NAG-3', which in the absence of, prevents target sequence cleavage (Mali *et al.*, 2013). When used in the human genome, this hypothetically allows for the targeting and formation of a double strand DNA break every eight nucleotides, making it theoretically possible to target almost any coding locus (Tan *et al.*, 2015). Alternative CRISPR endonucleases have different PAM sequence requirements (Gasiunas *et al.*, 2012; Cong *et al.*, 2013). The identification of novel CRISPR effector proteins is constantly expanding our ability to target additional nucleotide sequences. Although many bacterial nucleases are known and have been characterised, not all are able to retain their functionality when expressed in mammalian cells, making their use in mammalian genome editing problematic (Cong *et al.*, 2013; Zetsche *et al.*, 2015).

While it is suggested that there is a low level of toxicity of the Cas9 at the organismal level, there is the risk of undesired double strand DNA cleavage at unwanted loci due to sgRNA target sequence similarity (Fu *et al.*, 2013; Ran *et al.*, 2013; Tan *et al.*, 2015). These such events are known as 'off-target effects' and are able to be significantly reduced if a Cas9 nickase is used instead of the fully active endonuclease (Chiang *et al.*, 2016). This Cas9 mutant has one of its catalytic domains (of which there are two – RuvC & HNH) inactivated by a point mutation; the D10A

missense mutation results in RuvC inactivation (Gasiunas *et al.*, 2012; Cong *et al.*, 2013; Chiang *et al.*, 2016). When using the Cas9 nickase approach to genome editing, two of these nickase endonucleases must be used in conjunction, each with a different sgRNA which target opposite DNA strands to each other and have target sequences in close proximity (within 0-30 nucleotides) (Ran *et al.*, 2014; Chiang *et al.*, 2016). This reasoning is due to either mutant endonuclease only being able to cause a single strand DNA break, however together their activity can be recognised by the endogenous DNA repair pathways as a double-strand break (Chiang *et al.*, 2016). Alternatively, the off-target activity of Cas9 can also be reduced by several orders of magnitude by reducing the sgRNA length from 20 to 17 nucleotides (Fu *et al.*, 2014). These shorter sequences, known as tru-gRNAs have been reported to be just as efficient at directing Cas9 activity as the longer sgRNAs. Furthermore, it has been demonstrated that if the Cas9 nuclease is engineered to make the interactions between this enzyme and DNA weaker, it must rely more heavily upon the binding of the DNA target with the sgRNA for its activity. This has the outcome of much fewer off-target cleavage sites, with many sgRNAs tested solely cleaving at the desired locus and no loss of Cas9 endonuclease efficiency (Kleinstiver *et al.*, 2016).

To engineer the formation of a specific mutation at a desired genomic locus, a DNA repair template must be provided for use by the cell's own DNA repair machinery (Ran *et al.*, 2013). The homology directed repair pathway (HDR) can utilise these repair templates to incorporate mutations desired by the researcher, for example a single nucleotide change or even the insertion of an affinity tag (Dalvai *et al.*, 2015). The template provided can be either a double-stranded plasmid DNA sequence, or a single-stranded oligonucleotide; both require flanking homology arms on either side of the desired nucleotide changes and the latter is often utilised for small genetic alterations, such as a single nucleotide substitutions (Chen *et al.*, 2011). Double-stranded repair templates are usually found to have a higher success rate than single-stranded oligonucleotides when making larger genetic alterations

(Chen *et al.*, 2011). However, if no repair template is provided/available, DNA repair mechanisms will attempt to repair the cleaved locus in an essentially random manner via the non-homologous end joining pathway (NHEJ). This is an error-prone process which frequently results in nucleotide insertions or deletions, disrupting the native codon sequence and often produces frameshift, missense or nonsense mutations (Yu *et al.*, 2015). NHEJ can therefore result in the translation of truncated proteins, or even completely silence gene expression if the transcription start site is removed. Although it has a higher fidelity than NHEJ, the HDR pathway is usually only active in dividing cells and its efficiency can be influenced by a variety of factors, such as the length of the homology arms in the repair template provided, the genomic locus and cell type (Saleh-Gohari and Helleday, 2004; Lin *et al.*, 2014). For these reasons, utilising the HDR pathway for the generation of specific sequence variants has a much lower success rate, compared to DNA repair by NHEJ. Cell cycle synchronisation of a cell population prior to transfection of genome editing reagents has been reported to increase the efficacy of HDR, as has the treatment of cells with small molecules thought to enhance HDR activity (Saleh-Gohari and Helleday, 2004; Lin *et al.*, 2014; Yu *et al.*, 2015).

The CRISPR-Cas9 system is easier to design than zinc-finger nucleases (ZFNs) or transcription activator-like effector nucleases (TALENs), both of which use the pairing of endonuclease catalytic domains and DNA-binding proteins to cause double strand DNA breaks at genomic target loci (Ran *et al.*, 2013; Kim and Kim, 2014). Alternatively, CRISPR-Cas9 uses base pairing of small RNAs with target DNA, allowing this technique to be efficient, specific and suitable for high-throughput genomic editing in a wide variety of cell types (Ran *et al.*, 2013). Indeed, we have successfully used this method before on the same gene, and others to generate MOZ, AR, PIP5K1A, BRPF1, PNR and PPARγ knockout/mutant cell lines (Monteiro 2018, PhD thesis - unpublished; Hernandez-quiles 2016, Masters thesis - unpublished; Clemente 2016, PhD thesis - unpublished).

I. X. Aims & objectives

With the sequencing of individuals with developmental conditions becoming more commonplace, the technologies cheaper and enhanced workforce training, there has been a great increase in the number of genes associated with clinical conditions (Johansen Taber, Dickinson and Wilson, 2014; Retterer *et al.*, 2016; Shashi *et al.*, 2016; Lundsgaard *et al.*, 2017; Schoch *et al.*, 2017). This has led to the identification of novel genetic variants and also the division of previously generic umbrella terms for types of syndromes such as global developmental delay, into more specific groups determined by the genetic cause (Wright *et al.*, 2015). While this may be good news for patients and their families to have an official diagnosis, it is still often the case that the actual molecular link between identified gene variants and the presented syndrome is not understood and no appropriate therapies exist. Such is true for the KAT6A GDD syndrome. A greater understanding could allow for the development of therapies which may help to relieve/lessen some clinical symptoms. Furthermore, it cannot be understated how little we really know about domains or structures within proteins, despite there being a highly active biochemical research community the world over. Studying genetic variants identified within developmental syndromes allows for deeper observations of the function or requirement for encoded species, e.g. protein domains, within a range of body tissues.

Most of the research regarding the MOZ protein has focussed on the N-terminal regions, with crystal structures and substrate binding well known. Conversely, the C-terminal half of the protein is less well understood and this is the part associated with genetic variants in KAT6A GDD. With a lack of knowledge regarding the transactivation function of the MOZ C-terminal domains, predicting the effects upon gene transcription when these regions are lost is near impossible. Furthermore, it is unknown what other biological properties the C-terminal domains of this HAT have, other than the ability to interact with RUNX1, PML or p53. Since these C-terminal domains are lost in individuals with the KAT6A GDD syndrome, there is now a clinical

developmental interest in determining the requirement of the MOZ C-terminus, clarifying exactly what the target genes are of this HAT at a transcriptome-wide level and whether or not/how the C-terminal MOZ regions regulate their normal transcription. While the causative genetic variants have been pinpointed for KAT6A GDD, there is only limited information regarding exactly how they result in the clinical phenotype. Different clinical reports have suggested either the activity of NMD, or dominant-negative MOZ protein, but currently there is no substantial evidence for either hypothesis.

Here, we aim to expand the scientific knowledge of MOZ in gene transcription by determining novel target genes which are regulated in their expression by the C-terminal regions of MOZ. We have recently identified a collection of genes which are reliant upon MOZ for their correct transcriptional regulation (Monteiro 2018, PhD thesis - unpublished). While this analysis revealed many novel target genes, it could not distinguish those reliant upon the C-terminal regions of this HAT for their regulation, compared to any genes for which these are dispensable. Therefore, unfortunately it did not reveal any functional information regarding the C-terminal regions; we aim to examine this aspect further during the current study. In addition, we will assess the requirement for these regions in the maintenance of metabolically active, viable cells, cell cycle progression and correct protein localisation. These points will be achieved by utilising the CRISPR-Cas9 nickase technique in the commonly used K562 cell line, as well as via the analysis of patient RNA-seq data. This will allow us to facilitate the generation of KAT6A variants in the ultimate and penultimate exons of the KAT6A gene, resembling those in patients with KAT6A GDD syndrome. Considering the nonsense variants in the KAT6A gene which are associated with KAT6A GDD, they are most frequently located in exons 17 and 18, the penultimate and terminal exons, respectively. It is therefore plausible that the nonsense variants clustered within exon 17 are at least partially subjected to NMD activity, whereas we may expect the ones in exon 18 to be unaffected. Use of these cell lines will allow us to determine the

susceptibility of KAT6A mRNA transcripts with PTCs to undergo NMD, helping us to elucidate the mechanism(s) behind this syndrome.

Elucidation of the secondary and tertiary structures of C-terminal regions of MOZ, would allow the revelation of other binding regions and perhaps how gene transactivation is further regulated or target genes defined. Here, we will utilise bioinformatic tools to predict the secondary structure of the MOZ protein C-terminus, to hypothesise regions that may have structural and functional importance. This may help to explain what molecular functions or interactions are lost/gained in the KAT6A variants associated with KAT6A GDD. It is hoped that by elucidating the link between the syndrome genotype and clinical phenotype, this will pave the way for symptom-relieving therapies in the future.

CHAPTER II

MATERIALS & METHODS

Chapter II - Materials & Methods

II. I. Sources of materials & their composition

i. General reagents

All general laboratory reagents were purchased as analytical grade quality from either Fisher Scientific or Sigma-Aldrich. Phosphate Buffered Saline (PBS) tablets from OXOID Ltd were used to prepare PBS solutions. All solutions were made in double deionised water; deionised water was purified by passage through a Millipore Synergy UV-R Ultrapure water purification system. The pH of solutions was maintained by checking the pH with a pH meter (Jenway 3510), calibrated against standards of known pH. Sterilisation was achieved either by filtration through 0.2 micron filters, or via autoclaving at 121°C for 15 minutes.

ii. Bacterial reagents

All bacteriological growth medium ingredients were purchased from OXOID Ltd. The *Escherichia coli* strain DH5α was purchased from Stratagene. X-Gal was from Promega, as was also the *E.coli* strain JM109, however all antibiotics and IPTG were purchased from VWR International or Sigma-Aldrich in powder form.

Luria-Bertani (LB) medium = 1% (w/v) tryptone, 0.5% (w/v) yeast extract and 1% (w/v) NaCl in diH_2O, made to pH 7.0 with NaOH. For solid medium, 2% (w/v) bacteriological agar was added. Autoclaved and stored at room temperature until required. Upon use, if required the desired antibiotic was added to make a final concentration of 1x.

Solid LB agar plates = solid LB agar medium was melted and the desired antibiotic was added to make a final concentration of 1x. Approximately 20 ml was poured into each petri dish to cover the surface and allowed to solidify. Plates were stored at 4ºC for up to 1 month.

Solid LB agar plates with IPTG and X-Gal = same as for solid LB agar plates, however with the addition of 100 µl of 100 mM IPTG and 20 µl of

50 mg/ml X-Gal spread over their surface and allowed to absorb for 30 minutes at room temperature prior to use.

Ampicillin = 1000x stock solution made as 100 mg/ml in 50% ethanol, 50% sterile diH_2O. Stored at -20°C.

Kanamycin = 1000x stock solution made as 50 mg/ml in sterile diH_2O. Stored at -20°C.

iii. Molecular biology reagents

Restriction digestion, kinase, phosphatase and ligase enzymes and their relevant buffers were purchased from either Roche or New England Biolabs (NEB). DNA polymerases, along with their reaction buffers, supplements and dNTPs were purchased from NEB, as were DNA ladders. Agarose gel purification kits were purchased from NEB, Nucleospin® plasmid kits (miniprep kits) from Macherey-Nagel and plasmid maxi kits (maxiprep kits) from QIAGEN Ltd, along with RNeasy mini kits. The GenElute™ mammalian genomic DNA miniprep kits were purchased from Sigma-Aldrich. Oligonucleotides were purchased from Sigma-Aldrich as lyophilised and desalted pellets. They were resuspended in sterile water to a 100 µM concentration and stored at -20 °C. cDNA synthesis kits were purchased from QIAGEN Ltd and SYBR Green reagent for qRT-PCR was from Agilent Technologies.

DNA loading dye (10x) = 3.9 ml glycerol, 500 µl 10% (w/v) SDS, 200 µl 0.5 M EDTA, 0.025 g bromophenol blue & made up to a total volume of 10 ml with diH_2O. Used at a final concentration of 1x.

Tris-borate-EDTA (TBE) = 40 mM Tris-base, 40 mM boric acid, 0.5 mM EDTA (pH8.0) and made up to 1 L with diH_2O.

iv. Biochemical reagents

The Color Prestained Protein Standard molecular weight protein marker and 0.2 µm nitrocellulose membrane were purchased from NEB and Bio-Rad, respectively. All other reagents such as 30% (w/v) acrylamide and

N,N,N,N-tetramethylethylenediamine (TEMED), were purchased from Sigma-Aldrich.

SDS-PAGE loading buffer (4x) = 2.0 ml 1 M Tris-HCl pH 6.8, 0.8 g SDS, 4.0 ml 100% glycerol, 0.4 ml 14.7 M β-mercaptoethanol, 1.0 ml 0.5 M EDTA and 8 mg bromophenol blue, made up to 10 ml with diH_2O.
Ponceau S solution = 0.1% (w/v) Ponceau S in 5% (v/v) acetic acid. Made up to a total volume of 1 L with diH_2O.
Developing solution = 1 ml Tris-HCl pH 8.5, 6 µl H_2O_2, 22 µl solution A (90 mM p-Coumari acid in DMSO), 50 µl solution B (250 mM luminol in DMSO) to a final volume of 10 ml with diH_2O.
5% milk blocking solution = 5% (w/v) Marvel 0% fat milk powder in 1x PBS.
Tris-glycine-SDS (TGS) buffer (10x) = 250 mM Tris base, 2 M glycine, 35 mM SDS, pH 8.3.
Transfer buffer = 48 mM Tris base, 39 mM glycine, 0.037% (w/v) SDS, 20% (v/v) methanol.
10% Ammonium persulphate (APS) solution = 10% (w/v) APS in diH_2O. Stored at -20°C until use.
10% Sodium dodecyl sulphate (SDS) solution = 10% (w/v) SDS in diH_2O.
Phosphate lysis buffer = 250 mM NaCl, 0.1% NP-40, 20 mM sodium phosphate (pH 7, made from mixing sodium monobasic and sodium dibasic), 30 mM sodium pyrophosphate, 10 mM sodium fluoride, 0.1 mM sodium orthovanadate (Na_3VO_4), 1 mM phenylmethylsulfonyl fluoride (PMSF) and 1x complete protease inhibitor tablet (EDTA-free, Roche), made up to a total volume of 10 ml with sterile diH_2O. Stored at 4˚C for up to 10 days.

v. Tissue culture reagents

All tissue culture reagents were purchased from Lonza Biowhittaker or Sigma-Aldrich and all plasticware was obtained from Nunc or Techno Plastic Products. Compounds and inhibitors used such as MTT, PI or NMDI14 were purchased from Sigma.

2 M $CaCl_2$ = 2 M final concentration with sterile diH_2O. Filter sterilised and stored at 4°C when in use. Stored at -20°C for long term storage.

2 X HEPES-buffered saline (HBS) = 50 mM HEPES, 280 mM NaCl & 1.5 mM Na_2HPO_4. Made up to a total volume of 100 ml with diH_2O and pH adjusted to 7.07 with 1 M HCl. Filter sterilised and stored at 4°C when in use, or at -20°C for long term storage.

MTT solution = 2 mg/ml MTT powder in 1 x PBS. Filter sterilised through a 0.2 µM filter and stored wrapped in foil at 4°C for up to two weeks.

Propidium iodide (PI) staining solution = 50 µg/ml made from a 50 mg/ml stock with 1 x PBS.

1 x Trypsin = 10X stock, diluted with PBS to 1X prior to use.

Cryopreservation medium = A mixture containing 90% FBS and 10% DMSO. Stored at 4°C for up to 1 month.

NMDI14 = Powder used at a concentration of 50 µM in DMSO. Stored at 4°C for up to 1 month.

Emetine = Powder used at a concentration of 100 µg/ml in DMSO. Stored at 4°C for up to 1 month.

Sodium butyrate = Powder used at a concentration of 1 mM in H_2O. Stored at 4°C for up to 1 week.

Carnitine = Powder used at a concentration of 45 µM in H_2O. Stored at 4°C for up to 1 week.

Pantothenic acid = Powder used at a concentration of 25 µM in H_2O. Stored at 4°C for up to 1 week.

II. II. Bacterial preparation and culture

i. Preparation of heat-competent *Escherichia coli*

Escherichia coli DH5α or JM109 cells were streaked on an antibiotic-free LB agar plate and incubated at 37°C overnight. A single colony was used to inoculate a 10 ml volume of LB medium (antibiotic-free) and cultured overnight at 37°C, with shaking. This was then used to set up a 300 ml total volume culture in the same medium and was incubated at 37°C, with shaking until the optical density (measured at 600 nm (OD_{600})) reached 0.5. From this point onwards, all work was performed at 4°C. The culture was centrifuged for 10 minutes, 3,200 x *g* at 4°C in separate 50 ml Falcon tubes and the supernatants removed. Each pellet was washed in 8.3 ml of 0.1 M $MgCl_2$, before the centrifugation was repeated. The supernatants were discarded and each pellet was resuspended in 8.3 ml of 0.1 M $CaCl_2$. These suspensions were incubated on ice for a minimum of 20 minutes, before repeating the centrifugation. The supernatants were removed and all of the pellets were resuspended together in a total of 8.6 ml of 0.1 M $CaCl_2$. 1.4 ml 100% glycerol was added to the suspension, inverted 5 times and 200 µl volumes were aliquot into microcentrifuge tubes. These were snap frozen immediately in liquid nitrogen and stored at -80°C.

ii. Bacterial transformations

Escherichia coli DH5α cells were utilised in all bacterial transformations, with the exception of pGEM-T cloning, where JM109 cells were used instead. In either case, aliquots of heat-competent cells were allowed to thaw on ice for a maximum of 10 minutes after removal from storage at -80°C. 50 ng of intact vector/plasmid, or 10 µl of ligation products (2 µl for pGEM-T cloning) was added to 50 µl of cells for each transformation. The mixtures were allowed to sit on ice for 30 minutes, then heat shocked at 42°C for 2 minutes, before being placed on ice for a further 5 minutes. 950 µl of Luria-Bertani medium (LB) was added and each sample was cultured at 37°C for 1 hour with shaking, before centrifugation for 2 minutes at 4,200 x *g*. 920 µl of each supernatant was removed, pellets

resuspended in the remaining liquid and the entire volume was plated on LB agar supplemented with the relevant antibiotic. Agar plates were incubated overnight at 37°C. For each set of transformations, a 50 µl volume of heat-competent cells was subjected to the same procedure, although without the addition of any DNA, to act as a negative transformation control. A positive control of undigested vector was also used. In the case of transforming ligation products, a negative ligation agar plate was set up in addition, which contained the transformed products of the negative ligation sample.

II. III. Molecular biology techniques

i. Annealing of oligonucleotides

100 µM CRISPR-Cas9 sgRNA oligonucleotides (20-30 nt, designed using the MIT CRISPR website but synthesised by Sigma-Aldrich), were placed separately at 37°C for 60 minutes, each in the mixture shown in Table 1 to phosphorylate the oligonucleotide, then incubated at 65°C for 20 minutes to inactivate the enzyme.

	Sample
100 µM Oligonucleotide	2 µl
10x T4 DNA ligase buffer	2 µl (to make 1x)
Polynucleotide kinase	10 Units (1 µl)
Sterile diH_2O	To a total volume of 20 µl

Table 1: Reagents and composition of oligonucleotide phosphorylation mixtures

Once complete, 5 µl of each complementary top and bottom oligonucleotide were mixed together and made up to a total volume of 50 µl with diH_2O. These mixtures were then placed in a thermocycler, which was used to heat the mixture to 95°C for 3 minutes, then reduce by 1°C each minute until a temperature of 26°C was reached, followed by 5 minutes at 25°C. The products were stored at -20°C until required.

ii. Restriction digestions

DNA was digested with the desired enzyme using the manufacturer's recommended buffer at the recommended temperature for 2 hours. Typical restriction enzyme digestions are shown in Table 2.

	Vector/plasmid	PCR product
10 x Buffer	2 µl (to make 1x)	3 µl (to make 1x)
DNA	1 µg	20 µl PCR product
Restriction enzyme	10 Units (1 µl)	
Sterile diH_2O	To a total volume of 20 µl	To a total volume of 30 µl

Table 2: Reagents used and composition of restriction digestion reactions of vectors, plasmids and PCR products

If two restriction enzymes were to be used upon the same sample, however their activity levels were not compatible in the same buffer, the procedure above was followed but then after the first digestion they were ethanol precipitated (see below). The sample was then digested with the second enzyme for 2 hours at the recommended temperature.

If digestion of a vector/plasmid resulted in compatible ends, it was dephosphorylated at the 5' end by incubation with 1 µl TSAP (1 Unit) for 1 hour @ 37°C to prevent re-ligation without the insert. The products were stored at -20°C until required.

iii. Ethanol precipitation of DNA

To concentrate DNA or precipitate it for restriction enzyme digestion in another buffer, 200% and 10% of the original sample volume of 100% ethanol and 3 M sodium acetate was added respectively. The sample was briefly vortexed and then placed at -80°C for 10 minutes. Centrifugation at 18,000 x *g* for 30 minutes at 4°C was performed, then the supernatant was discarded and the pellet resuspended in 100 µl 75% ethanol. The previous centrifugation step was repeated, the supernatant removed and excess ethanol evaporated by air drying for 10 minutes in a Savant DNA Speed Vac DNA 120. The pellet of DNA was then resuspended in a desired volume of diH_2O (17 µl for sequential restriction digests).

iv. Ligation of DNA fragments

Vectors or plasmids were ligated to linear DNA fragments after restriction enzyme digestion to produce new circular constructs containing a genetic region of interest. Typical ligation reactions are depicted in Table 3.

	Sample
Linear DNA Insert	A molar ratio of 1:3/3:1/1:1 with respect to the vector
Digested Vector	100 ng
10x T4 DNA ligase buffer	2 µl (to make 1 x)
T4 DNA ligase	3 Units (1 µl)
Sterile diH_2O	To a total volume of 20 µl

Table 3: Reagents and composition of ligation reactions of vectors with linear DNA fragments

$$\frac{\boldsymbol{ng\ of\ vector\ \times size\ of\ insert\ (kb)}}{\boldsymbol{size\ of\ vector\ (kb)}} \times \boldsymbol{molar\ ratio\ of\ \frac{insert}{vector}} = \boldsymbol{ng\ of\ insert\ to\ use}$$

Equation 1: Calculation of the necessary mass of linear DNA insert to be used in a ligation reaction with a vector

For each set of ligation reactions, a negative control was included to indicate the level of background undigested/re-ligating vector. This negative control contained all the reagents, with the exception of the linear DNA insert. The optimal molar ratio of vector : insert varied between ligation reactions and the required mass of DNA insert for each ligation reaction was calculated using Equation 1. Each sample was vortexed, centrifuged and allowed to ligate overnight at 4°C. The products were stored at -20°C until required.

v. Agarose gel electrophoresis

Analysis and visualisation of DNA was performed via agarose gel electrophoresis. Agarose gels were made by mixing the relevant mass of agarose (w/v) required for the desired percentage agarose gel with 1x TBE buffer, then heating until a clear liquid was formed. 2 µl of a 10 mg/ml ethidium bromide solution was added per 50 ml of liquid agarose, before pouring and allowing to solidify at room temperature with a well comb. 10 µl of the desired DNA ladder was loaded and all samples to analyse were mixed with the necessary volume of 10 x DNA loading dye to make a final concentration of 1 x DNA loading dye, prior to loading in the gel. All agarose gels were ran at 90 V in 1 x TBE buffer for the necessary length of time that would ensure good DNA band separation of the desired size. DNA bands were visualised and imaged using a BIO-RAD Universal Hood II UV transilluminator.

vi. Purification of DNA from an agarose gel

Agarose gels were made and ran following the procedure described above, with the exception that the samples loaded were separated by an empty lane on either side, to help prevent cross-contamination. Once complete, the desired band(s) of DNA were excised from the agarose gel and purified using the NEB Monarch DNA Gel Extraction kit, following the manufacturer's instructions. The concentration of DNA in the eluate was determined using a NanoDrop® 1000 Spectrophotometer.

vii. Polymerase chain reaction (PCR)

PCR was utilised to synthesise fragments of DNA for ligation into vectors, pGEM-T cloning of cell lines modified by CRISPR-Cas9, screening of bacterial colonies for plasmids of interest, or for generating mutant plasmids. A DNA polymerase enzyme with 3' to 5' exonuclease activity for proofreading was utilised to generate DNA fragments for cloning and plasmid mutagenesis (NEB Phusion® High-Fidelity Polymerase). However, Taq DNA polymerase (NEB) was used for screening of bacterial colonies and pGEM-T cloning. Typical reactions consisted of

the mixtures and cycles shown in Table 4 & Table 5, and were performed in the Applied Biosystems 2720 thermal cycler. Upon completion, 10% of the PCR products were separated by agarose gel electrophoresis (see above) to verify that DNA fragment(s) of the desired size had been produced. PCR products were stored at -20°C until required.

	Reagents	**Cycle**
DNA template	50 ng	95°C (30 secs) x 1
10x Standard Taq buffer	5 µl (1x)	
10 µM Forward primer	1 µl (0.2 µM)	*x 35 sequential cycles*
10 µM Reverse primer	1 µl (0.2 µM)	*95°C (30 secs)*
10 mM dNTP mix	1 µl (0.2 mM)	*55°C (1 min)*
Taq DNA polymerase	0.25 µl (1.25 units)	*68°C (1 min/Kb)*
Nuclease-free sterile diH_2O	To a final volume	
	of 50 µl	68°C (5 mins) x 1

Table 4: Reagents and composition of typical PCR reactions using Taq DNA polymerase, along with typical thermocycler settings

	Reagents	**Cycle**
DNA template	50 ng	98°C (30 secs) x 1
5x Phusion HF/GC buffer	10 µl (1x)	
10 µM Forward primer	2.5 µl (0.5 µM)	*x 35 sequential cycles*
10 µM Reverse primer	2.5 µl (0.5 µM)	*98°C (10 secs)*
10 mM dNTP mix	1 µl (0.2 mM)	*55°C (30 secs)*
Phusion® High-Fidelity DNA polymerase	0.5 µl (1 unit)	*72°C (30 secs/Kb)*
Nuclease-free sterile diH_2O	To a final volume of 50 µl	72°C (10 mins) x 1

Table 5: Reagents and composition of typical PCR reactions using Phusion DNA polymerase, along with typical thermocycler settings

viii. Generation of mutations/deletions by recombinant PCR (site-directed mutagenesis)

In order to generate a specific nucleotide mutation (or several adjacent ones) within a DNA sequence, a pair of oligonucleotides were designed that were complementary, with the exception of the desired nucleotide mutation(s). They were designed as such that the mutant nucleotide(s) were centrally located. PCR was performed as described above using Phusion DNA polymerase, however 500 ng of plasmid DNA template was used and the cycling extension time was increased to 20 minutes at 72°C (not the final extension time). This incorporated the mutation(s) and resulted in nicked, circular mutated plasmid DNA. PCR products were then treated with the Dpn1 restriction enzyme to digest the methylated, non-mutated parental DNA template for 2 hours at 37°C (Table 6). A positive control which contained the original template DNA, was included. The positive control and sample reactions were separated by agarose gel electrophoresis (see above), to check for complete digestion of the parental template. The nicked, circular dsDNA was then transformed into *E.coli* DH5α cells (see above), and the resulting single colonies were miniprepped (see below). Purified plasmid was sent for sequencing (see below) to check for incorporation of the mutated nucleotide sequence.

	Positive control	**Sample**
10x NEB Cutsmart buffer	3 µl (dilutes to 1x)	
Plasmid/PCR product	500 ng	25 µl
Dpn1 enzyme (10 units)	0.5 µl	
Sterile diH_2O	To a total volume of 30 µl	

Table 6: Composition of Dpn1 restriction digestion enzyme reactions for site-directed mutagenesis

ix. Screening of colonies via colony PCR

Bacterial colonies were picked from transformation plates and used to inoculate 5 ml volumes of LB with the relevant antibiotic, then incubated at 37°C until the culture was visibly turbid. These cultures were then used

instead of a DNA template in PCR reactions. Taq DNA polymerase was used to screen the bacterial cultures for the desired DNA insert. Typically, PCR was performed as described in Section X, however 5 µl of bacterial culture was used in the place of 50 ng DNA template. Clones positive for the desired DNA sequence were identified by agarose gel electrophoresis (Section II. III. v), performed upon 10% of the PCR products. These positive clones were then miniprepped and sent for sequencing with the relevant primer(s), to confirm the presence and orientation of the DNA sequence (Sections II. III. x & II. III. xiv, respectively).

x. Small-scale purification of vectors & plasmids from bacteria (miniprep)

For the small scale purification of a vector or plasmid from bacteria, a miniprep was performed. A 5 ml volume of LB supplemented with the relevant antibiotic was inoculated with a single bacterial colony from the relevant transformation plate. This was then allowed to grow at 37°C overnight, with shaking. 2 ml of the culture was then transferred to a 2 ml microcentrifuge tube and centrifuged at 3,200 x *g* for 10 minutes at 4°C. Once complete, the Macherey-Nagel NucleoSpin Plasmid protocol for isolation of low copy plasmids was followed and the eluate DNA concentration was determined using a NanoDrop® 1000 Spectrophotometer.

xi. Large-scale purification of vectors & plasmids from bacteria (maxiprep)

For the large scale purification of a vector or plasmid from bacteria, a maxiprep was performed. A 10 ml volume of LB supplemented with the relevant antibiotic was inoculated with a single bacterial colony from the relevant transformation plate. This was then allowed to grow at 37°C, with shaking until the culture visually became turbid. This was then used to set up a 300 ml total volume culture in the same medium and was incubated at 37°C overnight, with shaking. The Qiagen Plasmid Maxi Kit

protocol was followed, with the exception of the centrifugation steps, which were all performed at 9,000 x *g* for 50 minutes at 4°C in an ALC PK131R centrifuge. Purified DNA was re-dissolved in 200 µl diH_2O and the concentration quantified using a NanoDrop® 1000 Spectrophotometer.

xii. Purification of genomic DNA from mammalian cells

Genomic DNA was purified from mammalian cells using the Sigma GenElute™ Mammalian Genomic DNA Miniprep Kit, following the manufacturer's instructions. To harvest cells, cultures were transferred to 15 ml tubes and centrifuged at room temperature at 600 x *g* for 5 minutes. The supernatants were discarded and the pellets resuspended separately in 1 ml 1 x PBS, transferred to 2 ml microcentrifuge tubes, then the centrifugation repeated. Again, the supernatants were removed and the genomic DNA was then purified. Genomic DNA was stored at -20°C until required.

xiii. DNA concentration quantification

The quality of DNA, as well as its quantification was determined using a NanoDrop® 1000 Spectrophotometer. A ratio of sample absorbance at 260 : 280 nm was used to assess the purity of DNA; ~1.8 was accepted as pure. A secondary ratio of sample absorbance at 260 : 230 nm was also used as a measurement of DNA purity, and a range of 1.8 - 2.2 was considered as pure nucleotides. A lower value for either of these ratios was considered as an indication of the presence of contaminants in the sample and they were not used in further experiments.

xiv. Sequencing of DNA

DNA was sequenced by Source Bioscience Ltd via the Sanger Sequencing technique using an Applied BioSystems 3730 DNA Analyzer and PeakTrace™ software. The volumes and concentrations which were necessary for successful sequencing are shown in Table 7. Vectors or plasmids sequenced were taken from miniprep stocks (see above),

whereas PCR products to be sequenced were taken from agarose gel purified stocks (see above).

	Vector or plasmid	PCR product
DNA	5 µl (100 ng/µl)	5 µl (1 ng/µl per 100 bp)
Primer	5 µl (3.2 pmol/µl)	5 µl (3.2 pmol/µl)

Table 7: The volume and concentration of vectors, plasmids, PCR products and primers required for sequencing reactions.

xv. Cloning of allele PCR fragments for sequencing (pGEM-T cloning)

To sequence individual alleles from mammalian cell lines, 50 ng genomic DNA was subjected to PCR using primer pairs encompassing the genomic region of interest (see above). Upon completion, ligation reactions were set up as depicted below (Table 8) and incubated overnight at 4°C, using the pGEM-T Easy Vector kit from Promega.

	Negative control	Positive control	Sample
2x Rapid Ligation Buffer	5 µl (dilutes to 1x)		
T4 DNA Ligase (3 Weiss units/µl)	1 µl		
pGEM®-T Easy Vector (50ng)	1 µl		
PCR product	-	-	2 µl
Control Insert DNA		2 µl	-
Sterile diH_2O	To a total volume of 10 µl		

Table 8: Composition of ligation reactions for pGEM-T cloning of separate alleles from PCR products

2 µl of each ligation reaction was then transformed into 50 µl *E.coli* JM109 cells (see above) and plated onto LB-ampicillin agar plates containing 100mM IPTG and 50mg/ml X-Gal. After incubation, white transformants were subjected to colony PCR (see above) to screen for the presence of different alleles.

xvi. Purification of RNA from mammalian cells

The QIAGEN RNeasy Mini kit was used to remove genomic DNA and isolate RNA from mammalian cell pellets, ready for cDNA synthesis. Three separate cultures (biological triplicates) of K562 cells were seeded in volumes of 10 ml normal culture medium at a density of 1 x 10^5 cells/ml. After 48 hours in normal cell culture conditions, 1 x 10^6 cells were harvested from each culture by centrifugation at 600 x *g* for 5 minutes. Cell pellets were washed in 1 x PBS and the centrifugation repeated. Isolation of RNA was performed as described in the manufacturer's protocol and RNA was stored at -80°C until required. Samples were kept on ice at all times to prevent degradation of RNA transcripts.

xvii. Reverse transcription of RNA to synthesise cDNA

In order to prepare extracted RNA for use as a template to synthesise cDNA, potential contaminating genomic DNA was removed using the QuantiTect Reverse Transcription kit (QIAGEN Ltd), following the manufacturer's instructions. Samples were set up as shown in Table 9, mixed and incubated for 2 minutes at 42°C, followed by immediate storage on ice.

	Volume per sample
7x gDNA Wipeout buffer	2 µl (1x)
Template RNA	500 ng
Sterile RNase-free water	To a total volume of 14 µl

Table 9: Composition of genomic DNA removal reactions from purified RNA

To synthesise cDNA, reverse transcription reactions were prepared using the genomic DNA-free RNA samples, as shown in Table 10. These were incubated for 15 minutes at 42°C, then for 3 mins at 95°C to inactivate the Quantiscript Reverse Transcriptase enzyme. 180 µl of sterile diH_2O

was added to the products of this reaction (1:20 dilution), thoroughly mixed and samples were stored at -20˚C until required.

	Volume per sample
Quantiscript Reverse Transcriptase	1 µl
5x Quantiscript RT buffer	4 µl (1x)
RT Primer mix	1 µl
Genomic DNA removal reaction	14 µl
Total volume	20 µl

Table 10: Composition of reverse transcription reactions for cDNA synthesis

xviii. Quantitative reverse transcription PCR (qRT-PCR)

Primers to be utilised for qRT-PCR were designed via the NCBI Primer BLAST software. Parameters were set as: 18-25 nt primer length, 70-200 nt product size, optimal Tm of 59°C, at least one primer spanning an exon-exon junction (except for the analysis of KAT6A pre-mRNA), no unintended targets via mismatches and the Human Refseq mRNA database was searched. All other parameters were set as their default values. All primer sequences can be seen in Appendix III - Oligonucleotides. For each different mRNA transcript to be assessed, mastermixes were prepared as noted in Table 11. qRT-PCR analysis was performed in technical triplicates upon biological triplicates, so a total of 9 readings were obtained for each gene. Analysis of the housekeeping gene, GAPDH was also performed for every biological repeat in the same manner. Furthermore, a negative control which contained no cDNA template was also set up for each pair of primers used. All samples and reagents were kept on ice until the qRT-PCR reactions were ready.

	Volume per well
Sterile diH_2O	8.9 µl
100 µM Forward primer	0.05 µl
100 µM Reverse primer	0.05 µl
Brilliant II SYBR Green qPCR master mix	10 µl
Total volume	19 µl

Table 11: Composition of master mix for qRT-PCR reactions

After thorough mixing and distribution of the mastermixes into the reaction tubes, 1 µl of the cDNA to be analysed was added to each well. Caps were placed on the wells prior to a very brief centrifugation and then placed into a Stratagene Mx3005P qPCR machine. The reaction cycles were ran as depicted in Table 12. Analysis of the relative gene expression was performed using Equation 2 with the output Ct values.

No. Cycles	Time	Temperature
x 1	10 mins (denaturation)	95°C
x 40	30 secs (denaturation)	95°C
	1 min (primer annealing)	57°C
	1 min (extension)	72°C
x 1	1 min (melt curve analysis)	95°C
	30 secs	57°C
	30 secs	95°C

Table 12: Typical quantitative thermocycler setting for qRT-PCR experiments

$$Normalised\ mRNA\ expression\ level = 2\text{^}\Delta\text{-}Ct$$

Equation 2: Calculation used to analyse the relative gene expression of qRT-PCR targets

This formula can be explained as: 2 to the power of the negative value of (Ct of gene of interest - Ct of housekeeping gene). All primer amplification efficiencies were assumed to be 100%.

qRT-PCR products were analysed by agarose gel electrophoresis (see above) to ensure the correct sized product (and only one product) was produced in the reaction. The dissociation curves and amplification plots of each reaction were also viewed in the MXPRO software to further confirm this.

II. IV. Mammalian cell culture

iii. Culturing of adherent cell lines

The HEK293 human embryonic kidney cell line was maintained in Dulbecco's Modified Eagle Medium (DMEM). This medium was supplemented with 10% (v/v) heat-inactivated fetal calf serum (FCS), 1% (v/v) 0.2 M L-glutamine and 1% Penicillin/Streptomycin. Cells were cultured in sterile plastics at 37°C, 5% CO_2 and passaged when at approximately 80% confluency. When splitting cells, the medium was aspirated, cells washed with 1 x PBS twice, before incubation with trypsin (1X) for 3 minutes at 37°C, 5% CO_2. After incubation, fresh medium was added to neutralise trypsin and cells were seeded into fresh sterile plasticware at a density of 5 x 10^4 cells per T25 flask, or 1 x 10^5 cells per T75 flask.

iv. Culturing of suspension cell lines

The K562 human lymphoblastoid cell line was maintained in Roswell Park Memorial Institute 1640 (RPMI-1640) medium. This medium was supplemented with 10% (v/v) heat-inactivated fetal calf serum (FCS), 1% (v/v) 0.2 M L-glutamine and 1% Penicillin/Streptomycin. Cells were cultured in sterile plastics at 37°C, 5% CO_2 and passaged when cell density reached 1 x 10^6 cells/ml. When splitting cells, the medium was centrifuged at 600 x *g*, cells washed with 1 x PBS twice and resuspended in fresh medium. At this point, the cell suspension was diluted and split into other sterile plasticware at a density of 5 x 10^4 cells/ml.

v. Calcium phosphate-mediated transfection of HEK293 cells

HEK293 cells were seeded at a density of 1 x 10^6 cells per 10 cm diameter dish in 10 ml normal culture medium, or 2 x 10^4 cells in 300 µl of normal culture medium per well in an Ibidi µ-Slide 8 well chamber. In either case, cells were allowed to adhere for 24 hours after seeding. The medium was replaced 3-4 hours before transfection. For each well/dish to be transfected, a mixture was set up as in Table 13; a single positive control of an EGFP expression plasmid was also used.

	Per 10 cm diameter dish	Per well of an Ibidi µ-Slide 8 well chamber
2 M $CaCl_2$	100 µl	3.75 µl
Plasmid DNA	5 µg	500 ng
Sterile diH_2O	To a total volume of 800 µl	To a total volume of 30 µl

Table 13: Composition of transfection mixtures for calcium phosphate-mediated transfection of HEK293 cells

Transfection mixtures were added separately in a dropwise fashion to equal volumes of 2x HEPES Buffered Saline (HBS), whilst gently vortexing the aliquots of HBS. These new mixtures were allowed to stand at room temperature for 20 minutes, before being added to the relevant wells in a dropwise fashion. Cells were incubated for 16 hours at 37°C, 5% CO_2, after which they were washed three times with 1 x phosphate buffered saline (PBS). Fresh medium was then added. Cells were analysed for expression of the plasmid construct via fluorescent microscopy after a further 24 hours, then either harvested and subjected to western blotting, flow cytometry or fixed *in situ* for immunostaining.

vi. Nucleofection of K562 cells

K562 cells were transfected using a Lonza 4D-Nucleofector and the Lonza SF cell line Nucleofector kit. Two days prior to transfection, K562 cells were seeded at a density of 1 x 10^5 cells/ml in normal medium and cultured under normal conditions. For each transfection, 1 x 10^6 cells were harvested by centrifugation at 600 x *g* for 5 minutes. Cell pellets were washed in 1 x PBS and the centrifugation repeated, before being resuspended in 100 µl transfection reagent mixture and placed into a Nucleocuvette. Transfections were then performed following the Amaxa™ 4D-Nucleofector™ protocol for K562 cells. A single positive control of an EGFP expression plasmid was also performed. Once complete, cells were cultured in a total volume of 2 ml in 6-well plates

under normal conditions and transfection efficiency was observed by fluorescence microscopy after 16-24 hours.

vii. Fluorescence-activated cell sorting of mammalian cells (FACS)

Transfected K562 or HEK293 cells were isolated based on their EGFP/YFP expression (therefore successful transfection), using the Beckman Coulter MoFlo XDP 48 hours post-transfection. Cell emission at 510 nm upon excitation using a 488 nm laser was used for analysis of EGFP/YFP expression. To prepare transfected cells for FACS, they were harvested and centrifuged at 600 x *g* for 5 minutes, prior to pellet resuspension in 500 µl of fresh normal culture medium. These volumes were then transferred to separate BD Falcon FACS tubes with cell-strainer caps to remove clumps of cells and were kept on ice until ready for sorting. EGFP positive K562 cells were first separated from their non-fluorescent counterparts (bulk-sorting), then seeded as single cells into 96-well plates, with each well containing 100 µl normal cell culture medium. An extra 100 µl of normal cell culture medium was added 4-6 hours after single-cell seeding. The remaining EGFP positive K562 cells (bulk population) were seeded into 6-well plates and cultured under normal conditions until ready for analysis. On the other hand, YFP positive K562/HEK293 cells were separated from their non-fluorescent counterparts (bulk-sorted) and subjected to RNA extraction for qRT-PCR analysis. To check that EGFP/YFP enrichment had been successful, flow cytometry data were analysed using the WEASEL software (http://www.frankbattye.com.au/Weasel/).

viii. MTT assays

5×10^3 K562 cells were seeded per well in quintuplicate in a 96-well plate, in a final volume of 200 µl of normal culture medium per well. This was repeated for four 96-well plates. Five wells of normal cell culture medium only were also included as background readings. Plates were incubated under normal conditions and at 0, 24, 48 and 72 hours post-seeding, 50 µl of MTT solution was added per well (final concentration of 0.4 mg/ml)

to a single plate (one plate per time point). After a further incubation of 3 hours, the relevant plate was centrifuged at 800 x *g* for 5 minutes. The medium was aspirated from each well and 150 µl of 100% DMSO added. The plate was covered with foil and placed on a rocker for 10 minutes at room temperature, then absorbances were recorded for each well at 570 nm using a BioTek Synergy HT spectrophotometer. The average of three of the quintuplicate absorbance readings, minus the average medium only background absorbance was used for analysis.

ix. Cell cycle analysis via propidium iodide (PI)

Two days prior to analysis, K562 cells were seeded in triplicate at a density of 1 x 10^5 cells/ml in 10 ml normal medium and cultured under normal conditions. For each replicate, 1 x 10^6 cells were harvested by centrifugation at 600 x *g* for 5 minutes. Cell pellets were washed in 1 x PBS and the centrifugation repeated, before being resuspended in 500 µl 1 x PBS. Cell suspensions were gently vortexed for 5 seconds, then fixed by the addition of 4.5 ml of ice cold 70% ethanol. Samples were then incubated at 4˚C for 24 hours, prior to centrifugation at 300 x *g* for 5 minutes. Care was taken to remove as much ethanol as possible and pellets were resuspended in 500 µl of 1 x PBS. The centrifugation was repeated and pellets washed again with 1 x PBS. After repeat centrifugation and removal of supernatants, pellets were resuspended in 20 µl of Sigma RNase A solution (catalogue #R6148), followed by incubation for 2 minutes at room temperature. Finally, 400 µl of PI staining solution was added to each sample and thoroughly mixed by gently vortexing for 15 seconds. Samples were transferred to 5 ml FACS tubes and allowed to sit for 3.5 hours in the dark at 4°C. The fluorescence of each cell suspension was quantified using a Beckman Coulter FC500 flow cytometer with excitation at 536 nm and emission measured at 617 nm wavelengths.

x. Cryopreservation of mammalian cells

For adherent cell lines, the cell culture medium was aspirated and cells washed with 5 ml of sterile 1 x PBS. PBS was then removed and cells incubated with 2 ml of 1 x trypsin for 5 minutes at 37˚C. 8 ml of normal cell culture medium was added and the cell density calculated using a haemocytometer; the cryopreservation of suspension cell lines began at this stage with the calculation of cell density. 1 x 10^6 cells were centrifuged at 600 x *g* for 5 minutes at room temperature, followed by resuspension of the cell pellet in 10 ml of 1 x PBS. Centrifugation was repeated and the supernatant discarded. The cell pellet was resuspended in 1 ml of cryopreservation medium (90% FBS, 10% DMSO) and transferred to a cryovial. Cryovials were placed at -80˚C immediately in an isopropanol freezing container, then 24-72 hours later were transferred to liquid nitrogen.

xi. Resuscitation of mammalian cells after cryopreservation

Cryovials were removed from liquid nitrogen and thawed swiftly at 37°C for approximately 2-3 minutes. The entire volume of the thawed cell suspensions were transferred into T75 flasks containing pre-warmed normal cell culture medium and mixed carefully. Cells were then allowed to adhere (if adherent) and/or recover for 16-24 hours in normal culture conditions. For adherent cell lines, dead and unattached cells/debris were removed by changing the culture medium with fresh equivalent. Alternatively, suspension cells were centrifuged at 600 x *g* for 5 minutes at room temperature, then the cell pellet resuspended in 10 ml of fresh culture medium. All cell lines were passaged a minimum of three times before using them in any experiments.

xii. Treatment of K562 cells with inhibitory compounds: emetine, NMDI14, MG-132

In all experiments, biological triplicates were performed. For each triplicate, 1 x 10^6 K562 cells were harvested by centrifugation at 600 x *g* for 5 minutes at room temperature, resuspended in 5 ml 1 x PBS and the

centrifugation repeated. Cell pellets were resuspended in 2 ml of normal culture medium and transferred to 6-well plates. The required volume of the relevant compound was added to each well and incubated for the necessary amount of time, in normal culture conditions. For each set of conditions, identical samples were set up containing cells incubated with vehicle only. Once complete, cells were harvested by centrifugation at 600 x *g* for 5 minutes, then washed with 5 ml 1 x PBS and the centrifugation repeated. The supernatant was discarded and cells subjected to downstream analysis.

xiii. Fixation and nuclear staining of mammalian cells

48 hours post-transfection, HEK293 cells in an Ibidi µ-Slide 8 well chamber were washed twice with 1x PBS, then fixed in 4% PFA (in PBS) for 10 minutes, RT. Cells were again washed twice with 1x PBS and stained with DRAQ5 for 5 minutes, RT (concentration of 1/5,000 in PBS). Excess DRAQ5 was removed by washing cells twice with 1x PBS and fluorescent cells were imaged in PBS using a Zeiss LSM 510 confocal microscope. A 40x objective was used, accompanied by a digital zoom of either 0 or 2.5.

II. V. Biochemical techniques

xiv. Acquisition of protein extracts from mammalian cells

2 x 10^6 K562 cells were centrifuged at 600 x *g* for 5 minutes at room temperature and pellet resuspended in 1 x PBS. Centrifugation was repeated and the supernatant discarded. 100 µl of phosphate lysis buffer was used to resuspend the cell pellet prior to rotation at 4˚C for 1 hour. This cell lysis mixture was freeze-thawed once at -80°C, allowing it to thaw on ice for 1 hour after. Sonication for 5 minutes in a Diagenode Bioruptor Standard sonicator, using cycles of 30 seconds on, 30 seconds off on the high power setting was then performed. The resulting cell lysate was centrifuged for 1 minute at 10,600 x *g*, and the supernatant transferred to a fresh microcentrifuge tube. Protein extracts were stored at -80˚C until required, or Bradford assays were performed, followed by mixing the desired amount of protein with the relevant volume of 4 x SDS-PAGE loading buffer, and boiling for 5 minutes at 95˚C. Boiled extracts were stored at -20°C.

i. Purification of histones from mammalian cells

The Epigentek EpiQuik total histone extraction kit was used to extract and purify histones from K562 cells. 1 x 10^6 cells were harvested by centrifugation at 600 x *g* for 5 minutes at room temperature, then resuspended in 5 ml 1 x PBS and centrifuged again. The supernatant was discarded and histones extracted following the manufacturer's protocol, keeping samples on ice at all times. Once isolated, histone extracts were stored at -80°C until required, or Bradford assays were performed, followed by mixing the desired amount of protein with the relevant volume of 4 x SDS-PAGE loading buffer, and boiling for 5 minutes at 95˚C. Boiled extracts were stored at -20°C.

xv. Protein concentration determination

Protein concentrations were determined using the Bradford assay method. Reactions were set up consisting of 2 µl of protein extract, 800 µl diH_2O & 200 µl of Sigma-Aldrich Bradford dye. The samples were

mixed thoroughly and left to stand at room temperature for 15 minutes. The absorbance of each sample was then read at 595 nm and the concentration calculated relative to values of diluted Bovine Serum Albumin (BSA) standards of known concentrations (0, 2, 4, 6, 8, 10, 20 µg/µl).

xvi. Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

The resolving section of SDS-PAGE gels was prepared as shown in Table 14 (makes 2); the composition of the stacking section is shown in Table 15 (also makes 2). The relevant volume of each protein extract was loaded and ran at 150 V with 1 x TGS buffer. 3 µl of Color Prestained Protein Standard, Broad Range (NEB) was also ran alongside samples as a reference for molecular weights of proteins of interest.

	6%	**12%**	**15%**
Sterile diH_2O	10.6 ml	6.60 ml	4.6 ml
30% (w/v) Acrylamide	4.0 ml	8.00 ml	10.0 ml
1.5 M Tris pH 8.8	5.0 ml	5.00 ml	5.0 ml
10% SDS solution	0.2 ml	0.20 ml	0.2 ml
10% APS solution	0.2 ml	0.20 ml	0.2 ml
TEMED	0.016 ml	0.008 ml	0.008 ml

Table 14: Reagents and composition of SDS-PAGE gels (resolving section)

	5%
Sterile diH_2O	3.40 ml
30% (w/v) Acrylamide	0.83 ml
1.5 M Tris pH 8.8	0.63 ml
10% SDS solution	0.05 ml
10% APS solution	0.05 ml
TEMED	0.005 ml

Table 15: Reagents and composition of SDS-PAGE gels (stacking section)

xvii. Western blotting

Proteins were transferred from SDS-PAGE gels onto a nitrocellulose membrane overnight at 30 V, 4°C, immersed in high glycine transfer buffer. 1% Ponceau S solution was applied to the membrane for 1 minute, washed 3 times with diH_2O and checked by eye for successful protein transfer. Non-specific antibody binding was blocked by adding 5% Marvel milk to the membrane for 1 hour at room temperature, with agitation. The membrane was then incubated overnight at 4°C, with agitation in 5 ml 5% Marvel milk containing the primary antibody. To remove excess antibody, the membrane was washed 3 times for 5 minutes each in PBST, with agitation at room temperature, then incubated for 1 hour with agitation at room temperature with the secondary antibody. The washes with PBST were repeated to remove excess secondary antibody. Developing solution (Thermo Scientific Pierce™ ECL Western Blotting Substrate) was added to the nitrocellulose membrane and incubated at room temperature for 1 minute. The membrane was wrapped in clear plastic and excess developing solution removed. A Fujifilm LAS-4000 system was used immediately to detect and image the chemiluminescent signal. Exposure times were typically 30 seconds, 5 minutes or 15 minutes. The protein marker was also imaged as a reference of molecular weight. For the concentrations of antibodies used, see Appendix III.

II. VI. Bioinformatics

i. Design of sgRNAs

Homo sapiens KAT6A mRNA isoform 1 (NM_006766.4) was used as an input for the MIT CRISPR Design Tool (http://crispr.mit.edu/), specifically 250 nucleotides centred on the desired target codon. Outputs were then assessed based on their score values, number of predicted off target effects and binding locations. This was repeated independently for both targeted KAT6A exons.

ii. qRT-PCR primer design

Either the gene accession number, or the specific region of mRNA sequence desired to be quantified was used as an input for the NCBI Primer BLAST tool (Ye *et al.*, 2012). The *Homo sapiens* Refseq mRNA database was specified and results selected so that at least one primer span an exon-exon junction, to assess spliced mRNA levels. To assess pre-mRNA levels, criteria were changed to specify that both primers must not span any exon-exon junctions. Outputs of the tool were judged by Tm value, self-complementarity ability and target gene specificity. Upon first using each pair of primers, qRT-PCR products were assessed via agarose gel electrophoresis to visually confirm amplification had occurred (or not), as well as amplicons of a single expected size. Melt curve analysis was also performed.

iii. Splice site analysis

The NNSPLICE tool (Reese *et al.*, 1997) was used with inputs of the mutant K562 KAT6A alleles, in each case incorporating both the mutant locus and 150 nucleotides on either side. Both 5’ and 3’ splice sites were assessed and all other setting left as default. No new significant donor or acceptor sites were predicted.

iv. RNA sequencing data analysis

Raw datasets were downloaded from the Sequence Read Archive (SRA) or Gene Expression Omnibus (GEO) databases, then adapters and low

quality reads were removed using the TrimGalore script, retaining those with a phred score of >30. RNA reads were aligned to the human genome build GRCh38.83 using STAR and the FeatureCounts tool was utilised to quantify gene expression. To identify differentially expressed genes, we used the RobinA function of EdgeR with a false discovery rate (FDR) of <0.05, comparing KAT6A GDD patient gene expression to that observed in control samples.

v. Gene Ontology analysis

Gene Ontologies (GO) were identified by Overrepresentation Enrichment Analysis (ORA) using the WebGestalt tool (Wang *et al.*, 2017). For this, our RNA-seq-derived gene lists were compared to the predefined reference set of “genome_protein-coding”, utilising the genoeontology, pathway and phenotype databases within WebGestalt. Biological_Process, Molecular_Function, Cellular_Component, KEGG, Panther and Human_Phenotype_Ontology functional analyses were used within these databases. Outputs from these were accrued in the form of smaller gene lists associated with the relevant GO term. An FDR cut off of <0.05 was used.

vi. **Protein structure prediction**

The amino acid sequence of the C-terminal region of MOZ which incorporates the acidic, S, PQ and M domains, namely amino acids 788-2004, was provided as an input for the Phyre2 (Kelley *et al.*, 2015) and RaptorX (Wang *et al.*, 2016) protein structure prediction software. For Phyre2, the intensive modelling option was selected, however all other options were left as default, as was the same for RaptorX.

CHAPTER III

RESULTS

Chapter III - Results

III. I Structure prediction of MOZ C-terminal regions

There is a distinct lack of scientific knowledge regarding the functional importance of the MOZ C-terminal regions. One way in which to partially elucidate this, is to determine its secondary & tertiary structure, allowing the prediction of interaction interfaces. To begin, we compared MOZ to MORF, its paralogue with which it shares a high degree of amino acid sequence identity. We utilised the NCBI Protein BLAST tool to align both human MOZ and MORF amino acid sequences for homology determination. Their homology is noted to be higher within the N-terminal domains of these HATs, such as in the DPF and MYST domains (89% & 88%, respectively, Figure 14), although the C-terminal regions do have substantial homology. For example, the S- and M-rich domains have a sequence identity of 78% and 77%, respectively, despite the acidic domains sharing only 43% (Figure 14). This suggests some potential functional overlap between the higher sequence identity sharing domains and it is possible they have the same/similar binding partners. Indeed, this appears to be true for their ability to bind with RUNX1 and RUNX2 to regulate target gene transcription (Pelletier *et al.*, 2002). Despite these two MYST HATs having such large C-termini that appear to lack domains with known functions, it is highly likely these regions do indeed possess biological functions, as they have been somewhat conserved through evolution (Yang and Ullah, 2007). Intriguingly, MORF does not possess a P/Q-rich domain, suggesting that whatever function this confers to MOZ, it is unique to this paralogue.

Assessment of the MOZ C-terminal amino acid residues mutated in patients with KAT6A GDD syndrome, revealed that many of these residues are conserved in vertebrates, indicating they may have an important role, perhaps as part of an interaction interface or facilitating protein conformation (Table 16). Amino acid residue conservation between species was performed using the primary MOZ isoform for each indicated species and the Multalin tool (Corpet, 1988).

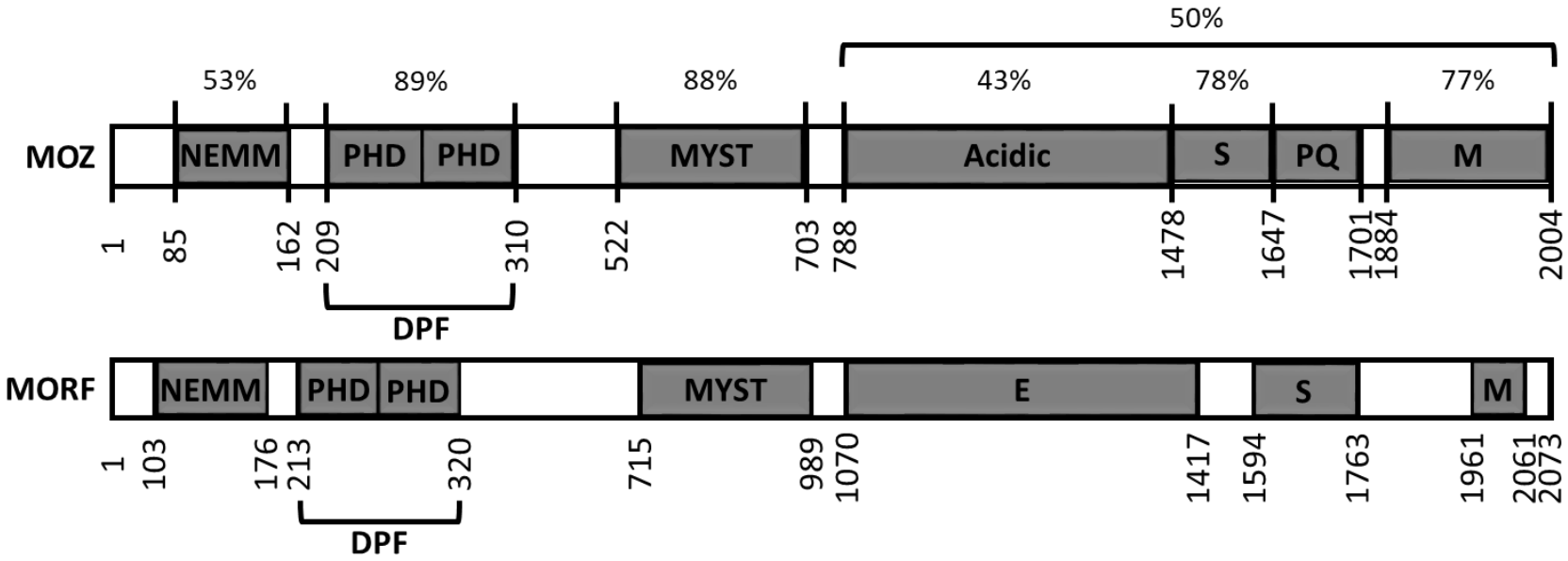


Figure 14: Shared primary amino acid identity between the paralogues MOZ & MORF

Schematic showing the relative amino acid identity shared between each domain of human MOZ and MORF. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2. MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. Acidic residue-rich (Acidic), serine-rich (S), glutamic acid-rich (E), proline and glutamine-rich (PQ) and methionine-rich (M) regions are also indicated. Also shown is the identity shared between the two whole C-termini. Note that MORF does not possess a PQ-rich domain.

Human	Chimpanzee	Cat	Dog	Horse	Mouse	Rat	Chicken	Sheep	Guinea Pig	Pig	Zebrafish
Arg1024 exon 16	Arg1024 exon 15	Arg1023 exon 15	Arg1023 exon 15	Arg1026 exon 17	Arg1024 exon 17	Arg1023 exon 15	Arg1028 exon 16	Arg1040 exon 16	Arg1022 exon 15	Arg910 exon 16	Arg1079 exon 15
Ser1039 exon 16	Ser1039 exon 15	Ser1038 exon 15	Ser1038 exon 15	Ser1041 exon 17	Ser1039 exon 17	Ser1038 exon 15	Ser1043 exon 16	Ser1055 exon 16	Ser1037 exon 15	Ser925 exon 16	Ser1094 exon 15
Asn1077 exon 16	Asn1077 exon 15	Asn1076 exon 15	Asn1076 exon 15	Asn1079 exon 17	Ser1078 exon 17	Asn1076 exon 15	Asn1086 exon 16	Ser1093 exon 16	Asn1075 exon 15	Region missing	Region missing
Cys1096 exon 16	Cys1096 exon 15	Cys1095 exon 15	Cys1095 exon 15	Cys1098 exon 17	Cys1097 exon 17	Cys1095 exon 15	Asn1105 exon 16	Cys1112 exon 16	Cys1094 exon 15	Cys966 exon 16	Lys1134 exon 15
Arg1129 exon 17	Arg1129 exon 16	Arg1128 exon 16	Arg1128 exon 16	Arg1131 exon 18	Arg1131 exon 18	Arg1130 exon 16	Arg1135 exon 17	Arg1145 exon 17	Arg1127 exon 16	Arg999 exon 17	Arg1167 exon 16
Gly1155 exon 17	Gly1155 exon 16	Gly1154 exon 16	Gly1154 exon 16	Gly1157 exon 18	Gly1157 exon 18	Gly1156 exon 16	Gly1161 exon 17	Gly1171 exon 17	Gly1153 exon 16	Gly1025 exon 17	Gly119, exon 16
Glu1370 exon 17	Glu1370 exon 16	Glu1368 exon 16	Glu1368 exon 16	Glu1370 exon 18	Glu1364 exon 18	Glu1363 exon 16	Glu1378 exon 17	Glu1386 exon 17	Glu1292 exon 16	Gln1240 exon 18	Asp1516 exon 16
Leu1431 exon 17	Leu1431 exon 16	Leu1429 exon 16	Leu1429 exon 16	Leu1431 exon 18	Leu1425 exon 18	Leu1424 exon 16	Leu1439 exon 17	Leu1447 exon 17	Leu1353 exon 16	Leu1294 exon 18	Leu1620 exon 16
Tyr1563 exon 17	Tyr1563 exon 16	Try1561 exon 16	Tyr1561 exon 16	Try1563 exon 18	Tyr1557 exon 18	Tyr1556 exon 16	Tyr1571 exon 17	Tyr1579 exon 17	Try1485 exon 16	Tyr1426 exon 18	Tyr1778 exon 16

Table 16: Amino acid identity shared between vertebrates at MOZ loci associated with KAT6A GDD syndrome

A selection of amino acid residue positions associated with KAT6A GDD syndrome in human patients are shown (black) and their corresponding residue and position in a selection of vertebrates are also displayed. Those in green represent a shared identity with the human residue, whereas those in red are either not homologous or the respective region is not encoded in these orthologues. Data for this analysis was acquired by alignment of primary isoform MOZ sequences using the *Multalin* online tool (Corpet, 1988).

Using the Phyre2 protein secondary structure prediction software to assess likely α-helical or β-sheet structures in the MOZ C-terminus (aa 788-2004), several possible α-helices were proposed towards the corresponding 5' end of the acidic region, with a medium confidence level (Kelley *et al.*, 2015). The more 3' regions of the MOZ C-terminus were predicted to be almost exclusively disordered and no regions were predicted to form β-sheets (Figure 15). The RaptorX software was also utilised to predict likely secondary structures using the same input sequence, with some α-helical regions overlapping that of the Phyre2 output, most notably amino acids ~1-15 and ~110-130 (Figure 15 & Figure 16) (Wang *et al.*, 2016). Alternatively, RaptorX also predicted several α-helical structures at the 3' end of the acidic domain with high confidence, which Phyre2 was unable to do (Figure 16). Furthermore, in those regions reported in the literature to correspond to RUNX1/2 and PML interaction sites, RaptorX was able to predict the presence of α-helical and β-sheet structures, suggesting these smaller structural regions may form part of the interaction interfaces. Despite a higher prediction/more frequent prediction of secondary structures by RaptorX, compared to Phyre2, there were still many large stretches of amino acid sequences that could not be successfully predicted to form either α-helices or β-sheets with any high degree of confidence, for either software. This caused a large proportion of the C-terminal region of MOZ to be classified as disordered in secondary structure by both tools. Indeed, they suggest that overall, the C-terminal half of the MOZ protein is disordered in nature.


Sequence
Secondary structure
SS confidence
Disorder
Disorder confidence


Sequence
Secondary structure
SS confidence
Disorder
Disorder confidence



Figure 15: Phyre2 secondary structure & disorder prediction

Secondary structure prediction for the human MOZ C-terminal amino acid residues 788-2004, using the Phyre2 software (Kelley *et al.*, 2015). Green spirals represent α-helices, blue arrows represent β-sheets, coloured ribbon below these denote confidence of prediction. A yellow, orange or red colouration is indicative of a high confidence of the secondary structure prediction at those loci. Prediction of the disorder of the structure is also shown below secondary structure confidence, with regions of disorder indicated by "?". Confidence of disorder prediction is denoted by a coloured ribbon, following the same key as that for secondary structure prediction. Note that >99% of the MOZ C-terminus is predicted to be disordered in nature, despite some regions of potential secondary structures.

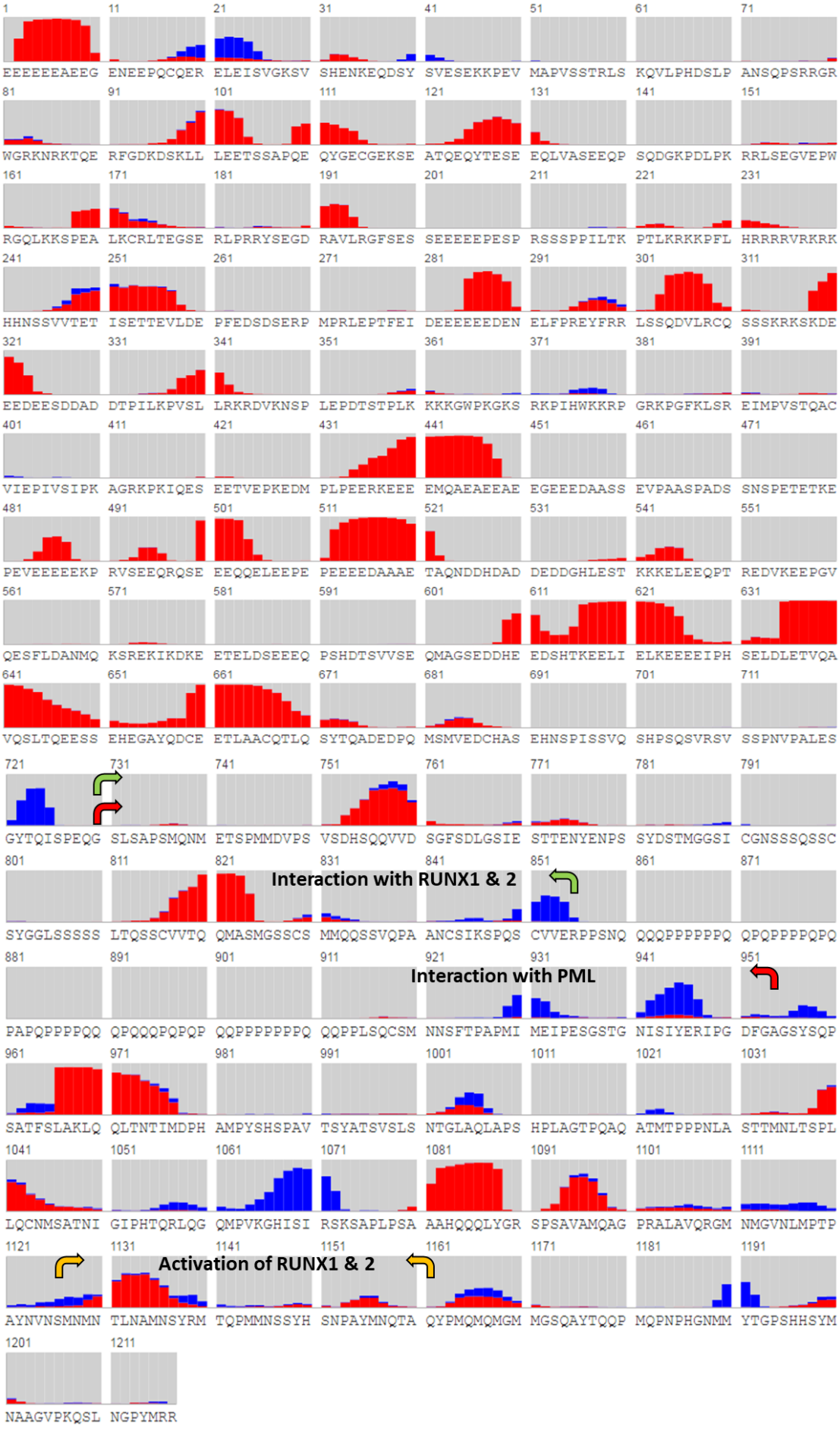

1
11
21
31
41
51
61
71
EEEEEEAEEG
ENEEPQCQER
ELEISVGKSV
SHENKEQDSY
SVESEKKPEV
MAPVSSTRLS
KQVLPHDSLP
ANSQPSRRGR
81
91
101
111
121
131
141
151
WGRKNRKTQE
RFGDKDSKLL
LEETSSAPQE
QYGECGEKSE
ATQEQYTESE
EQLVASEEQP
SQDGKPDLPK
RRLSEGVEPW
161
171
181
191
201
211
221
231
RGQLKKSPEA
LKCRLTEGSE
RLPRRYSEGD
RAVLRGFSES
SEEEEEPESP
RSSSPPILTK
PTLKRKKPFL
HRRRRVRKRK
241
251
261
271
281
291
301
311
HHNSSVVTET
ISETTEVLDE
PFEDSDSERP
MPRLEPTFEI
DEEEEEEDEN
ELFPREYFRR
LSSQDVLRCQ
SSSKRKSKDE
321
331
341
351
361
371
381
391
EEDEESDDAD
DTPILKPVSL
LRKRDVKNSP
LEPDTSTPLK
KKKGWPKGKS
RKPIHWKKRP
GRKPGFKLSR
EIMPVSTQAC
401
411
421
431
441
451
461
471
VIEPIVSIPK
AGRKPKIQES
EETVEPKEDM
PLPEERKEEE
EMQAEAEEAE
EGEEEDAASS
EVPAASPADS
SNSPETETKE
481
491
501
511
521
531
541
551
PEVEEEEEKP
RVSEEQRQSE
EEQQELEEPE
PEEEEDAAAE
TAQNDDHDAD
DEDDGHLEST
KKKELEEQPT
REDVKEEPGV
561
571
581
591
601
611
621
631
QESFLDANMQ
KSREKIKDKE
ETELDSEEEQ
PSHDTSVVSE
QMAGSEDDHE
EDSHTKEELI
ELKEEEEIPH
SELDLETVQA
641
651
661
671
681
691
701
711
VQSLTQEESS
EHEGAYQDCE
ETLAACQTLQ
SYTQADEDPQ
MSMVEDCHAS
EHNSPISSVQ
SHPSQSVRSV
SSPNVPALES
721
731
741
751
761
771
781
791
GYTQISPEQG
SLSAPSMQNM
ETSPMMDVPS
VSDHSQQVVD
SGFSDLGSIE
STTENYENPS
SYDSTMGGSI
CGNSSSQSSC
801
811
821
831
841
851
861
871
Interaction with RUNX1 & 2
SYGGLSSSSS
LTQSSCVVTQ
QMASMGSSCS
MMQQSSVQPA
ANCSIKSPQS
CVVERPPSNQ
QQQPPPPPPQ
QPQPPPPQPQ
881
891
901
911
921
931
941
951
Interaction with PML
PAPQPPPPQQ
QPQQQPQPQP
QQPPPPPPPQ
QQPPLSQCSM
NNSFTPAPMI
MEIPESGSTG
NISIYERIPG
DFGAGSYSQP
961
971
981
991
1001
1011
1021
1031
SATFSLAKLQ
QLTNTIMDPH
AMPYSHSPAV
TSYATSVSLS
NTGLAQLAPS
HPLAGTPQAQ
ATMTPPPNLA
STTMNLTSPL
1041
1051
1061
1071
1081
1091
1101
1111
LQCNMSATNI
GIPHTQRLQG
QMPVKGHISI
RSKSAPLPSA
AAHQQQLYGR
SPSAVAMQAG
PRALAVQRGM
NMGVNLMPTP
1121
1131
1141
1151
1161
1171
1181
1191
Activation of RUNX1 & 2
AYNVNSMNMN
TLNAMNSYRM
TQPMMNSSYH
SNPAYMNQTA
QYPMQMQMGM
MGSQAYTQQP
MQPNPHGNMM
YTGPSHHSYM
1201
1211
NAAGVPKQSL
NGPYMRR

Figure 16: Secondary structure prediction using RaptorX

Secondary structure prediction for the human MOZ C-terminal amino acid residues 788-2004, using the RaptorX software (Wang et al. 2016). Red columns represent predicted α-helices, blue columns depict predicted β-sheets and grey regions are unknown structures. The height of these columns shows the confidence of this predicted structure at this position. Higher columns = higher confidence of the predicted secondary structure. The reported interaction region with RUNX1 & RUNX2 is within green arrows, the PML interaction region is encased by red arrows and the yellow arrows represent the region of MOZ required for RUNX1 & RUNX2-mediated gene regulation, as documented in the literature. Note that there are several regions which have very high confidence in the prediction of α-helical structures; these correspond to regions of the MOZ acidic domain.

III. II Design and cloning of sgRNAs to target exons 17/18

The K562 cell line is a human erythroleukemia suspension cell line which was originally isolated from a 53 year old Caucasian female with chronic myelogenous leukemia, and which have the morphology of lymphoblasts (Lozzio and Lozzio, 1975). Some K562 cells are found to be triploid, however their model chromosome number is 67; therefore, they have a hypotriploid karyotype. MOZ resides in the cytogenetic band 8p11.21, for which the vast majority of K562 cells are reported to be triploid (3 copies of the complete chromosome 8) (Naumann *et al.*, 2001; Zhou *et al.*, 2018). We have also confirmed in previous work that this cell line is indeed triploid for the KAT6A gene (Monteiro 2018, PhD thesis - unpublished). Due to their amenability to genome editing by CRISPR-Cas9, the ease with which they can be studied and cultured and their well characterised genotype/gene expression profiles, we selected the K562 cell line for use in our studies of KAT6A C-terminal function and KAT6A GDD molecular pathogenicity (Lloyd, 2018). Moreover, we opted to utilise the CRISPR-Cas9 nickase approach discussed previously, to reduce the likelihood of “off-target effects” during genome editing.

Following DNA cleavage by Cas9/Cas9n, and the recognition of a dsDNA break by cellular machinery, DNA repair occurs by one of two mechanisms: non-homologous end joining (NHEJ), or homology-directed repair (HDR), with the latter occurring at a much lower frequency compared to the former (Figure 17C) (Ran *et al.*, 2013; Kim and Kim, 2014; Renaud *et al.*, 2016). To generate cell lines harbouring KAT6A variants associated with KAT6A GDD, we sought to utilise the HDR pathway in our CRISPR-Cas9n approach. Firstly, we designed pairs of sgRNAs targeting either exon 17 or exon 18 of the KAT6A gene (NM_006766.4), using the MIT CRISPR Design Tool and inputs of 250 nucleotides centred around the KAT6A variants 1039* (exon 17) or 1129* (exon 18). From the output of this search, we chose a single pair of sgRNAs for targeting each exon which had a high quality combined score, as well as high individual sgRNA scores and predicted off target effects of zero in the human genome (Figure 18A & B). Henceforth, the

sgRNAs chosen are referred to as sgRNAs E & F for the pair which target KAT6A exon 17, whereas those which target exon 18 are I & J.

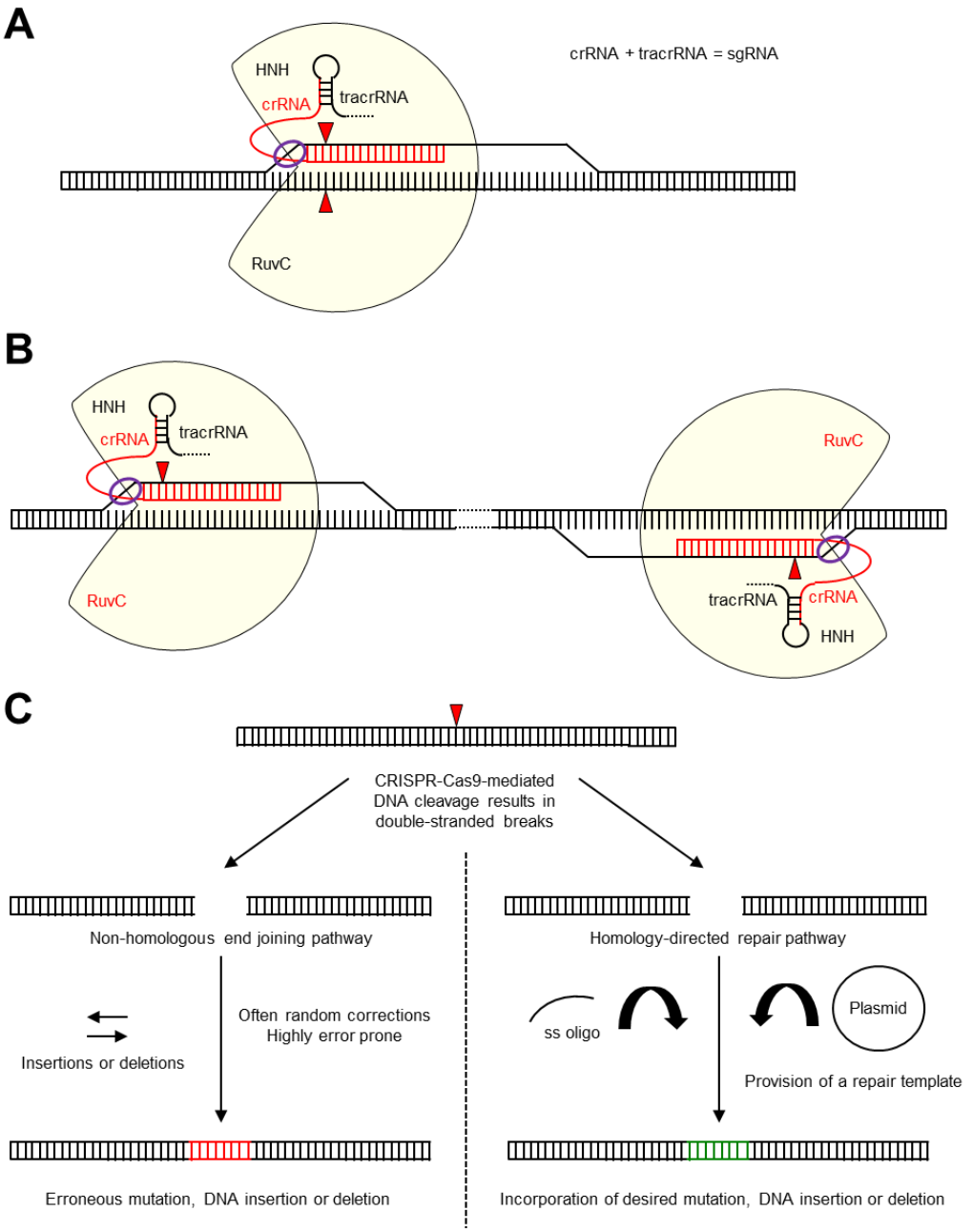
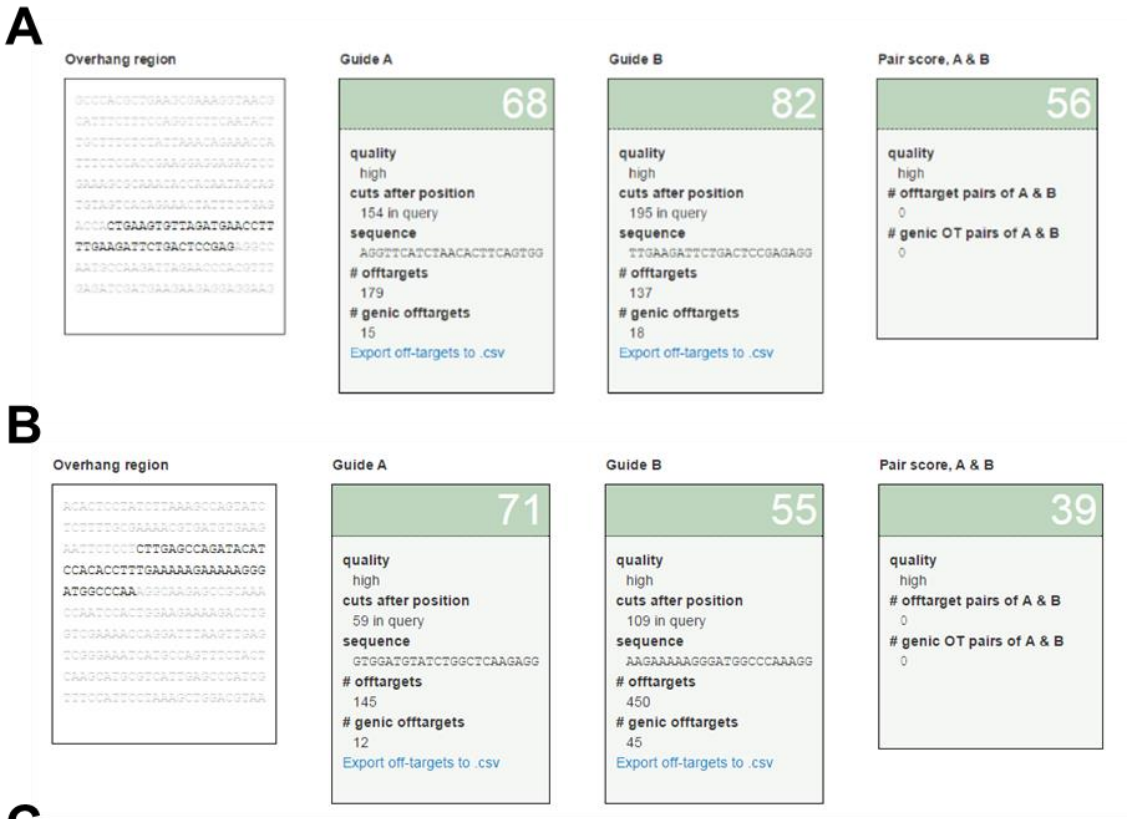


Figure 17: CRISPR-Cas9 and Cas9n-mediated genome editing

CRISPR-Cas9 **(A)** and CRISPR-Cas9 nickase (Cas9n) **(B)** approaches for genome editing at a locus of interest. Note that the Cas9n RuvC endonuclease domain is inactive. Red crRNA and black tracrRNA combined represent a complete sgRNA. Purple ovals indicate the position of the PAM. Red arrows show the position of endonuclease activity by the respective HNH or RuvC catalytic domain. **(C)** NHEJ and HDR DNA repair mechanisms after engineered dsDNA cleavage.



C

KAT6A Guide E

Top 5′-CACCGAGGTTCATCTAACACTTCAG-3′
Bottom 3′-CTCCAAGTAGATTGTGAAGTCCAAA-5′

KAT6A Guide F

Top 5′-CACCGTTGAAGATTCTGACTCCGAG-3′
Bottom 3′-CAACTTCTAAGACTGAGGCTCCAAA-5′

KAT6A Guide I

Top 5′-CACCGTGGATGTATCTGGCTCAAG-3′
Bottom 3′-CACCTACATAGACCGAGTTCCAAA-5′

KAT6A Guide J

Top 5′-CACCGAAGAAAAAGGGATGGCCCAA-3′
Bottom 3′-CTTCTTTTTCCCTACCGGGTTCAAA-5′

Figure 18: Selection of sgRNAs for targeting KAT6A exons 17/18

Information for sgRNA pairs chosen to target KAT6A **(A)** exon 17, or **(B)** exon 18, including individual and combined scores as well as number of predicted off target binding sites. **(C)** Chosen sgRNA sequences converted to annealed oligonucleotides for cloning into the Cas9n-encoding plasmid. Green represents the designed sgRNA, yellow is an additional “G” added to enhance transcription from the U6 promoter and underlined regions are overhangs required for ligation into the BbsI-digested pX461 vector.

sgRNAs E & I target the forward DNA strand at their designated loci, and F & J target the reverse DNA strand. These 20 nucleotide sgRNAs were converted into oligonucleotides with complementary reverse strands and for those whose 1st nucleotide was not guanine, a single guanine was added to the 5' end (Figure 18C). This was done due to reports that a guanine at the 5' end of a sgRNA enhances its transcription when under the control of a U6 promoter, which the Cas9n plasmid pX461 contains (Iijima *et al.*, 2006; Chiang *et al.*, 2016). In addition to this, "5'-CACC-3'" and "5'-AAAC-3'" sequences were added to the 5' ends of the respective forward and reverse sgRNA oligonucleotide sequences, to allow for cloning into a BbsI-generated overhang in the Cas9n plasmid (Figure 18C). These complementary oligonucleotides were commercially synthesised and individually phosphorylated, then mixed and allowed to anneal in a thermocycler. Simultaneously, the pSpCas9n(BB)-2A-GFP (pX461, (Ran *et al.*, 2013)) vector (Figure 19) was digested with the BbsI restriction enzyme, purified from an agarose gel and then ligated separately with each double-stranded overhang sgRNA duplex, as shown in Figure 20. Ligation products were transformed into *Escherichia coli* DH5α cells, plasmid DNA was purified from transformants and then sequenced with a human U6 forward primer to confirm the presence of the desired sgRNA (Figure 21). These plasmids were then planned to be transfected into K562 cells, in their sgRNA pairs to facilitate mutagenesis at the desired loci and single cells isolated for clonal expansion (Figure 20). The position of these sgRNA pairs for targeting each exon in relation to KAT6A GDD-associated variants and the overall KAT6A gene architecture is depicted in Figure 22A & B.

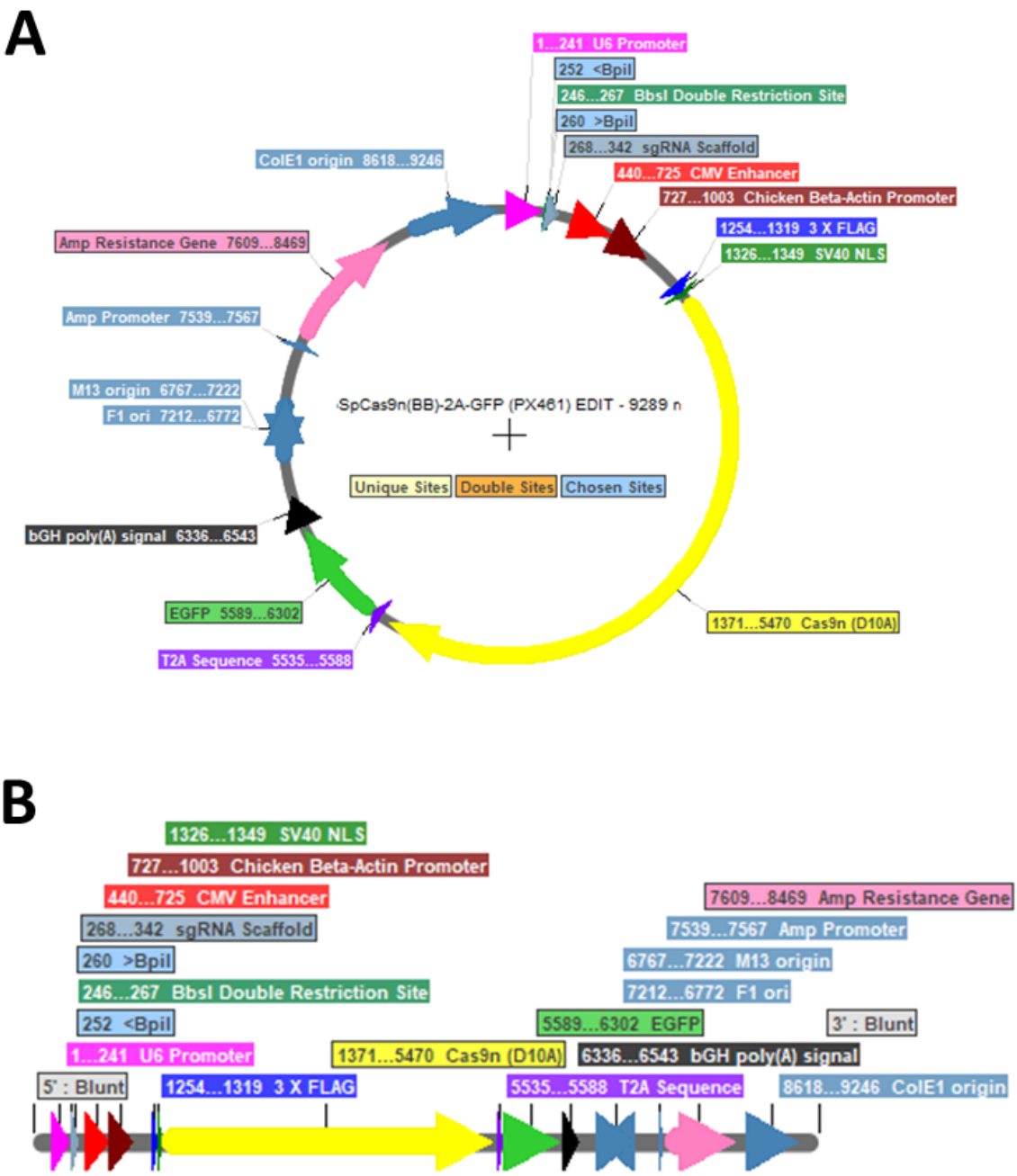


Figure 19: Map of the CRISPR-Cas9n vector pX461

Schematic of the pSpCas9n(BB)-2A-GFP (pX461) vector in **(A)** circular or **(B)** linear form. The direction of transcription is depicted by arrows, with important features highlighted in different colours; relative nucleotides encoding these features are numbered. The two sequential BbsI (BpiI) restriction sites for sgRNA cloning are also labelled. Note the FLAG-tagged Cas9n enzyme and EGFP are encoded within a bicistronic construct (separated by a T2A sequence), and that the sgRNA is expressed from a separate promoter.

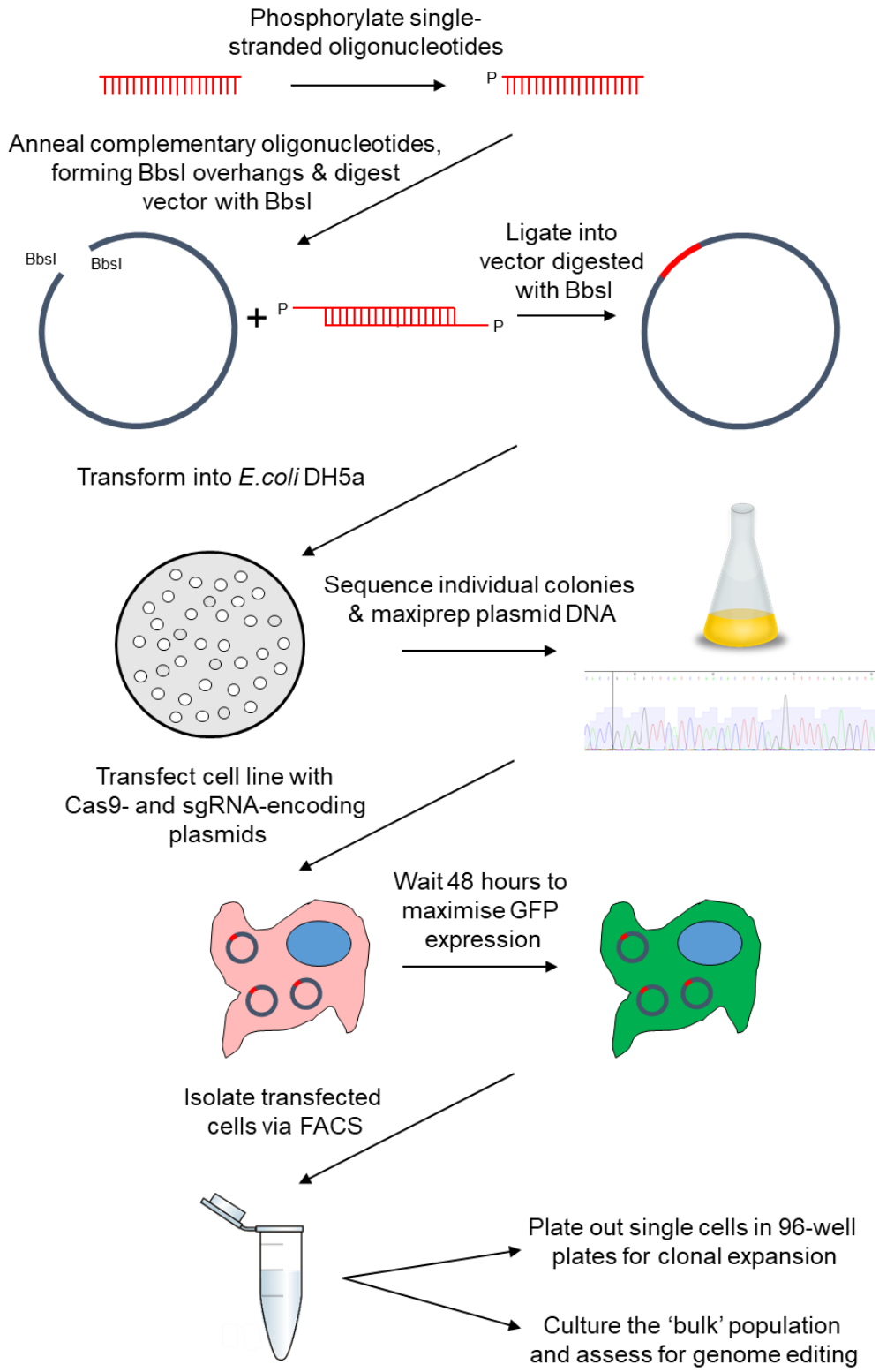


Figure 20: Workflow to generate KAT6A variant cell lines using CRISPR-Cas9n

Planned experimental workflow from phosphorylation, annealing and cloning of Cas9n sgRNAs which target KAT6A exon 17 or 18 into the Cas9n-encoding vector, through to cell transfections, FACS and single cell clonal expansion of genome-edited K562 cells.

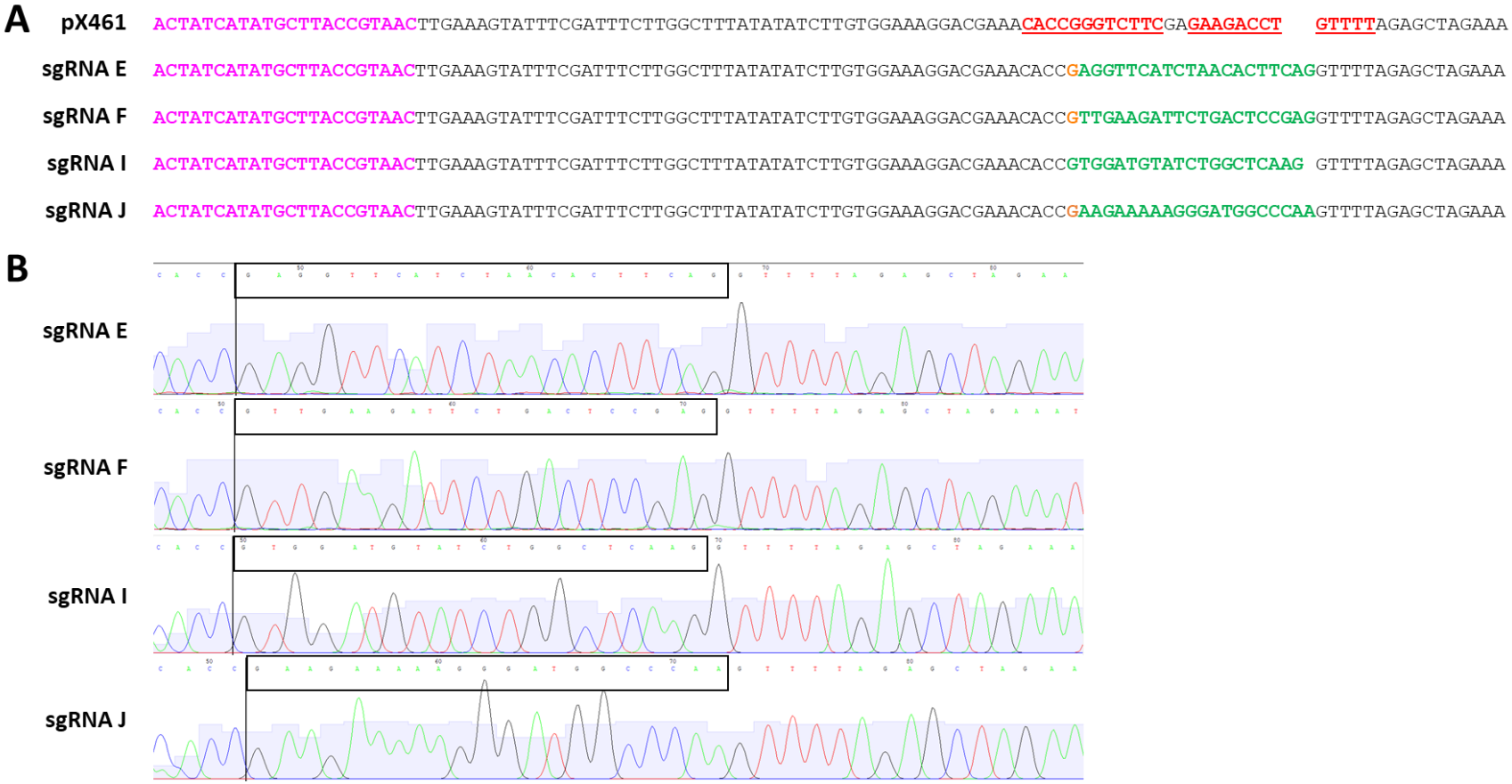


Figure 21: Sequencing results of the KAT6A exon 17 or 18-targeting sgRNAs cloned into the pX461 vector.

(A) Aligned nucleotide sequences of these plasmids, relative to the original vector backbone. Red is the double BbsI restriction enzyme site, green are sgRNAs with an extra “G” to enhance transcription from the U6 promoter shown in yellow. The U6 promoter sequence itself is in pink. **(B)** Electropherograms of sequencing results with the successfully cloned sgRNAs indicated in rectangles.

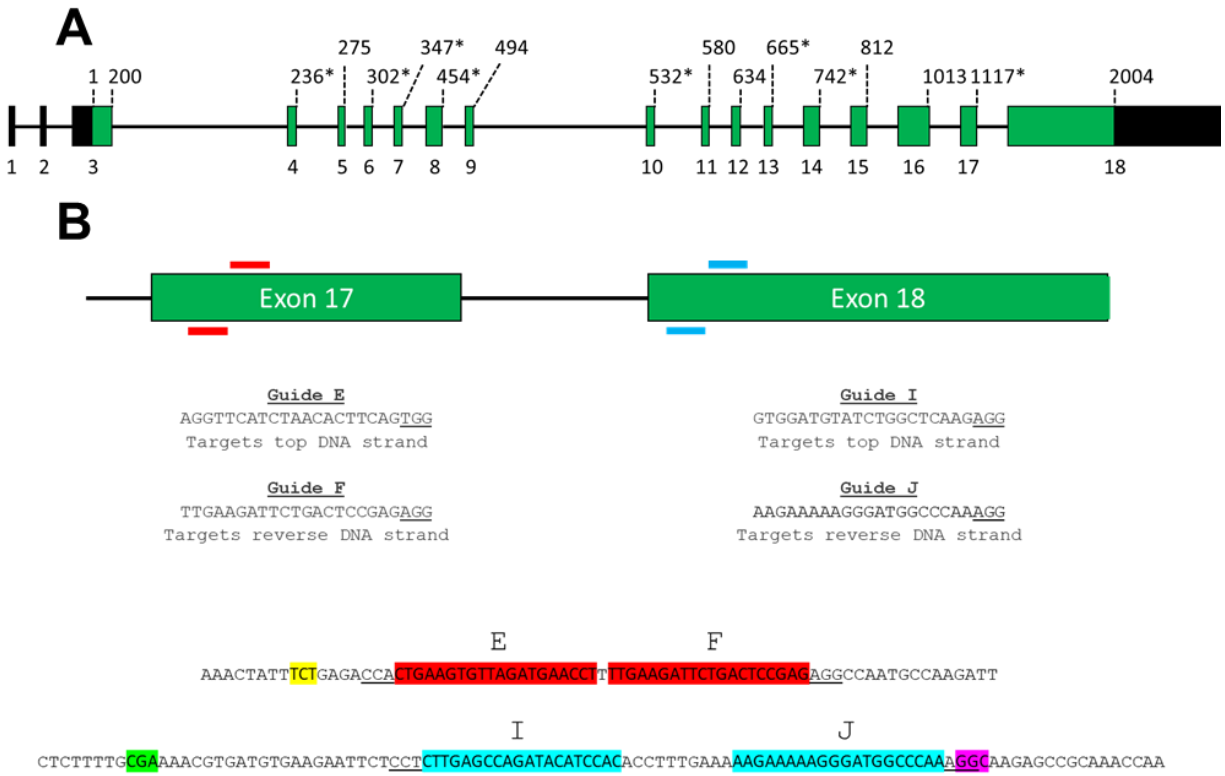


Figure 22: sgRNA-targeted loci within KAT6A exons 17/18 & design of DNA repair template

(A) Position of KAT6A introns/exons over the whole gene, with cumulative amino acid number shown. * = final codon spans two exons. **(B)** Regions of exons 17 and 18 targeted by sgRNAs (red/blue), relative to nearby KAT6A GDD variant positions. Yellow is Ser1039, green is Arg1129, pink is Gly1155. **(C)** Alignment of the DNA repair template for generation of the Ser1039* mutation via HDR. Red is the Ser1039/Ser1039* codon, pink is the non-native EcoRI restriction site and green/blue regions are the sgRNA E/F sequences, respectively. PAM sequences are underlined.

In addition to the above experimental design, to make use of the HDR pathway during CRISPR-mediated genome engineering, a single stranded oligonucleotide DNA repair template was created in the hope that transfected cells would use it to repair the targeted KAT6A locus. This was only done for exon 17, not exon 18 in order to try its efficiency. The 150 nucleotide repair template is depicted in Figure 22C and was commercially synthesised, according to our design based upon suggestions in the literature (Chen *et al.*, 2011; Ran *et al.*, 2013). It was constructed so that the codon encoding Ser1039 would be altered to a PTC and a single PAM sequence required by Cas9n for DNA cleavage was destroyed. This would thereby incorporate our desired mutation, creating the Ser1039* KAT6A GDD-associated variant, as well as also preventing continuous cleavage by Cas9n at the nearby locus, stopping the formation of dsDNA breaks and further mutagenesis. Furthermore, the PAM sequence in question was altered in such a way that it created a non-native EcoRI restriction site, allowing for screening of clones at a later stage and selection of those which carry at least one copy of this particular variant. In contrast to this approach, we chose to rely upon the higher frequency occurrence of DNA insertions or deletions (indels) when performing CRISPR-Cas9 genome editing of KAT6A exon 18, which results in a high chance of frameshift mutations and consequent incorporation of PTCs. This was primarily performed due to the reported low efficiency of success with HDR in conjunction with CRISPR-Cas9 (Ran *et al.*, 2013; Chiang *et al.*, 2016).

III. III Generation of KAT6A variant cell lines

To generate K562 clones which carry mutant KAT6A alleles, the plasmids encoding Cas9n and sgRNAs targeting either exon 17 or 18 were transfected into this cell line, using Lonza nucleofection and maxiprepped DNA stocks as described in Section II. IV. vi. Plasmids were transfected in their corresponding sgRNA pairs, however a single sgRNA alone was also transfected as a control to check that mutagenesis cannot occur without both sgRNAs being present with Cas9n. For this control, sgRNA E was used, which targets KAT6A exon 17 on the forward DNA strand. Additionally, for those cells which were transfected with plasmids encoding sgRNAs targeting KAT6A exon 17, the DNA repair template was co-transfected alongside. This was also done for the single sgRNA control transfection. In all cases, the transfections followed the workflow depicted previously in Figure 20, whereby cells successfully expressing EGFP 48 hours post-transfection were isolated from the mixed cell population via fluorescence-activated cell sorting (FACS), then plating of these EGFP-enriched cell pools (referred to as the “bulk” population) as individual cells in a total of three 96-well plates per transfection population (288 single cells). As can be seen in Figure 23 & Figure 24, there was a clear increase in emission at the expected bandwidth 48 hours post-transfection, relative to WT untransfected K562 cells. This was drastically enriched after two rounds of cell sorting to isolate EGFP-positive cells and it can be seen that particles which were not expressing EGFP, or were only weakly expressing EGFP were mostly removed from the retained population. These analyses demonstrate that both the transfections with CRISPR-Cas9n plasmids targeting the KAT6A gene and the subsequent isolation of cells expressing EGFP, were successful.

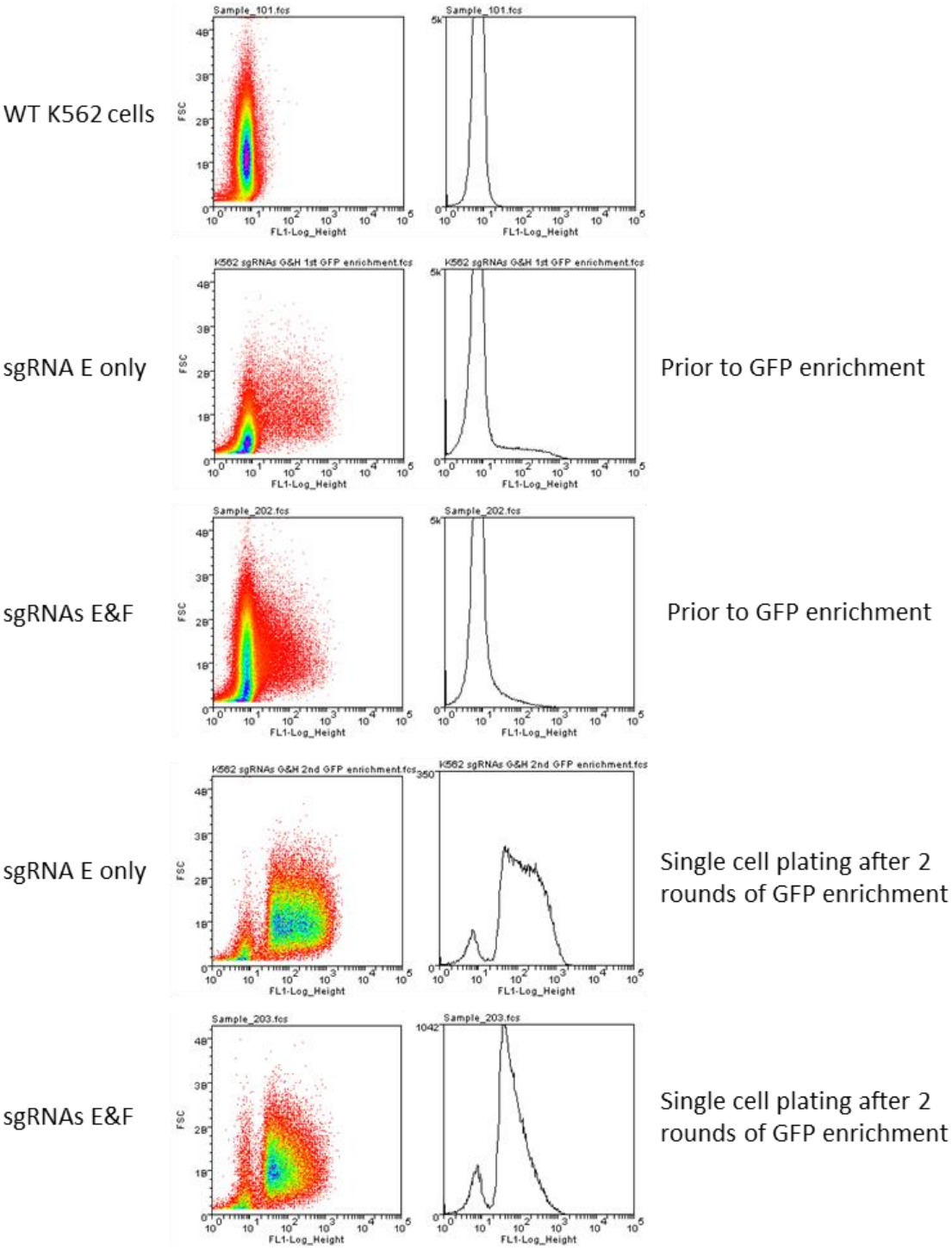


Figure 23: FACS of K562 cells transfected with CRISPR-Cas9n plasmids encoding KAT6A exon 17-targetting sgRNAs.

Dot plots and the corresponding histograms of K562 cells undergoing FACS after transfection with a plasmid encoding CRISPR-Cas9n and a single sgRNA targeting the forward DNA strand of KAT6A exon 17, or in combination with a second plasmid targeting the reverse DNA strand of the same exon. Transfected cells were isolated based upon their expression of EGFP from the transfected plasmid(s), relative to non-transfected K562 cells and subjected to two rounds of EGFP enrichment.

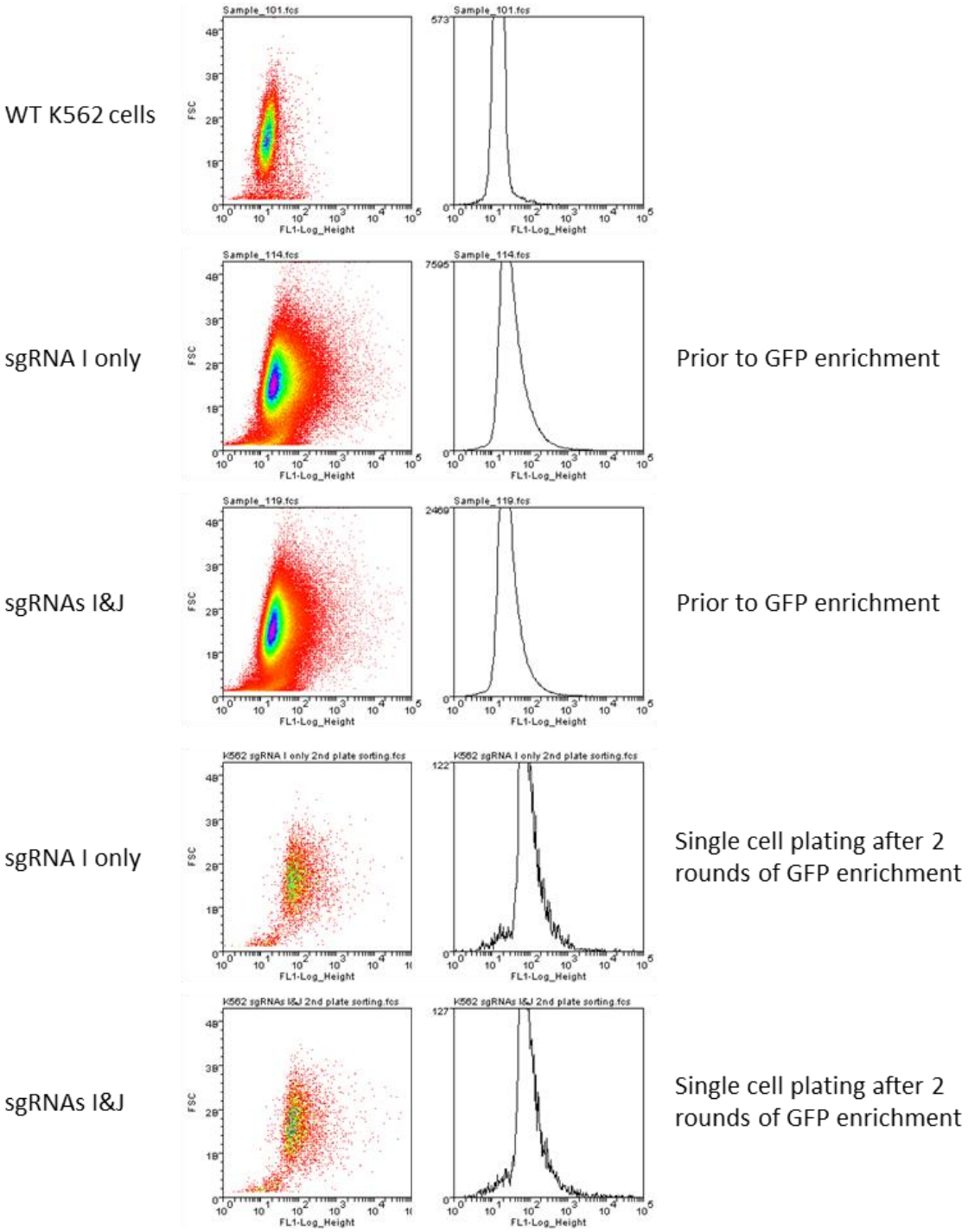


Figure 24: FACS of K562 cells transfected with CRISPR-Cas9n plasmids encoding KAT6A exon 18-targetting sgRNAs.

Dot plots and the corresponding histograms of K562 cells undergoing FACS after transfection with a plasmid encoding CRISPR-Cas9n and a single sgRNA targeting the forward DNA strand of KAT6A exon 18, or in combination with a second plasmid targeting the reverse DNA strand of the same exon. Transfected cells were isolated based upon their expression of EGFP from the transfected plasmid(s), relative to non-transfected K562 cells and subjected to two rounds of EGFP enrichment.

Excess EGFP-enriched cells in the bulk population were used as a quick indication of positive/negative genome editing by culturing for 72 hours prior to genomic DNA extraction. PCR upon these gDNA samples which amplified the regions targeted by Cas9, clearly showed that genome editing at the targeted loci in KAT6A exons 17 and 18 had been successful, by the visualisation of broad DNA bands around the expected size of amplicons using agarose gel electrophoresis (Figure 25). These broad DNA bands suggested the presence of DNA amplicons of various sizes, something which would be expected in a mixed population of cells harbouring KAT6A alleles with indels generated by the NHEJ DNA repair mechanism, as a result of Cas9 endonuclease activity at these loci. The appearance of a single, clear band of the expected WT size in the bulk population of EGFP-enriched single sgRNA-transfected control cells also indicated that indel formation had not occurred in these cells, as was expected.

To assess whether or not the HDR DNA repair pathway had successfully utilised the co-transfected DNA repair template, therefore generating KAT6A alleles possessing the Ser1039* mutation in those cells transfected with sgRNAs targeting exon 17, amplification of this targeted locus was performed via PCR and the resulting amplicons digested with the EcoRI restriction enzyme to reveal the presence/absence of the non-native restriction site conferred to KAT6A alleles by the DNA repair template. The presence of DNA bands of the expected sizes after EcoRI digestion (Figure 26A), led to the conclusion that in a subset of transfected cells, HDR using the DNA repair template had occurred. This was further indicated by sequencing of PCR amplicons of this locus, revealing the presence of both the WT Ser1039 and nearby PAM sequence, as well as the desired Ser1039* nonsense mutation and destroyed PAM, albeit at a low frequency (Figure 26B). Additionally, primers specific for the WT or Ser1039* DNA sequences were used in PCR upon bulk population gDNA; the visualisation of a DNA band of the expected size for those cells transfected with plasmids encoding sgRNAs targeting KAT6A exon 17 and the DNA repair

template, but not cells in the single sgRNA control population, also indicates that the Ser1039* mutation was successfully generated via HDR (Figure 26C). The presence of a DNA band in the same cell population of the expected size when using the WT-specific primers is indicative of non-mutated, WT KAT6A alleles in the bulk population. This, however does not mean that these particular alleles did not experience Cas9-mediated DNA cleavage, instead there is the possibility that they were repaired successfully back to their WT sequence. Alternatively, it may be that these alleles are indeed WT at the locus complementary to the primer used, allowing for PCR amplification, however due to the tendency of Cas9 to facilitate the formation of indels via NHEJ, they may deviate from the WT sequence at a locus 5' or 3' to the WT primer binding site.

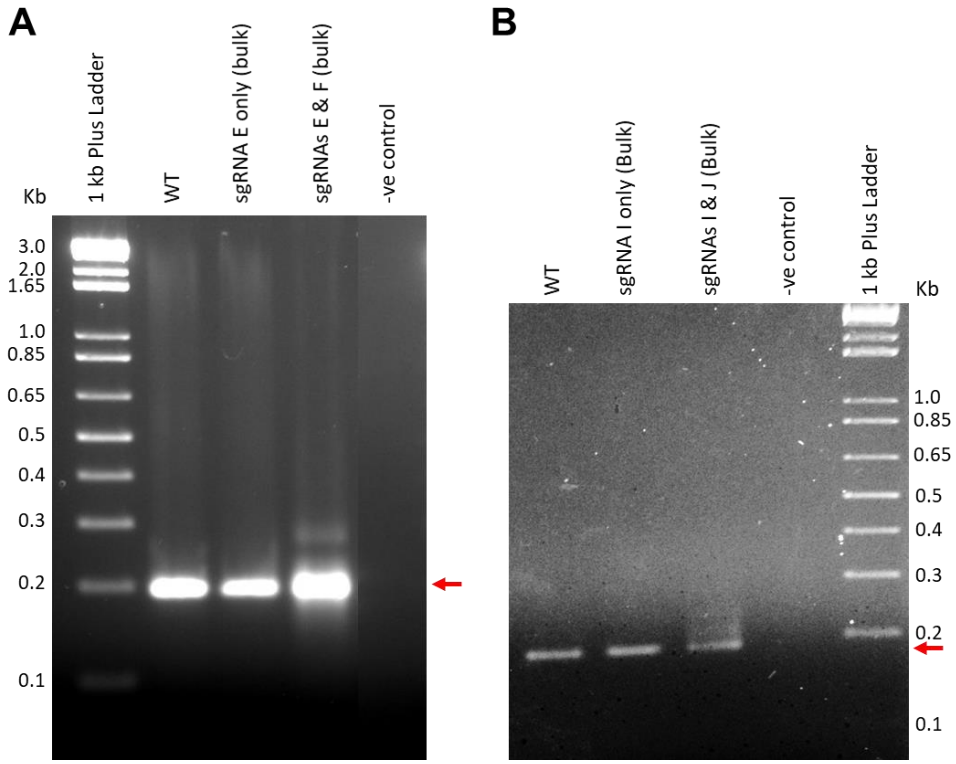


Figure 25: Assessment of genome editing in bulk FACS populations

Agarose gel electrophoresis of DNA fragments amplified from gDNA corresponding to bulk K562 cells isolated via FACS from populations transfected with sgRNAs targeting **(A)** KAT6A exon 17, or **(B)** KAT6A exon 18. The presence of broad DNA bands in those samples transfected with pairs of sgRNAs, is indicative of successful Cas9-mediated indel formation. Red arrows indicate the expected size of WT KAT6A amplicons.

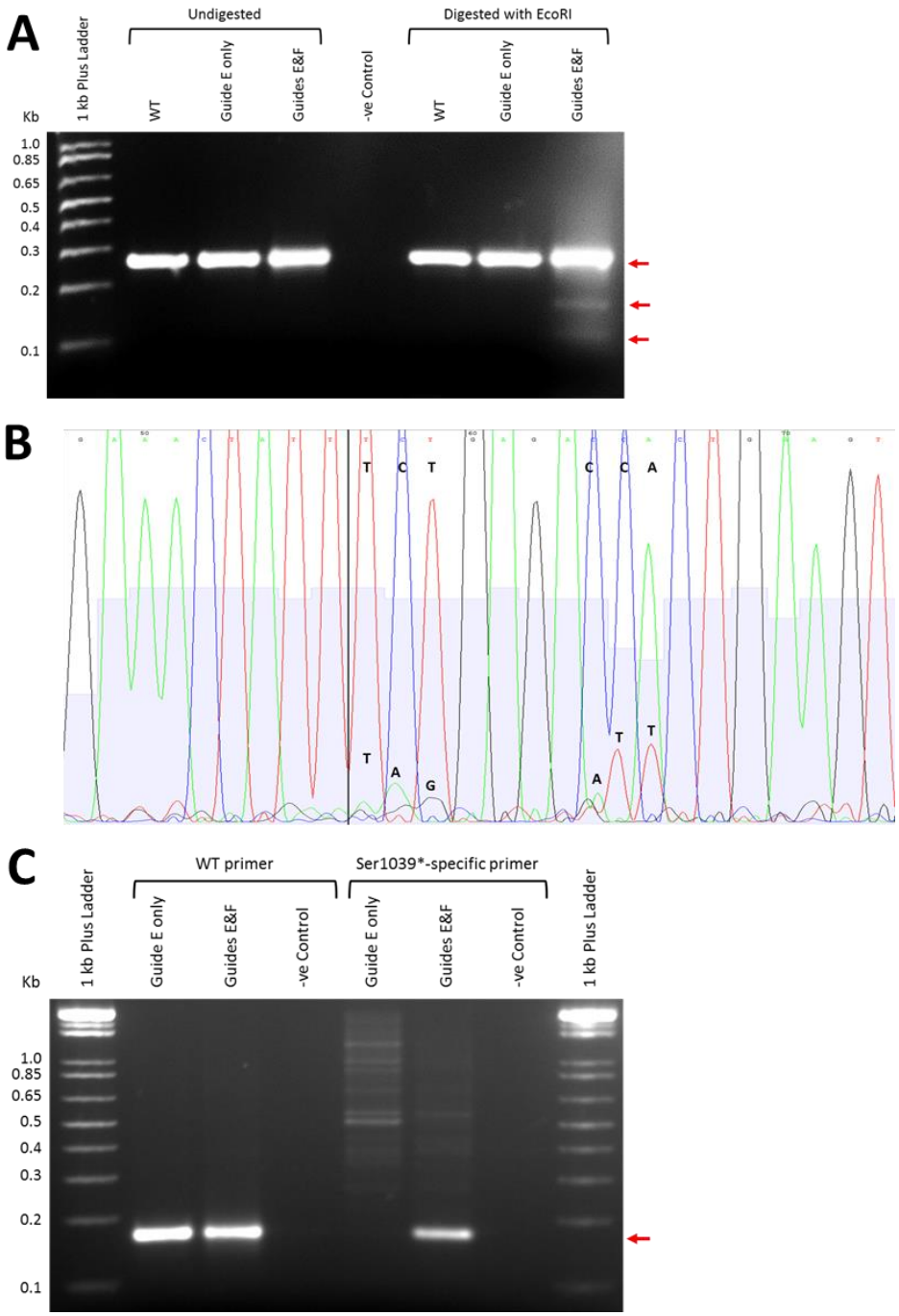


Figure 26: Ser1039* mutation in the bulk FACS population

(A) Agarose DNA electrophoresis of PCR products from gDNA isolated from the FACS K562 bulk population targeted by Cas9n in KAT6A exon 17. Products were digested with EcoRI to detect whether HDR had occurred, utilising our DNA repair template. Red arrows indicate the expected size of WT and desired EcoRI-digested KAT6A amplicons. **(B)** Sequencing electropherogram of the aforementioned undigested PCR products. TCT is the Ser1039 codon, whereas TAG is the Ser1039* mutation; CCA is the nearby PAM sequence and ATT is the destroyed PAM sequence. **(C)** PCR upon the same gDNA samples, utilising primers specific for the Ser1039* mutation, or the WT sequence at this locus. Red arrow indicates the expected size of DNA fragments without indels.

III. IV Validation of mutant clones

To identify K562 single cell-derived colonies which harbour KAT6A exon 17/18 variants that encode PTCs, techniques were utilised to assess mutagenic effects at both the DNA and protein levels. Of particular interest would be any clones that carry both WT and PTC-harbouring KAT6A alleles, as these would be ideal models to study the KAT6A GDD syndrome. It was noted that >80% of single cells plated out after FACS did not survive and undergo clonal expansion, even after frequent medium replacement and culturing for more than two months. This is likely due to the preference of K562 cells to grow as part of a multi-cell population, something which is generally accepted to be common to many cell lines (Andarawewa *et al.*, 2003). 48 single cell-derived colonies (from here on referred to as numbered clones) which successfully grew in the 96-well plates after FACS, were cultured and genomic DNA extracted. For those clones which had been targeted within KAT6A exon 17 by Cas9n, gDNA samples were subjected to PCR to amplify the targeted regions and the amplicons underwent restriction digestion with the EcoRI enzyme. DNA products were assessed by agarose gel electrophoresis to determine whether the DNA repair template had been utilised in HDR to generate the Ser1039* mutation and destroy the nearby Cas9 PAM site. If any cells were positive for this mutagenic event, a non-native EcoRI restriction site would have been created within the PCR-amplified DNA fragment and this would be detected by the presence of two new DNA bands compared to the undigested sample. A representation of this analysis is shown in Figure 27A. Unfortunately, despite this being observed in the EcoRI digestion of the KAT6A exon 17-targeted bulk population (Figure 26A), none of the single-cell derived clones appeared to harbour the Ser1039* variant, at least not utilising our DNA repair template and the novel EcoRI site.

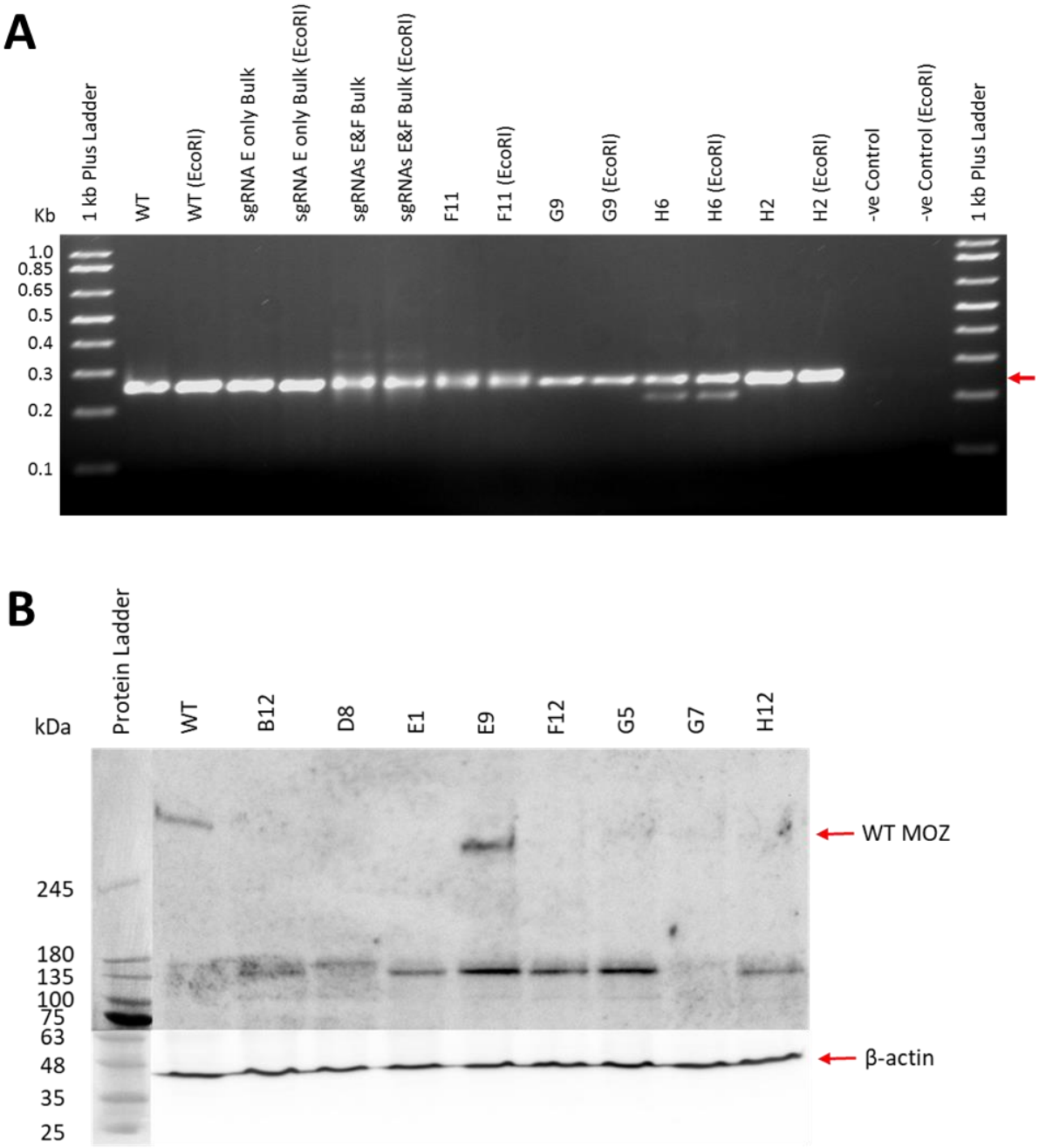


Figure 27: Screening of K562 single cell-derived colonies

(A) Agarose gel electrophoresis of PCR products from gDNA isolated from single cell-derived K562 colonies targeted by Cas9n in KAT6A exon 17. Products were digested with EcoRI to detect whether HDR had occurred, utilising our DNA repair template. **(B)** Western blot to detect MOZ polypeptides expressed in single cell-derived K562 colonies targeted by Cas9n in KAT6A exon 17. Image shown is a representative example of 48 Cas9n-targeted K562 colonies screened. Arrows in (A) indicate the expected size of DNA fragments without indels. The western blot is representative of n = 2 for each K562 colony.

In addition to EcoRI restriction digestion analysis, clones were also screened for MOZ protein expression via western blotting (Figure 27B); this was performed for all clones, regardless of which exon of the KAT6A gene was targeted by Cas9n. The majority of clones did not appear to express MOZ polypeptides of the expect WT size, suggesting that most KAT6A alleles had been successfully targeted by Cas9n and repaired in a manner which did not encode the original DNA sequence. Actually, it was observed that many clones did not express any MOZ protein close to its WT molecular weight of 225 kDa (runs at a higher molecular weight of ~250 kDa when analysed via SDS-PAGE), consistent with the presence of PTCs within the exon regions targeted (expected to result in a loss of ~1/3 – 1/2 of amino acids).

T-A cloning was performed upon clones which had been targeted by Cas9n in exon 17 or 18 (Figure 28 & Figure 29). Several clones were assessed for each exon region targeted, chosen after MOZ protein expression analysis via western blotting to ensure mutagenesis had occurred. Genomic DNA from each clone was extracted and the targeted locus was amplified by PCR using Taq DNA polymerase to introduce A-overhangs at the 3' end of each DNA strand. This allowed for ligation of amplicons via T-A cloning into a commercial vector, transformation into *E.coli* JM109 cells and colony PCR, followed by sequencing (Figure 28). A range of indels generated by the NHEJ DNA repair pathway were detected in the different clones using this allele sequencing approach, however noticeably there seemed to be a bias towards the deletion of nucleotides, compared to insertions. Any clones which were found to harbour inserted DNA sequences at their respective targeted locus that matched to known functional genomic regions via BLAST, were not chosen for downstream experimental analysis, however this was not a common occurrence.

To validate that targeting of KAT6A with a single sgRNA using the Cas9 nickase approach did not result in mutagenesis, a single clone was chosen from the FACS-plated single cell colonies that had been targeted by Cas9n in such a manner. All transformants screened for this K562

clone appeared to carry plasmids encoding KAT6A allele loci of the same size (Figure 29E). Sequencing of plasmids from transformants with a T7 forward primer and alignment of these to the WT KAT6A DNA sequence, confirmed all alleles corresponded to the WT sequence. Subsequently, this K562 clone was used as a WT control in downstream experiments and became known as clone G4.

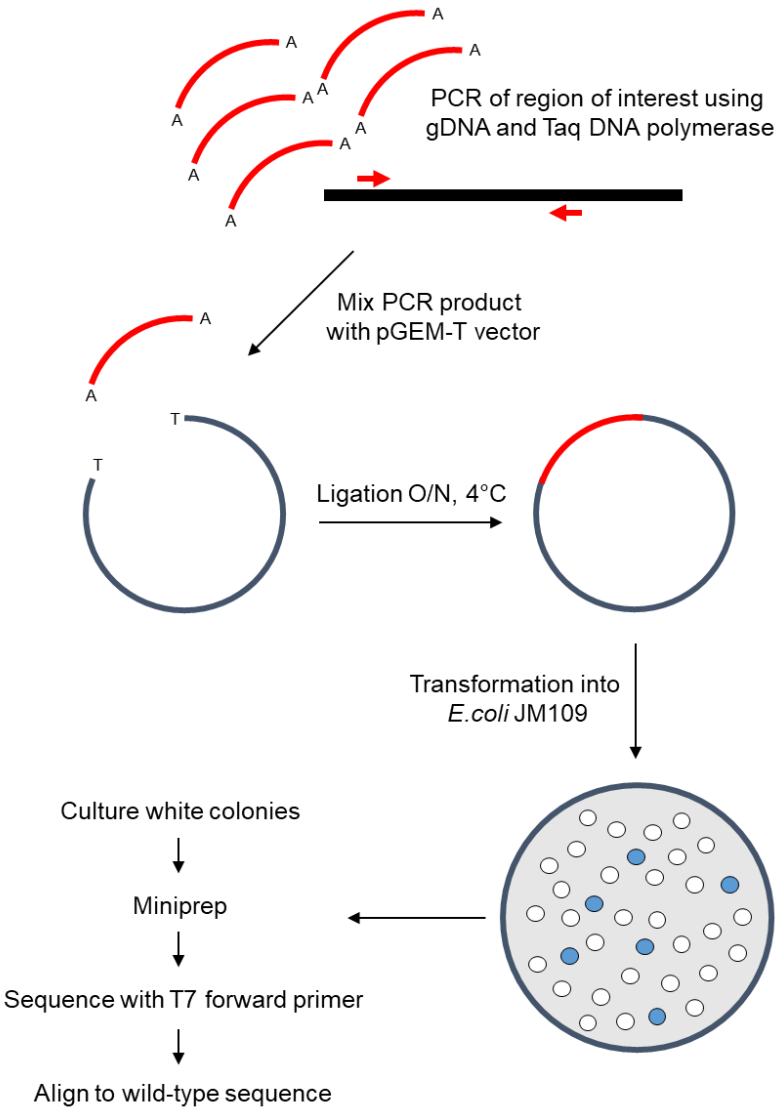


Figure 28: Workflow of T-A cloning using the pGEM-T vector

Schematic of the experimental process of T-A cloning for determination of individual KAT6A allele sequences in single cell-derived K562 colonies. Essentially, each plasmid should encode a single PCR-amplified KAT6A locus and *E.coli* JM109 cells allow for blue-white screening, with white colonies indicating the presence of a cloned DNA fragment. Selection of these, followed by sequencing allows the DNA sequence of individual alleles from a heterogeneous pool to be elucidated.

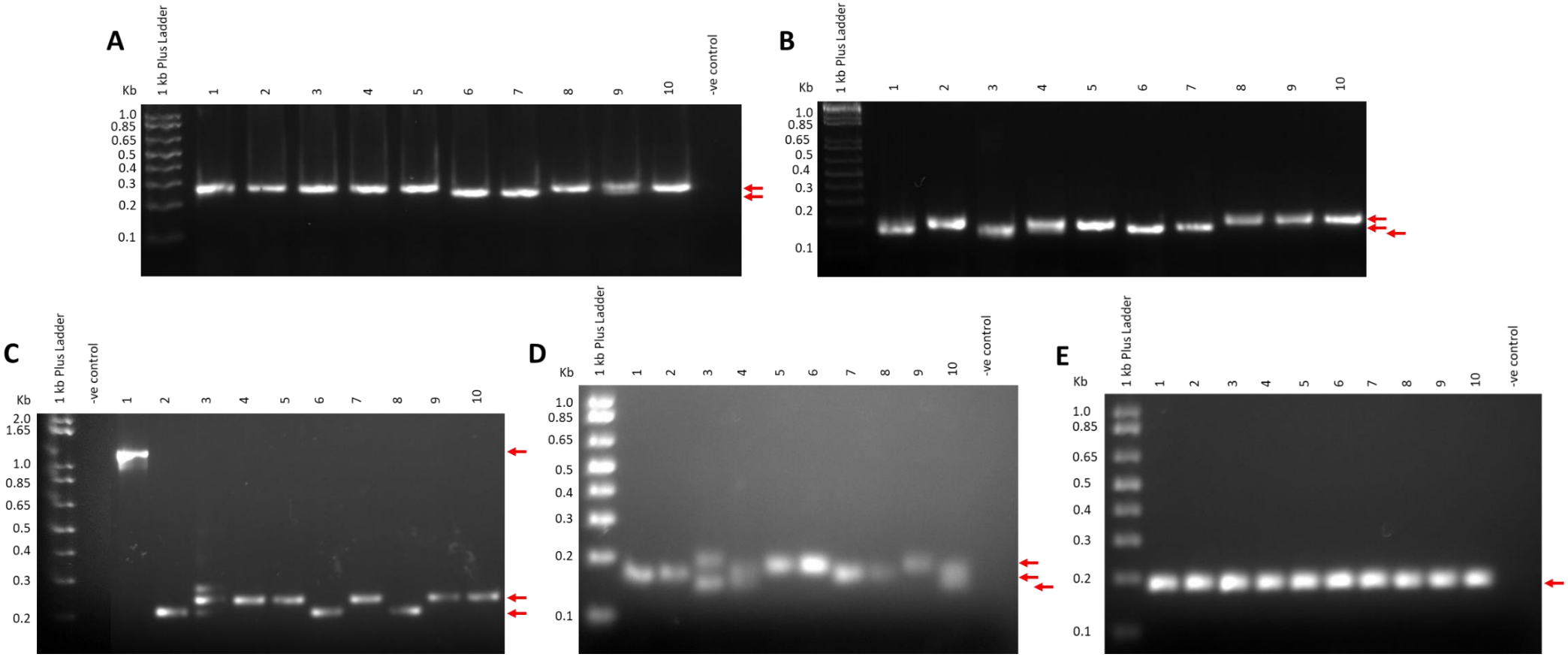


Figure 29: KAT6A allele screening of K562 clones E1 (17), E9 (17), E1 (18), H9 (18) & G4

Agarose gel electrophoresis of colony PCR products from T-A cloning experiments to determine KAT6A allele sequences in single cell-derived K562 clones E1 (17) **(A)** and E9 (17) **(B)**, which had been transfected with sgRNAs targeting KAT6A exon 17. **(C)** and **(D)** represent the same analysis for clones E1 (18) and H9 (18) respectively, both of which had been transfected with sgRNAs targeting KAT6A exon 18. **(E)** Clone G4, which had been transfected with the sgRNA E plasmid only and therefore did not undergo Cas9-mediated genome editing; note that for this colony, all PCR products are of the same size. Red arrows represent the readily detectable different KAT6A allele amplicons. These agarose gel images are representative of n = 20 for each K562 colony.

A large number of the indels present in targeted K562 clones resulted in frameshift mutations in the respective KAT6A exon, generating predicted PTCs, however in some cases, the indels were in-frame deletions, suggesting that the clones in question should express MOZ proteins lacking some amino acids at these loci. Interestingly, despite this the MOZ proteins predicted to lack in-frame amino acids could not be detected by western blotting (data not shown). Since the epitope for the primary antibody used is situated within the N-terminal part of the MOZ protein (amino acids 441-460), the failure to detect these polypeptides is not due to disruption of the epitope. This suggests that these novel in-frame deletion MOZ polypeptides are unstable, or that they aggregate and are degraded.

Of the K562 clones generated, we selected four for downstream experimental analysis. Two had been targeted within KAT6A exon 17 by Cas9n and the other two in exon 18. Henceforth, these clones are referred to by their number and the KAT6A exon targeted by Cas9n is shown in brackets, e.g. E1 (17). The KAT6A alleles harboured by these clones are detailed in Figure 29, Figure 30 & Figure 31, although all carry at least one allele encoding a PTC. Only one clone, namely E9 (17) had retained a WT allele of KAT6A. To assess whether our mutagenesis affected the predicted splicing of KAT6A alleles, we utilised the NNSPLICE tool (Reese *et al.*, 1997). No new donor or acceptor sites were predicted in any of the mutant KAT6A alleles in the chosen K562 clones, compared to the WT allele sequence (data not shown).

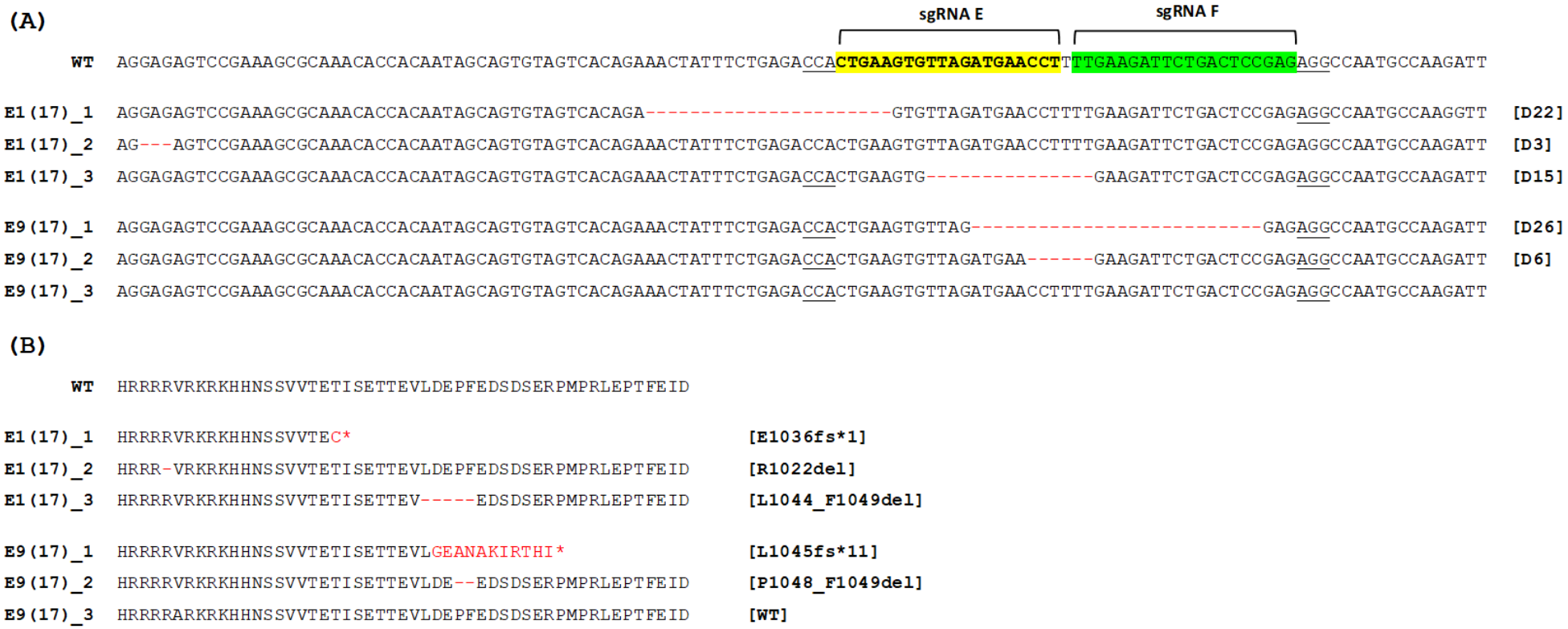


Figure 30: Alignment of KAT6A alleles in K562 KAT6A exon 17 mutant clones E1 and E9

Each allele is shown both in terms of **(A)** nucleotide sequence and **(B)** the translated protein sequence. Nucleotide I, insertions or D, deletions are shown in square brackets to the right of each allele, as are the consequential alterations in protein sequence. The genomic regions targeted with sgRNAs are also depicted by highlighted text; underlined regions represent the Cas9 PAM sequences. Red hyphens represent deleted nucleotides or amino acids. Red amino acids are mutant protein sequences.

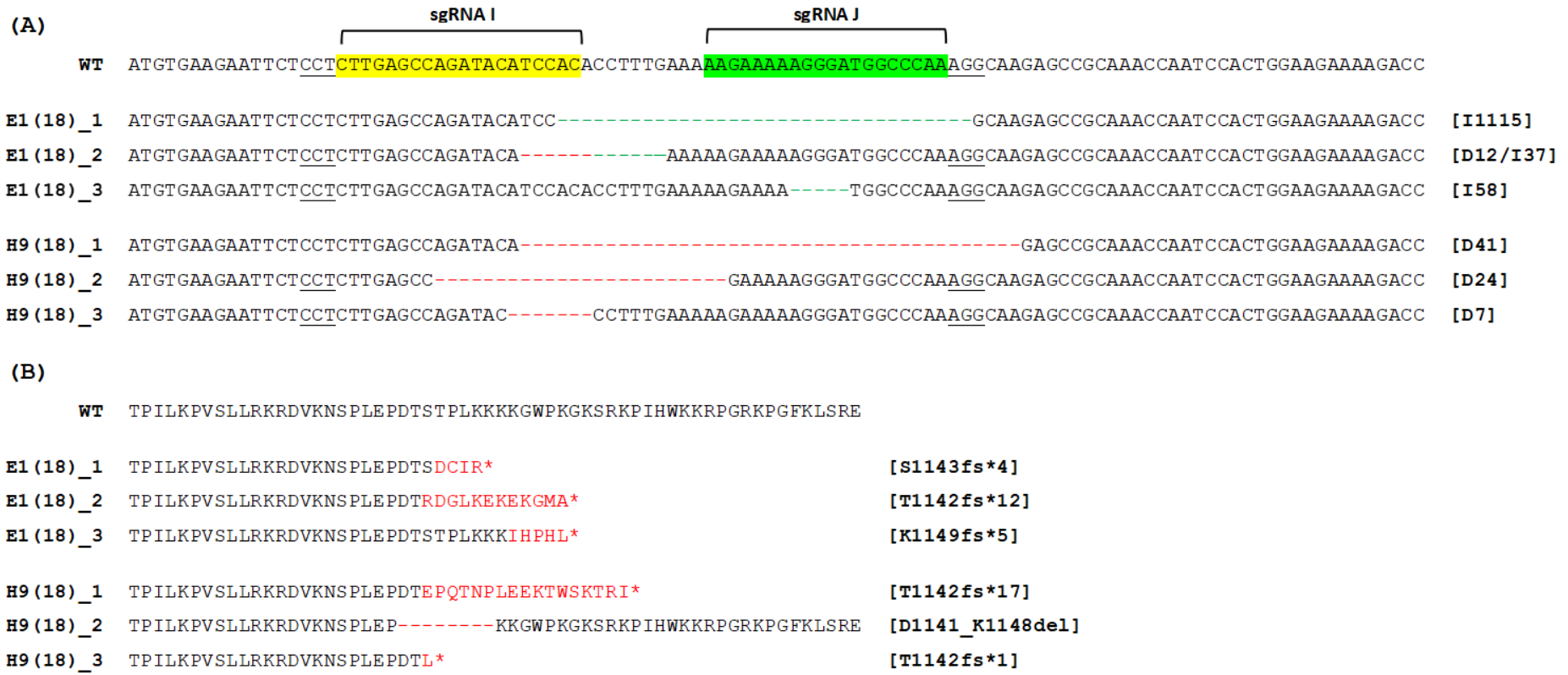
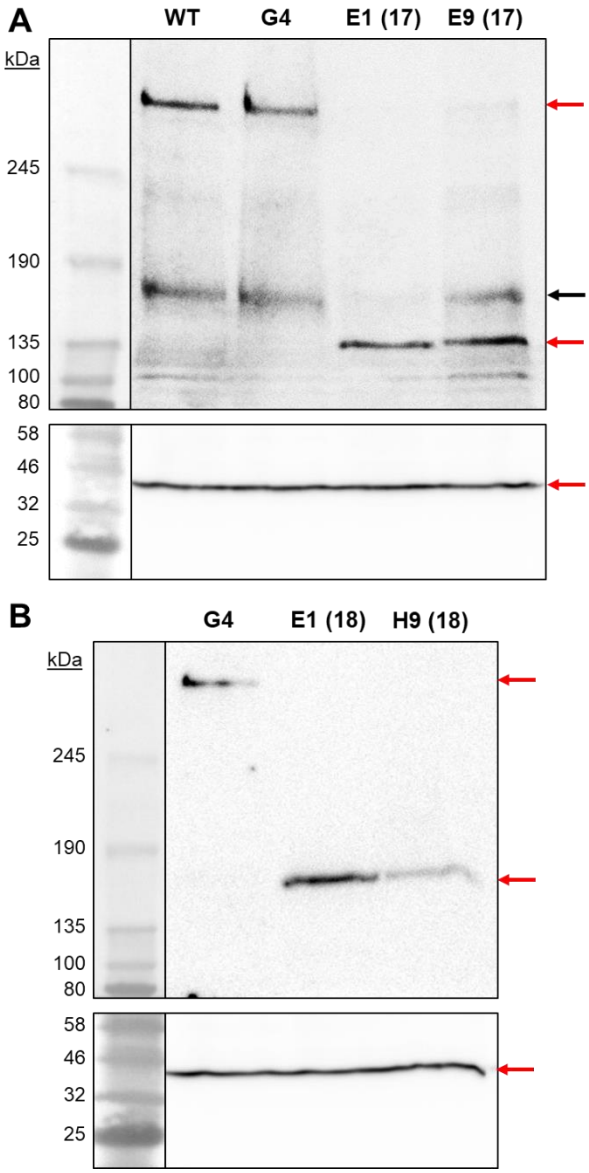


Figure 31: Alignment of KAT6A alleles in K562 KAT6A exon 18 mutant clones E1 and H9

Each allele is shown both in terms of **(A)** nucleotide sequence and **(B)** the translated protein sequence. Nucleotide I, insertions or D, deletions are shown in square brackets to the right of each allele, as are the consequential alterations in protein sequence. The genomic regions targeted with sgRNAs are also depicted by highlighted text; underlined regions represent the Cas9 PAM sequences. Red hyphens represent deleted nucleotides or amino acids, insertions are in green. Red amino acids are mutant protein sequences.

The MOZ protein expression of the chosen K562 KAT6A mutant clones can be observed in Figure 32, along with densitometry analysis, where it can also be seen that the predicted in-frame deletion MOZ polypeptides are not able to be detected by western blotting for clones E1 (17) and H9 (18). In the case of E9 (17), the expression of the in-frame deletion MOZ polypeptide could not be determined, due to its predicted similar molecular weight to that of the also-expressed WT MOZ protein. Note that the reduced intensity of the WT MOZ band for colony E9 (17) may be due to partial degradation of the protein (compare to band in WT and G4 samples marked with a black arrow, that is not present in E1 (17) - the latter colony does not possess a WT KAT6A allele unlike the other three colonies on the same western blot). In addition, the single band for H9 (18) may be fainter than that of E1 (18) due to one less KAT6A allele encoding a MOZ protein of the detected ~130 kDa size (2 alleles, rather than 3). These clones were chosen for further analysis due to the presence of their KAT6A PTC alleles and the lack of Cas9n-facilitated DNA insertions corresponding to other functional genomic regions. These characteristics made them suitable to assess the functional requirements of the MOZ C-terminus, as well as the association between PTCs in exons 17 or 18 of KAT6A and the KAT6A GDD syndrome at a molecular level. Summaries of the KAT6A variant alleles and their corresponding predicted translated proteins present in these chosen K562 clones are shown in Figure 33 & Figure 34, revealing similarities to those reported in the literature which are associated with KAT6A GDD.



Colony & Protein Band	Relative density
WT, WT band	1.11
G4, WT band	1.00
E1 (17), PTC band	0.93
E9 (17), WT band	0.15
E9 (17), PTC band	1.06

Colony & Protein Band	Relative density
G4, WT band	1.00
E1 (18), PTC band	2.00
H9 (18), PTC band	0.51

Figure 32: MOZ protein expression in K562 KAT6A mutants

Western blot detection of MOZ proteins in our chosen K562 KAT6A exon 17 **(A)**, or exon 18 **(B)** variant clones. Top panels show MOZ protein detection, bottom panels display β-actin as a loading control. The red arrows indicate bands of the expected sizes for full length or C-terminally truncated MOZ proteins in these clones, or alternatively β-actin. Black arrow indicates bands corresponding to WT MOZ protein thought to be partially degraded. Clone G4 is a negative K562 control, which was transfected with a single sgRNA and therefore did not undergo Cas9-mediated genome editing. Table to the right of each western blot displays the quantitative densitometry analysis of each of the indicated MOZ protein bands, relative to that of clone G4 on the same blots and taking β-actin expression into account. Quantification was performed using the ImageJ software. Note that the two boxed images in each of A and B are part of the same nitrocellulose membrane, but that A and B are two individual membranes. These westerns are representative of n = 3 for each K562 colony.

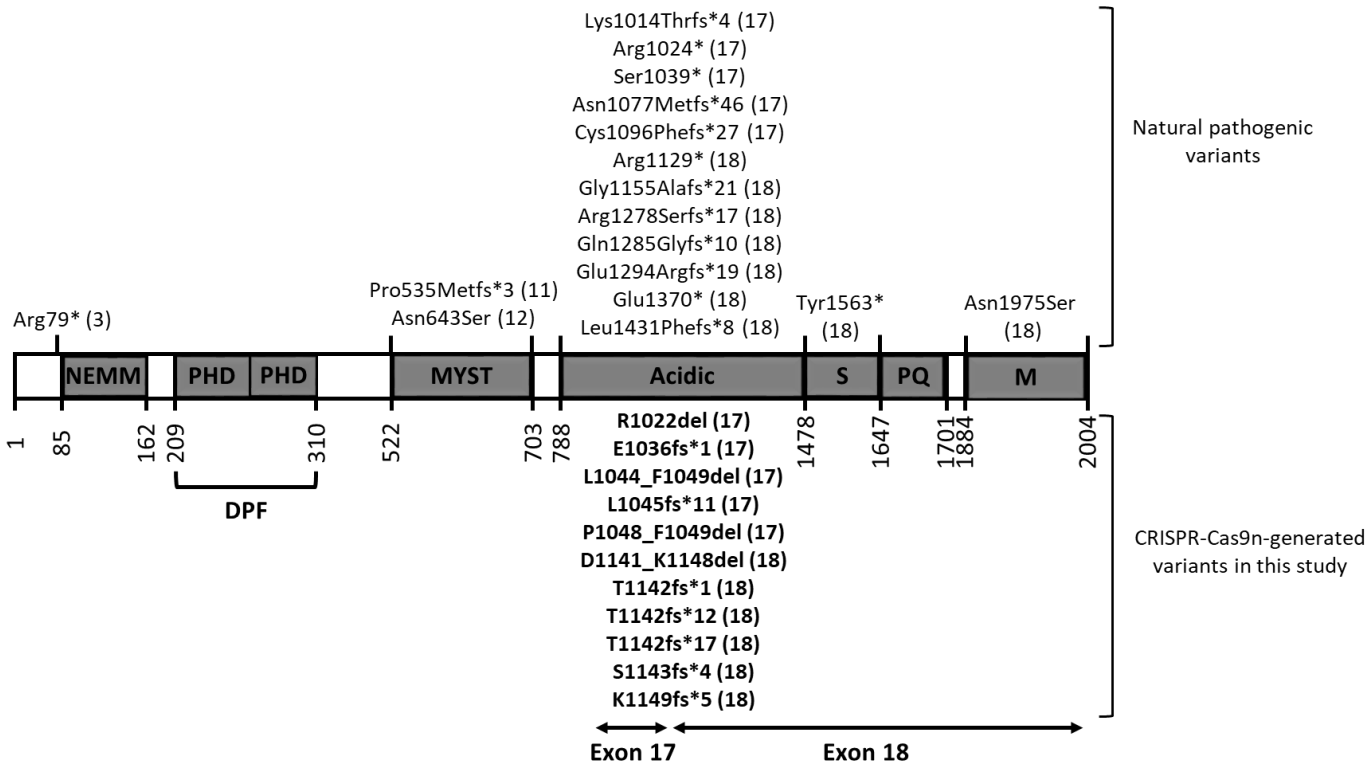


Figure 33: K562 KAT6A exon 17 and 18 variants generated by CRISPR-Cas9 genome editing

Location of K652 KAT6A variants in exons 17 and 18, which were generated by CRISPR-Cas9 genome editing, relative to MOZ protein amino acid number and domains (below protein, bold text). Also shown are published KAT6A variants associated with KAT6A GDD syndrome (above protein). Exon location of variants shown in brackets. fs = frameshift. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2. MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. Acidic residue-rich (Acidic), serine-rich (S), glutamic acid-rich (E), proline and glutamine-rich (PQ) and methionine-rich (M) regions are also indicated.

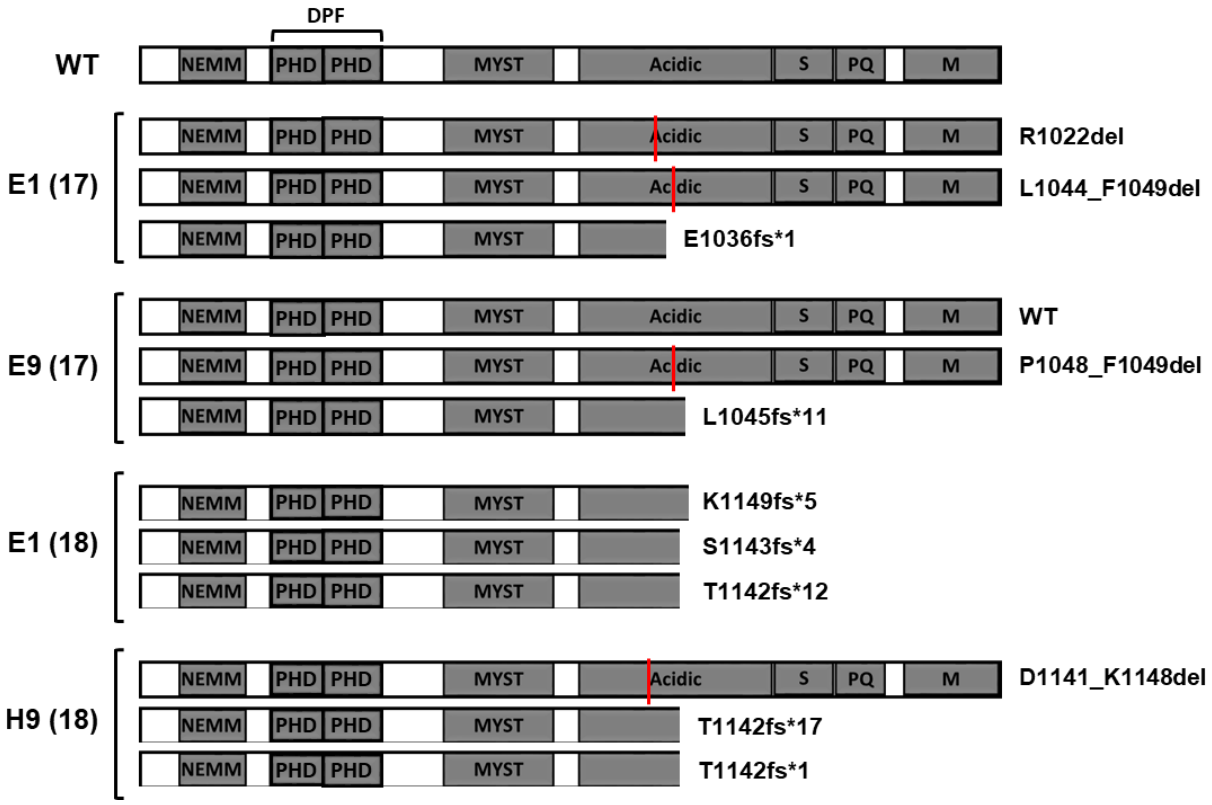
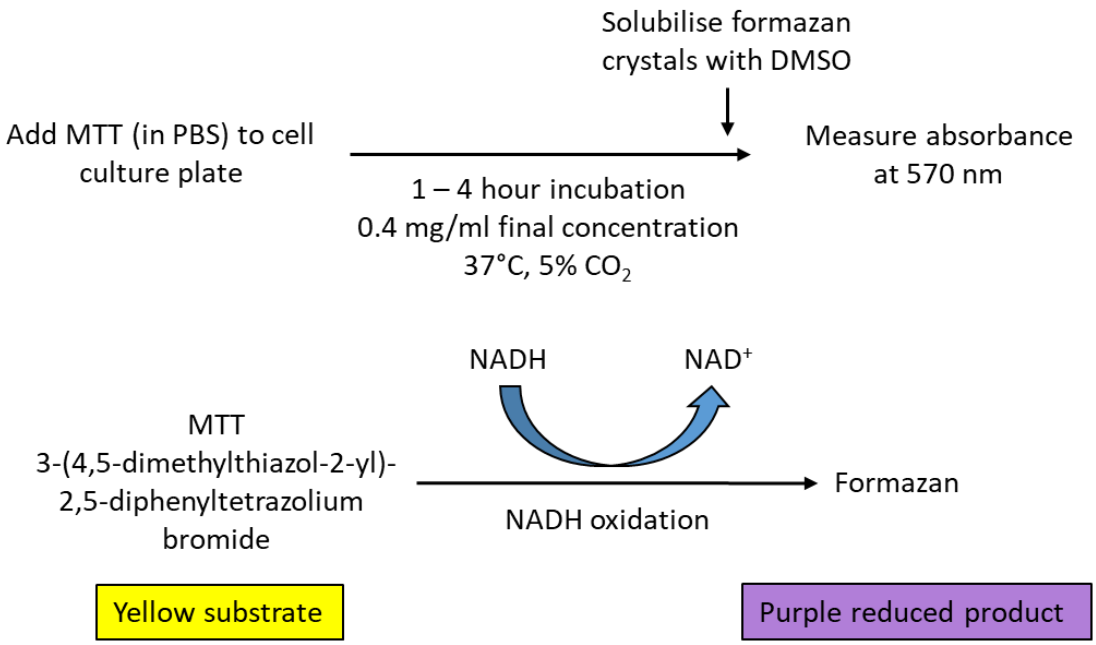


Figure 34: Predicted MOZ proteins expressed in our K562 KAT6A mutant cell lines

An alignment of the predicted MOZ proteins expressed in our chosen genome edited K562 clones, showing the presence or absence of domains, relative to WT MOZ. Red lines indicate the position of in-frame deletions. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2. MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. Acidic residue-rich (Acidic), serine-rich (S), glutamic acid-rich (E), proline and glutamine-rich (PQ) and methionine-rich (M) regions are also indicated.

III. V Cell cycle progression and number of viable metabolically active cells is affected by KAT6A PTCs

MOZ is known to have a role in both the inhibition and promotion of senescence via its interaction with the tumour suppressor protein p53 and the protein PML, as well as regulating the transcription of inhibitors of the cyclin-dependent kinase CDKN2A (Rokudai *et al.*, 2009, 2013; Sheikh, Phipson, *et al.*, 2015). Furthermore, this HAT is necessary for the enhanced proliferation abilities and self-renewal capacity which characterises stem cells (Katsumoto *et al.*, 2006; Sheikh *et al.*, 2016). We have previously demonstrated that both knockout of MOZ expression in HEK293 cells, as well as knockdown in K562 cells, results in greatly reduced fold changes in the number of viable cells over time (Monteiro 2018, PhD thesis - unpublished) (see previous Figure 11A). In order to assess whether loss of the MOZ C-terminus has a similar effect, we performed MTT assays upon our K562 KAT6A mutant clones. This technique measures the metabolic rate of a cell population. By measuring the conversion of a coloured substrate over a fixed period of time and comparing to this to cell populations at different times after seeding, MTT assays are frequently used as an indirect measurement of proliferation rate. However, the output is really only a direct indication of the change in the relative number of viable, metabolising cells, or the metabolism rate itself. Here, we have taken the output of our MTT assays to imply the number of metabolically active, viable cells in the population. The principle of how this technique works is depicted in Figure 35.



Measures metabolic activity, which can be correlated to total cell number

Increased purple product = increased number of viable, metabolically active cells
OR increased metabolic activity within cells

Figure 35: Principle of MTT assays

Schematic showing the key steps in the MTT protocol (top), along with the chemical conversion of the MTT reagent to formazan by NADH oxidation during cellular metabolism (bottom). MTT is a yellow substrate which is incubated with the cells under normal culture conditions for 1 to 4 hours. During this time in viable, metabolically active cells, NADH is oxidised to NAD^+ and the MTT reagent is reduced to formazan, which is purple in colour. After incubation is complete, the cell medium is removed and the insoluble formazan crystals are solubilised by the addition of DMSO, upon which time the absorbance is measured at 570 nm. Repeating this procedure upon cells which have been seeded over a period of time/in different conditions/mutant cells, allows differences in metabolism or cell number to be calculated. Full details can be found in the Methods section of this thesis. This assay is not a direct measurement of the rate of cell proliferation, without independently assessing the rate of apoptosis. Background information regarding MTT assays was acquired from Riss *et al.* (2004)

PTCs in either exon 17 or exon 18 of KAT6A result in a reduced increase in the number of viable, metabolically active cells compared to WT K562 cells, which is apparent after 48 hours (Figure 36A). This would suggest that a single, WT allele of KAT6A (clone E9 (17)) and the expression of MOZ protein truncated in the C-terminus (all other clones), is not enough to sustain a normal rate of increase in viable cell number. Furthermore, for clone E1 (17) this reduction was more pronounced (Figure 36A). In addition to this clone not expressing WT MOZ protein (no WT alleles), it also doesn't appear to express protein from the alleles in which we have identified in-frame deletions (see previous Figure 32). It does however express the C-terminally truncated MOZ protein predicted from an allele harbouring a PTC in exon 17 (see previous Figure 32). This suggests that MOZ protein truncated within this region alone, cannot promote a normal rate of increase in the number of viable cells within the population.

The yeast homologue of MOZ is Sas3, which has been shown to regulate the progression of cells through the G_2/M checkpoint during mitosis (Howe, 2001). This regulatory function has not been reported for MOZ, however we have previously shown that MOZ expression is necessary for normal cell cycle progression from G_1 to S phase and preventing exit from the cell cycle into G_0 (Monteiro 2018, PhD thesis - unpublished) (see previous Figure 11B). To assess whether the C-terminal regions of this HAT influence cell cycle progression, we analysed populations of our CRISPR-Cas9-generated KAT6A mutant clones for their cell cycle distribution. This was performed utilising the nucleotide-binding dye propidium iodide (PI), which can be used alongside flow cytometry to determine the DNA content of a cell, directly correlating with the stages of the cell cycle (Figure 36B). Representative plots of flow cytometry data output using this technique upon our K562 KAT6A mutant clones is depicted in Figure 37.

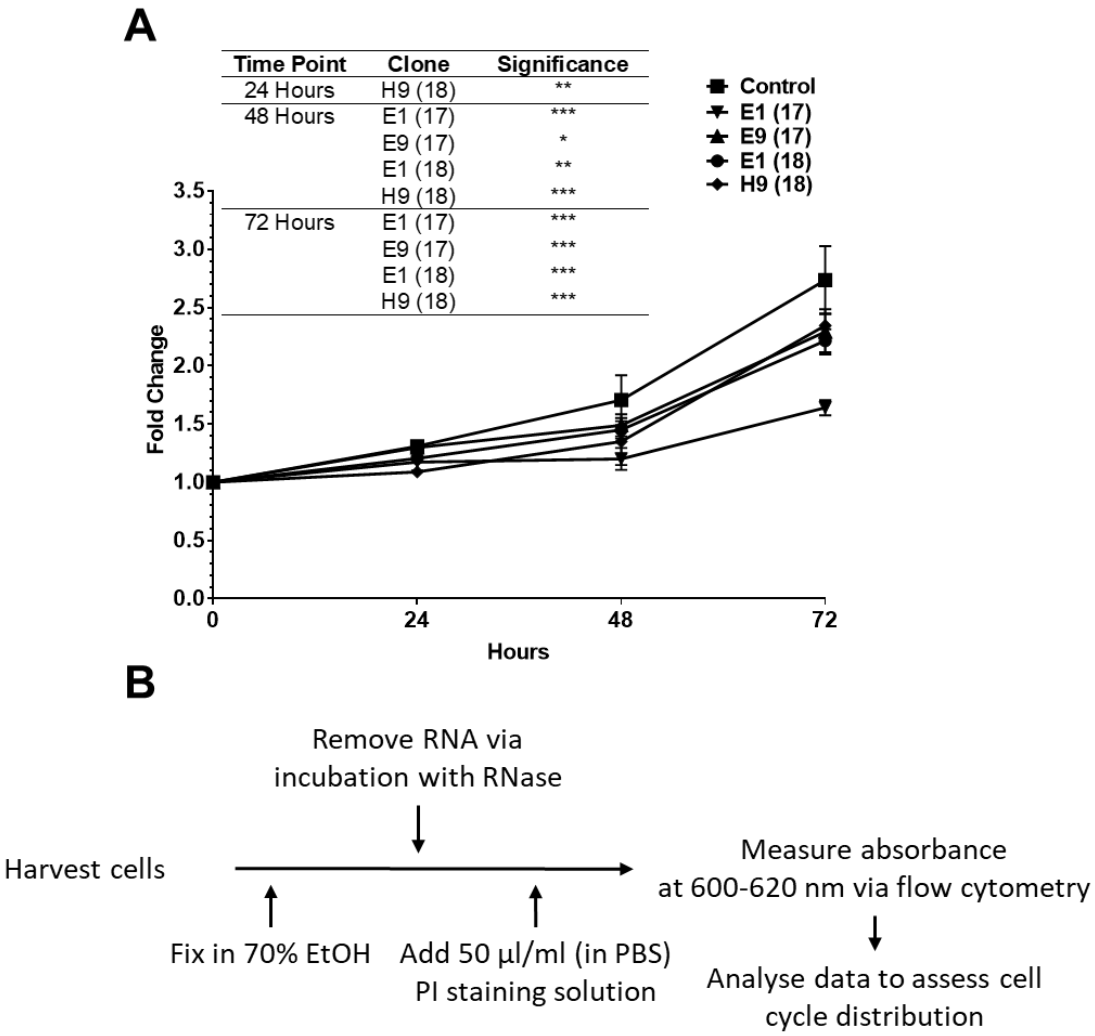
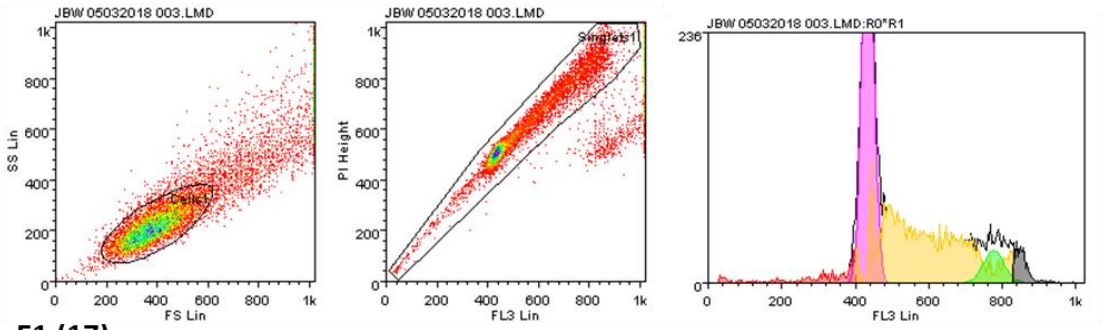


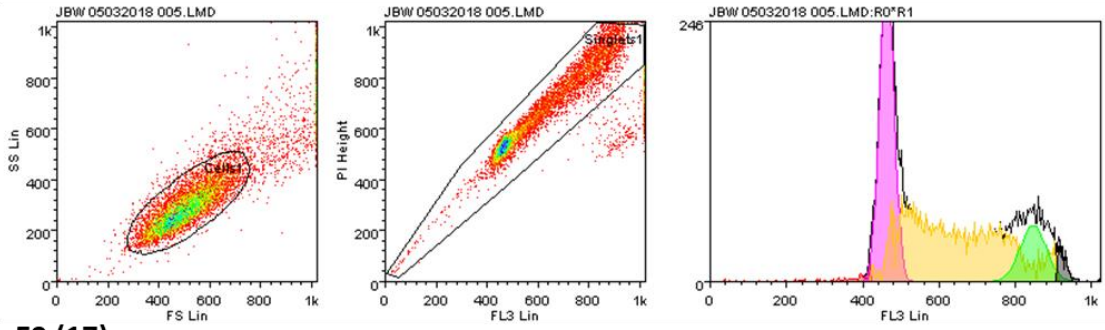
Figure 36: Total number of viable metabolically active cells is affected in K562 cells harbouring KAT6A exon 17/18 PTC variants

(A) Fold change in total metabolically active, viable cell number over time assessed by MTT assays of K562 CRISPR-Cas9-generated cells harbouring KAT6A exon 17/18 PTC variants. Error bars represent standard deviation of the mean. Statistical analysis performed using one-way ANOVA. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$. n = 5 for all data shown here. **(B)** Summary of the key steps in the propidium iodide staining protocol. Note that the wash steps with PBS are not shown, however full details can be found in the Materials & Methods section of this thesis. Cells are harvested and fixed in ethanol, treated with RNase and stained with PI. Propidium iodide binds to both RNA & DNA, therefore RNA is removed prior to staining. Sample absorbance spectra are acquired via flow cytometry and data analysis can be used to assess the proportion of the cell population in each of the main stages of the cell cycle (G_0/G_1, S and G_2).

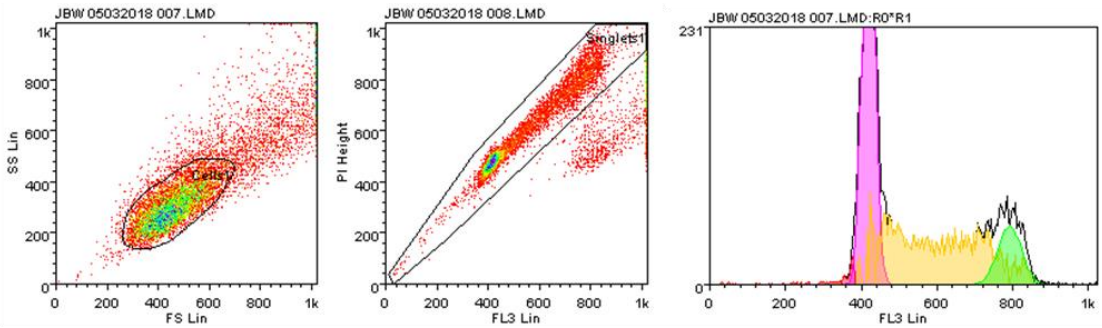
Control (clone G4)



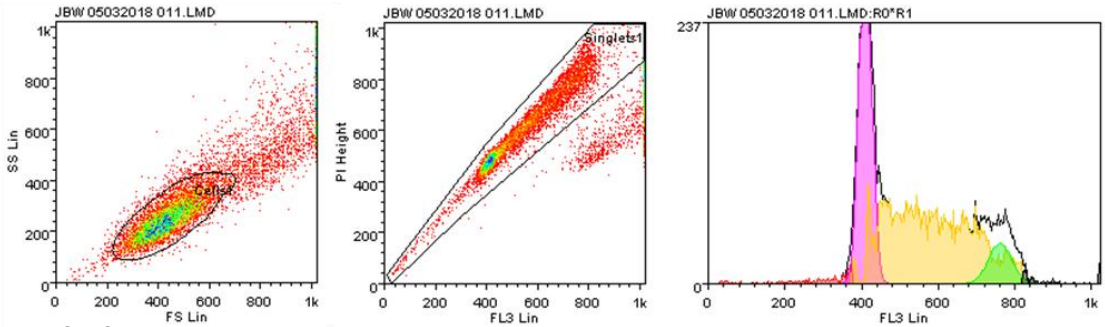
E1 (17)



E9 (17)



E1 (18)



H9 (18)

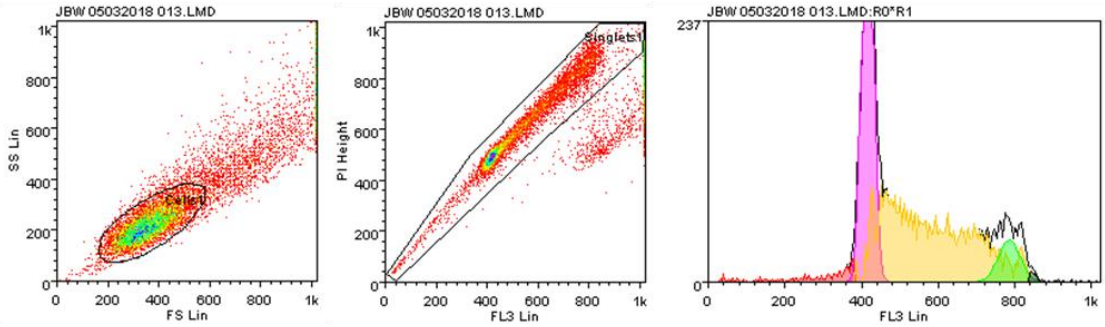


Figure 37: Flow cytometry analysis of PI staining of K562 KAT6A mutant clones
Dot plots and the corresponding histograms of PI-stained K562 control and KAT6A exon 17/18 mutant cell lines. Left panel shows gating of data points for selection of intact cells and middle panel shows gating for single cells to exclude doublets or clusters. Right panel shows histograms of PI staining, which indicates DNA content and position in the cell cycle in this context. Pink represents G_0/G_1, yellow is S phase and green is G_2. Percentage of the population in each phase of the cell cycle was computationally acquired from these divisions using the WEASEL flow cytometry software.

Cell cycle analysis performed upon the E9 (17) clone which harbours a PTC in exon 17 and also expresses WT MOZ, did not reveal any significant difference to WT cells in the proportion of the population at each stage of the cell cycle (Figure 38). Similar to what we observed for the rate of increase in number of viable, metabolically active cells, however is that clone E1 (17) (which does not express any WT MOZ and in western blotting only displays truncated MOZ protein), had a small but statistically significant reduction in the population of cells in S phase. Clones harbouring PTCs in exon 18 displayed similar cell cycle profiles to WT populations, with the exception that there was a reduction in the number of cells in G_2, raising the possibly that MOZ proteins truncated in the C-terminus may have heightened activity in regards to promoting progression through the G_2/M checkpoint (Figure 38).

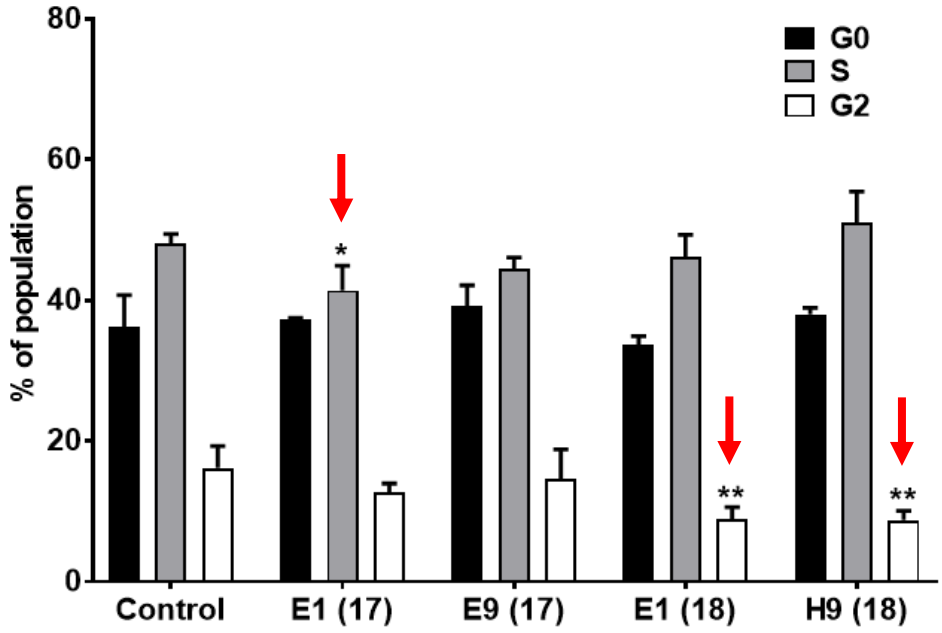


Figure 38: Cell cycle progression is affected in K562 cells harbouring KAT6A exon 17/18 PTC variants

Cell cycle analysis assessed by PI staining of K562 CRISPR-Cas9-generated cells harbouring KAT6A exon 17/18 PTC variants. Note there is a reduced proportion of the cell population in S phase for clone E1 (17), while those in G_2 are reduced for clones E1 (18) and H9 (18). Error bars represent standard deviation of the mean. Red arrows in highlight significant variations from the control population. Statistical analysis performed using one-way ANOVA. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$. n = 5 for all data shown here.

III. VI C-terminally truncated MOZ protein does not have aberrant cellular localisation properties

Protein mislocalisation has been shown to have a role in a range of diseases and can be due to a variety of factors, such as the loss of a NLS or loss of protein-protein interaction sites that would normally result in nuclear translocation when in an active complex (Cutress *et al.*, 2008; Laue *et al.*, 2008; Kanda and Swaroop, 2009). The protein region which has been reported to be primarily responsible for the nuclear localisation ability of MOZ, is situated within the NEMM domain, very close to the N-terminus. Since this domain is not affected by the presence of PTCs within the KAT6A gene in the majority of reported KAT6A GDD patients (with the exception of Arg79*), protein mislocalisation as a mode of pathogenesis would not be expected to be the case for KAT6A exon 17/18 variants.

We performed computational analysis of the full-length WT MOZ primary sequence using three NLS prediction tools, to assess whether there are any additional putative NLS within the C-terminal domains of this protein. As can be seen in Figure 39 & Table 17, there are four clustered regions of the full-length protein predicted to harbour putative NLS, two of which are within the C-terminal acidic domain. If these are indeed true NLS, then some individuals with KAT6A GDD would be expected to express MOZ protein lacking C-terminal sequences that promote import to the nucleus (primarily those with nonsense/frameshift mutations upstream to exon 18). However as mentioned above, while these proteins would be truncated and lack one or perhaps two putative NLS, they would still possess the region of the NEMM domain that is known to have a role in MOZ nuclear localisation. As such, we hypothesised that MOZ protein truncated within the C-terminal domain would still be able to effectively localise to the cell nucleus. To test this hypothesis, we transfected WT, Arg1024* or Arg1129* MOZ expression constructs into a HEK293 MOZ knockout cell line and visualised their cellular localisation by fluorescence microscopy (cell line generation data not shown, unpublished). Both the WT and C-terminally truncated

variants localised exclusively to nuclei and no obvious differences in their ability to do so were observed (Figure 40), indicating that MOZ protein mislocalisation does not contribute to KAT6A GDD pathogenesis in individuals harbouring these variants.

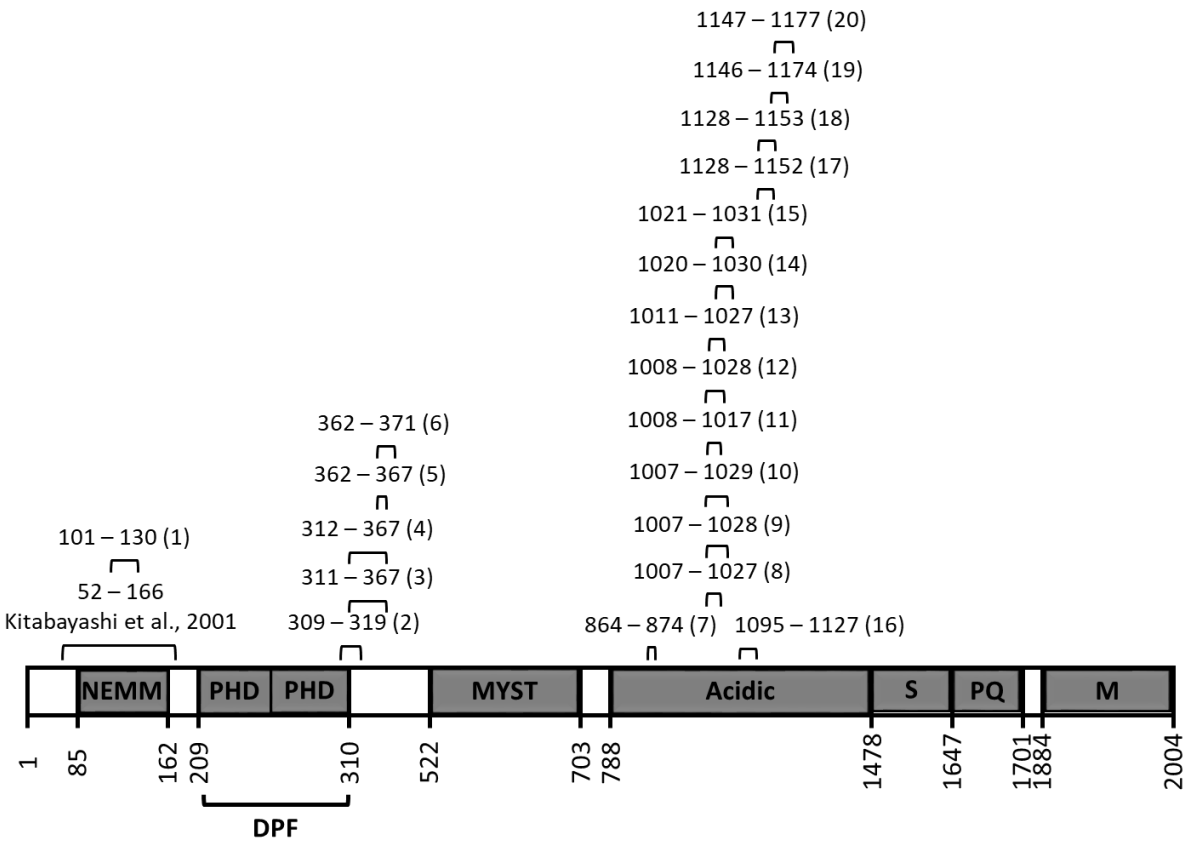


Figure 39: Regions within the MOZ protein computationally predicted to be nuclear localisation sequences

Amino acid numbers of each indicated region computationally predicted to form part of a nuclear localisation sequence within the linear MOZ protein are shown, along with the NLS region experimentally assessed by Kitabayashi et al. (2001), which resides within the NEMM domain. Numbers within parentheses relate to the NLS ID number in Table 17. A clustering of 3 / 4 regions of the MOZ protein suggested to harbour putative NLS can clearly be observed. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2. MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. Acidic residue-rich (Acidic), serine-rich (S), glutamic acid-rich (E), proline and glutamine-rich (PQ) and methionine-rich (M) regions are also indicated.

ID	Sequence	Position within MOZ	Mono-/bipartite	Prediction tool
1	KLIKRAVEGLAESGGSTLKSIERFLKGQKD	101-130	Bipartite	NLS Mapper
2	ICRPRKKGRKL	309-319	Monopartite	NLS Mapper
3	RPRKKGRKLLQKKAAQIKRRYTNPIGRPKNRLKKQNTVSKGPFSKVRTGPGRGRKRK	311-367	Bipartite	NLStradamus
4	PRKKGRKLLQKKAAQIKRRYTNPIGRPKNRLKKQNTVSKGPFSKVRTGPGRGRKRK	312-367	Bipartite	NLStradamus
5	RGRKRK	362-367	Monopartite	NucPred
6	RGRKRKITLS	362-371	Monopartite	NLS Mapper
7	RRGRWGRKNRK	864-874	Monopartite	NLStradamus
8	KPTLKRKKPFLHRRRRVRKRK	1007-1027	Bipartite	NLStradamus
9	KPTLKRKKPFLHRRRRVRKRKH	1007-1028	Bipartite	NLS Mapper
10	KPTLKRKKPFLHRRRRVRKRKHH	1007-1029	Bipartite	NLS Mapper
11	PTLKRKKPFL	1008-1017	Monopartite	NLS Mapper
12	PTLKRKKPFLHRRRRVRKRKH	1008-1028	Bipartite	NLStradamus
13	KRKKPFLHRRRRVRKRK	1011-1027	Bipartite	NucPred
14	RRRVRKRKHHN	1020-1030	Monopartite	NLS Mapper
15	RRVRKRKHHNS	1021-1031	Monopartite	NLS Mapper
16	RCQSSSKRKSKDEEEDEESDDADDTPILKPVSL	1095-1127	Bipartite	NLS Mapper
17	LRKRDVKNSPLEPDTSTPLKKKKGW	1128-1152	Bipartite	NLS Mapper
18	LRKRDVKNSPLEPDTSTPLKKKKGWP	1128-1153	Bipartite	NLS Mapper
19	LKKKKGWPKGKSRKPIHWKKRPGRKPGFK	1146-1174	Bipartite	NLStradamus
20	KKKKGWPKGKSRKPIHWKKRPGRKPGFKLSR	1147-1177	Bipartite	NLStradamus

Table 17: Details of mono- and bipartite nuclear localisation sequences with the full-length WT MOZ sequence, as predicted by computational prediction tools

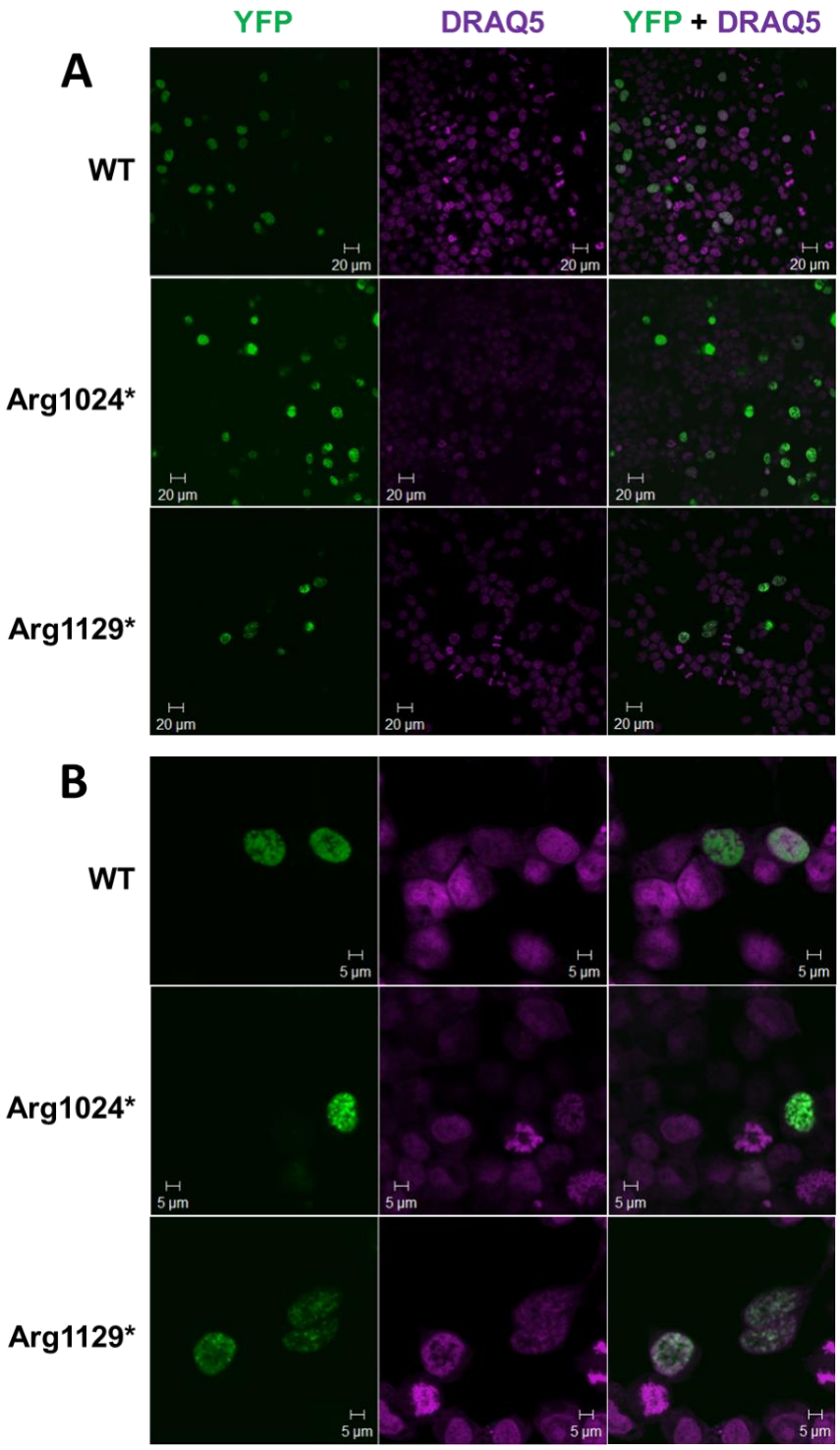


Figure 40: Cellular localisation of C-terminally truncated MOZ proteins in HEK293 MOZ KO cells

YFP-tagged MOZ protein variants (green, left panel) and nuclear staining (purple, middle panel), overlaid (right panel) in our HEK293 MOZ KO cell line after transfection with plasmids encoding the relevant MOZ transcripts. Imaged using fluorescence confocal microscopy. **(A)** 40x (no zoom) field view showing many cells, **(B)** 40x (2.5 zoom) field view showing several cells and clear nuclear staining. n = 100 cells assessed per condition.

III. VII Target gene expression is altered in cells with KAT6A GDD-like mutations

Target genes of MOZ are not well characterised, with only a handful identified in the literature. Previous work from our laboratory has identified and validated a range of MOZ target genes in both HEK293 and K562 cells, via CRISPR-Cas9-mediated MOZ knockout or siRNA-mediated knockdown, followed by RNA-seq and qRT-PCR analysis (Monteiro 2018, PhD thesis - unpublished, NCBI GEO Accession GSE100518 & GSE100513, see previous Figure 10). We sought to analyse the mRNA expression of both known and some of our newly identified MOZ target genes in our K562 KAT6A mutant cell lines, to determine the requirement for the MOZ C-terminus in their transcriptional regulation, via qRT-PCR.

All qRT-PCR primer sequences were designed using the NCBI Primer BLAST tool (Ye *et al.*, 2012), to ensure sequence specificity in the human genome. In each qRT-PCR experiment, dissociation curves were determined and assessed for indications of the presence of more than one amplicon sequence (Figure 41A). Amplification plots were also generated to check for technical issues during the qRT-PCR cycles (Figure 41B). Upon first using each pair of primers, qRT-PCR products were assessed via agarose gel electrophoresis to visually confirm amplification had occurred (or not), as well as amplicons of a single expected size. A representative example of this is shown in Figure 42A. All primer amplification efficiencies were assumed to be 100% and cycling conditions were assumed to work equally well for each gene assessed.

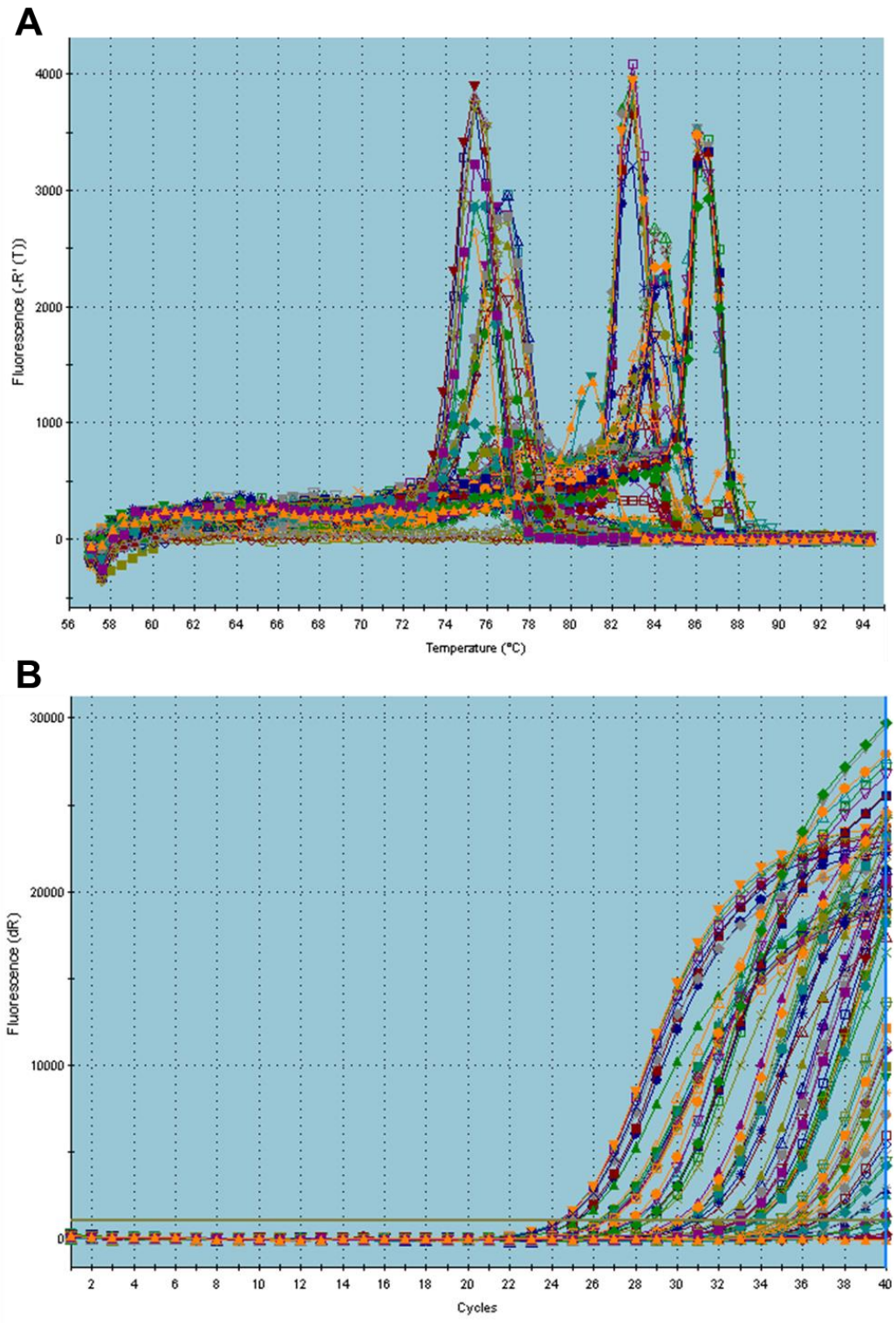


Figure 41: qRT-PCR amplification dissociation curves

(A) Representative dissociation curves plotted for several pairs of qRT-PCR primers used in a gene expression experiment. **(B)** Amplification plots of the same experiment, showing the maximum amplification has occurred for some samples.

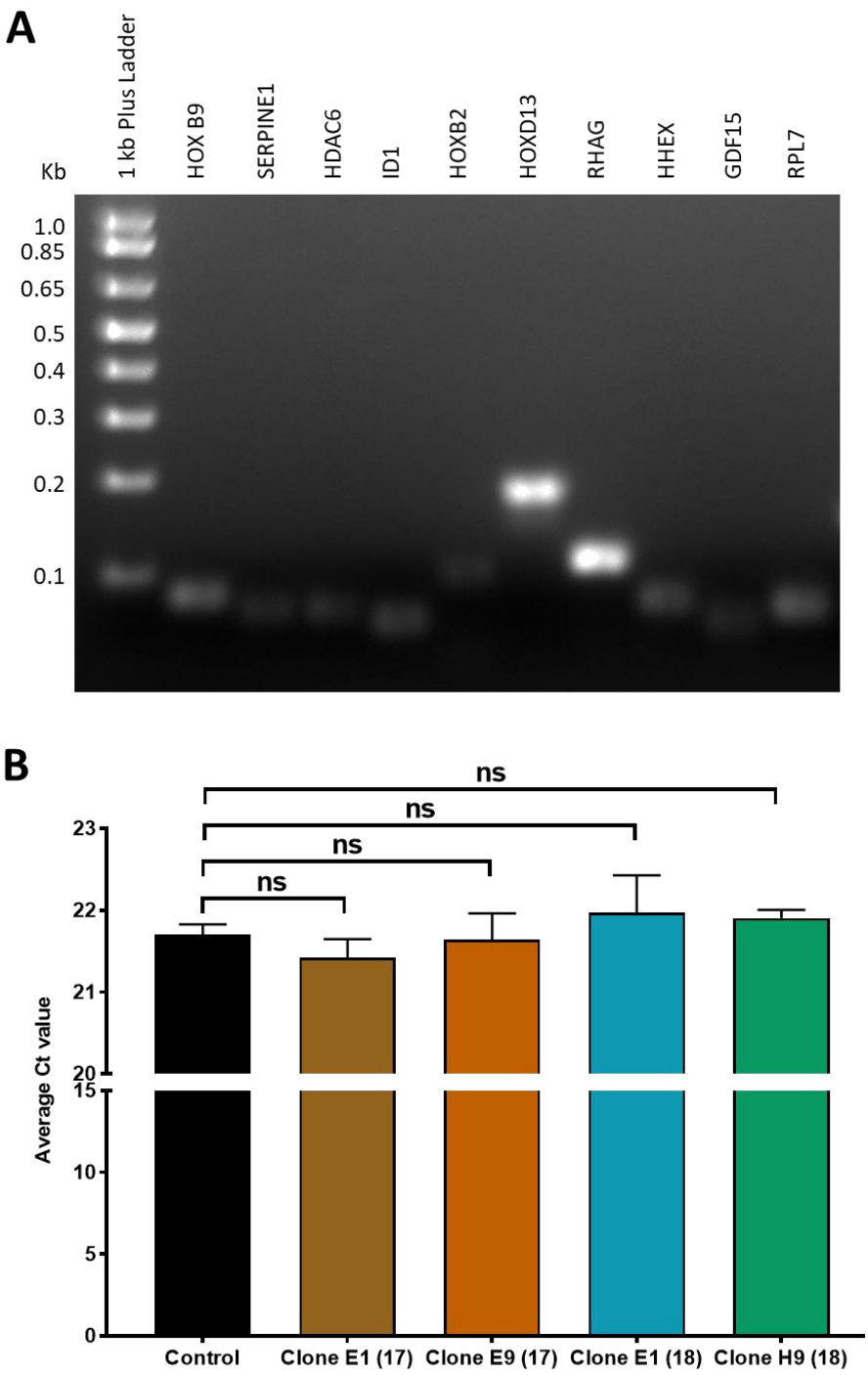


Figure 42: qRT-PCR product assessment and GAPDH expression

(A) Agarose gel electrophoresis of qRT-PCR products for the indicated genes in control K562 cells (clone G4), showing a single amplicon of the expected size was generated in each case. **(B)** mRNA expression levels of GAPDH in CRISPR-Cas9-generated cell lines harbouring KAT6A exon 17/18 PTCs. Error bars represent standard deviation of the mean. Statistical analysis performed using one-way ANOVA. ns = not significant. n = 3 for all data shown in (B).

After the extraction of RNA and cDNA synthesis had been performed for each K562 KAT6A mutant clone, the expression of our reference gene for normalisation (GAPDH) was assessed to ensure that there was no variation between samples. As can be seen in Figure 42B, the expression of GAPDH was unaffected by our KAT6A mutations, making it a suitable gene for expression normalisation. Subsequent qRT-PCR analysis revealed many of our selected MOZ target genes were downregulated in the presence of KAT6A exon 17 and 18 PTCs, such as GDF15 and PROX1, however the expression of others such as RUNX1 were only affected by exon 17 variants and not when PTCs in exon 18 were present (Figure 43). GDF15 is a secreted ligand which has important roles in regulating TGFβ-mediated gene expression and also activation of specialised neurons (de Jager *et al.*, 2011; Liu *et al.*, 2016; Hsu *et al.*, 2017). PROX1 is a transcription factor known to be crucial in embryonic development, with important roles in neurogenesis and cardiogenesis (Risebro *et al.*, 2009; Lavado *et al.*, 2010). These results suggest that in individuals with KAT6A GDD syndrome, there may be some MOZ target genes that are differentially affected by heterozygous PTCs in the penultimate, or ultimate exon of KAT6A. The effects upon the expression of some of these target genes in our previous K562 MOZ KD (siRNA) experiments can be viewed in Figure 44 for comparison, as assessed via qRT-PCR, demonstrating that they are true target genes of MOZ (Monteiro 2018, PhD thesis - unpublished).

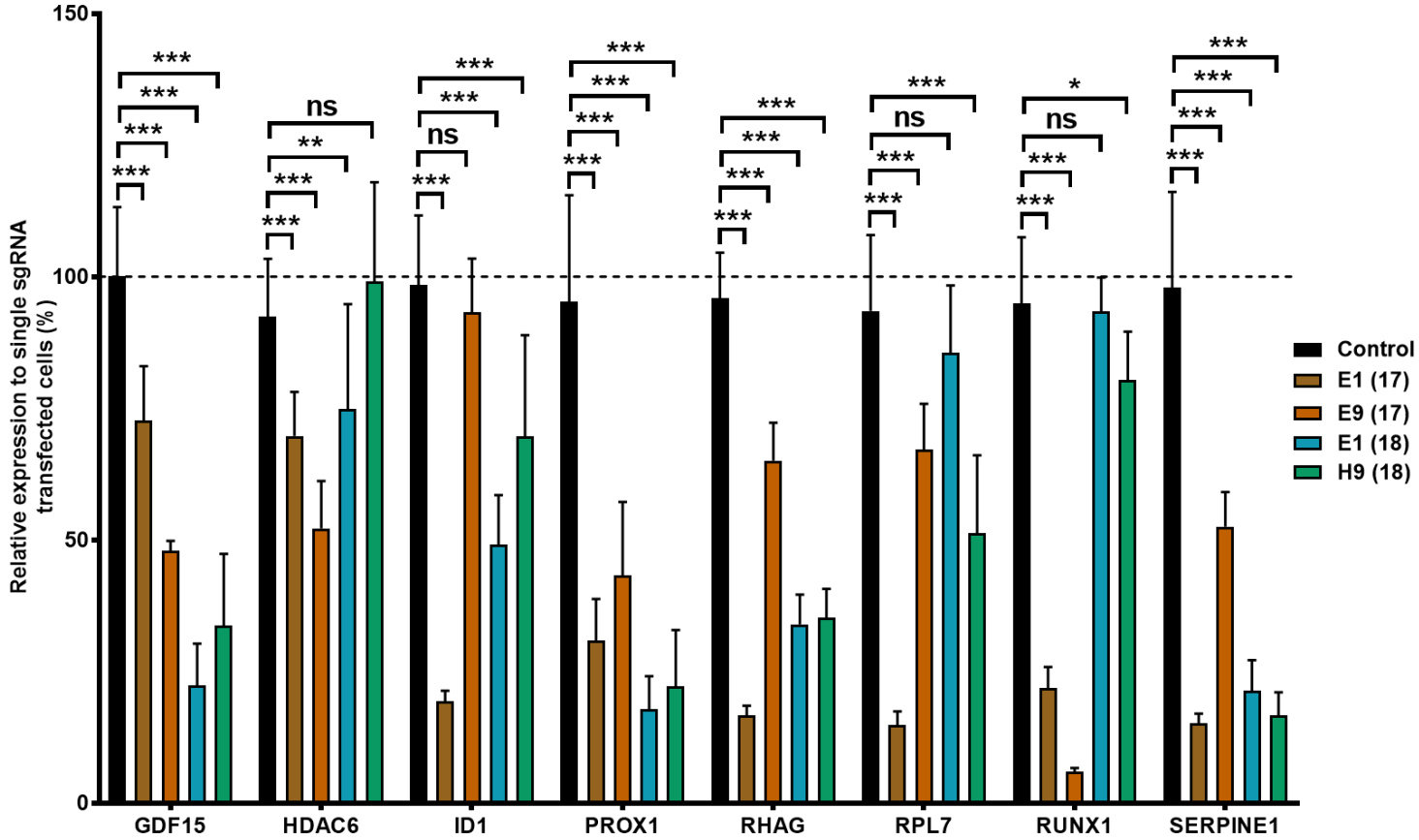


Figure 43: MOZ target gene expression in K562 cells harbouring PTCs in KAT6A exons 17 or 18

mRNA expression levels of 8 MOZ target genes in K562 CRISPR-Cas9-generated cell lines harbouring KAT6A exon 17/18 PTCs, relative to control cells (clone G4). Clone G4 is a negative K562 control, which was transfected with a single sgRNA and therefore did not undergo Cas9-mediated genome editing. Expression was normalised to GAPDH levels. Error bars represent standard deviation of the mean. Statistical analysis performed using one-way ANOVA. ns = not significant, * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$. n = 3 for all data shown here.

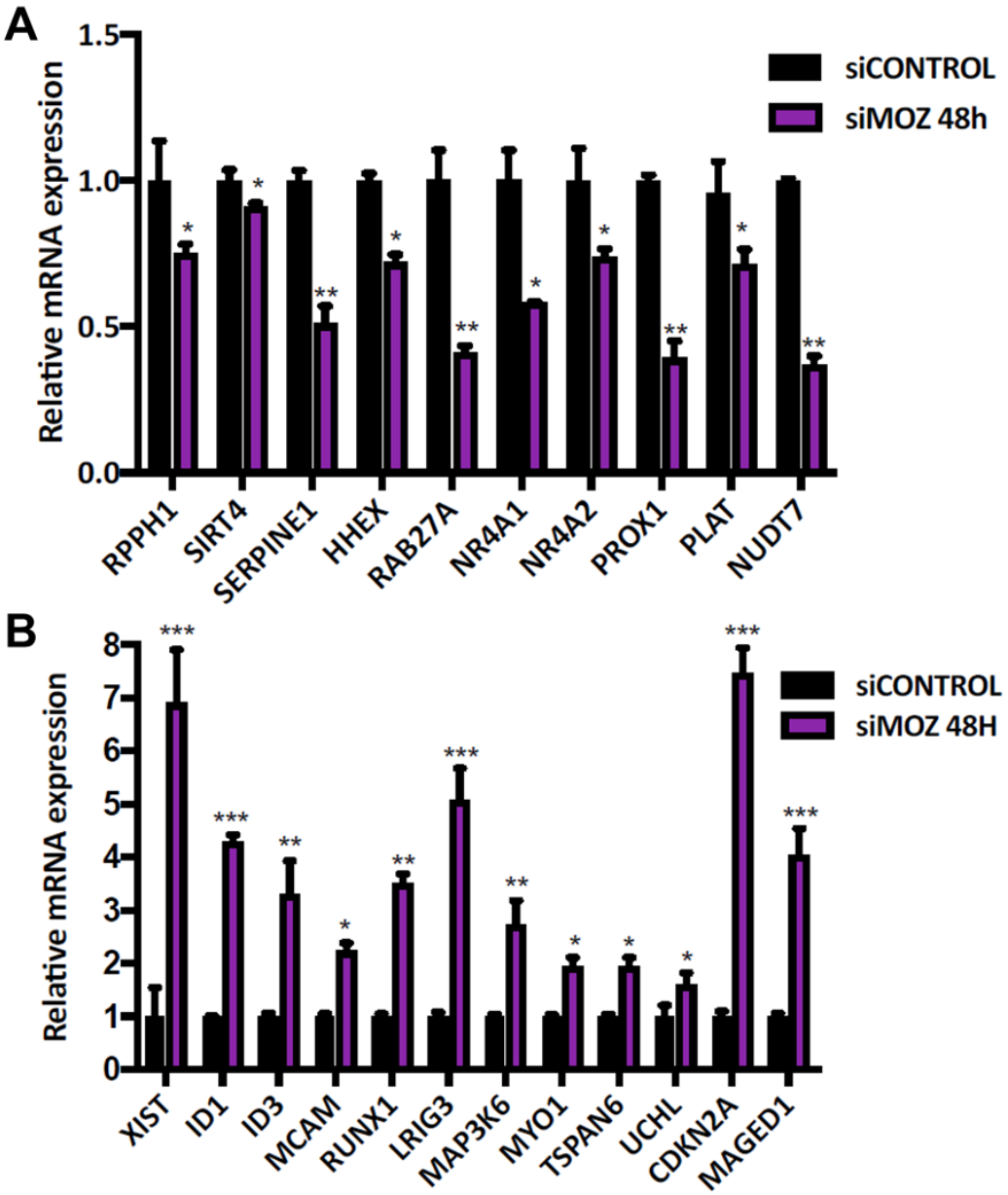


Figure 44: Validation of MOZ target genes identified by RNA-seq

qRT-PCR validation of a selection of downregulated **(A)** and upregulated **(B)** genes in K562 cells upon siRNA-mediated KAT6A KD, relative to siRNA control cells. The genes were chosen based upon RNA-seq data and assessed by qRT-PCR in order to validate changes in gene expression. Expression was normalised to GAPDH levels. Error bars represent standard error of the mean. Statistical analysis performed using Student's t-test. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$. n = 3 for all data shown here. Taken from Monteiro (2018), PhD thesis – unpublished.

Of particular interest is the downregulation of RUNX1, a transcription factor well known to be highly important in normal tissue development and specification, including that of the brain (Theriault, Roy and Stifani, 2004; Growney *et al.*, 2014; Logan, Rusnak and Symes, 2015). A differential effect upon transcription of this gene is observed between the presence of KAT6A exon 17 PTCs and those in exon 18, with a significant downregulation in the presence of the former, an effect not seen for the latter (Figure 43).

Due to the crucial role of the RUNX1 transcription factor in development, as well as in a range of biological processes, we investigated the potential for “knock-on” effects on RUNX1 target gene transcription in our cell lines. Five RUNX1 target genes were chosen, based upon both their expression in K562 cells and association with important developmental processes (Michaud *et al.*, 2008). For example, HHEX is known to have a role in haematopoietic differentiation, angiogenesis and cell survival, while CALR has been linked to neuronal differentiation (Noy *et al.*, 2010; Paz *et al.*, 2010; Shih *et al.*, 2012; Gauvrit *et al.*, 2018). Indeed, consequently several RUNX1 target genes were affected by KAT6A exon 17 PTCs, while the impact upon those same genes was generally less significant or pronounced in the presence of KAT6A exon 18 PTCs (Figure 45).

MOZ has been reported previously to regulate the expression of Hox genes in both zebrafish and mice, although it is unknown whether this is a direct or indirect effect of MOZ activity (Miller, Maves and Kimmel, 2004; Voss *et al.*, 2009). To examine the requirement for MOZ and its C-terminus in the regulation of HOX gene expression, we assessed their relative mRNA expression levels in our KAT6A PTC-harbouring K562 cell lines. Five HOX genes were chosen, all of which had been identified as transcriptionally downregulated in an ongoing K562 MOZ KD study in our lab, indicative of being regulated by MOZ HAT activity (NCBI GEO Accession GSE100518, data not shown, personal communication). The expression of all five HOX genes, namely HOXB2, HOXB3, HOXB8, HOXB9 and HOXD13 were found to be significantly downregulated,

although again, we observed a differential effect between the exon variants, with KAT6A exon 18 PTCs able to sustain a higher, or normal level of transcription for HOXB2, HOXB9 and HOXD13 (Figure 46). Alternatively, this was not the case for HOXB3 and HOXB8, whereby in all K562 mutant clones, the transcription of these genes was downregulated (Figure 46).

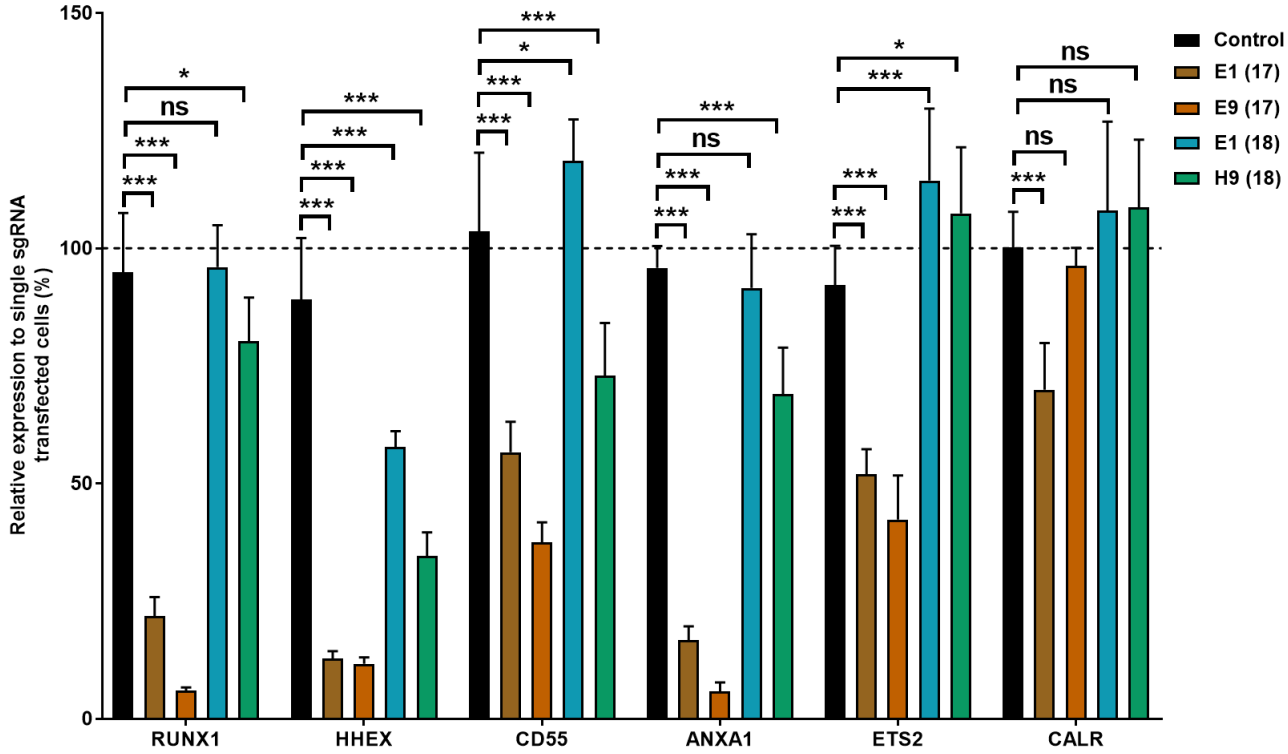


Figure 45: Expression of RUNX1 target genes in K562 cells harbouring PTCs in KAT6A exons 17 or 18

mRNA expression levels of RUNX1 target genes in K562 CRISPR-Cas9-generated cell lines harbouring KAT6A exon 17/18 PTCs, relative to control cells (clone G4). Clone G4 is a negative K562 control, which was transfected with a single sgRNA and therefore did not undergo Cas9-mediated genome editing. Expression was normalised to GAPDH levels. Error bars represent standard deviation of the mean. Statistical analysis performed using one-way ANOVA. ns = not significant, * = $p<0.05$, *** = $p<0.001$. n = 3 for all data shown here.

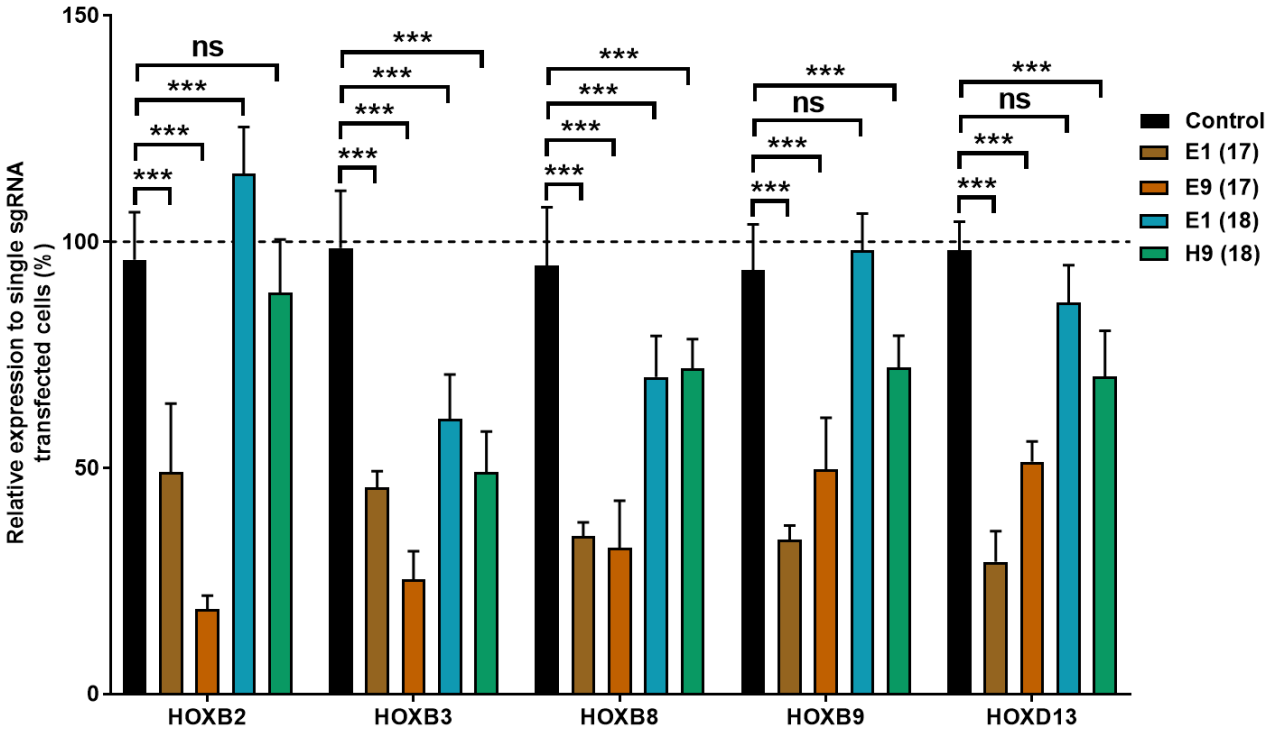


Figure 46: HOX gene expression in K562 cells harbouring PTCs in KAT6A exons 17 or 18

mRNA expression levels of HOX genes in K562 CRISPR-Cas9-generated cell lines harbouring KAT6A exon 17/18 PTCs, relative to control cells (clone G4). Clone G4 is a negative K562 control, which was transfected with a single sgRNA and therefore did not undergo Cas9-mediated genome editing. Expression was normalised to GAPDH levels. Error bars represent standard deviation of the mean. Statistical analysis performed using one-way ANOVA. ns = not significant, *** = $p<0.001$. n = 3 for all data shown here.

III. VIII Truncated MOZ protein has reduced gene transcription regulation abilities in a HEK293 MOZ KO cell line

Despite our mutant K562 cell lines generated via CRISPR-Cas9n-mediated genome editing harbouring PTCs in exons 17 or 18 of KAT6A, it is difficult to elucidate the exact contribution of these PTCs to the dysregulation of gene transcription in these cells. This is due to the heterogeneous nature of mutations present in the KAT6A alleles within each clone. For example, clone E9 (17) harbours a WT KAT6A allele, as well as an allele encoding a PTC within exon 17 and a third allele with an in-frame deletion of two amino acids in the same exon (Figure 30B). In order to definitively assess the impact of PTCs in KAT6A exons 17 and 18 upon gene transcription, and therefore investigate the ability of MOZ protein truncated in the C-terminus to regulate expression of MOZ target genes and HOX genes, we transfected a HEK293 MOZ KO cell line previously generated in our lab with plasmids encoding YFP-FLAG-WT, -Arg1024* or -Arg1129* MOZ protein (cell line generation data not shown, unpublished). These MOZ variant plasmids were generated by site-directed mutagenesis, using the YFP-FLAG-WT MOZ plasmid as a DNA template in PCR. After transformation of *E.coli* DH5α cells with newly synthesised plasmid DNA, transformants were sequenced with reverse primers specific for exon 17 or 18 of KAT6A, to check for successful incorporation of the nonsense mutations (Figure 47). Arg1024* and Arg1129* are two KAT6A variants which have both been reported in several unrelated individuals with KAT6A GDD; these two variants harbour PTCs corresponding to exon 17 and exon 18, respectively.

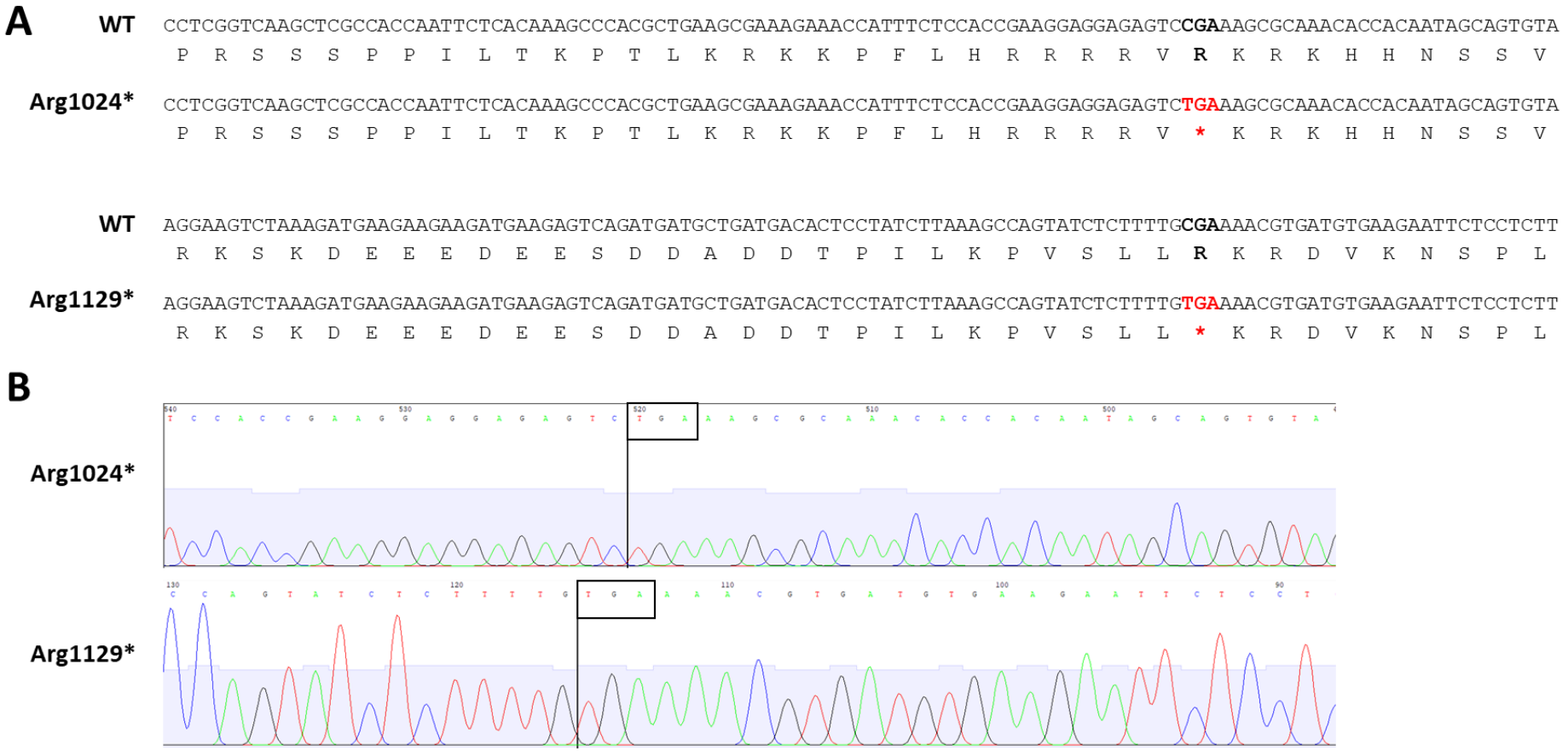


Figure 47: Sequencing results of YFP-FLAG-MOZ Arg1024* and Arg1129* variants

(A) Aligned nucleotide and translated amino acid sequences of the YFP-FLAG-MOZ variant plasmids, relative to the corresponding WT MOZ plasmid sequence. Red is the PTC, bold black is the original arginine codon. **(B)** Electropherograms of the same plasmid sequencing results with the successfully generated PTCs indicated in rectangles.

YFP-expressing cells were isolated by FACS (Figure 48) and RNA extracted, followed by cDNA synthesis. As can be seen in Figure 48, there was a clear increase in emission at the expected 515-543 nm bandwidth 48 hours post-transfection, relative to untransfected HEK293 MOZ KO cells upon excitation using a 488 nm laser. This demonstrates that the transfections with YFP-FLAG-MOZ or MOZ variants had been successful.

qRT-PCR analysis of the relative mRNA expression of WT, Arg1024* and Arg1129* MOZ transcripts did not reveal any significant differences in their mRNA levels (Figure 49A). Our previous work in HEK293 cells identified genes such as SPARC and RUNX3 as MOZ target genes that are upregulated in its absence, therefore this HAT functions to repress their transcription (Monteiro 2018, PhD thesis - unpublished). To determine whether the C-terminus of the MOZ protein is required for their transcriptional repression, we analysed the relative mRNA levels of the corresponding transcripts in our transfected HEK293 cells. In the context of HEK293 cells, expression of C-terminally truncated MOZ proteins are unable to correctly repress the expression of these two genes (Figure 49B & Figure 49C). In addition, the same truncated MOZ proteins could not promote the expression of the known MOZ target gene HOXB9 to the same extent as that of wild-type MOZ, albeit a low significant difference (Figure 49D).

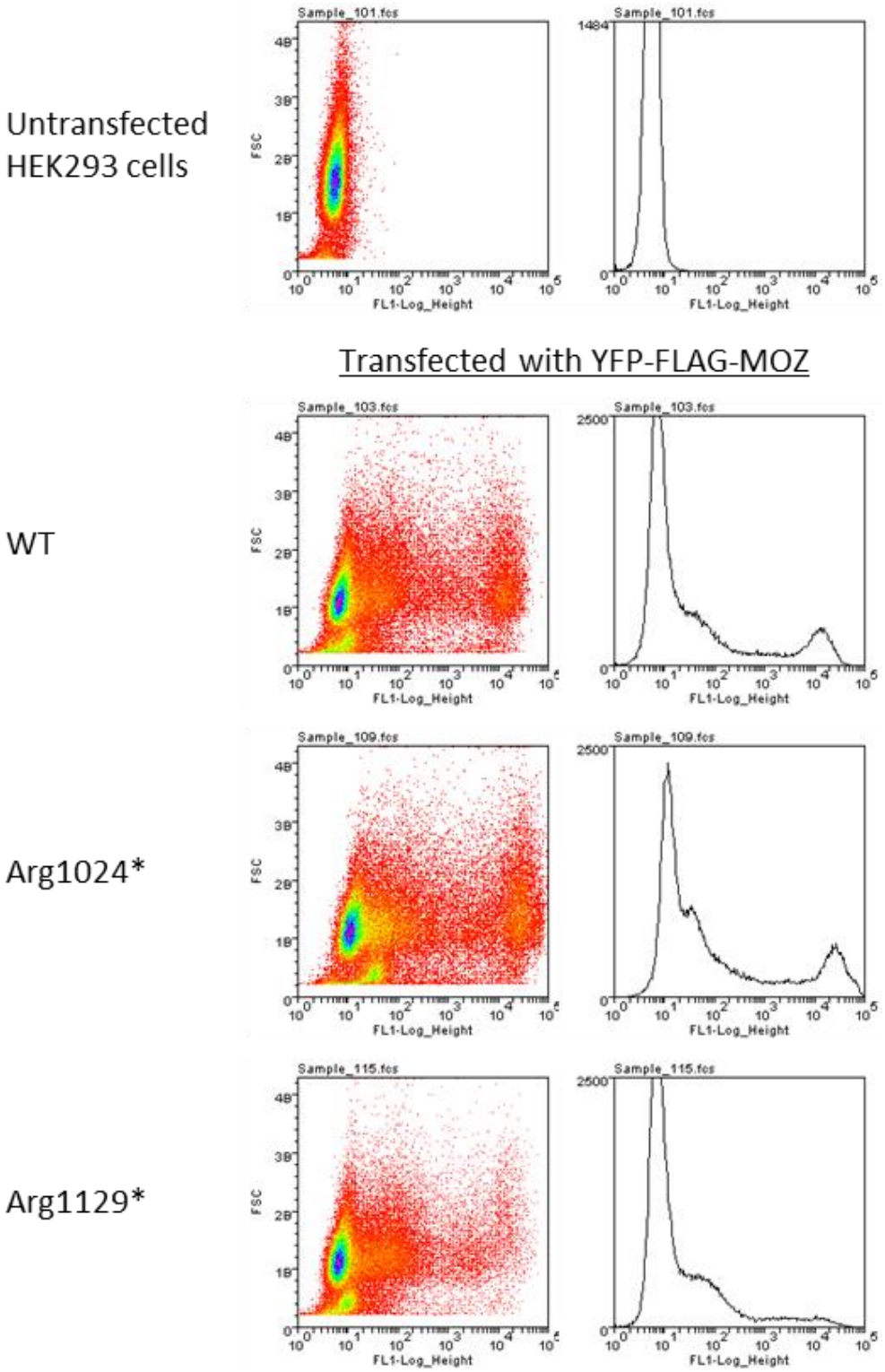


Figure 48: FACS of MOZ KO HEK293 cells transfected with plasmids encoding YFP-FLAG-MOZ WT, Arg1024* or Arg1129*

Dot plots and the corresponding histograms of MOZ KO HEK293 cells undergoing FACS after transfection with a plasmid encoding YFP-FLAG-MOZ WT, Arg1024* or Arg1129*. Transfected cells were isolated based upon their expression of YFP, relative to untransfected MOZ KO HEK293 cells via a single round of YFP enrichment.

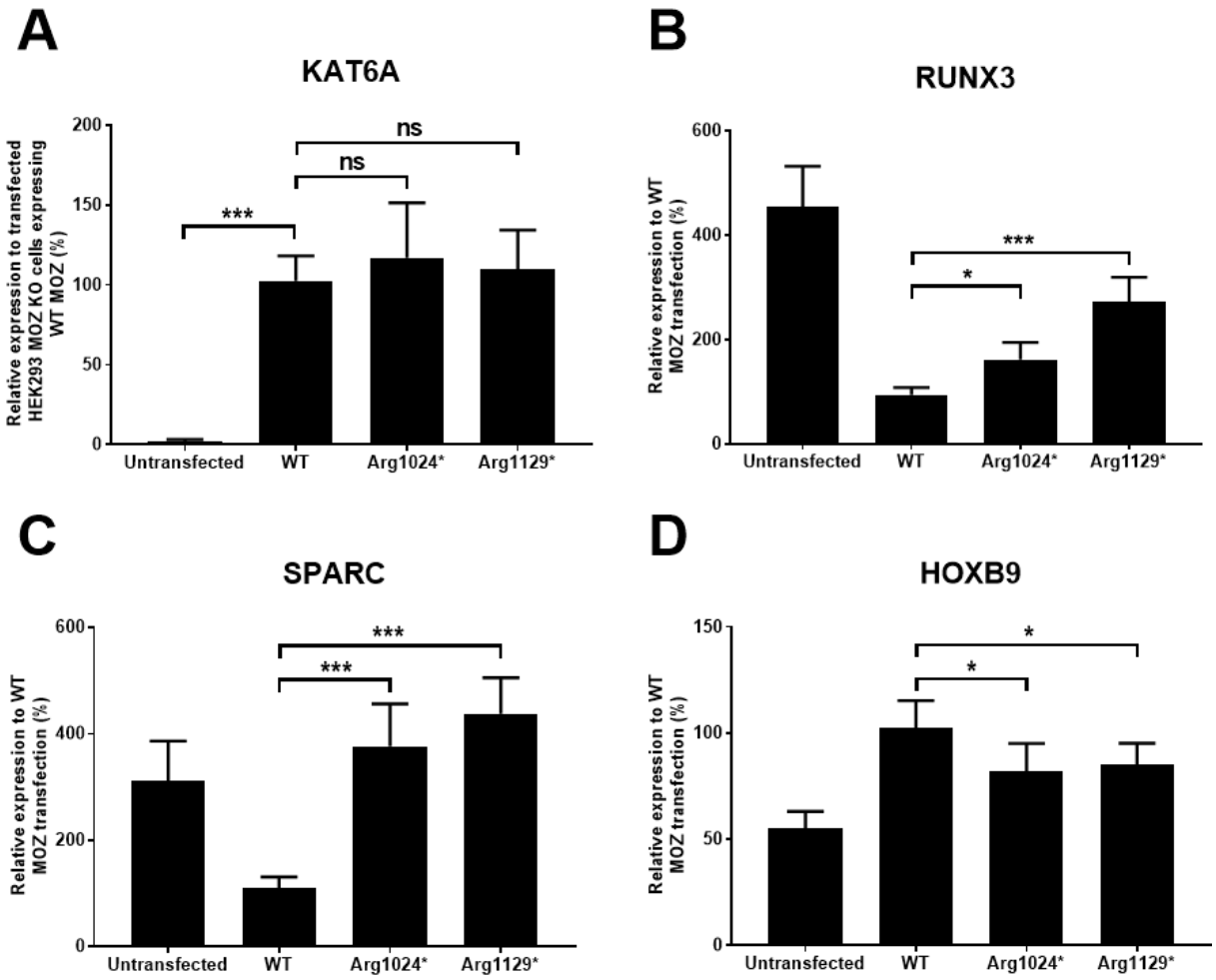


Figure 49: Ability of C-terminally truncated MOZ protein to regulate target gene expression in HEK293 MOZ KO cells

Relative mRNA expression levels of the KAT6A gene itself **(A)**, or known MOZ target genes **(B-D)** in a HEK293 MOZ knockout cell line transfected with plasmids encoding the indicated MOZ variants (WT, Arg1024* or Arg1129*). All genes were normalised for expression using the reference gene GAPDH and primers were designed as detailed in Section II. VI. ii. Note that under normal conditions in HEK293 cells, RUNX3 and SPARC are downregulated by MOZ, whereas HOXB9 is upregulated. Error bars represent standard deviation of the mean. Statistical analysis performed using one-way ANOVA. ns = not significant, * = $p<0.05$, *** = $p<0.001$. n = 3 for all data shown here.

III. IX Individuals with KAT6A GDD display abnormal gene expression

Independently-generated cell line models of a medical condition are a good way to investigate causes of pathogenesis or potential treatments, particularly if patient samples are unavailable or difficult to acquire. However, they should have a good genotypic match to those cells *in vivo* in order to truly recapitulate a condition within an artificial environment, i.e. a laboratory. Partially due to the rarity of the KAT6A GDD syndrome (~100 identified individuals to date), or at least the identification of it at present, patient samples are not easily available. There is, however RNA sequencing data from dermal fibroblasts derived from patient skin biopsies taken from two individuals with this syndrome, which we acquired (NCBI SRA Accession SRR6262221 & SRR6262222 (Arboleda *et al.*, 2015)). To assess the extent of altered gene expression in KAT6A GDD patients, as well as pathways and biological processes disrupted, we analysed this RNA-seq data.

We made use of the TrimGalore bioinformatic script to remove those RNA reads which did not have a Phred score of at least 30, as well as trim the adapter sequences from the end of reads. Alignment of these processed RNA reads to the human hg38 genome with the RNA STAR program followed, then gene expression of genomic loci was quantified via the FeatureCounts script. Subsequent calculation of log(2) changes in detectable expression levels on a whole genome basis by EdgeR with a false discovery rate of <0.05, allowed us to compare alterations in the expression of genes of interest. Rather than comparing both individuals combined to a control dataset (as was done in the previous study), we performed individual-control comparisons as well as to each other. This allows us to view any differences in the expression of genes between patients with PTCs in different exons of KAT6A, i.e. 17 vs 18. The corresponding patients were heterozygous for either the Arg1024* (Patient X) or Arg1129* (Patient Y) KAT6A variants and both were male aged 5 and 3 years of age, respectively. Regretfully, the control datasets utilised by the authors of the original study were not detailed within the

published article (Arboleda *et al.*, 2015); when contacted, the authors declined to indicate from where they had acquired the control dermal fibroblast RNA-seq data. We searched sequencing archives such as NCBI SRA (https://www.ncbi.nlm.nih.gov/sra), NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/) and ENCODE (https://www.encodeproject.org/) to find male dermal fibroblast RNA-seq datasets that would be suitable as WT controls for these comparisons. Unfortunately, none were found. The control dataset which we opted to use was RNA-seq data generated from foreskin fibroblasts of a newborn male, selected due to being gender-matched to the patient samples and also of a paediatric age (NCBI GEO Accession GSE90221).

A total of 1744 / 4014 of all the differentially expressed genes were shared between both patients, although the other 2270 genes were uniquely differentially expressed in each individual (Figure 50A). 1175 genes were found to be significantly downregulated in both patients, however 510 and 871 genes were additionally uniquely downregulated in exon 17 (Arg1024*, Patient X) or 18 (Arg1129*, Patient Y) variants, respectively (Figure 50B). A similar effect was observed for upregulated genes, whereby 561 were shared in both patients, but 386 (exon 17, Patient X) and 521 (exon 18, Patient Y) were differentially upregulated (Figure 50C). This suggests that while there is a large proportion of differentially expressed genes relative to unaffected individuals that are shared by KAT6A GDD patients, there are additional genes that are uniquely up- or downregulated for KAT6A exon 17 or exon 18 PTCs.

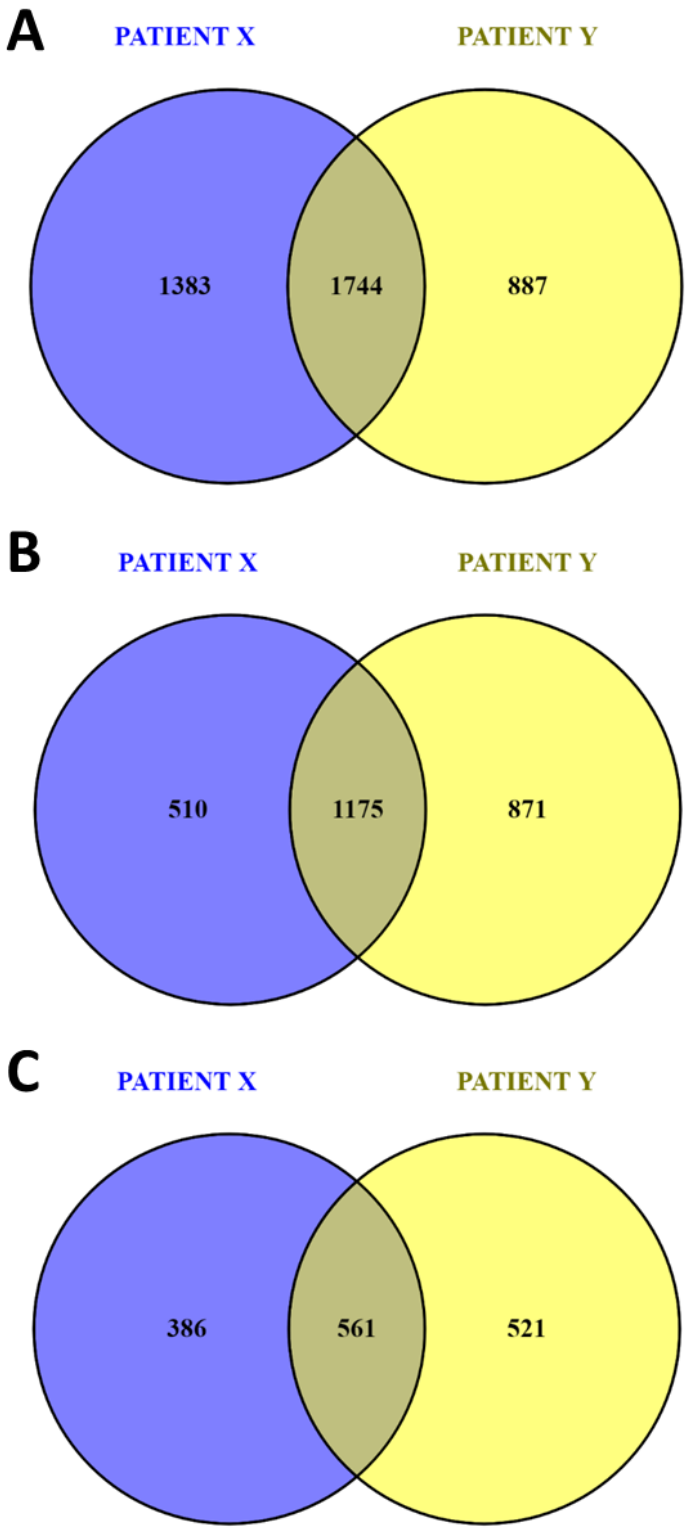


Figure 50: DEGs in individuals with KAT6A GDD syndrome

(A) Assessment of all differentially expressed genes in dermal fibroblasts derived from KAT6A GDD patients with heterozygous KAT6A exon 17 (Patient X) or 18 (Patient Y) PTC variants. RNA sequencing data revealed downregulated **(B)** and upregulated **(C)** genes compared to a control dataset. n = 1 for each of patient X & Y, n = 2 for control dataset.

Although currently, there is only data available for single individuals with a PTC in either exon 17 (Arg1024*) or 18 (Arg1129*) of KAT6A, and therefore biological replicates are lacking, several genes we identified in our cell line models are similarly affected in these patient samples relative to the control (Table 18). GDF15, PROX1, HOXB3 and HOXB8 are examples of such genes which were downregulated in the KAT6A GDD patient samples, as well as in our mutant KAT6A PTC K562 cell lines. Of note was also the reduced expression of KAT6A itself in the patient carrying a PTC in exon 17, which was not observed in the case of the exon 18 variant (Table 18). However, this did not reach statistical significance (this shall be returned to later). A comparison of the differentially expressed genes relative to the control for Patients X, Y, and our previous HEK293 MOZ knockout / K562 MOZ knockdown cells was performed ((Monteiro 2018, PhD thesis - unpublished) Figure 51). This allowed us to identify and independently validate MOZ target genes across cell lines and both *in vivo* and *in vitro (*Appendix VII – DEGs shared between individuals with KAT6A GDD, HEK293 MOZ KO and K562 MOZ siRNA cells). Only a handful of genes were differentially expressed in all four samples, however those which overlapped in Patient X, Y and the HEK293 MOZ KO cells included many HOXB cluster genes, as well as TBX1, RUNX3 and fibroblast growth factors (FGFs). This confirms these genes as MOZ transcriptional targets and also suggests they require the C-terminal domains of MOZ for the correct regulation of their expression.

Gene	KAT6A GDD 17 vs control	KAT6A GDD 18 vs control	KAT6A GDD 17 vs 18
KAT6A	-0.46	0.06	-0.76
GDF15	-0.88	-0.14	-1.08
HDAC6	-0.73	-0.65	-0.40
PROX1	-1.09	**-2.17**	0.74
ID1	1.57	**2.99**	-1.66
SERPINE1	**2.24**	1.53	0.43
HHEX	1.55	**3.58**	-2.32
CD55	-0.44	-0.20	-0.52
ANXA1	0.30	0.64	-0.61
ETS2	0.69	0.29	0.09
CALR	1.12	1.01	-0.17
RUNX1	-1.35	-0.61	-1.01
HOXB2	**-2.80**	**-1.94**	-1.24
HOXB3	**-6.67**	**-4.09**	-2.72
HOXB8	**-10.98**	**-6.46**	-4.92
HOXB9	**-11.95**	**-6.46**	-5.98
HOXD13	**8.54**	-0.79	5.17

Table 18: Affected genes in both KAT6A GDD patients and our K562 CRISPR-Cas9-generated cell lines

Fold change (log2) in expression of genes in KAT6A GDD RNA-seq data, which were also assessed in our K562 CRISPR-Cas9-generated cell lines. Patients were compared individually to a control dataset, as well as to each other. Genes underlined showed similar expression changes in both the patients and our K562 cell lines. Bold indicates those genes which reached a FDR significance of <0.05. KAT6A GDD 17 = KAT6A PTC variant in exon 17, KAT6A GDD 18 = KAT6A PTC variant in exon 18. n = 1 for each KAT6A exon variant, n = 2 for control dataset.

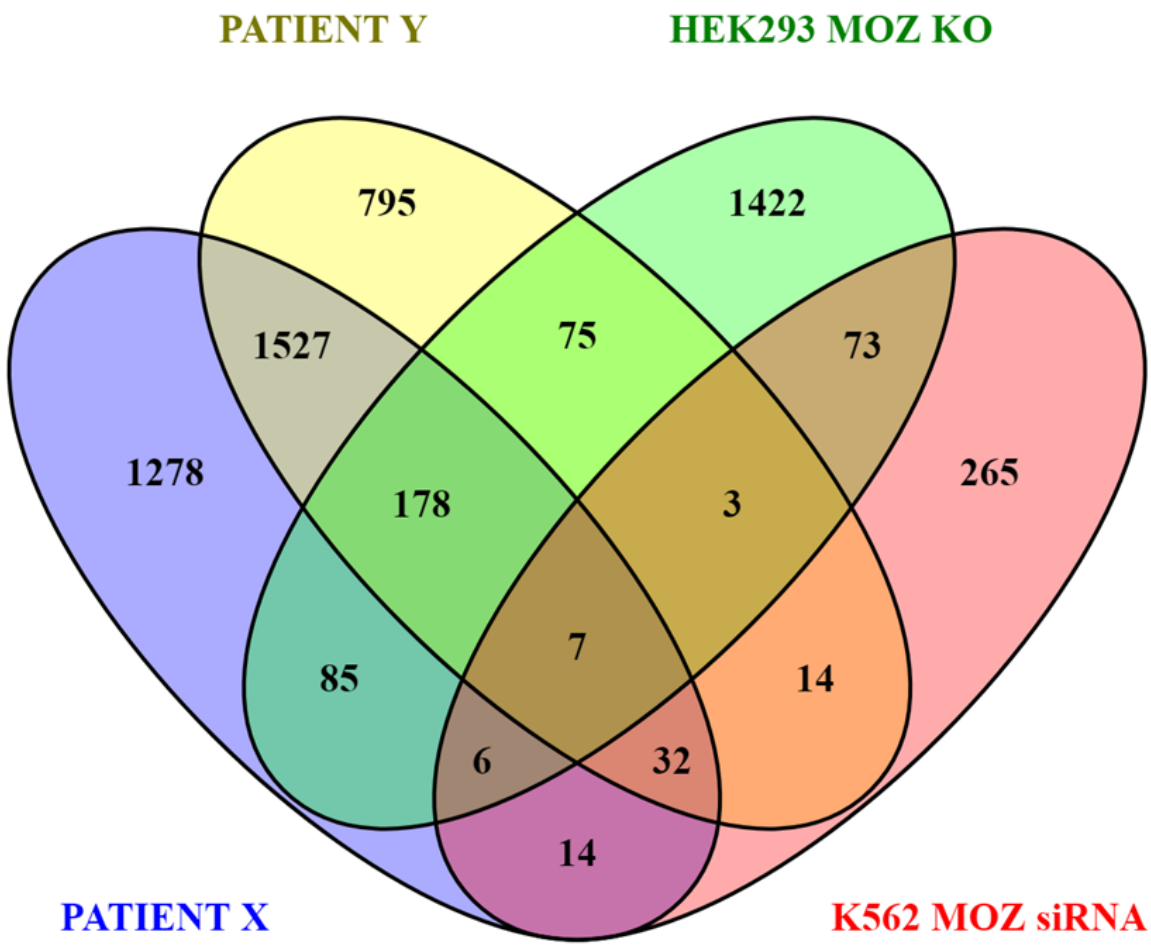


Figure 51: DEGs shared between MOZ KO, KD and C-terminal PTCs

Comparison of differentially expressed genes in four datasets, namely our two individuals with KAT6A GDD, HEK293 cells in which MOZ has been knocked out via CRISPR-Cas9n-mediated genome editing and K562 cells where MOZ expression has been knocked down using siRNA. Note that only 7 genes are differentially expressed in all four datasets. n = 1 for each of patient X & Y, n = 2 for each of HEK293 & K562 cells and in all cases n = 2 for the respective control datasets.

A clear observation from the KAT6A GDD RNA-seq data when comparing the two exon variants was the expression of the HOTAIR gene, an lncRNA product from the HOXC cluster that consequentially represses HOXD cluster expression (Li *et al.*, 2013). This gene was upregulated relative to the control almost 13-fold in the patient harbouring the KAT6A Arg1129* variant, but was unaffected in the case of the Arg1024* variant, with HOXC/HOXD expression affected accordingly (Table 19). Another noticeable change in HOX gene expression is the substantial and significant downregulation of the HOXB cluster in both KAT6A GDD patients, relative to the control (Table 19).

Despite no significant changes in RUNX1 expression in the RNA-seq data, its paralogues RUNX2 and RUNX3 were significantly downregulated (Table 20). Similar to RUNX1, these two transcription factors are associated with cell proliferation and differentiation (Torshabi *et al.*, 2011; Lucero *et al.*, 2013); additionally, RUNX2 is linked to circadian rhythm in the brain and RUNX3 to regulation of Wnt signalling and the expression of Wnt target genes (Reale *et al.*, 2013; Ju *et al.*, 2014).

The transcription factors TBX1 & TBX5 are known to be MOZ target genes, however it is not known whether other members of the TBX family are similarly regulated. In both patients, TBX1 expression was drastically reduced and TBX4 was significantly upregulated, while TBX2 and TBX5 were differentially affected (Table 20); all four of these genes are important for normal heart development (Harrelson, 2004; Krause *et al.*, 2004; Voss *et al.*, 2012). Furthermore, TBX2 has been shown in zebrafish to have a role in cell migration during neural plate formation and TBX2, TBX4 and TBX5 have also been implicated in the Wnt signalling pathway (Takeuchi *et al.*, 2003; Fong *et al.*, 2005; Lüdtke *et al.*, 2016). These findings are of great interest due to the presence of cardiac defects and clear intellectual disability in KAT6A GDD patients.

Gene	KAT6A GDD 17 vs control	KAT6A GDD 18 vs control	KAT6A GDD 17 vs 18
HOXA2	-1.64	**-2.69**	0.72
HOXA3	**-3.74**	**-4.69**	0.75
HOXA5	**-1.77**	**-3.25**	1.15
HOXA6	-1.33	**-2.45**	0.78
HOXA7	-0.90	**-2.06**	1.16
HOXA13	1.56	**3.16**	-2.39
HOXB2	**-2.80**	**-1.94**	-1.24
HOXB3	**-6.67**	**-4.09**	-2.72
HOXB4	**-11.66**	**-9.99**	-1.96
HOXB5	**-11.20**	**-6.94**	-4.45
HOXB6	**-12.15**	**-7.59**	-4.65
HOXB7	**-10.67**	**-7.16**	-3.80
HOXB8	**-10.98**	**-6.46**	-4.92
HOXB9	**-11.95**	**-6.46**	-5.98
HOXC10	**-2.44**	**4.84**	-7.87
HOXC11	-1.20	**11.85**	-8.83
HOXC12	**-4.54**	5.63	-6.16
HOXC13	**-1.92**	**9.41**	-7.38
HOTAIR	0.40	**12.86**	-8.48
HOXD1	0.00	**-6.15**	1.58
HOXD4	-3.95	**-7.46**	2.81
HOXD8	-1.53	**-2.93**	1.05
HOXD9	1.00	**-1.99**	2.51

Table 19: HOX gene expression in KAT6A GDD patients

Fold change (log2) in expression of HOX genes in KAT6A GDD RNA-seq data. Patients were compared individually to a control dataset, as well as to each other. Bold indicates those genes which reached a FDR significance of <0.05. Underlined is the HOTAIR gene. KAT6A GDD 17 = KAT6A PTC variant in exon 17, KAT6A GDD 18 = KAT6A PTC variant in exon 18. n = 1 for each KAT6A exon variant, n = 2 for control dataset.

Gene	KAT6A GDD 17 vs control	KAT6A GDD 18 vs control	KAT6A GDD 17 vs 18	Gene function
RUNX1	-1.35	-0.61	-1.01	Cell proliferation Differentiation
RUNX2	**-2.28**	**-1.80**	-0.78	
RUNX3	**-5.22**	**-5.16**	-0.30	
TBX1	**-8.97**	**-4.04**	-5.17	Heart development Neurogenesis (TBX2)
TBX2	1.80	**2.03**	-0.52	
TBX4	**2.62**	**10.38**	-3.81	
TBX5	**2.32**	1.85	0.19	
WNT4	**-6.69**	**-4.57**	-2.46	Wnt signalling
WNT11	**-3.97**	**-3.58**	-0.66	
FZD8	**-4.74**	**-3.26**	-1.80	
SFRP1	**4.86**	**3.54**	1.04	
DVL2	**-3.88**	**-3.44**	-0.71	

Table 20: Differentially expressed genes of interest in KAT6A GDD patients

Fold change (log2) in expression of genes of interest in KAT6A GDD RNA-seq data, as well as their associated function(s). Patients were compared individually to a control dataset, as well as to each other. Bold indicates those genes which reached a FDR significance of <0.05. KAT6A GDD 17 = KAT6A PTC variant in exon 17, KAT6A GDD 18 = KAT6A PTC variant in exon 18. n = 1 for each KAT6A exon variant, n = 2 for control dataset.

III. X Gene Ontologies reveal potential for disrupted developmental processes and transcriptional control

The output of differentially expressed genes from RNA-seq analysis is highly valuable, particularly in terms of whether or not the observed changes in gene expression are statistically significant, however this information alone is not enough to determine overall alterations in particular signalling pathways or associations with biological processes and disease. This limitation can be overcome by the use of tools such as DAVID or WebGestalt, which assess lists of genes for enriched associations with various biological processes, molecular functions, cellular components, pathways or phenotypes, and more (Huang, Sherman and Lempicki, 2009; Wang *et al.*, 2017). These associations can highlight the role of genes at a transcriptome-wide level. One such type of analysis using these tools is Overrepresentation Enrichment Analysis (ORA), which looks at a chosen list of genes and determines if there are any common associations shared between members of the list. ORA of differentially expressed genes shared by both KAT6A GDD patients using the WebGestalt tool, revealed the altered expression of genes associated with animal organ morphogenesis, regulation of cell proliferation, regulation of cell differentiation and tissue development (Figure 52). Gene ontology analysis of our previous K562 MOZ KD RNA-seq results revealed similar results (Table 21).

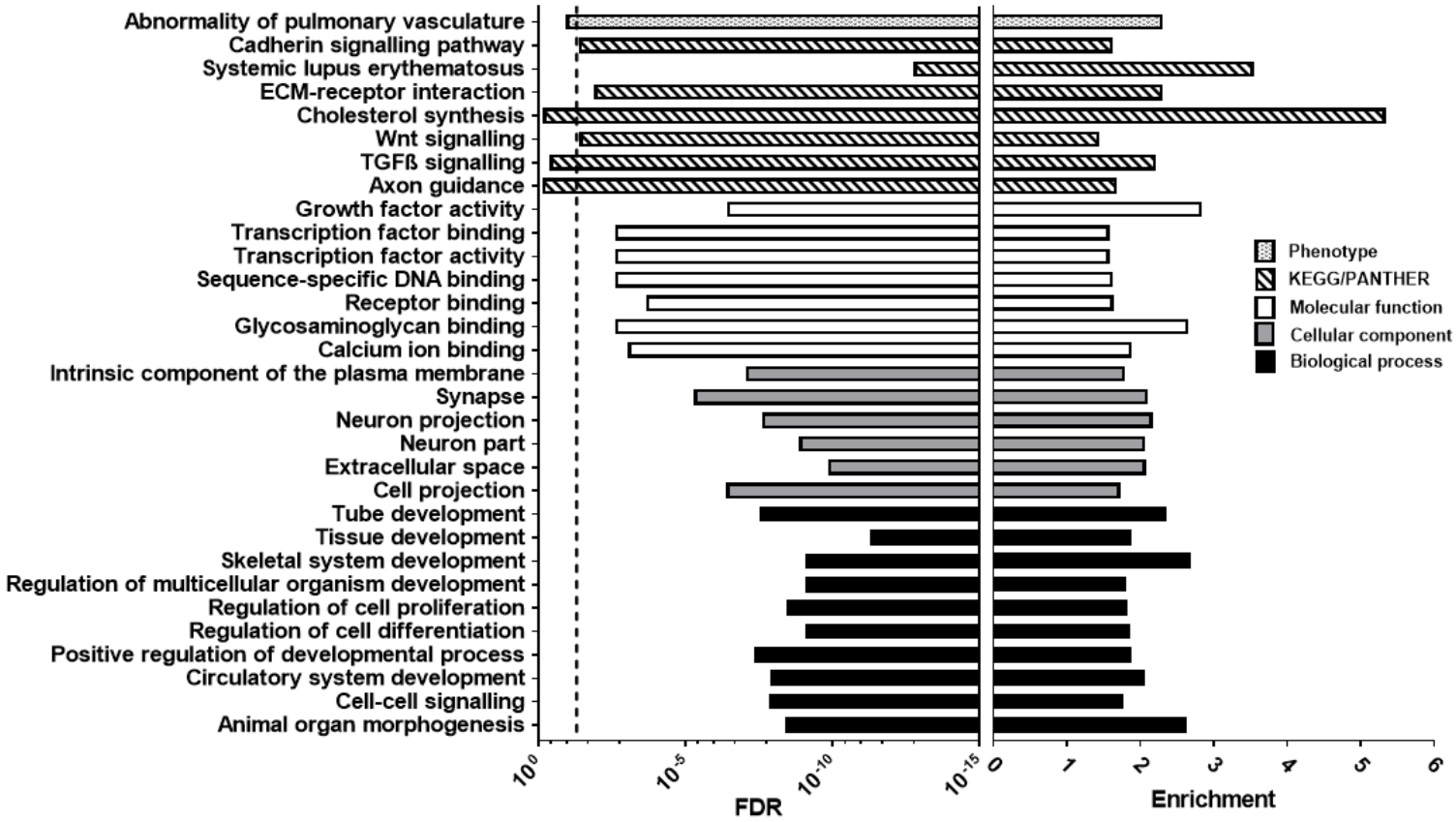


Figure 52: Gene Ontology terms for DEGs identified in KAT6A GDD RNA-seq data

Gene Ontology (molecular function, cellular component and biological process), KEGG, PANTHER and Human Phenotype Ontology terms identified from the list of differentially expressed genes determined from the KAT6A GDD patient RNA-seq data (DEGs shared by patient X & patient Y). Both the degree of enrichment (R) and false discovery rate (FDR) are shown for each term. Dotted vertical line represents a FDR of <0.05. n = 2 each for KAT6A GDD patients (combined) & controls.

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

Table 21: KAT6A target genes are associated with key biological processes
Significant biological processes associated with downregulated genes identified from RNA-seq data from siRNA-mediated KAT6A KD in K562 cells. A large proportion of the genes are involved in cell proliferation, the regulation of biological processes and/or organism development. Taken from Monteiro (2018), PhD thesis – unpublished.

In the KAT6A GDD patients, altered expression of genes associated with cell proliferation include CDKN1C and CDKN2B, two cyclin-dependent kinases associated with negative regulation of cell proliferation, as well as the fibroblast growth factors FGF9 and FGF17, both of which are linked to promoting cell growth and proliferation (Samuelsson *et al.*, 1999; Xu, Liu and Ornitz, 2000; Schreiber and Spicer, 2012; Leeper *et al.*, 2013). All four genes were found to be significantly downregulated (Table 22). Several neural-associated terms were also revealed by ORA, such as synapse and neuron projection, which is of interest due to the clinical phenotypes of intellectual disability and poor speech development in the majority of patients with this syndrome. Genes which appeared downregulated under these terms were SYP, a membrane protein found to have a role in synaptic vesicle transport and synapse formation and SNCA, a synuclein family member thought to be involved in presynaptic signalling and membrane trafficking in neurons (Table 22) (Tarsa and Goda, 2002; Kwon and Chapman, 2011; Snead and Eliezer, 2014; Phan *et al.*, 2017). Conversely, the expression of another member of the synuclein family, namely SNCG (a paralogue of SNGA) was found to be upregulated in KAT6A GDD patients. KCNB1 and KCNN2, both of which are potassium ion channels associated with synaptic functions, were significantly up- and downregulated, respectively compared to the control RNA-seq dataset (Table 22).

Gene	KAT6A GDD 17 vs control	KAT6A GDD 18 vs control	KAT6A GDD 17 vs 18	Gene function
CDKN1C	**-3.11**	**-2.41**	-0.97	Cell proliferation
CDKN2B	**-3.10**	**-2.62**	-0.73	
FGF9	**-3.62**	**-2.35**	-1.39	Cell proliferation
FGF17	**-3.37**	**-2.65**	-1.13	Differentiation / Migration
				Brain development
SYP	**-2.84**	**-2.78**	-0.37	Synapse formation
				Synaptic vesicle transport
SNCA	**-2.28**	**-3.32**	1.28	Presynaptic signalling
				Neurotransmitter transport
SNCG	**6.42**	**2.91**	3.00	Axon architecture
KCNB1	**-1.98**	**-2.34**	0.09	Action potential repolarisation
KCNN2	**11.40**	**7.20**	0.00	Regulates neuronal excitability

Table 22: Differentially expressed genes associated with cell proliferation and neural processes in KAT6A GDD patients

Fold change (log2) in expression of several differentially expressed genes associated with cell proliferation or neural processes in KAT6A GDD RNA-seq data, as well as their associated function(s). Patients were compared individually to a control dataset, as well as to each other. Bold indicates those genes which reached a FDR significance of <0.05. KAT6A GDD 17 = KAT6A PTC variant in exon 17, KAT6A GDD 18 = KAT6A PTC variant in exon 18. n = 1 for each KAT6A exon variant, n = 2 for control dataset.

Prominent molecular function terms from our ORA included transcription factor binding, sequence-specific DNA binding and transcription factor activity (Figure 52); the genes identified in these groups included many HOX, FOX and Zinc Finger genes such as HOXA5, HOXB2, FOXL2, FOXP2, ZNF213 & ZNF454 (Table 19 & Table 23). There were several Zinc Finger Proteins we found to be differentially expressed in the RNA-seq data, with the majority being downregulated (Table 23).

Also aberrantly expressed were genes involved in the Wnt and TGFβ signalling pathways, however the false discovery rate of the latter was not enough to reach significance (see previous Figure 52). Looking at the genes associated with Wnt signalling that were differentially expressed in both KAT6A GDD patients compared to the control dataset, it can be seen that some of the Wnt glycoproteins themselves are downregulated in terms of mRNA expression (WNT4, WNT11), as well as a Frizzled receptor (FZD8) and Dishevelled-like protein (DVL2) (see previous Table 20). Alternatively, the Soluble Frizzled-like Protein (SFRP1) is significantly upregulated in both KAT6A GDD individuals, although the class of proteins to which this belongs is reported to regulate Wnt signalling in a negative manner (see previous Table 20) (Dennis *et al.*, 1999; Ezan *et al.*, 2004; Wang *et al.*, 2018). From our analysis of this data, we suggest that Wnt signalling could be downregulated in individuals who harbour PTCs in exons 17 or 18 of the KAT6A gene. However, we stress that because Wnt signalling only just reached the statistical significance required for us to categorise it as definitively affected in KAT6A GDD patients, further studies are necessary to validate this.

Gene	KAT6A GDD 17 vs control	KAT6A GDD 18 vs control	KAT6A GDD 17 vs 18	Gene function
FOXD3	**-9.46**	**-9.48**	-0.09	Neural stem cell differentiation
FOXD4	**-2.83**	**-2.90**	0.13	Heart development
FOXL2	**-6.19**	**-5.56**	-0.90	Ovarian development, Wnt signalling
FOXP2	**-3.00**	**-2.19**	-1.09	Neuronal differentiation Wnt signalling, synapse formation
FOXE1	**8.43**	**3.85**	4.09	Thyroid morphogenesis Wnt signalling
ZNF213	**-2.84**	**-3.32**	0.09	Transcriptional regulation
ZNF454	**-2.47**	**-2.06**	-0.73	
ZNF628	**-2.73**	**-2.44**	-0.54	
ZFPM1	**-2.34**	**-2.20**	-0.50	Cardiac development
ZFPM2	**2.37**	**4.47**	-2.25	Haematopoiesis

Table 23: Differentially expressed FOX and Zinc Finger Proteins in KAT6A GDD patients

Fold change (log2) in expression of differentially expressed FOX and Zinc Finger Protein genes in KAT6A GDD RNA-seq data, as well as their associated function(s). Patients were compared individually to a control dataset, as well as to each other. Bold indicates those genes which reached a FDR significance of <0.05. KAT6A GDD 17 = KAT6A PTC variant in exon 17, KAT6A GDD 18 = KAT6A PTC variant in exon 18. n = 1 for each KAT6A exon variant, n = 2 for control dataset.

III. XI Nonsense-mediated decay as a contributory mechanism in KAT6A GDD

The RNA degradation mechanism known as NMD is an important surveillance system active in all cells of the human body. It has been reported to have a role in homeotic gene expression, however has probably been best characterised in terms of the degradation of mRNA transcripts which contain PTCs (Lykke-Andersen and Jensen, 2015; Colombo *et al.*, 2017). The stalling of ribosomes at PTCs 5' to the terminal exon of spliced mRNAs triggers a signalling cascade that results in the recruitment of decapping, deadenylation and endonuclease enzymes. Their activity swiftly degrades the transcript, thereby preventing further rounds of translation and stops the synthesis of a potentially undesired protein. While this is generally a beneficial occurrence that can degrade mRNAs aberrantly transcribed and do not possess an open reading frame, in some cases it has been associated with diseases like cancer and developmental syndromes that arise from either nonsense or frameshift mutations (Miller and Pearce, 2014; Lindeboom, Supek and Lehner, 2016).

To determine whether KAT6A mRNA expression levels were affected in our CRISPR-Cas9-generated cell lines, we extracted RNA, synthesised cDNA and performed qRT-PCR utilising two pairs of primers which would allow the detection of all of our mutant and WT KAT6A mRNA transcripts. For those clones which harbour PTCs within exon 17 of the KAT6A gene, we observed a significant decrease in the abundance of overall KAT6A mRNA, something which was either not, or less affected for exon 18 PTCs, indicative of NMD (Figure 53). At the protein level, however there did not appear to be a corresponding reduction in the levels of MOZ proteins translated (see previous Figure 32).

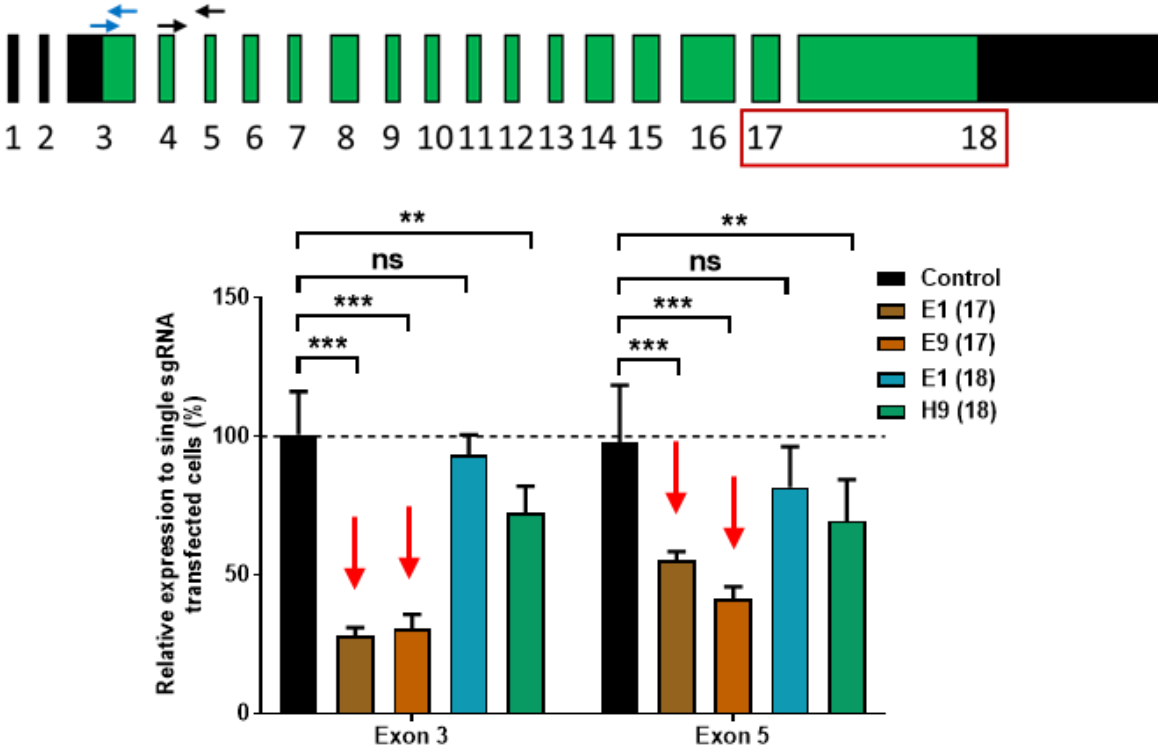


Figure 53: Total KAT6A mRNA in our K562 CRISPR-Cas9-generated cell lines

Total spliced KAT6A mRNA determined using two different pairs of primers in K562 CRISPR-Cas9-generated cell lines harbouring KAT6A PTC variants. Blue and black arrows indicate the location of KAT6A qPCR primer binding sites (exon 3 & exon 4/5 respectively), relative to numbered KAT6A mRNA exons in. mRNA expression was normalised to GAPDH levels. Black boxes represent untranslated exon regions, green boxes are translated mRNA regions. Red box shows the exons targeted by CRISPR-Cas9 mutagenesis. Red arrows indicate significantly reduced levels of detectable mRNA. Error bars represent standard deviation of the mean. Statistical analysis performed using one-way ANOVA. ns = not significant, ** = $p<0.01$, *** = $p<0.001$. n = 3 for all data.

To further investigate this, we designed primers for use in qRT-PCR to allow the detection of each specific KAT6A allele variant, or WT. These primers were designed in such a manner that we could detect either the unspliced, or spliced forms of each mRNA transcript in each clone (Figure 54 & Figure 55). To try and validate the allele specificity of these, PCR was performed using the respective clone genomic DNA as a template, prior to analysing amplicons by agarose gel electrophoresis (Figure 56).

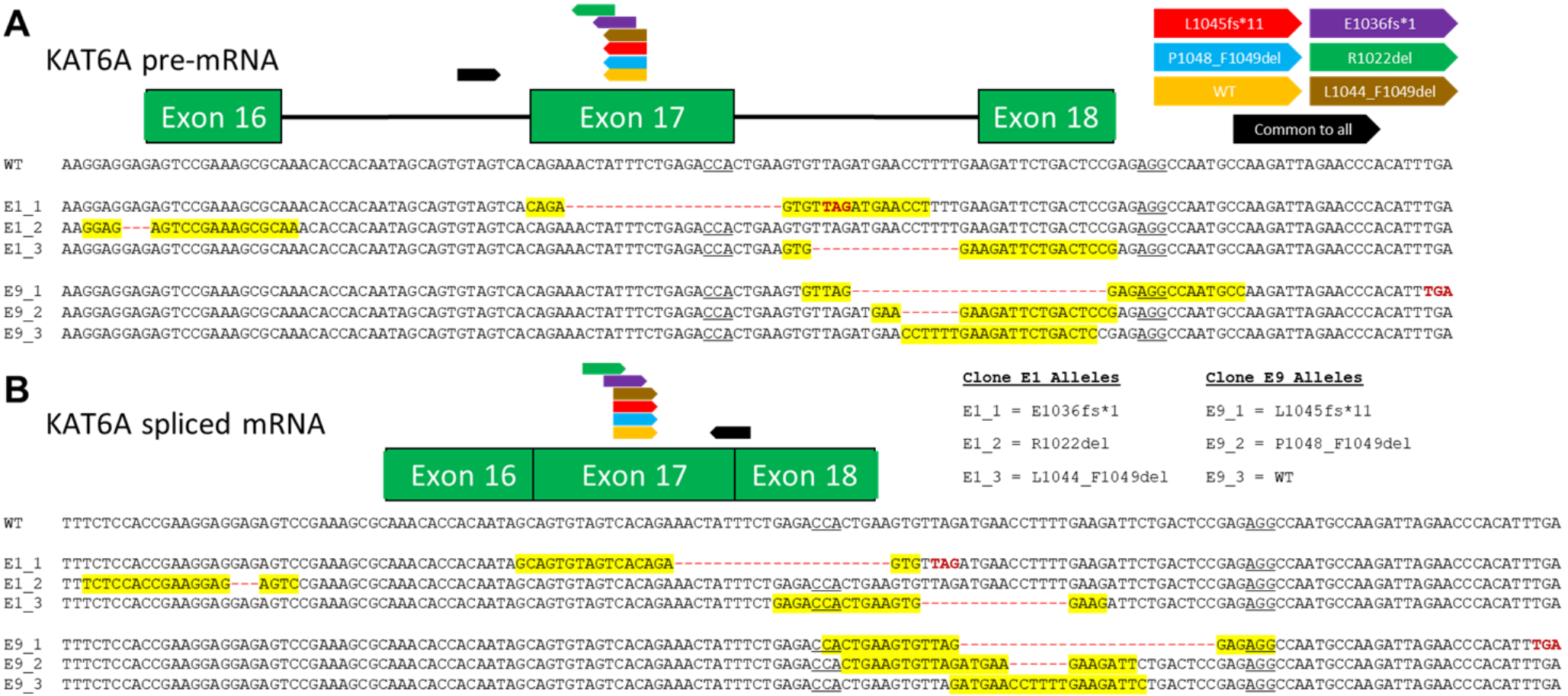


Figure 54: qRT-PCR KAT6A allele-specific primer designs for individual alleles in KAT6A exon 17 mutant clones

KAT6A pre-mRNA **(A)** and spliced mRNA **(B)** primer designs for the specific detection of each allele in K562 clones E1 (17) and E9 (17). Primer sequences are highlighted in yellow and PTCs are shown as brown text in the relevant alleles. 5' to 3' orientations of primer binding are also depicted, along with allele genotype. Variant alleles are aligned to the WT KAT6A sequence.

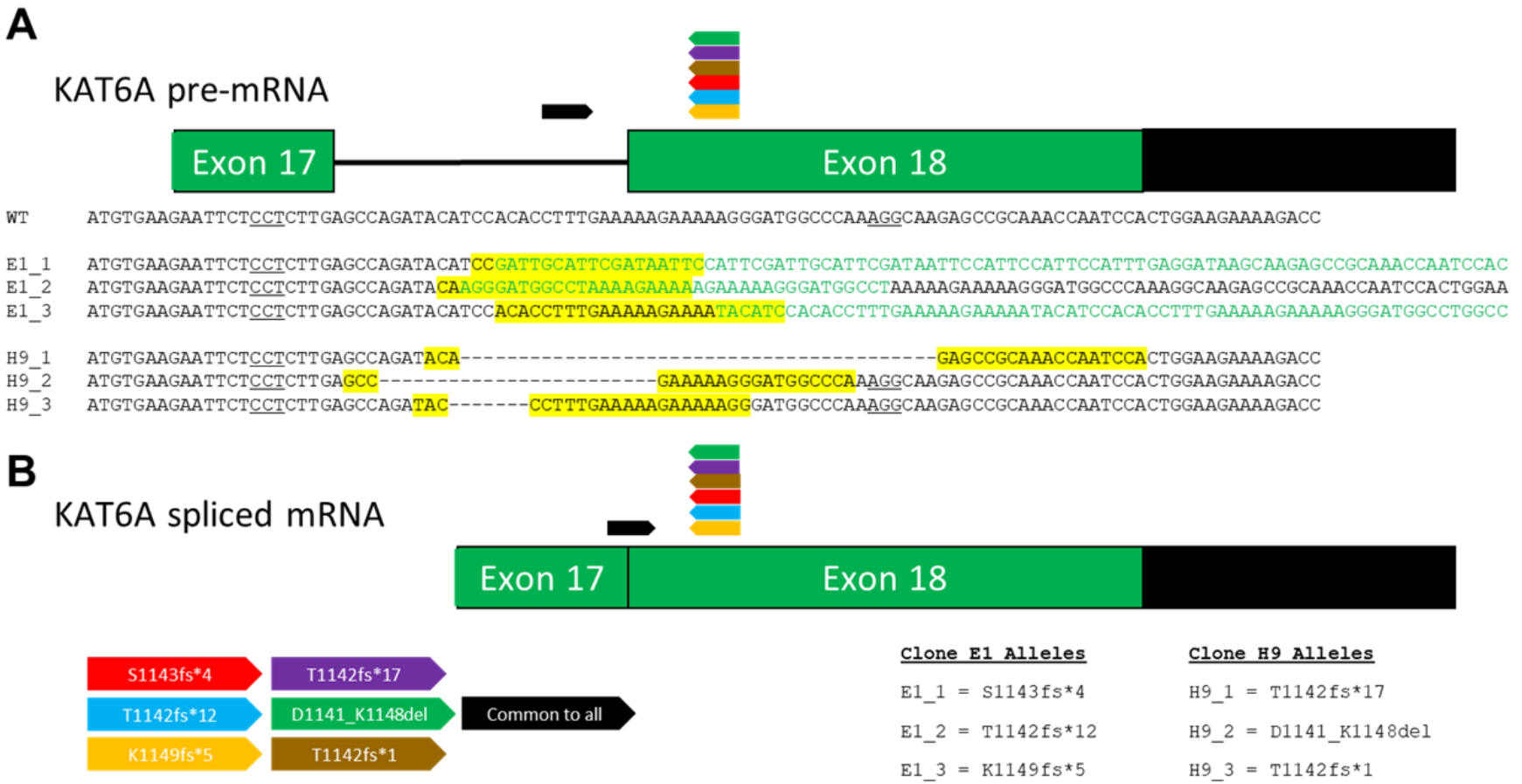


Figure 55: qRT-PCR KAT6A allele-specific primer designs for individual alleles in KAT6A exon 18 mutant clones

KAT6A pre-mRNA **(A)** and spliced mRNA **(B)** primer designs for the specific detection of each allele in K562 clones E1 (18) and H9 (18). The same reverse primers were used in both cases. Primer sequences are highlighted in yellow. 5' to 3' orientations of primer binding are also depicted, along with allele genotype. Variant alleles are aligned to the WT KAT6A sequence.

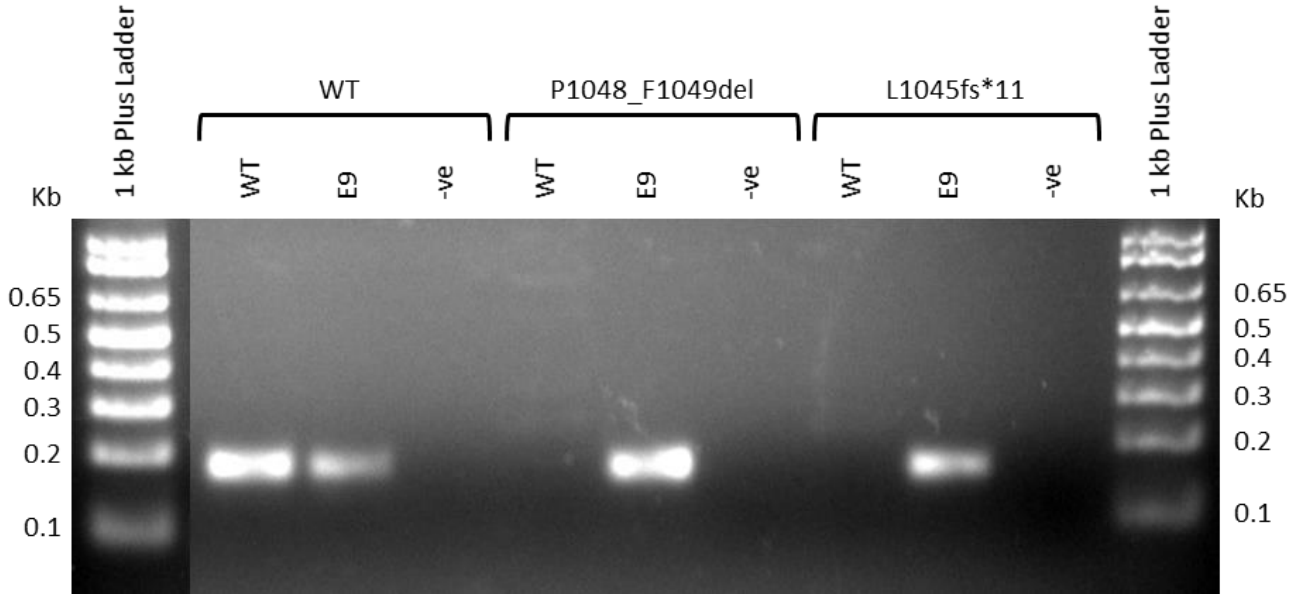


Figure 56: Amplification using KAT6A allele-specific primers

Agarose gel electrophoresis of PCR products generated from the amplification of K562 KAT6A mutant clone E9 (17) or WT K562 gDNA, using our designed KAT6A variant allele-specific primers. The primers were designed to detect pre-mRNA expressed from each KAT6A variant allele within this mutant clone; shown is a representative example of their ability to do so. Note the presence of a single band in each sample. Expected size of DNA products is ~180-190 nucleotides. This process was also repeated for all other KAT6A allele-specific primers designed to amplify either pre- or spiced mRNA transcripts, utilising gDNA from the respective K562 KAT6A mutant clone.

qRT-PCR revealed significantly lower abundances of spliced KAT6A variant mRNA transcripts harbouring PTCs in exon 17, relative to a single WT allele, compared to clone H9 (18) which has PTCs in KAT6A exon 18 (Figure 57A). This is in agreement with our previous result of the overall KAT6A mRNA levels. Contrastingly, no significant differences in unspliced mRNA abundances were detected for clones E1 (17), E9 (17) and H9 (18), which indicates the effects we have observed are not due to differences in transcription. Furthermore, it is consistent with NMD activity upon KAT6A PTC exon 17 transcripts, as NMD can only affect the degradation of spliced mRNAs, not unspliced.

Significantly increased abundance of any mRNA transcripts relative to the WT allele, indicated in our qRT-PCR experiments is likely be an artefact of primer binding to more than one KAT6A variant mRNA.

Since the amplification efficiencies of our designed KAT6A mutant allele-specific primers have not been experimentally determined, their ability to bind to more than one allele within each cell line cannot be ruled out. This may explain the results observed for colony E1 (18) in Figure 57A & B, whereby at least one KAT6A PTC allele has a detected expression level 150 – 200% of that of the WT; in-frame deletion alleles for both E1 (17) and H9 (18) also show similar heightened expression.

Analysis of read alignments for the KAT6A gene in the two KAT6A GDD patient RNA-seq samples, supports the concept of differential NMD activity between exon 17 and exon 18 PTC variants; the abundance of reads corresponding to Arg1024* are greatly reduced, almost non-existent, in comparison to those for Arg1129* (Figure 58A) (Figure 58B). This indicates that NMD may contribute to KAT6A GDD pathogenesis for those individuals who harbour PTCs in exon 17 of the KAT6A gene.

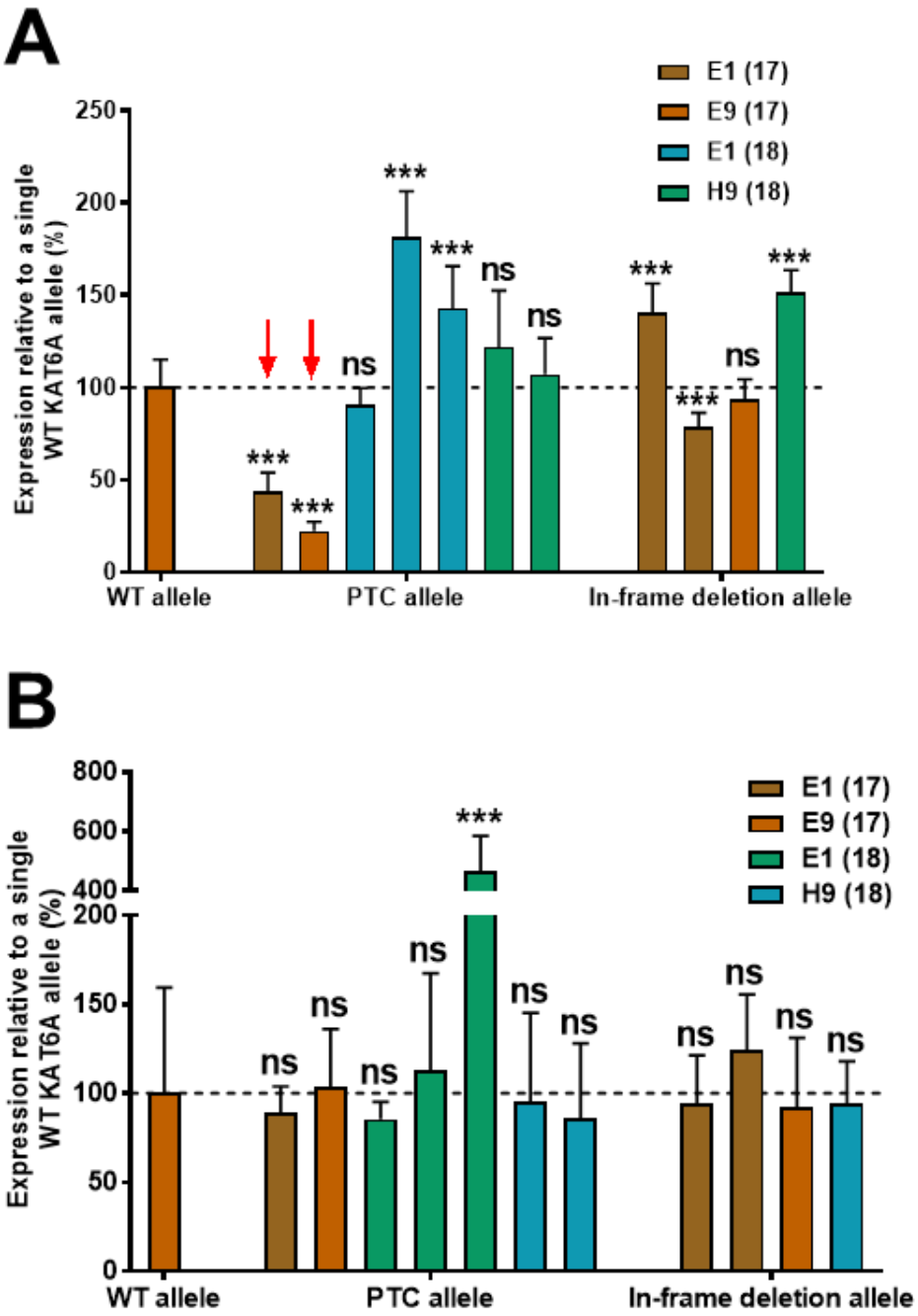


Figure 57: KAT6A pre- and spliced mRNA levels expressed from variant alleles in our K562 CRISPR-Cas9-generated cell lines

(A) Allele-specific spliced KAT6A mRNA levels and **(B)** allele-specific KAT6A pre-mRNA levels in K562 CRISPR-Cas9-generated cell lines harbouring KAT6A PTC variants. Red arrows highlight significantly decreased detectable levels of KAT6A exon 17 PTC variant spliced mRNA. Normalised to GAPDH expression. Error bars represent standard deviation of the mean. Statistical analysis performed using one-way ANOVA. ns = not significant, *** = $p<0.001$. n = 3 for all data shown here.

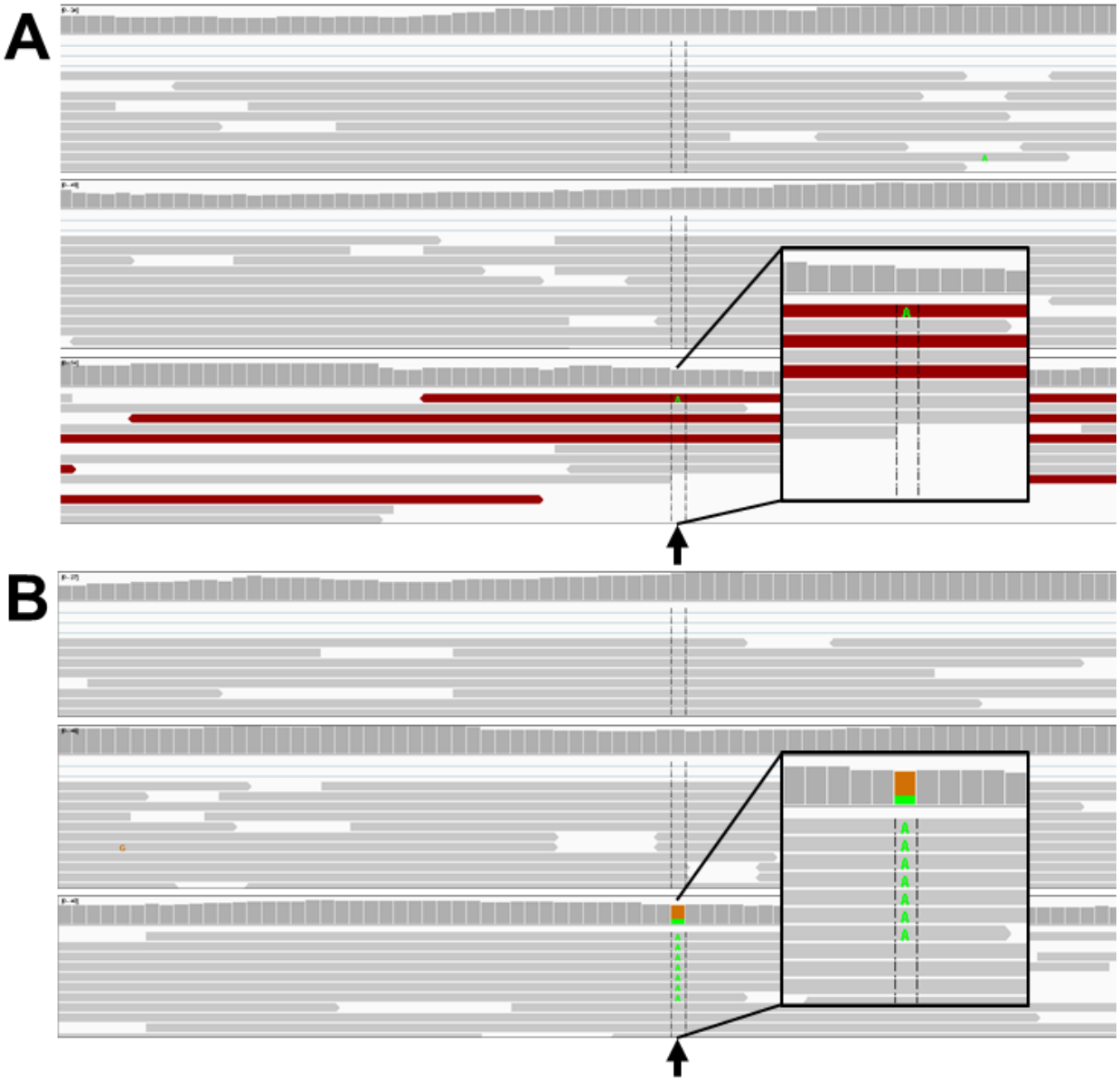


Figure 58: Abundance of KAT6A exon 17/18 variant reads in KAT6A GDD RNA-seq data

Human hg38-aligned RNA-seq reads from the two KAT6A GDD patients and control datasets, showing the **(A)** Arg1024* or **(B)** Arg1129* KAT6A variants. Top two panels in both (A) and (B) represent control datasets at the same genomic locus. Green "A" represent the C>T variants in RNA transcripts from each KAT6A GDD-associated allele (opposite strand sequenced) and arrows highlight the position of expected variant nucleotides corresponding to these. Orange and green rectangles show the relative ratio of G : A nucleotides present in RNA transcripts at that genomic position, respectively. Note that in (A), this is not visible due to a very low presence of adenine at that position, most likely due to NMD activity upon these variant transcripts. Detailed views are shown within magnified boxes. n = 1 for both KAT6A GDD variants, n = 2 for controls.

Since inhibition of NMD would be expected to increase the abundance of KAT6A mRNA transcripts harbouring PTCs in exon 17, we utilised the inhibitors NMDI14 and emetine to confirm our hypothesis of NMD activity. Emetine is a compound derived from the root of the Ipecacuanha plant and is primarily utilised as a vomit-inducing and anti-protozoal drug, however it has been found to prevent translation elongation (Carrasco, Jimenez and Vazquez, 1976; Datta, Singh and Chhuttani, 2017). Due to the halting of translation, emetine also prevents the activity of NMD, which requires at least one round of mRNA translation for its activation. Alternatively, NMDI14 is a specific synthetic inhibitor of the SMG7-UPF1 interaction necessary for NMD activation (Martin *et al.*, 2015). This consequentially prevents the signalling of stalled ribosomes to NMD machinery.

We treated our CRISPR-Cas9-generated KAT6A exon 17 and 18 mutant cell lines for 6 hours with either 50 µM NMDI14, or 100 µg/ml emetine, prior to RNA extraction, cDNA synthesis and qRT-PCR. Use of either compound significantly increased the abundance of PTC-harbouring KAT6A mRNA for exon 17 mutant clones, an effect that was not observed in exon 18 mutant clones (Figure 59A & B). This validates our model of KAT6A GDD pathogenesis, showing that NMD degrades KAT6A exon 17 PTC-harbouring mRNA, but not equivalent transcripts with PTCs in exon 18. Assessment of MOZ target gene expression after treatment of exon 17 variant clones with NMDI14, demonstrated that inhibition of NMD and therefore likely increased levels of C-terminally truncated MOZ protein, was able to partially rescue their mRNA expression levels (Figure 59C-F). Conversely, target gene and MOZ protein expression in KAT6A exon 18 PTC mutant clones was unaffected, again demonstrating that NMD selectively degrades KAT6A mRNA transcripts harbouring PTCs in exon 17, but not those in exon 18, suggestive of NMD involvement in KAT6A GDD.

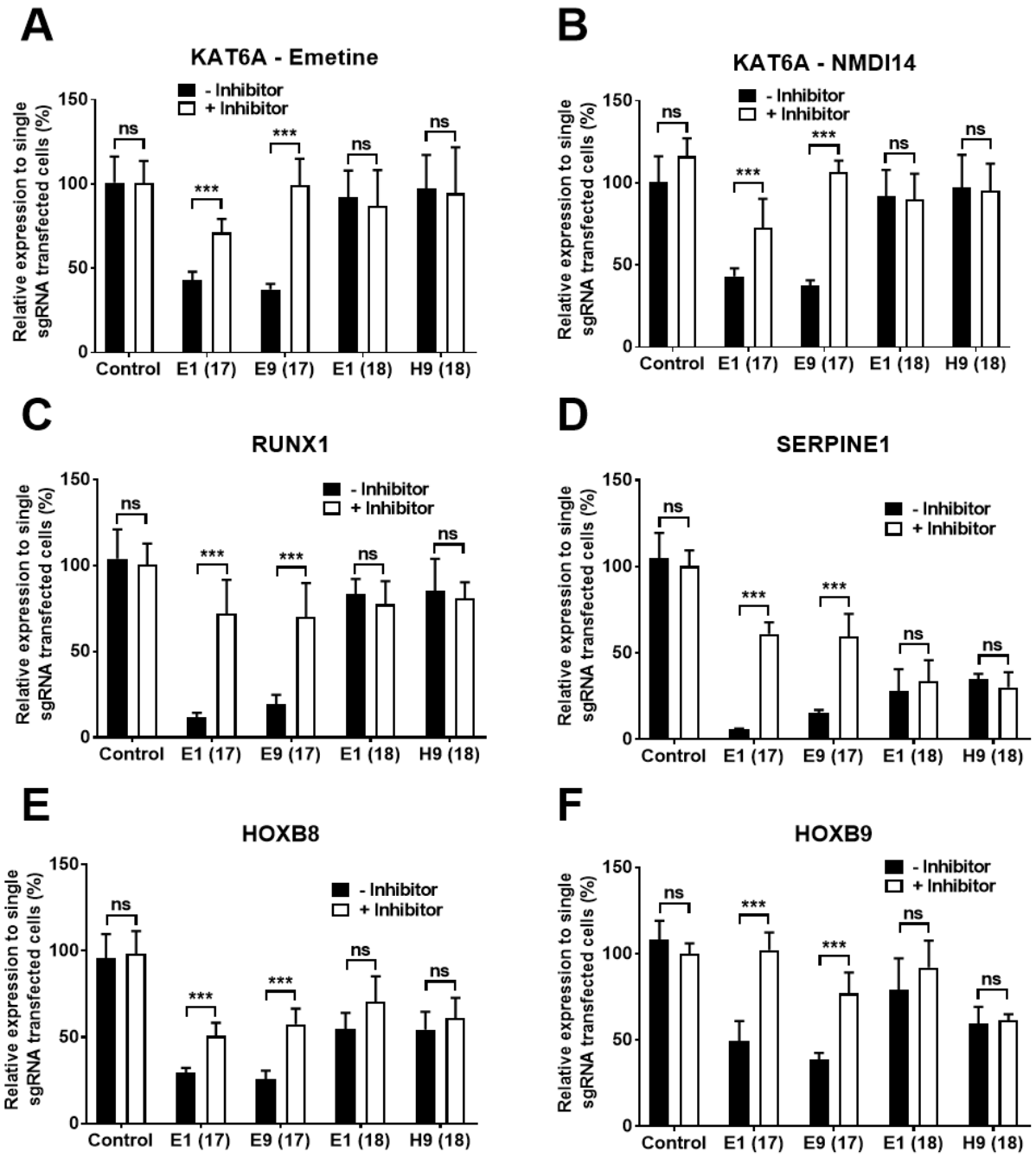


Figure 59: Inhibition of NMD partially restores KAT6A mRNA levels and MOZ target gene expression

Total spliced KAT6A mRNA levels in K562 CRISPR-Cas9-generated cell lines harbouring KAT6A PTC variants, before and after treatment with either 100 µg/ml emetine **(A)**, or 50 µM NMDI14 **(B)**. **(C-F)** Assessment of MOZ target gene expression after treatment with 50 µM NMDI14 in the same cell lines. Expression was normalised to GAPDH levels. Error bars represent standard deviation of the mean. Statistical analysis performed using unpaired t-tests. ns = not significant, *** = $p<0.001$. n = 3 for all data shown here.

CHAPTER V

DISCUSSION

Chapter V - Discussion

V. I. Protein structure prediction

In the 22 years since MOZ was first discovered and characterised, the vast majority of research upon this histone acetyltransferase has focussed upon the N-terminal domains, namely NEMM, DPF and MYST (Borrow *et al.*, 1996). A wealth of information is now known regarding the structure, function and regulation of these domains, however there is a clear lack of scientific knowledge surrounding the C-terminal regions. From a cancer perspective, there has been less interest in these due to their loss in cases of acute myeloid leukaemia that result from chromosomal translocations of the KAT6A gene. The generation of fusion genes and proteins such as MOZ-TIF2 or MOZ-CBP, always retains the MOZ N-terminal domains and substitutes the MOZ C-terminal regions for that of its corresponding fusion partner (Collins *et al.*, 2006; Serravalle *et al.*, 2010). Furthermore, efforts to express the C-terminal half of MOZ in recombinant protein expression systems such as *E.coli*, have been unsuccessful, or result in a very poor yield of recombinant protein. NCBI Protein BLAST searches of the C-terminus of MOZ do not reveal any other proteins which share a high degree of homology at the primary amino acid level (with the exception of MORF or predicted MOZ/MORF-like proteins), giving little indication of functional domains or novel information based upon similar protein sequences. For these reasons, there is no structural information available in the literature about the MOZ C-terminal domains.

MOZ has been reported to interact with RUNX1, RUNX2, p53 and also PML in its C-terminal region, with these interactions being narrowed down to the S- and PQ-rich regions, although the actual protein interaction interfaces have not yet been determined (Pelletier et al., 2002; Rokudai et al., 2009, 2013). These regions of interaction however, do not shed any light on the functionality of the acidic, S-, PQ- or M-rich regions, which together account for >800 amino acids, more than a third and almost half of the MOZ protein. Interestingly, there are >20 proteins

reported to interact with this full length HAT, such as the polyadenylation and gene expression regulating protein Symplekin and the chromatin-binding protein PHF14, but their specific interaction domains are as yet unknown (Largeot et al., 2013; Hein et al., 2015). It is therefore plausible that the undetermined interaction sites for these binding partners lie within the large C-terminal half of MOZ. Some of these interactions, however may not be direct associations with the MOZ protein, instead interacting through a protein complex, as some of the reported interactions were found using proximity-based techniques like co-immunoprecipitation (Co-IP). The vast majority of KAT6A variants associated with the KAT6A GDD syndrome lack the known C-terminal binding sites for RUNX1/2, PML and p53. This would consequently suggest that these KAT6A variants cannot function as normal to regulate signalling pathways or transcriptional programs in which these MOZ-binding partner interactions are necessary. This hypothesis is also applicable for any protein interactions not yet known, or not mapped to an absent C-terminal region of MOZ in KAT6A GDD variants.

We utilised two bioinformatics approaches to predict the presence of secondary structures within amino acids 788-2004 of MOZ (C-terminal half). Outputs suggested that the entire C-terminus of MOZ is disordered in structure. Several regions consisting of α-helices within known protein interaction sites/regions required for transactivation were predicted, although these varied between the two tools, as did their confidence level of the output. It may be the case that these regions of MOZ are indeed disordered in nature, but then take on a defined conformation when in a complex with other proteins (such as the N-terminus of p21 when Cdk2 binds (Kriwacki *et al.*, 1996). In addition, there are a wealth of amino acid residues located within the C-terminal half of MOZ known to be post-translationally modified, the most common modification of which is phosphorylation. These PTMs cluster more highly in the acidic region and may assist in stabilisation or protein folding, as has been demonstrated in other cases (such as phosphorylation of tyrosine hydroxylase by MAP

kinase (Sakaguchi *et al.*, 1997; Stultz, Levin and Edelman, 2002; Chen *et al.*, 2006; Nishi, Hashimoto and Panchenko, 2011)).

The histone acetyltransferase, CBP is known to have regions of its structure which take the conformation of a “molten globule”, which is a form of protein structure that has minimal conformation under mildly denaturing conditions until binding of a ligand or other protein (Kjaergaard, Teilum and Poulsen, 2010). Therefore, it may be that MOZ also has a similar molten globule structure. Charged regions of the MOZ C-terminal domains may engage negatively charged DNA, or it could be that there are regions which bind to the positively charged histone tails/other proteins, strengthening the interaction of this HAT with its substrates. We have determined *in silico* several potential regions of secondary structure in the MOZ C-terminus, however these require experimental investigation both in terms of tertiary structure arrangement and facilitating protein-protein interactions. This knowledge will help to characterise the molecular interactions which are lacking in patients with KAT6A GDD and elucidate the mechanisms and pathways behind the syndrome.

V. II. CRISPR-mediated generation of model cell lines

Recent advancements in genome editing techniques like the famous CRISPR-Cas9 tools, now make it possible to relatively quickly, easily and precisely edit genomic loci of interest in mammalian cells. Use of these techniques allows the study of genetic variants or indeed whole genes in biological processes, such as transcription, translation, cell proliferation, differentiation, or migration, to name just a few. Furthermore, researchers can now generate cell lines carrying genetic variants of their choice to produce disease models without the need for patient samples initially. While this has its limitations and there will always be the caveat of an *in vitro* system if applied in this way, use of CRISPR-derived cell lines have the potential to facilitate the initial mechanistic studies or screening of treatments for practically any genetic condition. This is of particular value when patient-derived cell lines or samples are not available, either due to the rarity of a disease, issues with tissue collection or consent. Moreover, the use of CRISPR tools for generating *in vivo* models is increasing, facilitating the ability of scientists to further study genetic conditions.

Here, we have chosen to utilise the advances in CRISPR technology to allow us to study the biological role of the MOZ C-terminal domains *in vitro* in the K562 cell line. By targeting the two terminal exons of the KAT6A gene which encodes this histone acetyltransferase with the Cas9n endonuclease, we have generated variant alleles that translate into C-terminally truncated MOZ proteins similar to those present in patients with KAT6A GDD. These targeted exons encode approximately half of the entire protein. Initial screening of colonies derived from single mutant cells by western blotting and PCR, revealed a range of mutations had been generated within the targeted KAT6A loci. Unfortunately, the Ser1039* variant we specifically attempted to engineer was not present in the single cell-derived colonies, despite us detecting it in the total mixed population. Clones were chosen for further study based upon the presence of predicted premature termination codons (PTCs) in MOZ C-termini, and we confirmed the expression of these proteins by western

blotting and validated the mutations in each KAT6A allele via sequencing. Of the four clones carried forwards for experimental analysis, two possessed PTCs within exon 17 of KAT6A, whereas the remaining two had PTCs within exon 18. A single clone carried a WT KAT6A allele, however all other alleles in the selected clones were deemed to have been edited by CRISPR-Cas9 and repaired via the NHEJ DNA-repair pathway. These clones harboured a mixture of mutations within other KAT6A alleles, such as in-frame deletions, however in these cases no MOZ protein could be detected by western blotting.

For the clones which were found to harbour KAT6A alleles with in-frame mutations, but no MOZ protein could be detected corresponding to the reduced molecular weight, there is the possibility that the amino acids lost within these deletions are crucial for secondary and tertiary structure. This may increase the chances of these MOZ mutant polypeptides aggregating or being degraded due to improper folding; indeed there are α-helices predicted within the targeted regions and these may be fundamental for tertiary protein structure formation or stability (see Figure 16 amino acids ~250 & ~340, corresponding to MOZ amino acids ~1040 & 1130, respectively). Alternative reasons for their lack of detection by western blotting include the possibility that their corresponding alleles are not transcribed, perhaps as a result of epigenetic modification to silence gene expression after Cas9-mediated dsDNA break formation. It may also be that the in-frame deletions alter the stability of KAT6A mRNA transcripts, reducing their abundance and consequentially preventing their translation. Further analysis will be necessary to fully explain this phenomenon.

Aside from the well documented use of various CRISPR effector proteins to cleave DNA targets, there is an increasing range of alternative uses being developed such as transcriptional repression or activation (Harrison *et al.*, 2014; Nihongaki *et al.*, 2017). Many of these utilise catalytically inactivated endonucleases like Cas9, fused to a functional protein. This has the effect of being able to utilise the endonuclease to direct active proteins to desired genomic loci upon provision of a sgRNA,

without causing DNA cleavage. Kuscu et al. (2017) demonstrated that catalytically-inactive Cas9 fused to a uracil glycosylase inhibitor and the cytidine deaminase enzyme APOBEC1, can be used to target specific codons in the human genome, resulting in C>T nucleotide conversions without the formation of a dsDNA break. These C>T conversions allow the generation of nonsense mutations at the desired genomic locus. This method does not have the risk of causing unwanted nucleotide insertions or deletions and facilitates the efficient and specific modification of codons of interest. It is estimated that approximately 17,000 human genes can be targeted in this manner, at roughly 216,000 positions (Kuscu *et al.*, 2017). As an alternative to the approach we took when trying to generate specific nonsense mutations in the KAT6A gene, this method should be considered for any future model cell line genome engineering. It does not require any DNA repair, therefore does not run the risk of indel formation via NHEJ activity. This makes it an attractive tool when considering applications such as what we aimed to do here.

V. III. Viability & cell cycle progression

MOZ and its orthologues have been well documented to influence cell proliferation and cell cycle progression. We discovered that in the presence of C-terminally truncated MOZ protein, the rate of increase in viable, metabolically active cells was significantly reduced. This indicates that MOZ protein truncated within this region, cannot promote a normal rate of increase in the number of viable cells within the population. While there appeared to be differences in the reduced rates of increase in the number of metabolically active, viable cells between our K562 mutant cell lines, we suggest that may be due to a MOZ dose-dependent effect. For example, clone E1 (18) expresses truncated MOZ protein from 3 alleles and displays a reduction in this rate, however the reduction is not as extreme as for E1 (17), which expresses truncated protein from only a single KAT6A allele (no other MOZ protein detected by western blotting). Alternatively, it is possible that the additional amino acids present in MOZ proteins translated from transcripts harbouring PTCs in exon 18, compared to exon 17, confer a heightened ability to promote an increase in the number of metabolically active, viable cells, albeit less so than the full length MOZ protein.

In addition to cell viability, we also suggest that MOZ C-terminal domains are likely necessary for the regulation of normal cell cycle progression, as we detected a reduced number of cells in the G_2/M stage for those clones with PTCs in exon 18 of the KAT6A gene, suggesting a role for MOZ at this checkpoint. It may therefore be that loss of the C-terminus of MOZ results in a gain-of-function ability to promote progression at this cell cycle checkpoint. Additionally, a complete lack of full length MOZ expression combined with the expression of truncated MOZ from a single allele, reduced the number of cells in S phase, a similar effect to that of knocking out MOZ (Monteiro 2018, PhD thesis - unpublished) (Figure 60). We hypothesise that this is due to reduced proliferation within the cell population, however our analysis using MTT assays is limited to measuring viable, metabolically active cells and does not take into account the potential for altered rates of apoptosis within the

population. For this reason, we cannot conclusively say that the rate of proliferation itself is directly reduced in K562 cells which harbour PTCs in exons 17/18 of KAT6A. Furthermore, we have not directly assessed the proportion of each cell population in either senescence or quiescence, so are unable to determine whether these states are affected in cell lines harbouring PTCs within the KAT6A gene. Despite these limitations, alterations in the expression of genes associated with the regulation of cell proliferation, were observed in our analysis of RNA sequencing data from KAT6A GDD patients with the 1024* and Arg1129* variants; this was a key GO term identified and significantly affected genes included cyclin dependent kinases, growth factors and cell division cycle proteins. Therefore, these findings demonstrate that the C-terminus of MOZ is important for this HAT's role in maintaining normal maintenance of viable cells and likely, cell proliferation.

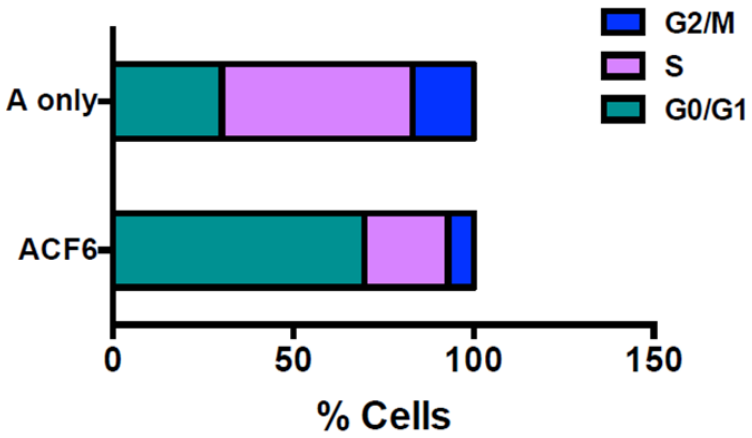


Figure 60: Cell cycle progression is affected in KAT6A KO cells

Cell cycle analysis assessed by PI staining of K562 KAT6A KO cells. A only = K562 CRISPR-Cas9 control cells, ACF6 = K562 KAT6A KO cells. Note the reduction in the proportion of the cell population in S phase and increase in G_0/G_1 for KAT6A KO cells. n = 3 for all data shown here. Taken from Monteiro (2018), PhD thesis - unpublished.

V. IV. MOZ localisation

Loss of protein-protein interactions, or specific localisation sequences can lead to the incorrect distribution and mislocalisation of proteins. An example of this is BRPF1, which is distributed between the nucleus and cytoplasm, but then becomes exclusively nuclear once bound to MOZ (Laue *et al.*, 2008). MOZ itself does not have an experimentally-defined mono- or bipartite NLS, however from our computational analysis, there are several amino acid sequences which meet these criteria. While it is known that the N-terminal NEMM domain is sufficient for nuclear localisation, it has also been shown that MOZ lacking this domain is still able to localise in the nucleus, suggesting that the NEMM domain is not the only amino acid sequence to possess nuclear targeting properties (Kitabayashi *et al.*, 2001). This can potentially be explained by our computational NLS prediction results, whereby there are four clustered regions of the MOZ protein downstream of amino acids 1-310 predicted to harbour putative NLS. If any, or all of these predicted NLS are indeed involved in the import of MOZ into the nucleus, then this explains why MOZ protein lacking both the NEMM and DPF domains can still at least partially localise to the nucleus. The fact that deletion of these two domains results in both nuclear and cytoplasmic localisation of MOZ, rather than just nuclear, suggests that a NLS upstream of amino acid 310 is the predominant dictator of import into the nucleus. From our predictions, only one putative NLS resides within the NEMM domain, from position 101-130 (Table 17, ID 1 & Figure 39). Therefore, we suggest that this is the strongest putative NLS within MOZ, while other putative NLS convey a weaker nuclear import signal.

We assessed the possibility that MOZ proteins truncated in their C-terminus could result in mislocalisation of this HAT and therefore they would not be readily available to undertake essential functions as a regulator of gene transcription at target loci. To do this, we expressed YFP-tagged MOZ variants in a HEK293 MOZ knockout cell line and visualised YFP localisation. Consistent with expectations, we found that loss of the C-terminal domains of MOZ do not affect its nuclear

localisation ability, thought to be due to the retention of the nuclear-targeting NEMM domain. Therefore, it is improbable that protein mislocalisation has a role in the pathogenesis of KAT6A variants which lack the C-terminal domains. However, we did not test this for MOZ proteins with more N-terminal PTCs, such as Arg79*, which we would predict is likely to suffer reduced nuclear localisation ability due to loss of the important NEMM domain.

V. V. MOZ target gene analysis

The MOZ histone acetyltransferase is essential for the correct normal development of vertebrates via its regulatory function in the transcription of HOX genes, inhibitors of senescence and other genes involved in tissue-specific functions. It does this by modulating the acetylation of histone lysine residues associated with genomic regions of target genes. Furthermore, MOZ is likely to regulate the expression of particular genomic loci via its interaction with other proteins that may confer specificity, such as the bromodomain protein BRPF1, or transcription factors RUNX1 or RUNX2 (Pelletier *et al.*, 2002; Bristow and Shore, 2003; Laue *et al.*, 2008). This is in addition to the DPF domain of MOZ, which can bind to acetylated H3K14, providing specificity and directing catalytic activity (Dreveny *et al.*, 2014). Interestingly, the MYST domain of MOZ has been shown to directly bind DNA, however there have not been any reports of consensus DNA sequences, therefore it is unknown as to whether this domain confers any target specificity, or merely binds to nucleosomal DNA during histone tail modification (Holbert *et al.*, 2007).

Even though a small collection of target genes of this HAT have been reported, transcriptome-wide analyses have not been performed in a developmental context. Furthermore, the studies to date which have generated MOZ differentially expressed gene lists, have all done so based upon the expression of full-length MOZ protein. While this knowledge is invaluable, particularly in cases of MOZ up- or downregulation like those observed in a subset of breast cancer and glioblastoma patients, it does little to provide information regarding the requirement of different protein domains for the subsequent regulation of these genes. Not much is known about the function or requirement of the C-terminal regions of MOZ, although they have been linked to the transcriptional activation of several HOX genes, as well as RUNX1-bound genes, suggested to be due to the physical interaction between these two proteins. The revelation that individuals who harbour nonsense mutations in the KAT6A gene, which are predicted to result in the

expression of C-terminally truncated MOZ proteins, display a global developmental delay syndrome, indicates an important role of the MOZ C-terminus in developmental processes.

HOX genes are transcription factors well documented to be expressed throughout development and have crucial roles in specifying cell identities, differentiation, body patterning and limb determination in animals from flies to humans (Mallo and Alonso, 2013; Pineault and Wellik, 2014; Seifert *et al.*, 2015). Assessment of MOZ target gene transcription in our CRISPR-Cas9-generated KAT6A mutant cell lines, revealed extensive downregulation of HOX genes. MOZ has been reported in both mice and zebrafish to regulate the expression of several Hox genes, particularly the HoxA and HoxB clusters, which is in agreement with our results (Voss *et al.*, 2009). Analysis of our KAT6A mutant cell lines indicates that the MOZ protein C-terminus is necessary for HOXB3 and HOXB8 transcription, however it is at least in part dispensable for the transcription of other HOX genes such as HOXB2, HOXB9 and HOXD13. Noticeably, in the KAT6A GDD patient RNA-seq data, we observed dysregulation in the expression of HOXA, B, C & D clusters. We suggest that the HOXC cluster is directly upregulated in its transcription by MOZ, which in turn downregulates expression of the HOXD cluster via the action of the HOTAIR lncRNA. This lncRNA is transcribed within the HOXC cluster and its function has been associated with cell proliferation, adhesion and development, although this appears to vary by cell and tissue type, suggesting that the HOX gene correlation observed here may not be the case in all human tissues (Heubach *et al.*, 2015).

We have demonstrated that the C-terminal domains of MOZ are essential for both the promotion and repression of gene transcription, as shown by our assessment of impaired SPARC & RUNX3 transcription repression / HOXB9 transcription activation, in MOZ KO HEK293 cells transiently expressing C-terminally truncated MOZ proteins. However, at the present moment, we do not have protein expression data (western blots) to confirm equal expression or stability of the respective YFP-

FLAG-MOZ protein variants utilised. Therefore, it cannot be ruled out that the differential transcriptional effects observed between WT MOZ, Arg1024* and Arg1129*, are not due to altered protein stability, folding or degradation. SPARC is a secreted matrix-associated protein necessary for adipogenesis and influences cell shape, as well as extracellular matrix (ECM) synthesis (Bradshaw, 2009; Nie and Sage, 2009; Oh *et al.*, 2013); it has also been linked to Wnt signalling (Nie and Sage, 2009). RUNX3 is a transcription factor implicated in neurogenesis and growth. It has also been reported to regulate the expression of its paralogue, RUNX1 in B cells (Spender *et al.*, 2005). This information from the scientific literature, in conjunction with our data here, potentially suggests that Wnt signalling, extracellular matrix and cell-cell signalling may be altered when MOZ is not present, or is truncated in its C-terminus. If correct, this is likely to have a detrimental impact in the context of developing brain tissue, such as in individuals with KAT6A GDD syndrome. It should however be noted that in our experiments here, we have utilised a MOZ KO HEK293 cell line, therefore our results may not be directly comparable to that observed in patient tissues, or other cell lines (such as K562), which will naturally have different transcriptional profiles.

Interestingly, we observed differential effects between the KAT6A exon 17 and the exon 18 K562 mutant cell lines for some affected genes, such as RUNX1 and HOXB2. From this, we conclude that the C-terminus of MOZ is not completely necessary for the normal expression of RUNX1, as even clone E1 (18) (which does not express WT MOZ) had unaffected RUNX1 expression (Figure 43). Alternatively, the expression of RUNX1 may instead be susceptible to a MOZ dose-dependent effect, as clones harbouring KAT6A alleles we believe to be subjected to NMD (those with PTCs in exon 17), displayed significantly downregulated RUNX1 transcription.

Crucially, while HHEX expression is reported to be upregulated by RUNX1 in the scientific literature, our data indicates that RUNX1 is not simply enough to ensure a normal level of HHEX transcription, as all K562 KAT6A mutant clones showed significant downregulation of this

transcription factor, even in the presence of RUNX1 mRNA expression (Figure 45) (Behrens *et al.*, 2017). The MOZ C-terminus (S, P/Q & M domains) is known to interact with RUNX1 and increased MOZ expression has previously been shown to enhance RUNX1-mediated transcription (Kitabayashi *et al.*, 2001). Therefore, it can be hypothesised from our data that the interaction between MOZ and RUNX1, mediated by the interaction region within the MOZ C-terminus, is essential for the correct upregulation of HHEX transcription and that without the MOZ C-terminus, RUNX1 cannot independently regulate expression of the HHEX locus (or has a poor ability to do so). Aside from HHEX, additional RUNX1 target genes were also found to be downregulated and we propose that the downregulation of these RUNX1 target genes is due to the downregulation of RUNX1 itself, demonstrating that this transcription factor is a target gene of MOZ. This confirms our findings from another ongoing research project, whereby MOZ knockout resulted in a significant downregulation in the expression of RUNX1 (data not shown, personal communication, unpublished).

Overall, our analysis indicates that MOZ protein truncated in the C-terminus can still function to regulate the expression of some target genes, with the C-terminal region lost being dispensable for this purpose, although this is not the same for all genes we assessed. For all genes, the stable expression of MOZ protein from more than a single PTC allele appears to be important, as clone E1 (17) which does not express detectable MOZ protein from 2 / 3 alleles, was frequently more severely affected in terms of target gene expression. This is suggestive of a dose-dependent effect. Notably, for many of the genes we analysed, downregulation in expression was observed even in the clone expressing WT MOZ in combination with a truncated MOZ protein. This implies that reduced full length MOZ protein (e.g. haploinsufficiency) is not sufficient for the maintenance of MOZ target gene expression. In mice haploinsufficient for Moz, defects in B cell progenitor population maintenance have been reported (Sheikh, Lee, *et al.*, 2015); it may be that patients with KAT6A GDD (who only have a single WT KAT6A allele

and a PTC allele), also have reduced stem cell populations, although this is only likely to be noticed with increasing age.

Through the use of publically available RNA-seq data, we detected a significant enrichment of genes associated with growth factor activity, transcription factor activity, regulation of cell proliferation, cell differentiation and many other interesting GO terms like tissue development, synapse formation and Wnt signalling in individuals with KAT6A GDD syndrome. Our analysis of the RNA-seq data available for these individuals who are predicted to lack expression of the C-terminal domains of MOZ, indicates that the C-terminal half of this HAT is essential for correct transcriptome-wide gene expression. Surprisingly, while there were many genes similarly transcriptionally affected in both patients, we observed genes which were differentially affected in their transcription between the two patients. This could be hypothesised to be due to the presence, or lack of additional amino acids between KAT6A variants, which may confer additional target specificity, interactions with binding partners that regulate the function of this HAT, or structural conformations. Alternatively / in conjunction, it is likely at least partially to be due to the relative abundance of these truncated proteins, as a result of NMD activity upon the KAT6A exon 17 PTC transcripts, which KAT6A exon 18 PTC transcripts are exempt from. We have shown in our K562 cell lines that this mechanism affects MOZ target gene expression and suggest this may also be the case in patients with KAT6A GDD. We propose that Patient X (Arg1024*) has a lower abundance of KAT6A mRNA than Patient Y (Arg1129*), due to selective degradation of the KAT6A PTC transcripts via NMD; our data for this aspect is not conclusive, however it does not exclude this possibility. As such, both NMD and MOZ dose-dependent mechanisms are implicated in the KAT6A GDD syndrome. Whether or not the truncated MOZ proteins in KAT6A GDD patients act in a dominant-negative manner, perhaps by having an enhanced repression activity at target loci, or rather are non-functional, thereby conferring functional protein haploinsufficiency, is unknown. Additional studies will be required to functionally characterise

C-terminally truncated MOZ proteins *in vitro,* to assess their ability to correctly assume their tertiary structure, bind targets, catalyse acetylation, recruit the correct binding partners to genomic loci and regulate transcription.

The downregulation of FGF gene transcription was also detected in KAT6A GDD patient RNA-seq data. FGF proteins are known to have essential roles in many developmental processes, with links to embryonic development, correct formation of the brain, morphogenesis and tissue repair, therefore downregulation of their expression in KAT6A GDD may impair these processes (Neugebauer and Joseph Yost, 2014; Maddaluno, Urwyler and Werner, 2017). Furthermore, FOX gene transcription was also downregulated in these individuals. FOX (Forkhead box) genes are a diverse family of transcription factors conserved throughout eukaryotes from yeast to humans, with over 40 identified in the human genome alone (Hannenhalli and Kaestner, 2009; Jackson *et al.*, 2010). They have been linked to cell cycle regulation and T-cell development, as well as embryonic, heart and thymus development. FOXP2, which was downregulated in both patients, is associated with the development of the cerebellum, a neural structure thought to coordinate voluntary motor movements and speech (Table 23) (Devanna, Middelbeek and Vernes, 2014)(Lai *et al.*, 2003; Devanna, Middelbeek and Vernes, 2014). Indeed, mutations in or deletion of this gene results in a syndrome known as FOXP2-related speech and language disorder, whereby individuals have delayed speech development (Lai *et al.*, 2001). Interestingly, individuals with this speech and language disorder also display delayed motor development and autism spectrum disorder, similar to those with KAT6A GDD, thereby impaired FOX gene transcription, perhaps directly via FOXP2 may be a mechanism by which these clinical features arise. In addition to FOXP2, FOXL2 was also found to be downregulated in both KAT6A GDD patients, a gene in which mutations have been reported to cause ptosis, a “droopy eyelid” appearance (Table 23) (Crisponi *et al.*, 2001). This characteristic has been noted in several KAT6A GDD patient

phenotypes. (Arboleda *et al.*, 2015; Tham *et al.*, 2015; Millan *et al.*, 2016; Zwaveling-Soonawala *et al.*, 2017). We suggest that this clinical feature may arise through reduced MOZ-mediated FOXL2 expression, although the potential for MOZ to regulate the transcription of FOX genes and their biological consequences, clearly requires further investigation.

Patients with KAT6A GDD display altered transcription of genes associated with neurological processes or diseases. Mutations in SYP have been associated with intellectual disability, while those in SNCA are linked to Parkinson disease (Tarpey *et al.*, 2009; Stefanis, 2012). KCNB1 is a voltage-gated K^+ membrane channel with huge importance in both the brain and cardiovascular systems, facilitating action potential repolarisation and smooth muscle contractions (Du *et al.*, 2000; Amberg *et al.*, 2004). On the other hand, KCNN2 is a calcium-activated K^+ membrane channel thought to regulate neuronal excitability (Ohtsuki *et al.*, 2012). Therefore, the alterations in expression revealed by log(2) changes for these genes, implies they are likely to be regulated in their transcription via the activity of MOZ and require its C-terminal regions for correct expression control (Table 20). Aberrant expression of these genes may impair neural function and signalling, contributing to the intellectual disability, poor speech development and seizures which are characteristics of KAT6A GDD.

Zinc Finger Proteins are a large family of proteins that are well known to bind DNA or RNA via their various zinc-finger motifs that confer target specificity. For some members of this family, little is known about their specific biological function, other than involvement in transcriptional regulation, although some have been associated with chromatin remodelling, cell cycle progression and focal adhesions, most likely through the transcriptional regulation of their target genes (Cassandri *et al.*, 2017). ZFPM1 and ZFPM2, two Zinc Finger Proteins we identified as being significantly up- and downregulated respectively in both KAT6A GDD patients (Table 23), are important in cardiogenesis due to their activity in the regulation of genes essential for the development of this organ (Walton *et al.*, 2006). Since individuals with KAT6A GDD have

developmental cardiac defects, this is of interest and along with the TBX genes, indicates a disrupted network of transcription factors/co-factors necessary for the correct development of the cardiovascular system. Additionally, in both mice and zebrafish, the orthologues of ZFPM1 and ZFPM2 are expressed in the brain, therefore it is possible they may also have a role in neurogenesis, although this remains to be seen (Walton *et al.*, 2006). MOZ itself harbours a zinc-finger motif, positioned within its MYST domain and our analysis suggests that the C-terminus of MOZ is required for the transcriptional activation of a range of other Zinc Finger Protein genes.

Within the same RNA-seq dataset, we detected significant downregulation in the transcription of genes which encode inhibitors of cell proliferation, such as CDKN1C & CDKN2B (Table 22). Looking at those genes alone, it could be expected that we would consequently observe an increase in the rate of proliferation in patient tissues. On the other hand, we also found genes known to promote proliferation to be downregulated, such as the RUNX paralogues. Therefore, it may be that downregulation of both these inhibitory and promotional proliferative genes results in a cancelling out of either effect, depending on the relative contribution of each affected gene.

Wnt signalling is an essential evolutionarily conserved developmental pathway that has crucial roles in organogenesis, cell migration and the differentiation of stem cells, to name a few processes (Komiya and Habas, 2008). Wnt proteins themselves are secreted glycoproteins that act as signalling molecules and bind to their receptors on the cell surface, known as Frizzled receptors (Fz). Their binding results in a transduction of this signal to the cytoplasmic side of the plasma membrane by Dishevelled (Dsh), or Dishevelled-like (DVL) proteins (MacDonald, Tamai and He, 2009). There are three signalling pathways which can respond to such a signal: canonical, Planar Cell Polarity and Wnt/Ca^{2+}, however there is some degree of overlap/convergence between them (Kestler and Kühl, 2008). In our ORA of the KAT6A GDD RNA-seq data, Wnt signalling marginally reached

statistical significance. This may be due to sample size limitations, however since several members of the Wnt protein family (WNT4, WNT11, FZD8 etc.) were downregulated in patient samples and Wnt signalling is of high importance in the regulation of organismal development, this provides a warrant for further future investigation in the context of KAT6A GDD.

Our results of the RNA-seq analysis have confirmed previously suggested MOZ target genes like the TBX transcription factors, and identified interesting new ones, such as WNT4, WNT11 & FZD8, which are involved in Wnt signalling. This leads us to a hypothesis that MOZ is a “master regulator” of a variety of transcription factors, as we have shown that HOX, FOX, TBX, Zinc Finger Proteins and RUNX genes are all affected in their transcription by the loss of the C-terminal regions of this HAT. Developmentally, this implies that the biological processes in which these target genes are involved, may consequentially be affected in cells harbouring heterozygous KAT6A exon 17/18 PTC variants, which would undoubtedly disrupt normal organogenesis and organ function.

The differences in affected genes between the KAT6A GDD RNA-seq data and our cell lines are likely due to analysis of different tissue types, i.e. fibroblasts from the patients and erythroleukemia K562 cells. Furthermore, our K562 KAT6A GDD model is derived from a known cancer cell line with a partially unstable genotype that differs from the normal human population. In addition, there may also be differences in gene expression due to differing diploid or triploid states of the KAT6A gene in these patients and K562 cells, respectively. For example, SERPINE1, a secreted serpin protease inhibitor involved in the inhibition of tissue plasminogen activator (tPA) and urokinase (uPA), thereby inhibiting fibrinolysis, as well as the genes ID1 and HHEX are either downregulated or unchanged in their mRNA expression in our CRISPR-Cas9-generated cell lines, whereas all three of them appear to be upregulated in KAT6A GDD patients (Table 18) (Ehrlich *et al.*, 1990; Planus *et al.*, 1997; Jiang *et al.*, 2017). Furthermore, our K562 KAT6A mutant cell lines harbour a variety of mutations, including both PTCs and

in-frame deletions. Determining the relative contribution of each mutant allele to the disrupted gene expression landscape is not straightforward. In contrast to this, individuals with KAT6A GDD have only a single KAT6A PTC allele, with the other allele being WT, allowing for an easier interpretation of gene expression changes when comparing data to normal, healthy individuals.

Caveats to our RNA-seq analysis unfortunately include an extremely small sample size and the use of a control dataset which is not perfectly matched to the KAT6A GDD patients. For these reasons, it is likely that significance and average log(2) values of the genes of interest we identified will change with the inclusion of more KAT6A GDD samples and the use of a better matched control dataset. An example of an ideal control sample would be an unaffected sibling of someone with this syndrome, of the same gender and using tissue isolated from a similar site of the body. Future RNA-seq analysis of individuals with the KAT6A GDD syndrome will assist to validate our findings.

V. VI. NMD of KAT6A mRNA transcripts

It has been hypothesised that the phenotype of patients with KAT6A GDD is a result of aberrant MOZ transactivation activity, however there has also been some suggestions of the potential for NMD. Evidence for both, or either mechanisms in the manifestation of the KAT6A GDD syndrome has, until now, been very limited. Crucially, we detected in both our CRISPR-Cas9-generated cell lines and the KAT6A GDD RNA-seq data, a reduction in the abundance of spliced KAT6A mRNA transcribed from KAT6A alleles which harbour PTC variants in the penultimate exon (exon 17). Consequently, the total pool of KAT6A mRNA was reduced. While this was significant in our K562 KAT6A mutant clones, it was not deemed a significant decrease in the patient tissues, although this may be due to a small sample size. Assessment visually by eye based upon the number of KAT6A WT / mutant mapped RNA-seq reads, however did appear to follow the same pattern. In contrast to these findings for PTCs in exon 17, we found those in exon 18 (the terminal exon) are unaffected in their abundance; a larger KAT6A GDD patient sample size will be necessary to validate our KAT6A mRNA expression findings.

The reduced abundance of KAT6A mRNA transcripts in clones harbouring PTCs in exon 17, would suggest that either reduced transcription at KAT6A promoters, or an RNA degradation mechanism is responsible. Because no significant decrease, or a lesser significant decrease in mRNA levels was observed for clones with PTCs in exon 18, we hypothesised that NMD is the likely cause. In addition, there was no significant difference in KAT6A pre-mRNA levels between the alleles with or without PTCs, indicating that the effects observed at the spliced mRNA level were not due to reduced transcription. Indeed, treatment of our K562 KAT6A PTC cell lines with one of two inhibitors that block NMD activity (emetine / NMDI14), resulted in the rescue of KAT6A mRNA expression levels in the presence of exon 17 PTCs. NMD inhibition also managed to partially rescue the expression of MOZ target genes in these cells. In agreement with a lack of NMD activity towards mRNA transcripts harbouring PTCs in exon 18 of KAT6A, NMD inhibition did not alter the

abundance of KAT6A mRNA, or the expression of MOZ target genes in KAT6A exon 18 PTC clones. However, further work to experimentally calculate KAT6A allele primer pair amplification efficiencies, followed by usage of the Pfaffl equation for qRT-PCR data is needed to confirm these findings. This will allow us to take into consideration any differences in their amplification efficiencies, which our use of the delta-delta Ct method here does not (Livak and Schmittgen, 2001; Pfaffl, 2001); it current assumes an amplification efficiency of 90 – 110%, which we suspect may not be the case for some of our primers.

V. VII. Origins of KAT6A GDD-associated variants & therapy potential

Pathogenic variants located within the KAT6A gene that are associated with the KAT6A GDD syndrome, arise *de novo* in almost all cases. Exactly how these arise and whether they are a result of germline mosaicism in the parents of affected individuals, is unknown. We chose to look into this by assessing the genetic characteristics of loci where these pathogenic variants reside.

DNA methylation has long been known to function in a regulatory manner in terms of the transcriptional activation/repression of genetic loci, however it is only in more recent times that the range of DNA epigenetic modifications are beginning to be understood. This is in a large part due to technological advances that now enable the identification of novel modifications and their subsequent study. 5-methylcytosine (5mC) was the first type of DNA methylation to be studied, however with 5-hydroxymethylcytosine (5hmC) and N6-methyladenine (6mA) now identified, along with their corresponding enzymes, this field is an active area of research (Breiling and Lyko, 2015). 5-methylcytosine is an evolutionarily conserved epigenetic modification of chromatin, which occurs at dinucleotide cytosine-guanine (CG) sequences called CpG sites and has an important role in transcriptional regulation and imprinting of the X-chromosome (Bird, 2002). It has been estimated that between 1 and 4% of all cytosine nucleotides in the human genome can be methylated in this manner, with this modification being symmetric, i.e. present on both the forward and reverse DNA strands (Ehrlich *et al.*, 1982; Breiling and Lyko, 2015). 80% of CpG dinucleotides are thought to be methylated in human cells, although methylation analyses have revealed that 5-methylcytosine is lacking in regions surrounding active enhancers and promoters (Breiling and Lyko, 2015). Suggestions have arisen that these genomic regions by default are methylated, requiring demethylation prior to becoming transcriptionally active (Bird, 2002; Breiling and Lyko, 2015).

5-methylcytosine CpG loci can spontaneously undergo deamination rather than demethylation, thereby converting cytosine to thymine. Deamination of 5-methylcytosine results in a thymine-guanine nucleotide pairing, which if the base-excision repair (BER) machinery does not correct, results in incorrect incorporation of thymine in the place of cytosine during replication (Duncan and Miller, 1980; Bellacosa and Drohat, 2015). Considering that 4% of the human genome can constitute 5-methylcytosine CpG sites and 80% of these are assumed to be methylated, there is a large potential for spontaneous deamination and C>T mutation incorporation at these loci. Therefore, CpG sites are well known to be mutational hotspots (Steinberg and Gorman, 1992; Denissenko *et al.*, 1997). Since these mutations are incorporated permanently during replication if not repaired, it can be hypothesised that cells undergoing high rates of proliferation such as stem cells during development, are more likely to harbour C>T mutations as a result. If this process were to occur during spermatogenesis or oogenesis, such mutations would be passed on to offspring and become *de novo* in the genome of the developing zygote, although they would be heterozygous. Applying this hypothesis to cases of KAT6A GDD syndrome, C>T mutations could arise in this manner at 5-methylcytosine CpG sites within the KAT6A gene, generating *de novo* PTCs if occurring at a particular arginine codon (CGA>TGA) to create nonsense mutations.

Using the iMETHYL database (Komaki *et al.*, 2018), we identified KAT6A GDD-associated codons that arise from C>T nonsense mutations in the KAT6A gene, as CpG sites of 5-cytosine methylation. Of the patients documented in the literature, 9/26 harbour C>T variants at CpG sites in this gene, all resulting in transitions from arginine to nonsense mutations (Figure 61). Arg79, Arg1024 and Arg1129 are all CpG sites that have been shown to have 5mC marks (Figure 62) (Komaki *et al.*, 2018). Similarly, all three codons are associated with C>T mutations in individuals with KAT6A GDD syndrome, producing PTCs. Interestingly, the frequency of each aforementioned locus bearing 5mC marks in populations of monocytes, neutrophils or $CD4^+$ T-cells, was reported to

be >94% (Figure 62). Taking all the KAT6A GDD-associated nonsense variants and excluding those which are not a single nucleotide substitution, 5/10 are at position Arg1024 and 3/10 are at Arg1129 (C>T). As such, 5-methylcytosine deamination may explain why pathogenic variants have been identified with a higher frequency at these two loci than other non-CpG sites. However, 5mC deamination has not previously been suggested as a mechanism behind their *de novo* occurrence in individuals with KAT6A GDD.

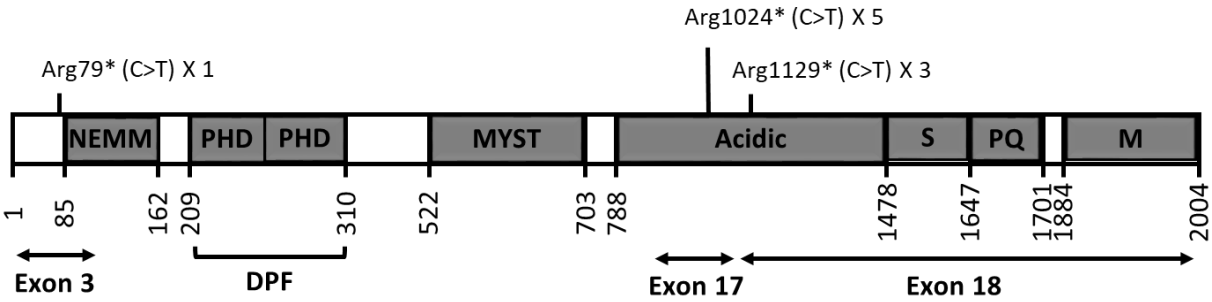


Figure 61: Nonsense KAT6A variants associated with KAT6A GDD are located at CpG sites

Location of KAT6A variants associated with KAT6A GDD syndrome which are nonsense mutations, relative to MOZ protein amino acid number and domains. DNA mutation shown in brackets and number of individuals with KAT6A GDD who harbour each variant are also indicated. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2. MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. Acidic residue-rich (Acidic), serine-rich (S), glutamic acid-rich (E), proline and glutamine-rich (PQ) and methionine-rich (M) regions are also indicated.

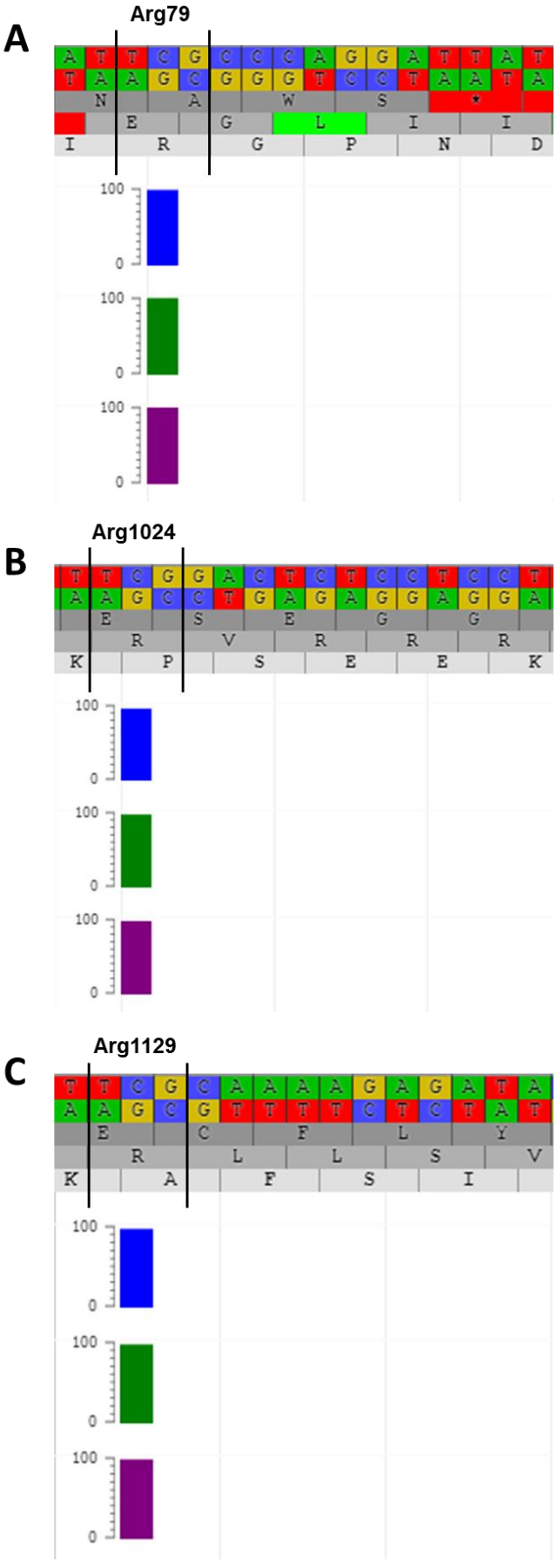

A
Arg79
CD4+ T cells – 97.0%
Monocytes - 98.5%
Neutrophils – 98.2%
B
Arg1024
CD4+ T cells – 94.3%
Monocytes - 95.6%
Neutrophils – 96.2%
C
Arg1129
CD4+ T cells – 95.3%
Monocytes - 95.5%
Neutrophils – 96.5%

Figure 62: Abundance of 5-methylcytosine at KAT6A GDD-associated CpG sites in the normal human population

Detected abundance of 5-methylcytosine at CpG sites in normal human $CD4^+$ T cells, monocytes and neutrophils. This data is taken from the iMETHYL database. Specifically shown are three of the corresponding CpG sites reported as C>T transitions which result in the formation of PTCs associated with KAT6A GDD, **(A)** Arg79, **(B)** Arg1024 and **(C)** Arg1129. Blue indicates 5mC abundance in CD4+ T cells, green is monocytes and purple is neutrophils on a scale from 0 – 100% in the normal human dataset. Note that all have values of 94% or higher and values are rounded to 1 dp. The nucleotide sequence encoding MOZ is on the reverse strand and the corresponding amino acid sequence is in the first (Arg79), or second reading frame (Arg1024 & Arg1129).

Ten-eleven translocation (TET) enzymes are methylcytosine dioxygenases, which function to demethylate methylcytosine species present in DNA (He *et al.*, 2011; Ito *et al.*, 2011). The activity of TET proteins converts 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) via a three-step reaction, resulting in the controlled removal of methylation from genomic loci such as CpG sites (An, Rao and Ko, 2017). TET-mediated oxidation tends to occur preferentially in regions with higher chromatin accessibility, suggesting it is active in regions of the genome that are being actively transcribed (An, Rao and Ko, 2017). In mouse ESCs, approximately 10% of 5mC marks are converted to 5hmC by TET enzymes, with only 10% of 5hmC marks oxidized further to 5fC/5caC (Ito *et al.*, 2011; Pfaffeneder *et al.*, 2011). This is due to a reduced processivity of TET enzymes for the further oxidised substrates. Indeed, TET proteins are less active on 5hmC and 5fC than on 5mC *in vitro*, suggestive of a substrate preference (Ito *et al.*, 2011; Hu *et al.*, 2015). Once CpG site methylcytosine species have been successfully oxidised to 5fC or 5caC, they are efficiently excised by thymine DNA glycosylase (TDG), which inserts an unmodified cytosine in its place (He *et al.*, 2011). It can be hypothesised that if TET enzyme activity was reduced (by mutation, or some other reason), this would reduce the conversion of 5mC to 5hmC and their derivatives. With an increased presence of 5mC, there may be a higher occurrence of spontaneous 5mC deamination

(resulting in T:G mismatches), and therefore an increased chance of thymine incorporation and C>T nucleotide substitutions if not corrected by the BER pathway prior to DNA replication. This can result in codon alterations at the affected locus and give rise to nonsense mutations. Should such events occur within germ cells, any mutations would be passed on to offspring and would appear to be *de novo* in nature, as the mutations would not be readily detectable from most tissues in the parents. Alternatively, if a mutation were to arise by the same principle shortly after fertilisation in embryonic stem cells, it would be *de novo* and present in every tissue of the developing embryo. While these hypotheses have not yet been investigated in any detail, it may be worthwhile considering the influence of TET activity upon the origin of nonsense mutations in cases of KAT6A GDD, when C>T nucleotide substitutions occur in the KAT6A gene.

Over the past few years, there has been a growing range of medical conditions which have been shown to have genetic causes associated with the presence of PTCs within coding regions of the genome (Hoischen *et al.*, 2011; Arboleda *et al.*, 2015; Shashi *et al.*, 2016; Li *et al.*, 2017). For many of these, it has been revealed that NMD has a role to play in their pathogenesis. Developments in so called PTC “readthrough”, therapies have been reported, with their successful application in some clinical trials (Kerem *et al.*, 2014; Roy *et al.*, 2015; Baradaran-Heravi *et al.*, 2017; Richardson *et al.*, 2017). Essentially, these PTC readthrough therapies are drug treatments (in many cases, low doses of aminoglycosides), that facilitate the ability of a ribosome to incorporate a random amino acid at the PTC codon during transcription, via disrupted base pairing and codon recognition at these sites. By preventing the early termination of translation and instead allowing ribosomes to proceed to the natural terminator codon, an almost wild-type protein can be synthesised. In some cases, the proteins translated in this manner have been demonstrated to be functional and alleviate clinical symptoms in disease models (Finkel *et al.*, 2013; Shulman *et al.*, 2014). It would therefore be of great benefit to determine whether such

therapeutic approaches would be successful in cases of the KAT6A GDD syndrome caused by nonsense variants. While it is hoped such treatments may have a positive impact upon restoring full length KAT6A protein translation, a concern is at which stage of life such a treatment would need to be delivered in order to be effective. It may be that treatment would need to be delivered during embryonic development, which raises many ethical and practical questions for future consideration.

CHAPTER VI

CONCLUDING REMARKS AND FUTURE DIRECTIONS

Chapter VI - Concluding remarks and future directions

Throughout the course of this research project, there have been five main achievements, which are indicated by bold text in parentheses in the following paragraph.

In our study, we have offered a deeper insight into the molecular mechanisms responsible for KAT6A nonsense variant pathogenicity (Figure 63). We did this through a combination of CRISPR-Cas9n-mediated generation of PTCs, reflecting similar mutations in the KAT6A gene which are reported in patients with KAT6A GDD, and the subsequent analysis of mRNA and protein expression levels of these variants **(ACHIEVEMENT 1)**. Furthermore, we have characterised the biological roles of the MOZ C-terminus in both maintenance of viable, metabolically active cell populations and in cell cycle progression **(ACHIEVEMENT 2)**. Transcriptome-wide changes were also assessed in RNA-seq data from two KAT6A GDD patients, revealing MOZ target genes that are dependent upon both MOZ expression and this HAT's C-terminal domains for their transcription **(ACHIEVEMENT 3)**. Our results have clearly indicated the high potential for NMD activity upon KAT6A mRNA transcripts which harbour PTCs within exon 17. Conversely, we have shown that this degradation mechanism does not act upon exon 18 PTC variants **(ACHIEVEMENT 4)**. We also determined that MOZ protein truncated within the 5' region of the acidic C-terminal domain is unaffected in its cellular localisation, being exclusively and consistently localised to the nucleus. This confirms that MOZ protein mislocalisation is not responsible for the development of KAT6A GDD in individuals who harbour PTCs in KAT6A exons 17 or 18 **(ACHIEVEMENT 5)**.

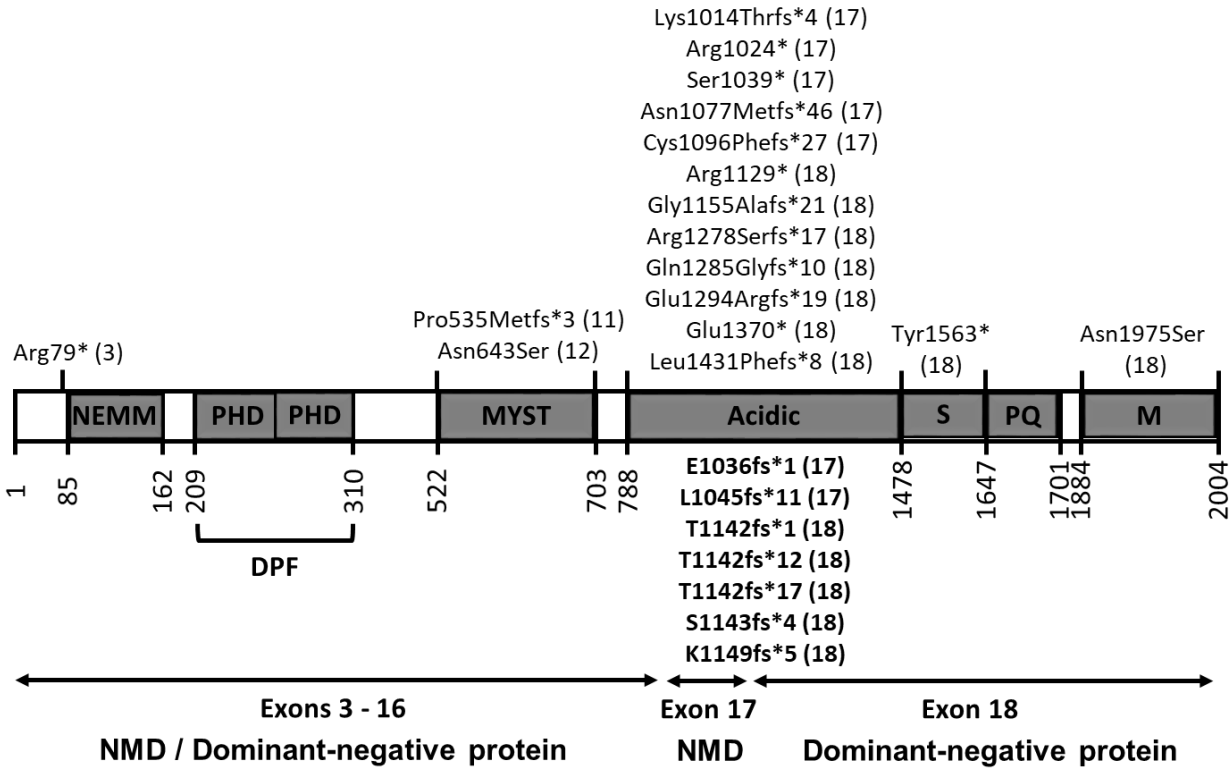


Figure 63: Different mechanisms of pathogenicity predicted for KAT6A GDD-associated PTC variants

Schematic showing our hypothesis of the molecular mechanisms which contribute to the pathogenicity of KAT6A GDD-associated PTC variants reported in the literature (above), and our CRISPR-Cas9-generated variants in K562 cells (below), relative to the linear MOZ protein structure. Different mechanisms are predicted to be responsible, depending on the exon location of the KAT6A variant. NEMM = N-terminal region of ENOK, MOZ or MORF domain which includes a linker H1- and H5-like nuclear localisation sequence. DPF = Double PHD finger domain comprised of Plant Homeodomains 1 and 2. MYST = A domain named after MOZ, Ybf2, Sas2, and Tip60. Acidic residue-rich (Acidic), serine-rich (S), glutamic acid-rich (E), proline and glutamine-rich (PQ) and methionine-rich (M) regions are also indicated.

We propose that these two differential effects of PTCs in KAT6A exons 17 and 18 upon their respective mRNA prevalence, differentially influences MOZ target gene transcription. Indeed, our analysis of RNA-seq data from patients with the KAT6A GDD syndrome is in agreement with this, although there are many genes similarly affected by both KAT6A exon 17 and 18 variants. Differentially expressed genes shared in both individuals included many HOX and other transcription factors, with prominent gene ontology terms such as transcription factor activity, regulation of cell proliferation, regulation of cell differentiation, growth factor activity and tissue development identified. Consistent with this, we determined a significant reduction in the number of metabolically active, viable cells in each of our MOZ C-terminally truncated K562 cell lines, regardless of whether the PTCs were in exon 17 or 18. In addition, cell cycle analysis of our mutant cell lines revealed a trend towards a decreased proportion of the KAT6A exon 18 PTC-harbouring cell population in phase G_2, however this was not observed for cells with PTCs in exon 17. There was a clear overlap of the differentially expressed genes we identified here in the patient RNA-seq data and our CRISPR-Cas9n-generated cells, as well as those reported in previous studies of MOZ expression, thereby validating our analyses. Our proposed model of the molecular pathogenesis of these nonsense variants is shown in Figure 64A & Figure 64B.

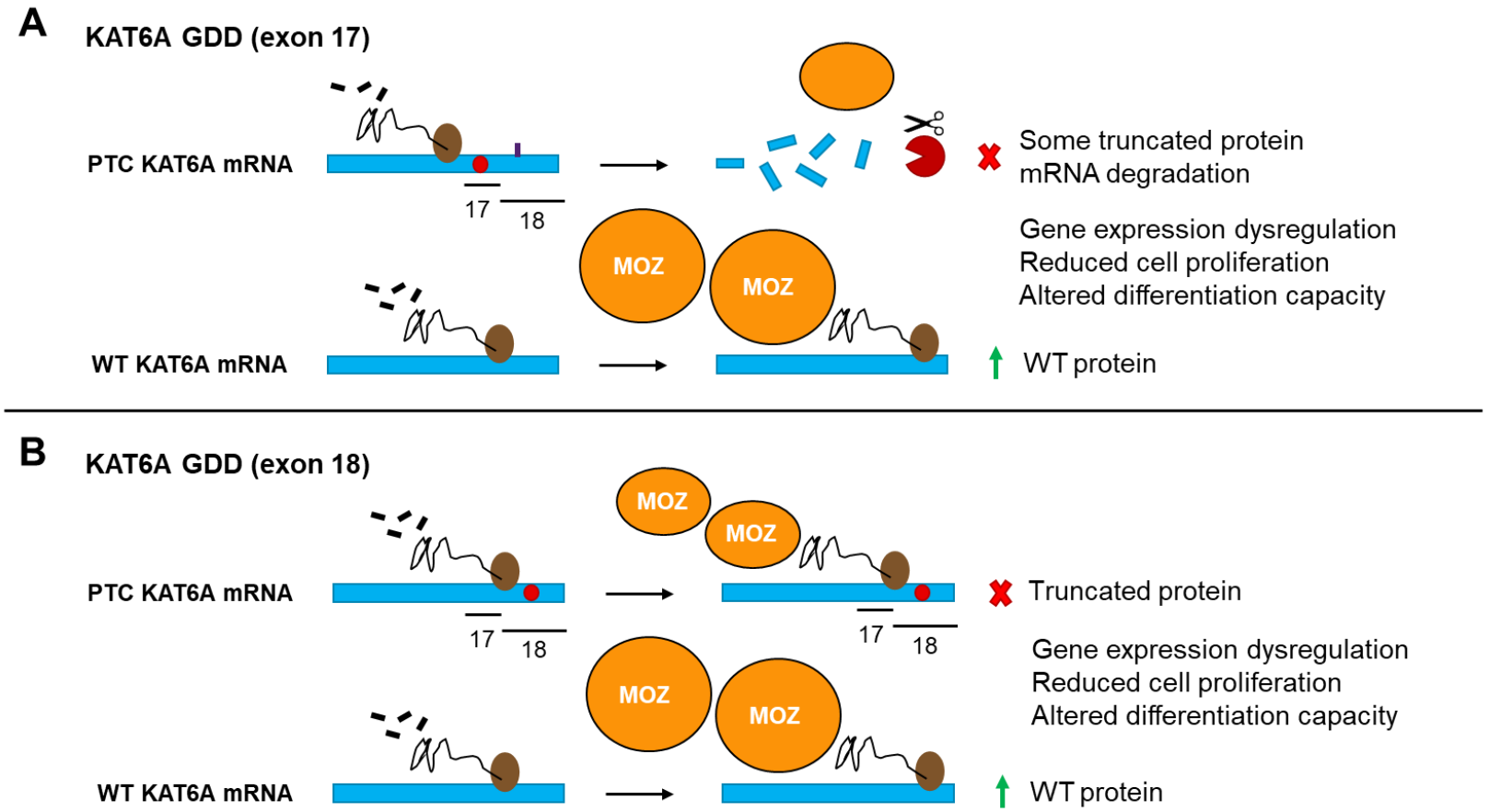


Figure 64: Proposed model of molecular pathogenesis for KAT6A GDD-associated PTC variants

Our model of how nonsense/frameshift mutations that produce in-frame PTCs in **(A)** KAT6A exon 17, or **(B)** KAT6A exon 18 result in the presentation of the KAT6A GDD syndrome. Blue rectangles = mRNA transcript, brown ovals = ribosomes, black rectangles = exon junction complexes, red circles = PTCs in the indicated exon, orange circles are MOZ polypeptides.

Despite our progress and novel findings surrounding MOZ function in the regulation of gene expression, cell cycle progression, cellular localisation and maintenance of viable, metabolically active cell populations, as well as its contribution in the KAT6A GDD syndrome, much work remains in order to fully characterise this HAT and its function in development. For example, we do not know whether C-terminally truncated MOZ protein can fold into its correct tertiary structure. If not, then this may help to explain why the expression of C-terminally truncated MOZ protein cannot regulate the transcription of its target genes, such as SPARC, RUNX3, HOXB3 and HHEX, to the same extent as WT MOZ in our KAT6A mutant cell lines. Furthermore, it should be investigated to what extent the failure of truncated MOZ protein to upregulate the expression of RUNX1, is directly responsible for the failed normal transcription of other affected genes we have identified. MOZ is known to regulate the expression of RUNX1, however how this consequentially controls downstream genes has not previously been thoroughly investigated. Experiments such as overexpression of RUNX1 in our KAT6A mutant cell lines, or knockdown in WT K562 cells, would allow us to determine target genes downstream of RUNX1 that are regulated by MOZ-mediated expression of this transcription factor, thereby identifying novel regulatory genetic pathways.

To assess the role of the MOZ C-terminus *in vivo* during development, mouse models harbouring mutant KAT6A alleles would be highly useful resources. With the recent advances in CRISPR technology allowing scientists to generate targeted genetic mutations in embryonic stem cells, it is now feasible to create mouse models harbouring specific desired allelic variants, such as those present in human patients with KAT6A GDD. Not only would these allow the assessment of the phenotypic and behavioural requirements, but they could also help to validate the findings we have presented here with regards to the molecular mechanisms behind the presentation of the KAT6A GDD syndrome. Additionally, aberrant developmental characteristics of this syndrome could be studied in more detail than what has been currently

reported in the literature; potential therapies such as the PTC readthrough treatments previously mentioned within this thesis, may also be demonstrated to be effective in the case of KAT6A GDD when utilising such models.

A major question regarding exactly how these nonsense or frameshift mutations arise to cause KAT6A PTCs and the subsequent development of KAT6A GDD, still remains unanswered. We have suggested that the position of the mutations within the KAT6A locus, are associated with a high level of 5-methylcytosine at CpG sites. Such sites are noted to be mutational hotspots due to spontaneous cytosine deamination and conversion to thymine, which is corrected by base-excision repair machinery. However, whether the efficiency of this repair mechanism is lower in individuals with KAT6A GDD, or indeed their parents (and therefore their germ cells) is unknown. If this were the case, it may explain how cases of KAT6A GDD arise and why several unrelated individuals possess nonsense mutations at the same genomic loci. Understanding how this syndrome arises will be key to the provision of family planning advice and genetic counselling for affected individuals and their families.

In our study, we utilised the K562 cell line in order to investigate the molecular requirement for the MOZ C-terminal region, however this is a cell line isolated from cancerous tissue and therefore has limitations with regards to studying processes like differentiation. For this reason, it would be desirable to investigate how truncated MOZ proteins affect proliferation, stem cell self-renewal capability and differentiation in a more biologically relevant cell line, such as mouse embryonic fibroblasts (MEFs), or neural stem cells that can form neurospheres. Genome editing to replicate the PTCs in KAT6A alleles present in individuals with KAT6A GDD, similar to what we performed here, would allow researchers to assess the potential for these stem cells to differentiate into their more specialised cell types and reveal whether MOZ has a role in lineage specification, something that may contribute to KAT6A GDD pathogenesis. In particular, experiments using neural stem cells may

reveal findings which help explain the intellectual disabilities and global developmental delay observed in these patients. Further to this, there is a need to investigate the potential for MOZ regulation of quiescence; this HAT is associated with having both positive and negative influences upon senescence, however to date, there has not been any research upon whether or not MOZ influences quiescence.

While we have implicated the RNA degradation mechanism NMD in the syndrome KAT6A GDD through our use of qRT-PCR and compounds which inhibit NMD, we have not conclusively validated this finding. The use of siRNA to knockdown a component of the NMD pathway, such as UPF1, then assessing the relative levels of spliced KAT6A mRNA harbouring PTCs in exon 17, would add weight to our hypothesis of the role of NMD in this syndrome. Applying this technique to MEFs isolated from KAT6A C-terminal mutant mice may possibly validate our findings here and demonstrate this *in vivo*. Furthermore, this would allow us to validate the MOZ target genes we have identified, or indeed identify additional target genes in a range of tissues, perhaps using techniques such as RNA-seq.

In summary, we have concluded that the KAT6A GDD syndrome is caused not only by C-terminally truncated MOZ protein, but also NMD of KAT6A mRNA transcripts, depending upon the exon location of the nonsense variant within the KAT6A gene. This knowledge raises the potential for PTC readthrough therapies, either alone or in conjunction with nonsense therapies, which may help to alleviate some clinical symptoms of this syndrome and improve the quality of life of affected individuals. Indeed, there are an increasing number of clinical disorders caused by PTCs within coding genomic loci, for which these therapies are being shown to benefit (Linde and Kerem, 2008; Keeling and Bedwell, 2011; Nagel-Wolfrum *et al.*, 2014; Matalonga *et al.*, 2015; Moosajee *et al.*, 2015). As such, these therapies should be tested in the context of KAT6A exon 17 and 18 PTC variants, utilising our CRISPR-Cas9n-generated cell lines or patient-derived tissues to determine their potential in rescuing the effects we have reported. Any success in these respects

would be of great importance and interest to not only those who present with the KAT6A GDD syndrome, but also their carers, family members, clinicians involved in their healthcare needs and the wider scientific community.

BIBLIOGRAPHY

Bibliography

Allfrey, V. G., Faulkner, R. and Mirsky, A. E. (1964) '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*. National Academy of Sciences, 51(7–8), pp. 786–94. doi: 10.1111/jocn.13120.

Amberg, G. C. *et al.* (2004) 'NFATc3 regulates Kv2.1 expression in arterial smooth muscle', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 279(45), pp. 47326–47334. doi: 10.1074/jbc.M408789200.

An, J., Rao, A. and Ko, M. (2017) 'TET family dioxygenases and DNA demethylation in stem cells and cancers', *Experimental & Molecular Medicine*. Nature Publishing Group, 49(4), pp. e323–e323. doi: 10.1038/emm.2017.5.

Andarawewa, K. L. *et al.* (2003) 'Dual stromelysin-3 function during natural mouse mammary tumor virus-ras tumor progression', *Cancer Research*, 63(18), pp. 5844–5849. doi: 10.1038/356397a0.

Arboleda, V. A. *et al.* (2015) 'De novo nonsense mutations in KAT6A, a lysine acetyl-transferase gene, cause a syndrome including microcephaly and global developmental delay', *Am J Hum Genet*, 96(3), pp. 498–506. doi: 10.1016/j.ajhg.2015.01.017.

Ariki, R. *et al.* (2014) 'Homeodomain transcription factor Meis1 is a critical regulator of adult bone marrow hematopoiesis.', *PloS one*, 9(2), p. e87646. doi: 10.1371/journal.pone.0087646.

Bannister, A. J. and Kouzarides, T. (1996) 'The CBP co-activator is a histone acetyltransferase.', *Nature*, 384(6610), pp. 641–3. doi: 10.1038/384641a0.

Bannister, A. J. and Kouzarides, T. (2005) 'Reversing histone methylation', *Nature*, 436(7054), pp. 1103–1106. doi: 10.1038/nature04048.

Baradaran-Heravi, A. *et al.* (2017) 'Gentamicin B1 is a minor gentamicin component with major nonsense mutation suppression activity',

Proceedings of the National Academy of Sciences, 114(13), pp. 3479–3484. doi: 10.1073/pnas.1620982114.

Barrangou, R. *et al.* (2007) 'CRISPR provides acquired resistance against viruses in prokaryotes.', *Science*, 315(5819), pp. 1709–1712. doi: 10.1126/science.1138140.

Basehoar, A. D., Zanton, S. J. and Pugh, B. F. (2004) 'Identification and distinct regulation of yeast TATA box-containing genes', *Cell*. Cell Press, 116(5), pp. 699–709. doi: 10.1016/S0092-8674(04)00205-3.

Bashir, R. A. *et al.* (2017) 'Lin-Gettig syndrome: Craniosynostosis expands the spectrum of the *KAT6B* related disorders', *American Journal of Medical Genetics Part A*, (January), pp. 1–9. doi: 10.1002/ajmg.a.38355.

Behrens, K. *et al.* (2017) 'RUNX1 cooperates with FLT3-ITD to induce leukemia', *The Journal of experimental medicine*, 214(3), pp. 737–752. doi: 10.1084/jem.20160927.

Bellacosa, A. and Drohat, A. C. (2015) 'Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites', *DNA Repair*. NIH Public Access, 32, pp. 33–42. doi: 10.1016/j.dnarep.2015.04.011.

Belsham, G. J. and Sonenberg, N. (1996) 'RNA-protein interactions in regulation of picornavirus RNA translation.', *Microbiological reviews*, 60(3), pp. 499–511

Bernstein, B. E. *et al.* (2006) 'A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells', *Cell*. Elsevier, 125(2), pp. 315–326. doi: 10.1016/j.cell.2006.02.041.

Bhattacharya, A. (2013) *Identification of MOZ/MYST3 binding sites using ChIP-Seq*. The University of Nottingham

Bird, A. (2002) 'DNA methylation patterns and epigenetic memory.', *Genes & development*. Cold Spring Harbor Laboratory Press, 16(1), pp. 6–21. doi: 10.1101/gad.947102.

Borrow, J. *et al.* (1996) 'The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-

binding protein.', *Nature genetics*, 14(1), pp. 33–41. doi: 10.1038/ng0996-33.

Bradshaw, A. D. (2009) 'The role of SPARC in extracellular matrix assembly', *Journal of Cell Communication and Signaling*. Springer, pp. 239–246. doi: 10.1007/s12079-009-0062-6.

Breiling, A. and Lyko, F. (2015) 'Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond', *Epigenetics and Chromatin*. BioMed Central, p. 24. doi: 10.1186/s13072-015-0016-6.

Brinkmann, A. O. *et al.* (1989) 'The human androgen receptor: domain structure, genomic organization and regulation of expression.', *Journal of Steroid Biochemistry*, 34(1–6), pp. 307–310. doi: 10.1016/0022-4731(89)90098-8.

Bristow, C. A. P. and Shore, P. (2003) 'Transcriptional regulation of the human MIP-1α promoter by RUNX1 and MOZ', *Nucleic Acids Research*, 31(11), pp. 2735–2744. doi: 10.1093/nar/gkg401.

Brownell, J. E. *et al.* (1996) 'Tetrahymena Histone Acetyltransferase A: A Homolog to Yeast Gcn5p Linking Histone Acetylation to Gene Activation', *Cell*. Elsevier, 84(6), pp. 843–851. doi: 10.1016/S0092-8674(00)81063-6.

Camós, M. *et al.* (2006) '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 Research*, 66(14), pp. 6947–6954. doi: 10.1158/0008-5472.CAN-05-4601.

Campeau, P. M. *et al.* (2012) 'Mutations in KAT6B, Encoding a Histone Acetyltransferase, Cause Genitopatellar Syndrome', *The American Journal of Human Genetics*, 90(2), pp. 282–289. doi: 10.1016/j.ajhg.2011.11.023.

Carrasco, L., Jimenez, A. and Vazquez, D. (1976) 'Specific Inhibition of Translocation by Tubulosine in Eukaryotic Polysomes', *European Journal of Biochemistry*, 64(1), pp. 1–5. doi: 10.1111/j.1432-1033.1976.tb10268.x.

Cassandri, M. *et al.* (2017) 'Zinc-finger proteins in health and disease', *Cell Death Discovery*. Nature Publishing Group, 3, p. 17071. doi: 10.1038/cddiscovery.2017.71.

Celeste, A. *et al.* (2002) 'Genomic instability in mice lacking histone H2AX', *Science*. American Association for the Advancement of Science, 296(5569), pp. 922–927. doi: 10.1126/science.1069398.

Chaffanet, M. *et al.* (2000) 'MOZ is fused top300 in an acute monocytic leukemia with t(8;22)', *Genes, Chromosomes and Cancer*, 28(2), pp. 138–144. doi: 10.1002/(SICI)1098-2264(200006)28:2<138::AID-GCC2>3.0.CO;2-2.

Champagne, N., Pelletier, N. and Yang, X. J. (2001) 'The monocytic leukemia zinc finger protein MOZ is a histone acetyltransferase.', *Oncogene*, 20(3), pp. 404–9. doi: 10.1038/sj.onc.1204114.

Chan, C. *et al.* (2004) 'eIF4A3 is a novel component of the exon junction complex', *RNA*, 10(2), pp. 200–209. doi: 10.1261/rna.5230104.

Chan, E. M. *et al.* (2007) 'MOZ and MOZ-CBP cooperate with NF-κB to activate transcription from NF-κB-dependent promoters', *Experimental Hematology*, 35(12), pp. 1782–1792. doi: 10.1016/j.exphem.2007.07.015.

Chan, H. M. and La Thangue, N. B. (2001) 'p300/CBP proteins: HATs for transcriptional bridges and scaffolds.', *Journal of cell science*, 114(Pt 13), pp. 2363–73

Chang, Y.-F., Imam, J. S. and Wilkinson, M. F. (2007) 'The Nonsense-Mediated Decay RNA Surveillance Pathway', *Annual Review of Biochemistry*, 76(1), pp. 51–74. doi: 10.1146/annurev.biochem.76.050106.093909.

Chen, F. *et al.* (2011) 'High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases', *Nature Methods*. NIH Public Access, 8(9), pp. 753–757. doi: 10.1038/nmeth.1653.

Chen, S. *et al.* (2006) 'Androgen receptor phosphorylation and stabilization in prostate cancer by cyclin-dependent kinase 1',

Proceedings of the National Academy of Sciences. National Academy of Sciences, 103(43), pp. 15969–15974. doi: 10.1073/pnas.0604193103.

Cheung, N. *et al.* (2016) 'Supplemental Information Targeting Aberrant Epigenetic Networks Mediated by PRMT1 and KDM4C in Acute Myeloid Leukemia', *Cancer Cell*. Elsevier Ltd, 29(1), pp. 32–48. doi: 10.1016/j.ccell.2015.12.007.

Cheung, P. and Lau, P. (2005) 'Epigenetic regulation by histone methylation and histone variants.', *Molecular endocrinology (Baltimore, Md.)*, 19(3), pp. 563–73. doi: 10.1210/me.2004-0496.

Chiang, T. W. *et al.* (2016) 'CRISPR-Cas9(D10A) nickase-based genotypic and phenotypic screening to enhance genome editing', *Sci Rep*. Nature Publishing Group, 6(January), p. 24356. doi: 10.1038/srep24356.

Chilcote, R. R. *et al.* (1987) 'Association of red cell spherocytosis with deletion of the short arm of chromosome 8.', *Blood*, 69(1), pp. 156–159

Chinen, Y. *et al.* (2014) '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*, 53(4), pp. 299–308. doi: 10.1002/gcc.22140.

Cho, H. I., Kim, M. S. and Jang, Y. K. (2016) 'The BRPF2/BRD1-MOZ complex is involved in retinoic acid-induced differentiation of embryonic stem cells', *Experimental Cell Research*. Elsevier, 346(1), pp. 30–39. doi: 10.1016/j.yexcr.2016.05.022.

Chuang, L. S. H., Ito, K. and Ito, Y. (2013) 'RUNX family: Regulation and diversification of roles through interacting proteins', *International Journal of Cancer*, 132(6), pp. 1260–1271. doi: 10.1002/ijc.27964.

Clemente, M. (2016) *Targeting BCL11A/nuclear receptor complexes in breast tumour cells using CRISPR-Cas9 genome editing tool*. The

University of Nottingham

Coenen, E. A. *et al.* (2013) 'Pediatric acute myeloid leukemia with t(8;16)(p11;p13), a distinct clinical and biological entity: a collaborative study by the International-Berlin-Frankfurt-Munster AML-study group', *Blood*, 122(15), pp. 2704–2713. doi: 10.1182/blood-2013-02-485524.

Cohen, H. *et al.* (1991) 'Congenital spherocytosis, B19 parvovirus infection and inherited interstitial deletion of the short arm of chromosome 8', *British Journal of Haematology*, 78(2), pp. 251–257. doi: 10.1111/j.1365-2141.1991.tb04425.x.

Colak, D. *et al.* (2013) 'Regulation of Axon Guidance by Compartmentalized Nonsense-Mediated mRNA Decay', *Cell*, 153(6), pp. 1252–1265. doi: 10.1016/j.cell.2013.04.056.

Collins, H. M. *et al.* (2006) 'MOZ-TIF2 alters cofactor recruitment and histone modification at the RARβ2 promoter: Differential effects of MOZ fusion proteins on CBP- and MOZ-dependent activators', *Journal of Biological Chemistry*, 281(25), pp. 17124–17133. doi: 10.1074/jbc.M602633200.

Colombo, M. *et al.* (2017) 'Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways', *RNA*. Cold Spring Harbor Laboratory Press, 23(2), pp. 189–201. doi: 10.1261/rna.059055.116.

Cong, L. *et al.* (2013) 'Multiplex genome engineering using CRISPR/Cas systems', *Science*, 339(6121), pp. 819–823. doi: 10.1126/science.1231143.

Consortium, T. E. P. (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature*. Nature Publishing Group, 489(7414), pp. 57–74. doi: 10.1038/nature11247.

Corpet, F. (1988) 'Multiple sequence alignment with hierarchical clustering.', *Nucleic acids research*. Oxford University Press, 16(22), pp. 10881–90

Crisponi, L. *et al.* (2001) 'The putative forkhead transcription factor

FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome', *Nature Genetics*, 27(2), pp. 159–166. doi: 10.1038/84781.

Crump, J. G. *et al.* (2006) 'Moz-dependent Hox expression controls segment-specific fate maps of skeletal precursors in the face', *Development*, 133(14), pp. 2661–2669. doi: dev.02435 [pii]\r10.1242/dev.02435.

Cutress, M. L. *et al.* (2008) 'Structural basis for the nuclear import of the human androgen receptor', *J Cell Sci*, 121(Pt 7), pp. 957–968. doi: 10.1242/jcs.022103.

Dalvai, M. *et al.* (2015) 'A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells', *Cell Reports*, 13(3), pp. 621–633. doi: 10.1016/j.celrep.2015.09.009.

Datta, D. V., Singh, S. A. K. and Chhuttani, P. N. (2017) 'Treatment of Amebic Liver Abscess with Emetine Hydrochloride, Niridazole, and Metronidazole', *The American Journal of Tropical Medicine and Hygiene*. The American Society of Tropical Medicine and Hygiene, 23(4), pp. 586–589. doi: 10.4269/ajtmh.1974.23.586.

Deguchi, K. *et al.* (2003) 'MOZ-TIF2-induced acute myeloid leukemia requires the MOZ nucleosome binding motif and TIF2-mediated recruitment of CBP', *Cancer Cell*, 3(3), pp. 259–271. doi: 10.1016/S1535-6108(03)00051-5.

Denissenko, M. F. *et al.* (1997) 'Cytosine methylation determines hot spots of DNA damage in the human P53 gene.', *Proceedings of the National Academy of Sciences of the United States of America*, 94(8), pp. 3893–8. doi: 10.1073/pnas.94.8.3893.

Dennis, S. *et al.* (1999) 'A secreted frizzled related protein, FrzA, selectively associates with Wnt-1 protein and regulates wnt-1 signaling.', *Journal of cell science*, 112 (Pt 2, pp. 3815–20

Devanna, P., Middelbeek, J. and Vernes, S. C. (2014) 'FOXP2 drives neuronal differentiation by interacting with retinoic acid signaling pathways', *Frontiers in Cellular Neuroscience*. Frontiers Media SA, 8, p. 305. doi: 10.3389/fncel.2014.00305.

Drazic, A. *et al.* (2016) ‘The world of protein acetylation’, *Biochimica et Biophysica Acta - Proteins and Proteomics*. The Authors, 1864(10), pp. 1372–1401. doi: 10.1016/j.bbapap.2016.06.007.

Dreveny, I. *et al.* (2014) ‘The double PHD finger domain of MOZ/MYST3 induces a-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification’, *Nucleic Acids Research*, 42(2), pp. 822–835. doi: 10.1093/nar/gkt931.

Dreyfuss, G., Kim, V. N. and Kataoka, N. (2002) ‘Messenger-RNA-binding proteins and the messages they carry’, *Nature Reviews Molecular Cell Biology*, pp. 195–205. doi: 10.1038/nrm760.

Du, J. *et al.* (2000) ‘Frequency-dependent regulation of rat hippocampal somato-dendritic excitability by the K+channel subunit Kv2.1’, *Journal of Physiology*, 522(1), pp. 19–31. doi: 10.1111/j.1469-7793.2000.t01-2-00019.xm.

Duncan, B. K. and Miller, J. H. (1980) ‘Mutagenic deamination of cytosine residues in DNA’, *Nature*. Nature Publishing Group, 287(5782), pp. 560–561. doi: 10.1038/287560a0.

Ehrlich, H. J. *et al.* (1990) ‘Alteration of serpin specificity by a protein cofactor. Vitronectin endows plasminogen activator inhibitor 1 with thrombin inhibitory properties.’, *The Journal of biological chemistry*, 265(22), pp. 13029–35. doi: 10.1016/j.agwat.2008.06.008.

Ehrlich, M. *et al.* (1982) ‘Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells’, *Nucleic Acids Research*, 10(8), pp. 2709–2721. doi: 10.1093/nar/10.8.2709.

Eising, E. *et al.* (2018) ‘A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development’, *Molecular Psychiatry*. Springer US, 2, pp. 1–14. doi: 10.1038/s41380-018-0020-x.

Eitoku, M. *et al.* (2008) ‘Histone chaperones: 30 Years from isolation to elucidation of the mechanisms of nucleosome assembly and disassembly’, *Cellular and Molecular Life Sciences*, pp. 414–444. doi: 10.1007/s00018-007-7305-6.

Elenius, V. *et al.* (2017) ‘Food allergy in a child with de novo KAT6A

mutation', *Clinical and Translational Allergy*. BioMed Central, 7(1), p. 19. doi: 10.1186/s13601-017-0155-x.

Ernst, P. *et al.* (2004) 'An Mll-dependent Hox program drives hematopoietic progenitor expansion', *Current Biology*, 14(22), pp. 2063–2069. doi: 10.1016/j.cub.2004.11.012.

Esteyries, S. *et al.* (2008) 'NCOA3, a new fusion partner for MOZ/MYST3 in M5 acute myeloid leukemia.', *Leukemia*, 22(3), pp. 663–665. doi: 10.1038/sj.leu.2404930.

Ezan, J. *et al.* (2004) 'FrzA/sFRP-1, a secreted antagonist of the Wnt-Frizzled pathway, controls vascular cell proliferation in vitro and in vivo', *Cardiovascular Research*. Oxford University Press, 63(4), pp. 731–738. doi: 10.1016/j.cardiores.2004.05.006.

Feller, C. *et al.* (2015) 'Global and specific responses of the histone acetylome to systematic perturbation', *Molecular Cell*. Elsevier, 57(3), pp. 559–572. doi: 10.1016/j.molcel.2014.12.008.

Feng, Y. *et al.* (2016) 'BRPF3-HBO1 regulates replication origin activation and histone H3K14 acetylation', *The EMBO Journal*, 35(2), pp. 176–192. doi: 10.15252/embj.201591293.

Filarsky, M. *et al.* (2015) 'The extended AT-hook is a novel RNA binding motif', *RNA Biology*, 12(8), pp. 864–876. doi: 10.1080/15476286.2015.1060394.

Finkel, R. S. *et al.* (2013) 'Phase 2a study of ataluren-mediated dystrophin production in patients with nonsense mutation Duchenne muscular dystrophy', *PLoS ONE*, 8(12). doi: 10.1371/journal.pone.0081302.

Fong, S. H. *et al.* (2005) 'Wnt signalling mediated by Tbx2b regulates cell migration during formation of the neural plate.', *Development*. The Company of Biologists Ltd, 132(16), pp. 3587–96. doi: 10.1242/dev.01933.

Frietze, S. and Farnham, P. J. (2011) 'Transcription Factor Effector Domains', in *Sub-Cellular Biochemistry*, pp. 261–277. doi: 10.1007/978-90-481-9069-0_12.

Fu, Y. *et al.* (2013) 'High-frequency off-target mutagenesis induced by

CRISPR-Cas nucleases in human cells.', *Nature biotechnology*. Nature Publishing Group, 31(9), pp. 822–6. doi: 10.1038/nbt.2623.

Fu, Y. *et al.* (2014) 'Improving CRISPR-Cas nuclease specificity using truncated guide RNAs', *Nature Biotechnology*. Nature Publishing Group, 32(3), pp. 279–284. doi: 10.1038/nbt.2808.

Gama-Carvalho, M. and Carmo-Fonseca, M. (2001) 'The rules and roles of nucleocytoplasmic shuttling proteins', *FEBS Letters*. No longer published by Elsevier, pp. 157–163. doi: 10.1016/S0014-5793(01)02487-5.

Gannon, T. *et al.* (2015) 'Further delineation of the KAT6B molecular and phenotypic spectrum', *European Journal of Human Genetics*, 23(9), pp. 1165–1170. doi: 10.1038/ejhg.2014.248.

Gao, Y. *et al.* (2014) 'Allele compensation in Tip60+/- mice rescues white adipose tissue function In Vivo', *PLoS ONE*, 9(5). doi: 10.1371/journal.pone.0098343.

Gasiunas, G. *et al.* (2012) 'Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria', *Proceedings of the National Academy of Sciences*, 109(39), pp. E2579–E2586. doi: 10.1073/pnas.1208507109.

Gauthier-Vasserot, A. *et al.* (2017) 'Application of whole-exome sequencing to unravel the molecular basis of undiagnosed syndromic congenital neutropenia with intellectual disability', *American Journal of Medical Genetics, Part A*, 173(1), pp. 62–71. doi: 10.1002/ajmg.a.37969.

Gauvrit, S. *et al.* (2018) 'HHEX is a transcriptional regulator of the VEGFC/FLT4/PROX1 signaling axis during vascular development.', *Nature communications*. Nature Publishing Group, 9(1), p. 2704. doi: 10.1038/s41467-018-05039-1.

Gervais, C. *et al.* (2008) 'Acute myeloid leukaemia with 8p11 (MYST3) rearrangement: an integrated cytologic, cytogenetic and molecular study by the groupe francophone de cytogénétique hématologique', *Leukemia*, 22(8), pp. 1567–1575. doi:

10.1038/leu.2008.128.

Ghosh, T. K. *et al.* (2018) 'Acetylation of TBX5 by KAT2B and KAT2A regulates heart and limb development', *Journal of Molecular and Cellular Cardiology*. Elsevier Ltd, 114, pp. 185–198. doi: 10.1016/j.yjmcc.2017.11.013.

Goodings, C. *et al.* (2015) 'Hhex is Required at Multiple Stages of Adult Hematopoietic Stem and Progenitor Cell Differentiation', *STEM CELLS*, 33(8), pp. 2628–2641. doi: 10.1002/stem.2049.

Gorgoni, B. and Gray, N. K. (2004) 'The roles of cytoplasmic poly(A)-binding proteins in regulating gene expression: a developmental perspective', *Briefings in functional genomics & proteomics*, 3(2), pp. 125–41

Grant, P. A. *et al.* (1999) 'Expanded lysine acetylation specificity of Gcn5 in native complexes', *Journal of Biological Chemistry*, 274(9), pp. 5895–5900. doi: 10.1074/jbc.274.9.5895.

Greer, E. L. and Shi, Y. (2012) 'Histone methylation: A dynamic mark in health, disease and inheritance', *Nature Reviews Genetics*, pp. 343–357. doi: 10.1038/nrg3173.

Growney, J. D. *et al.* (2014) 'Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype', *Blood*, 106(2), pp. 494–505. doi: 10.1182/blood-2004-08-3280.Supported.

Guccione, E. *et al.* (2007) 'Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive', *Nature*, 449(7164), pp. 933–937. doi: 10.1038/nature06166.

Gundry, M. C. *et al.* (2016) 'Highly Efficient Genome Editing of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9', *Cell Reports*. ElsevierCompany., 17(5), pp. 1453–1461. doi: 10.1016/j.celrep.2016.09.092.

Hannenhalli, S. and Kaestner, K. H. (2009) 'The evolution of Fox genes and their role in development and disease', *Nature Reviews Genetics*. NIH Public Access, pp. 233–240. doi: 10.1038/nrg2523.

Harrelson, Z. (2004) ‘Tbx2 is essential for patterning the atrioventricular canal and for morphogenesis of the outflow tract during heart development’, *Development*, 131(20), pp. 5041–5052. doi: 10.1242/dev.01378.

Harrison, M. M. *et al.* (2014) ‘A CRISPR view of development’, *Genes & Development*, 28(17), pp. 1859–1872. doi: 10.1101/gad.248252.114.

Hayflick, L. (1965) ‘The limited in vitro lifetime of human diploid cell strains’, *Experimental cell research*, 37, pp. 614–36

He, Y.-F. *et al.* (2011) ‘Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA’, *Science*, 333(6047), pp. 1303–1307. doi: 10.1126/science.1210944.

Heair, H. M. *et al.* (2015) ‘MicroRNA 665 Regulates Dentinogenesis through MicroRNA-Mediated Silencing and Epigenetic Mechanisms’, *Molecular and Cellular Biology*, 35(18), pp. 3116–3130. doi: 10.1128/MCB.00093-15.

Hedera, P. and Innis, J. W. (2002) ‘Possible third case of Lin-Gettig syndrome’, *American Journal of Medical Genetics*, 110(4), pp. 380–383. doi: 10.1002/ajmg.10460.

Heery, D. M. *et al.* (1997) ‘A signature motif in transcriptional co-activators mediates binding to nuclear receptors.’, *Nature*, 387(6634), pp. 733–736. doi: 10.1038/42750.

Hein, M. Y. *et al.* (2015) ‘A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances’, *Cell*, 163(3), pp. 712–723. doi: 10.1016/j.cell.2015.09.053.

Hendriks, I. A. *et al.* (2017) ‘Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation.’, *Nature structural & molecular biology*, 24(3), pp. 325–336. doi: 10.1038/nsmb.3366.

Hernandez-quiles, M. (2016) *PIP5K1A in prostate cancer cells using CRISPR-Cas9 genome editing tool*. The University of Nottingham

Hershko, A. *et al.* (1982) ‘Immunochemical analysis of the turnover of

ubiquitin-protein conjugates in intact cells. Relationship to the breakdown of abnormal proteins.', *The Journal of biological chemistry*, 257(23), pp. 13964–70

Heubach, J. *et al.* (2015) 'The long noncoding RNA HOTAIR has tissue and cell type-dependent effects on HOX gene expression and phenotype of urothelial cancer cells', *Molecular Cancer*. BioMed Central, 14(1), p. 108. doi: 10.1186/s12943-015-0371-8.

Le Hir, H. *et al.* (2000) 'The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions', *EMBO Journal*, 19(24), pp. 6860–6869. doi: 10.1093/emboj/19.24.6860.

Hoischen, A. *et al.* (2011) 'De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome', *Nature Genetics*. Nature Publishing Group, 43(8), pp. 729–731. doi: 10.1038/ng.868.

Holbert, M. A. *et al.* (2007) 'The human monocytic leukemia zinc finger histone acetyltransferase domain contains DNA-binding activity implicated in chromatin targeting', *Journal of Biological Chemistry*, 282(50), pp. 36603–36613. doi: 10.1074/jbc.M705812200.

Hornbeck, P. V. *et al.* (2015) 'PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations', *Nucleic Acids Research*, 43(D1), pp. D512–D520. doi: 10.1093/nar/gku1267.

Houben, A. *et al.* (2007) 'Phosphorylation of histone H3 in plants-A dynamic affair', *Biochimica et Biophysica Acta - Gene Structure and Expression*, pp. 308–315. doi: 10.1016/j.bbaexp.2007.01.002.

Howe, L. (2001) 'Histone H3 specific acetyltransferases are essential for cell cycle progression', *Genes & Development*, 15(23), pp. 3144–3154. doi: 10.1101/gad.931401.

Hsu, J.-Y. *et al.* (2017) 'Non-homeostatic body weight regulation through a brainstem-restricted receptor for GDF15', *Nature*, 550(7675), pp. 255–259. doi: 10.1038/nature24042.

Hu, L. *et al.* (2015) 'Structural insight into substrate preference for TET-mediated oxidation', *Nature*, 527(7576), pp. 118–122. doi:

10.1038/nature15713.

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009) 'Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.', *Nature protocols*, 4(1), pp. 44–57. doi: 10.1038/nprot.2008.211.

Huang, F., Abmayr, S. M. and Workman, J. L. (2016) 'Regulation of KAT6 Acetyltransferases and Their Roles in Cell Cycle Progression, Stem Cell Maintenance, and Human Disease', *Molecular and Cellular Biology*, 36(14), pp. 1900–1907. doi: 10.1128/MCB.00055-16.

Iijima, O. *et al.* (2006) 'A purine at +2 rather than +1 adjacent to the human U6 promoter is required to prepare effective short hairpin RNAs', *Biochemical and Biophysical Research Communications*, 350(4), pp. 809–817. doi: 10.1016/j.bbrc.2006.08.187.

Ikura, T. *et al.* (2000) 'Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis.', *Cell*, 102(4), pp. 463–473. doi: 10.1016/S0092-8674(00)00051-9.

Illingworth, R. S. and Bird, A. P. (2009) 'CpG islands - "A rough guide"', *FEBS Letters*, pp. 1713–1720. doi: 10.1016/j.febslet.2009.04.012.

Inoue, K., Shiga, T. and Ito, Y. (2008) 'Runx transcription factors in neuronal development', *Neural Development*, 3(1), p. 20. doi: 10.1186/1749-8104-3-20.

Ito, S. *et al.* (2011) 'Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine', *Science*, 333(6047), pp. 1300–1303. doi: 10.1126/science.1210597.

Jackson, B. C. *et al.* (2010) 'Update of human and mouse forkhead box (FOX) gene families.', *Human genomics*. BioMed Central, 4(5), pp. 345–52. doi: 10.1186/1479-7364-4-5-345.

Jackson, R. J., Hellen, C. U. T. and Pestova, T. V. (2010) 'The mechanism of eukaryotic translation initiation and principles of its regulation', *Nature Reviews Molecular Cell Biology*, 11(2), pp. 113–127. doi: 10.1038/nrm2838.

de Jager, S. C. A. *et al.* (2011) 'Growth differentiation factor 15 deficiency

protects against atherosclerosis by attenuating CCR2-mediated macrophage chemotaxis.', *The Journal of experimental medicine*, 208(2), pp. 217–25. doi: 10.1084/jem.20100370.

Jia, J. *et al.* (2017) 'Premature termination codon readthrough in human cells occurs in novel cytoplasmic foci and requires UPF proteins', *Journal of Cell Science*, 130(18), pp. 3009–3022. doi: 10.1242/jcs.198176.

Jiang, C. *et al.* (2017) 'Serpine 1 induces alveolar type II cell senescence through activating p53-p21-Rb pathway in fibrotic lung disease', *Aging Cell*. Wiley-Blackwell, 16(5), pp. 1114–1124. doi: 10.1111/acel.12643.

Jiang, H. Y. and Wek, R. C. (2005) 'Phosphorylation of the α-subunit of the eukaryotic initiation factor-2 (eIF2α) reduces protein synthesis and enhances apoptosis in response to proteasome inhibition', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 280(14), pp. 14189–14202. doi: 10.1074/jbc.M413660200.

Johansen, K. M. and Johansen, J. (2006) 'Regulation of chromatin structure by histone H3S10 phosphorylation', *Chromosome Research*, pp. 393–404. doi: 10.1007/s10577-006-1063-4.

Johansen Taber, K. A., Dickinson, B. D. and Wilson, M. (2014) 'The Promise and Challenges of Next-Generation Genome Sequencing for Clinical Care', *JAMA Internal Medicine*, 174(2), p. 275. doi: 10.1001/jamainternmed.2013.12048.

Joshi, B. *et al.* (1995) 'Phosphorylation of eukaryotic protein synthesis initiation factor 4E at Ser-209.', *The Journal of biological chemistry*, 270(24), pp. 14597–603

Ju, X. *et al.* (2014) 'Context-dependent activation of Wnt signaling by tumor suppressor RUNX3 in gastric cancer cells', *Cancer Science*, 105(4), pp. 418–424. doi: 10.1111/cas.12356.

Jufvas, Å., Strålfors, P. and Vener, A. V. (2011) 'Histone variants and their post-translational modifications in primary human fat cells', *PLoS ONE*. Edited by M. Federici. Public Library of Science,

6(1), p. e15960. doi: 10.1371/journal.pone.0015960.

Kamakaka, R. T. and Biggins, S. (2005) ‘Histone variants: Deviants?’, *Genes and Development*. Cold Spring Harbor Laboratory Press, pp. 295–310. doi: 10.1101/gad.1272805.

Kanda, A. and Swaroop, A. (2009) ‘A comprehensive analysis of sequence variants and putative disease-causing mutations in photoreceptor-specific nuclear receptor NR2E3.’, *Molecular vision*, 15, pp. 2174–218484

Kang, H. *et al.* (2017) ‘Bivalent complexes of PRC1 with orthologs of BRD4 and MOZ / MORF target developmental genes in Drosophila’, pp. 1–15. doi: 10.1101/gad.305987.117.5.

Katoh, M. (2013) ‘Functional and cancer genomics of ASXL family members’, *British Journal of Cancer*. Nature Publishing Group, pp. 299–306. doi: 10.1038/bjc.2013.281.

Katsumoto, T. *et al.* (2006) ‘MOZ is essential for maintenance of hematopoietic stem cells’, *Genes & Development*, 20(10), pp. 1321–1330. doi: 10.1101/gad.1393106.

Kaufman, L., Ayub, M. and Vincent, J. B. (2010) ‘The genetic basis of non-syndromic intellectual disability: A review’, *Journal of Neurodevelopmental Disorders*, pp. 182–209. doi: 10.1007/s11689-010-9055-2.

Keeling, K. (2016) ‘Nonsense Suppression as an Approach to Treat Lysosomal Storage Diseases’, *Diseases*, 4(4), p. 32. doi: 10.3390/diseases4040032.

Keeling, K. M. and Bedwell, D. M. (2011) ‘Suppression of nonsense mutations as a therapeutic approach to treat genetic diseases’, *Wiley Interdisciplinary Reviews: RNA*, pp. 837–852. doi: 10.1002/wrna.95.

Kelley, L. A. *et al.* (2015) ‘The Phyre2 web portal for protein modeling, prediction and analysis’, *Nature Protocols*. Nature Publishing Group, 10(6), pp. 845–858. doi: 10.1038/nprot.2015.053.

Kerem, E. *et al.* (2014) ‘Ataluren for the treatment of nonsense-mutation cystic fibrosis: A randomised, double-blind, placebo-controlled

phase 3 trial', *The Lancet Respiratory Medicine*, 2(7), pp. 539–547. doi: 10.1016/S2213-2600(14)70100-6.

Kestler, H. A. and Kühl, M. (2008) 'From individual Wnt pathways towards a Wnt signalling network', *Philosophical Transactions of the Royal Society B: Biological Sciences*. The Royal Society, pp. 1333–1347. doi: 10.1098/rstb.2007.2251.

Kim, H. and Kim, J. S. (2014) 'A guide to genome engineering with programmable nucleases', *Nature Reviews Genetics*. Nature Publishing Group, pp. 321–334. doi: 10.1038/nrg3686.

Kim, Y. R. *et al.* (2017) 'Identifying the KAT6B Mutation via Diagnostic Exome Sequencing to Diagnose Say-Barber- Biesecker-Young-Simpson Syndrome in Three Generations of a Family', 41(3), pp. 505–510

Kindle, K. B. *et al.* (2005) 'MOZ-TIF2 inhibits transcription by nuclear receptors and p53 by impairment of CBP function', *Molecular and Cellular Biology*, 25(3), pp. 988–1002. doi: 10.1128/MCB.25.3.988.

Kitabayashi, I. *et al.* (2001) 'Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ-CBP fusion protein.', *The EMBO journal*. EMBO Press, 20(24), pp. 7184–96. doi: 10.1093/emboj/20.24.7184.

Kitatani, M. *et al.* (1988) 'Localization of the spherocytosis gene to chromosome segment 8p11.22----8p21', *Human genetics*, 78(1), pp. 94–95

Kjaergaard, M., Teilum, K. and Poulsen, F. M. (2010) 'Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 107(28), pp. 12535–12540. doi: 10.1073/pnas.1001693107.

Kleff, S. *et al.* (1995) 'Identification of a Gene Encoding a Yeast Histone H4 Acetyltransferase', *Journal of Biological Chemistry*, 270(42), pp. 24674–24677. doi: 10.1074/jbc.270.42.24674.

Kleinstiver, B. P. *et al.* (2016) 'High-fidelity CRISPR–Cas9 nucleases with

no detectable genome-wide off-target effects', *Nature*. Nature Publishing Group, 529(7587), pp. 490–495. doi: 10.1038/nature16526.

Köhler, A. and Hurt, E. (2007) 'Exporting RNA from the nucleus to the cytoplasm', *Nature Reviews Molecular Cell Biology*, 8(10), pp. 761–773. doi: 10.1038/nrm2255.

Kohlmann, A. *et al.* (2005) 'New insights into MLL gene rearranged acute leukemias using gene expression profiling: shared pathways, lineage commitment, and partner genes', *Leukemia*, 19(6), pp. 953–964. doi: 10.1038/sj.leu.2403746.

Komaki, S. *et al.* (2018) 'iMETHYL: an integrative database of human DNA methylation, gene expression, and genomic variation', *Human Genome Variation*. Nature Publishing Group, 5, p. 18008. doi: 10.1038/hgv.2018.8.

Komander, D. (2009) 'The emerging complexity of protein ubiquitination', *Biochemical Society Transactions*, 37(5), pp. 937–953. doi: 10.1042/BST0370937.

Komiya, Y. and Habas, R. (2008) 'Wnt signal transduction pathways.', *Organogenesis*. Taylor & Francis, 4(2), pp. 68–75

Koonin, E. V, Makarova, K. S. and Zhang, F. (2017) 'Diversity, classification and evolution of CRISPR-Cas systems.', *Current opinion in microbiology*, 37, pp. 67–78. doi: 10.1016/j.mib.2017.05.008.

Kouzarides, T. (2007) 'Chromatin Modifications and Their Function', *Cell*, pp. 693–705. doi: 10.1016/j.cell.2007.02.005.

Krause, A. *et al.* (2004) 'Tbx5 and Tbx4 transcription factors interact with a new chicken PDZ-LIM protein in limb and heart development', *Developmental Biology*. Academic Press, 273(1), pp. 106–120. doi: 10.1016/j.ydbio.2004.05.024.

Kriwacki, R. W. *et al.* (1996) 'Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of

Sciences, 93(21), pp. 11504–9

Kueh, A. J. *et al.* (2011) 'HBO1 Is Required for H3K14 Acetylation and Normal Transcriptional Activity during Embryonic Development', *Molecular and Cellular Biology*. American Society for Microbiology, 31(4), pp. 845–860. doi: 10.1128/MCB.00159-10.

Kurosaki, T. and Maquat, L. E. (2016) 'Nonsense-mediated mRNA decay in humans at a glance', *Journal of Cell Science*, 129(3), pp. 461–467. doi: 10.1242/jcs.181008.

Kuscu, C. *et al.* (2017) 'CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations', *Nature Methods*. Nature Publishing Group, 14(7), pp. 710–712. doi: 10.1038/nmeth.4327.

Kwon, S. E. and Chapman, E. R. (2011) 'Synaptophysin regulates the kinetics of synaptic vesicle endocytosis in central neurons.', *Neuron*. NIH Public Access, 70(5), pp. 847–54. doi: 10.1016/j.neuron.2011.04.001.

Lai, C. S. L. *et al.* (2001) 'A forkhead-domain gene is mutated in a severe speech and language disorder', *Nature*, 413(6855), pp. 519–523. doi: 10.1038/35097076.

Lai, C. S. L. *et al.* (2003) 'FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder', *Brain*. Oxford University Press, 126(11), pp. 2455–2462. doi: 10.1093/brain/awg247.

Largeot, A. *et al.* (2013) 'Symplekin, a polyadenylation factor, prevents MOZ and MLL activity on HOXA9 in hematopoietic cells', *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier B.V., 1833(12), pp. 3054–3063. doi: 10.1016/j.bbamcr.2013.08.013.

Laue, K. *et al.* (2008) 'The multidomain protein Brpf1 binds histones and is required for Hox gene expression and segmental identity', *Development*, 135(11), pp. 1935–1946. doi: 10.1242/dev.017160.

Lavado, A. *et al.* (2010) 'Prox1 Is required for granule cell maturation and intermediate progenitor maintenance during brain neurogenesis',

PLoS Biology. Edited by T. D. Palmer. Public Library of Science, 8(8), pp. 43–44. doi: 10.1371/journal.pbio.1000460.

Lawrence, H. J. *et al.* (1997) 'Mice bearing a targeted interruption of the homeobox gene HOXA9 have defects in myeloid, erythroid, and lymphoid hematopoiesis.', *Blood*, 89(6), pp. 1922–1930

Lawrence, H. J. *et al.* (2005) 'Loss of expression of the Hoxa-9 homeobox gene impairs the proliferation and repopulating ability of hematopoietic stem cells', *Blood*, 106(12), pp. 3988–3994. doi: 10.1182/blood-2005-05-2003.

Lee, T. I. and Young, R. A. (2013) 'Transcriptional regulation and its misregulation in disease', *Cell*. NIH Public Access, pp. 1237–1251. doi: 10.1016/j.cell.2013.02.014.

Leeper, N. J. *et al.* (2013) 'Loss of CDKN2B promotes p53-dependent smooth muscle cell apoptosis and aneurysm formation', *Arteriosclerosis, Thrombosis, and Vascular Biology*. NIH Public Access, 33(1), pp. e1–e10. doi: 10.1161/ATVBAHA.112.300399.

Legube, G. and Trouche, D. (2003) 'Regulating histone acetyltransferases and deacetylases.', *EMBO reports*, 4(10), pp. 944–7. doi: 10.1038/sj.embor.embor941.

Lepack, A. E. *et al.* (2016) 'Aberrant H3.3 dynamics in NAc promote vulnerability to depressive-like behavior', *Proceedings of the National Academy of Sciences*, 113(44), pp. 12562–12567. doi: 10.1073/pnas.1608270113.

Lewin, B. (2008) *GENES IX*. 9th edn. Sudbury, Massachusetts: Jones & Bartlett Publishers

Li, G. *et al.* (2017) 'De Novo Mutation of KAT6B Gene Causing Atypical Say–Barber–Biesecker–Young–Simpson Syndrome or Genitopatellar Syndrome', *Fetal and Pediatric Pathology*. Taylor & Francis, 36(2), pp. 130–138. doi: 10.1080/15513815.2017.1281364.

Li, L. *et al.* (2013) 'Targeted Disruption of Hotair Leads to Homeotic Transformation and Gene Derepression', *Cell Reports*. Cell Press,

5(1), pp. 3–12. doi: 10.1016/j.celrep.2013.09.003.

Liang, J. *et al.* (1998) 'Acute mixed lineage leukemia with an inv(8)(p11q13) resulting in fusion of the genes for MOZ and TIF2.', *Blood*, 92(6), pp. 2118–2122

Lin, A. E. and Gettig, E. (1990) 'Craniosynostosis, Agenesis of the Corpus Callosum, Severe Mental Retardation, Distinctive Facies, Camptodactyly, and Hypogonadism', *American journal of medical genetics*, 25, pp. 582–585. doi: 10.1002/ajmg.1320350427.

Lin, S. *et al.* (2014) 'Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery', *eLife*. eLife Sciences Publications, Ltd, 3, p. e04766. doi: 10.7554/eLife.04766.

Linde, L. *et al.* (2007) 'Nonsense-mediated mRNA decay affects nonsense transcript levels and governs response of cystic fibrosis patients to gentamicin', *Journal of Clinical Investigation*, 117(3), pp. 683–692. doi: 10.1172/JCI28523.

Linde, L. and Kerem, B. (2008) 'Introducing sense into nonsense in treatments of human genetic diseases', *Trends in Genetics*, pp. 552–563. doi: 10.1016/j.tig.2008.08.010.

Lindeboom, R. G. H., Supek, F. and Lehner, B. (2016) 'The rules and impact of nonsense-mediated mRNA decay in human cancers', *Nature Genetics*, 48(10), pp. 1112–1118. doi: 10.1038/ng.3664.

Liu, D. D. *et al.* (2016) 'Growth differentiation factor-15 promotes glutamate release in medial prefrontal cortex of mice through upregulation of T-Type calcium channels', *Scientific Reports*. Nature Publishing Group, 6(1), p. 28653. doi: 10.1038/srep28653.

Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method', *Methods*. Academic Press, 25(4), pp. 402–408. doi: 10.1006/meth.2001.1262.

Lloyd, J. A. (2018) *Erythropoiesis*. 1st edn. Edited by J. A. Lloyd. New

York, NY: Springer New York (Methods in Molecular Biology). doi: 10.1007/978-1-4939-7428-3.

Lloyd, J. T. and Glass, K. C. (2018) 'Biological function and histone recognition of family IV bromodomain-containing proteins', *Journal of Cellular Physiology*, 233(3), pp. 1877–1886. doi: 10.1002/jcp.26010.

Logan, T. T., Rusnak, M. and Symes, A. J. (2015) 'Runx1 promotes proliferation and neuronal differentiation in adult mouse neurosphere cultures', *Stem Cell Research*. Elsevier, 15(3), pp. 554–564. doi: 10.1016/j.scr.2015.09.014.

Lonardo, F. *et al.* (2019) 'Say-Barber-Biesecker-Young-Simpson syndrome and Genitopatellar syndrome: Lumping or splitting?', *Clinical Genetics*, 95(2), pp. 253–261. doi: 10.1111/cge.13127.

Loomis, R. J. *et al.* (2009) 'Chromatin Binding of SRp20 and ASF/SF2 and Dissociation from Mitotic Chromosomes Is Modulated by Histone H3 Serine 10 Phosphorylation', *Molecular Cell*, 33(4), pp. 450–461. doi: 10.1016/j.molcel.2009.02.003.

Lou, C.-H. *et al.* (2016) 'Nonsense-Mediated RNA Decay Influences Human Embryonic Stem Cell Fate', *Stem Cell Reports*. Elsevier, 6(6), pp. 844–857. doi: 10.1016/j.stemcr.2016.05.008.

Lozzio, C. B. and Lozzio, B. B. (1975) 'Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome.', *Blood*, 45(3), pp. 321–34

Lucero, C. M. J. *et al.* (2013) 'The cancer-related transcription factor Runx2 modulates cell proliferation in human osteosarcoma cell lines', *Journal of Cellular Physiology*, 228(4), pp. 714–723. doi: 10.1002/jcp.24218.

Lüdtke, T. H. *et al.* (2016) 'Tbx2 and Tbx3 Act Downstream of Shh to Maintain Canonical Wnt Signaling during Branching Morphogenesis of the Murine Lung', *Developmental Cell*, 39(2), pp. 239–253. doi: 10.1016/j.devcel.2016.08.007.

Lundblad, J. R. *et al.* (1995) 'Adenoviral E1A-associated protein p300 as a functional homologue of the transcriptional co-activator CBP.',

Nature, pp. 85–88. doi: 10.1038/374085a0.

Lundsgaard, M. *et al.* (2017) 'De novo KAT6B Mutation Identified with Whole-Exome Sequencing in a Girl with Say-Barber/Biesecker/Young-Simpson Syndrome', *Molecular Syndromology*, 8(1), pp. 24–29. doi: 10.1159/000452258.

Lux, S. E. *et al.* (1990) 'Hereditary spherocytosis associated with deletion of human erythrocyte ankyrin gene on chromosome 8.', *Nature*, 345(6277), pp. 736–9. doi: 10.1038/345736a0.

Lv, D. *et al.* (2017) 'Histone acetyltransferase KAT6A upregulates PI3K/Akt signaling through TRIM24 binding', *Cancer Research*, p. canres.1388.2017. doi: 10.1158/0008-5472.CAN-17-1388.

Lykke-Andersen, S. and Jensen, T. H. (2015) 'Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 16(11), pp. 665–677. doi: 10.1038/nrm4063.

MacDonald, B. T., Tamai, K. and He, X. (2009) 'Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases', *Developmental Cell*. NIH Public Access, pp. 9–26. doi: 10.1016/j.devcel.2009.06.016.

Maddaluno, L., Urwyler, C. and Werner, S. (2017) 'Fibroblast growth factors: key players in regeneration and tissue repair', *Development*, 144(22), pp. 4047–4060. doi: 10.1242/dev.152587.

Makarova, K. S. *et al.* (2015) 'An updated evolutionary classification of CRISPR-Cas systems', *Nature Reviews Microbiology*, 13(11), pp. 722–736. doi: 10.1038/nrmicro3569.

Mali, P. *et al.* (2013) 'RNA-guided human genome engineering via Cas9', *Science*. NIH Public Access, 339(6121), pp. 823–826. doi: 10.1126/science.1232033.

Mallo, M. and Alonso, C. R. (2013) 'The regulation of Hox gene expression during animal development', *Development*. Oxford University Press for The Company of Biologists Limited, 140(19), pp. 3951–3963. doi: 10.1242/dev.068346.

Manroe, B. L. *et al.* (1979) 'The neonatal blood count in health and disease. I. Reference values for neutrophilic cells.', *The Journal of*

pediatrics, 95(1), pp. 89–98

Marmorstein, R. and Zhou, M.-M. (2014) 'Writers and Readers of Histone Acetylation: Structure, Mechanism, and Inhibition', *Cold Spring Harbor Perspectives in Biology*, 6(7), pp. a018762–a018762. doi: 10.1101/cshperspect.a018762.

Martin, C. and Zhang, Y. (2005) 'The diverse functions of histone lysine methylation', *Nature Reviews Molecular Cell Biology*, pp. 838–849. doi: 10.1038/nrm1761.

Martin, L. *et al.* (2015) 'Identification and characterization of small molecules that inhibit nonsense mediated RNA decay and suppress nonsense p53 mutations', *Cancer Research*, 74(11), pp. 3104–3113. doi: 10.1158/0008-5472.CAN-13-2235.Identification.

Matalonga, L. *et al.* (2015) 'Effect of Readthrough Treatment in Fibroblasts of Patients Affected by Lysosomal Diseases Caused by Premature Termination Codons', *Neurotherapeutics*, 12(4), pp. 874–886. doi: 10.1007/s13311-015-0368-4.

Mattioli, F. *et al.* (2017) 'Mutations in Histone Acetylase Modifier BRPF1 Cause an Autosomal-Dominant Form of Intellectual Disability with Associated Ptosis', *American Journal of Human Genetics*, 100(1), pp. 105–116. doi: 10.1016/j.ajhg.2016.11.010.

McRae, J. F. *et al.* (2017) 'Prevalence and architecture of de novo mutations in developmental disorders', *Nature*, 542(7642), pp. 433–438. doi: 10.1038/nature21062.

Melis, M. A. *et al.* (2005) 'New case of contiguous gene syndrome at chromosome 8p11.2p12 [3]', *American Journal of Medical Genetics*, 136 A(2), pp. 221–222. doi: 10.1002/ajmg.a.30814.

Meneghini, M. D., Wu, M. and Madhani, H. D. (2003) 'Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin', *Cell*. Cell Press, 112(5), pp. 725–736. doi: 10.1016/S0092-8674(03)00123-5.

Meyer, B. *et al.* (2016) 'Histone H3 Lysine 9 Acetylation Obstructs ATM Activation and Promotes Ionizing Radiation Sensitivity in Normal

Stem Cells', *Stem Cell Reports*. ElsevierCompany., 7(6), pp. 1013–1022. doi: 10.1016/j.stemcr.2016.11.004.

Michaud, J. *et al.* (2008) 'Integrative analysis of RUNX1 downstream pathways and target genes.', *BMC genomics*, 9(1), p. 363. doi: 10.1186/1471-2164-9-363.

von Mikecz, A. (2006) 'The nuclear ubiquitin-proteasome system', *Journal of Cell Science*. The Company of Biologists Ltd, 119(10), pp. 1977–1984. doi: 10.1242/jcs.03008.

Mikkelsen, T. S. *et al.* (2007) 'Genome-wide maps of chromatin state in pluripotent and lineage-committed cells', *Nature*. Nature Publishing Group, 448(7153), pp. 553–560. doi: 10.1038/nature06008.

Millan, F. *et al.* (2016) 'Whole exome sequencing reveals de novo pathogenic variants in KAT6A as a cause of a neurodevelopmental disorder', *American Journal of Medical Genetics Part A*, 170(7), pp. 1791–1798. doi: 10.1002/ajmg.a.37670.

Miller, C. T., Maves, L. and Kimmel, C. B. (2004) 'moz regulates Hox expression and pharyngeal segmental identity in zebrafish.', *Development*. The Company of Biologists Ltd, 131(10), pp. 2443–61. doi: 10.1242/dev.01134.

Miller, J. N. and Pearce, D. A. (2014) 'Nonsense-mediated decay in genetic disease: Friend or foe?', *Mutation Research/Reviews in Mutation Research*. NIH Public Access, 762(0), pp. 52–64. doi: 10.1016/j.mrrev.2014.05.001.

Milne, T. A. *et al.* (2005) 'MLL associates specifically with a subset of transcriptionally active target genes', *Proceedings of the National Academy of Sciences*, 102(41), pp. 14765–14770. doi: 10.1073/pnas.0503630102.

Minsky, N. *et al.* (2008) 'Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells.', *Nature cell biology*, 10(4), pp. 483–8. doi: 10.1038/ncb1712.

Mishima, Y. *et al.* (2011) 'The Hbo1-Brd1/Brpf2 complex is responsible for global acetylation of H3K14 and required for fetal liver erythropoiesis', *Blood*, 118(9), pp. 2443–2453. doi:

10.1182/blood-2011-01-331892.

Monteiro, C. (2018) *Function of MYST3 / MOZ: A Histone Acetyltransferase and Chromatin Regulator Implicated in Haematopoietic Stem Cell Development and Leukaemia*. University of Nottingham

Moore, J. D. (2015) 'The impact of CRISPR–Cas9 on target identification and validation', *Drug Discovery Today*. Elsevier Ltd, 20(4), pp. 450–457. doi: 10.1016/j.drudis.2014.12.016.

Moosajee, M. *et al.* (2015) 'Functional rescue of REP1 following treatment with PTC124 and novel derivative PTC-414 in human choroideremia fibroblasts and the nonsensemediated zebrafish model', *Human Molecular Genetics*, 25(16), pp. 3416–3431. doi: 10.1093/hmg/ddw184.

Murray, C. R. *et al.* (2017) 'Novel Causative Variants in DYRK1A, KARS, and KAT6A Associated with Intellectual Disability and Additional Phenotypic Features', *J Pediatr Genet*, 00, pp. 1–7. doi: 10.1055/s-0037-1598639.

Nagel-Wolfrum, K. *et al.* (2014) 'Translational read-through as an alternative approach for ocular gene therapy of retinal dystrophies caused by in-frame nonsense mutations', *Visual Neuroscience*, pp. 309–316. doi: 10.1017/S0952523814000194.

Naumann, S. *et al.* (2001) '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', *Leukemia Research*, 25(4), pp. 313–322. doi: 10.1016/S0145-2126(00)00125-9.

Nechaev, S. and Adelman, K. (2011) 'Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1809(1), pp. 34–45. doi: 10.1016/j.bbagrm.2010.11.001.

Neugebauer, J. M. and Joseph Yost, H. (2014) 'FGF signaling is required

for brain left-right asymmetry and brain midline formation', *Developmental Biology*. NIH Public Access, 386(1), pp. 123–134. doi: 10.1016/j.ydbio.2013.11.020.

Newman, D. M. *et al.* (2016) 'Acetylation of the Cd8 Locus by KAT6A Determines Memory T Cell Diversity', *Cell Reports*. ElsevierCompany., 16(12), pp. 3311–3321. doi: 10.1016/j.celrep.2016.08.056.

Nie, J. and Sage, E. H. (2009) 'SPARC inhibits adipogenesis by its enhancement of beta-catenin signaling.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 284(2), pp. 1279–90. doi: 10.1074/jbc.M808285200.

Nihongaki, Y. *et al.* (2017) 'CRISPR-Cas9-based photoactivatable transcription systems to induce neuronal differentiation', *Nature Methods*, 14(10), pp. 963–966. doi: 10.1038/nmeth.4430.

Niida, Y. *et al.* (2017) 'A Say-Barber-Biesecker-Young-Simpson variant of Ohdo syndrome with a KAT6B 10-base pair palindromic duplication: A recurrent mutation causing a severe phenotype mixed with genitopatellar syndrome', *Congenital Anomalies*, 57(3), pp. 86–88. doi: 10.1111/cga.12196.

Nishi, H., Hashimoto, K. and Panchenko, A. R. (2011) 'Phosphorylation in protein-protein binding: Effect on stability and function', *Structure*. NIH Public Access, 19(12), pp. 1807–1815. doi: 10.1016/j.str.2011.09.021.

Noy, P. *et al.* (2010) 'PRH/Hhex Controls Cell Survival through Coordinate Transcriptional Regulation of Vascular Endothelial Growth Factor Signaling', *Molecular and Cellular Biology*. American Society for Microbiology Journals, 30(9), pp. 2120–2134. doi: 10.1128/MCB.01511-09.

Ogryzko, V. V *et al.* (1996) 'The Transcriptional Coactivators p300 and CBP Are Histone Acetyltransferases', *Cell*. Elsevier, 87(5), pp. 953–959. doi: 10.1016/S0092-8674(00)82001-2.

Oh, D.-J. *et al.* (2013) 'Overexpression of SPARC in human trabecular

meshwork increases intraocular pressure and alters extracellular matrix.', *Investigative ophthalmology & visual science*, 54(5), pp. 3309–19. doi: 10.1167/iovs.12-11362.

Ohta, K. *et al.* (2005) 'Cloning and Characterization of a cDNA Encoding the Histone Acetyltransferase Monocytic Leukemia Zinc Finger Protein (MOZ) in the Rat', *Journal of health science*, 51(2), pp. 253–256. doi: 10.1248/jhs.51.253.

Ohtsuki, G. *et al.* (2012) 'SK2 channel modulation contributes to compartment-specific dendritic plasticity in cerebellar Purkinje cells', *Neuron*, 75(1), pp. 108–120. doi: 10.1016/j.neuron.2012.05.025.

Okano, S. *et al.* (2018) 'Genitopatellar syndrome: the first reported case in Japan', *Human Genome Variation*. Springer US, 5(1), p. 8. doi: 10.1038/s41439-018-0010-1.

Ono, H. *et al.* (2018) 'Partial androgen insensitivity syndrome caused by a deep intronic mutation creating an alternative splice acceptor site of the AR gene', *Scientific Reports*, 8(1), pp. 4–11. doi: 10.1038/s41598-018-20691-9.

Paggetti, J. *et al.* (2010) 'Crosstalk between leukemia-associated proteins MOZ and MLL regulates HOX gene expression in human cord blood CD34+ cells', *Oncogene*. Nature Publishing Group, 29(36), pp. 5019–5031. doi: 10.1038/onc.2010.254.

Paz, H. *et al.* (2010) 'The homeobox gene Hhex regulates the earliest stages of definitive hematopoiesis', *Blood*, 116(8), pp. 1254–1262. doi: 10.1182/blood-2009-11-254383.

Pelletier, N. *et al.* (2002) 'MOZ and MORF histone acetyltransferases interact with the Runt-domain transcription factor Runx2.', *Oncogene*. Nature Publishing Group, 21(17), pp. 2729–40. doi: 10.1038/sj.onc.1205367.

Peng, M. *et al.* (2018) 'Linking PHYTOCHROME-INTERACTING FACTOR to Histone Modification in Plant Shade Avoidance', *Plant Physiology*, 176(2), pp. 1341–1351. doi: 10.1104/pp.17.01189.

Perez-Campo, F. M. *et al.* (2009) 'The histone acetyl transferase activity

of monocytic leukemia zinc finger is critical for the proliferation of hematopoietic precursors', *Blood*, 113(20), pp. 4866–4874. doi: 10.1182/blood-2008-04-152017.

Perez-Campo, F. M. *et al.* (2014) 'MOZ-Mediated Repression of p16 INK4a Is Critical for the Self-Renewal of Neural and Hematopoietic Stem Cells', *Stem Cells*, 32, pp. 1591–1601

Peri, S. *et al.* (2004) 'Human protein reference database as a discovery resource for proteomics.', *Nucleic acids research*, 32(Database issue), pp. D497-501. doi: 10.1093/nar/gkh070.

Peterson, C. L. and Laniel, M.-A. (2004) 'Histones and histone modifications', *Current Biology*, 14(14), pp. R546–R551. doi: 10.1016/j.cub.2004.07.007.

Pfaffeneder, T. *et al.* (2011) 'The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA', *Angewandte Chemie International Edition*. John Wiley & Sons, Ltd, 50(31), pp. 7008–7012. doi: 10.1002/anie.201103899.

Pfaffl, M. W. (2001) 'A new mathematical model for relative quantification in real-time RT-PCR', *Nucleic Acids Research*. Oxford University Press, 29(9), pp. 45e – 45. doi: 10.1093/nar/29.9.e45.

Phan, J. A. *et al.* (2017) 'Early synaptic dysfunction induced by α-synuclein in a rat model of Parkinson's disease', *Scientific Reports*. Nature Publishing Group, 7(1), p. 6363. doi: 10.1038/s41598-017-06724-9.

Pineault, K. M. and Wellik, D. M. (2014) 'Hox genes and limb musculoskeletal development', *Current Osteoporosis Reports*. NIH Public Access, 12(4), pp. 420–427. doi: 10.1007/s11914-014-0241-0.

Planus, E. *et al.* (1997) 'Binding of urokinase to plasminogen activator inhibitor type-1 mediates cell adhesion and spreading.', *Journal of cell science*, 110 (Pt 9, pp. 1091–8

Plaschka, C. *et al.* (2015) 'Architecture of the RNA polymerase II-Mediator core initiation complex', *Nature*, 518(7539), pp. 376–380. doi: 10.1038/nature14229.

Qiu, Y. *et al.* (2012) ‘Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ reveals a regulatory mechanism for HOXA9 transcription’, *Genes and Development*, 26, pp. 1376–1391. doi: 10.1101/gad.188359.112.is.

Ran, F. A. *et al.* (2013) ‘Genome engineering using the CRISPR-Cas9 system’, *Nature Protocols*, 8(11), pp. 2281–2308. doi: 10.1038/nprot.2013.143.

Ran, F. A. *et al.* (2014) ‘Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity’, *Cell*, 154(6), pp. 1380–1389. doi: 10.1016/j.cell.2013.08.021.Double.

Reale, M. E. *et al.* (2013) ‘The transcription factor Runx2 is under circadian control in the suprachiasmatic nucleus and functions in the control of rhythmic behavior.’, *PloS one*. Edited by L. C. Lyons, 8(1), p. e54317. doi: 10.1371/journal.pone.0054317.

Reese, M. G. *et al.* (1997) ‘Improved splice site detection in Genie’, *Journal of Computational Biology*, 4(3), pp. 311–23. doi: 10.1089/cmb.1997.4.311.

Reifsnyder, C. *et al.* (1996) ‘Yeast SAS silencing genes and human genes associated with AML and HIV- 1 Tat interactions are homologous with acetyltransferases’, *Nature Genetics*, 14(1), pp. 42–49. doi: 10.1038/ng0996-42.

Renaud, J. B. *et al.* (2016) ‘Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases’, *Cell Reports*, 14(9), pp. 2263–2272. doi: 10.1016/j.celrep.2016.02.018.

Retterer, K. *et al.* (2016) ‘Clinical application of whole-exome sequencing across clinical indications’, *Genetics in Medicine*. IOP Publishing, 18(7), pp. 696–704. doi: 10.1038/gim.2015.148.

Richardson, R. *et al.* (2017) ‘Mechanism and evidence of nonsense suppression therapy for genetic eye disorders’, *Experimental Eye Research*. Elsevier Ltd, 155, pp. 24–37. doi: 10.1016/j.exer.2017.01.001.

Risebro, C. A. *et al.* (2009) 'Prox1 maintains muscle structure and growth in the developing heart', *Development*. Company of Biologists, 136(3), pp. 495–505. doi: 10.1242/dev.030007.

Riss, T. L. *et al.* (2004) *Cell Viability Assays*, *Assay Guidance Manual*. Eli Lilly & Company and the National Center for Advancing Translational Sciences

Rogakou, E. P. *et al.* (1998) 'DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 273(10), pp. 5858–5868. doi: 10.1074/jbc.273.10.5858.

Rokudai, S. *et al.* (2009) 'Monocytic leukemia zinc finger (MOZ) interacts with p53 to induce p21 expression and cell-cycle arrest.', *The Journal of biological chemistry*, 284(1), pp. 237–44. doi: 10.1074/jbc.M805101200.

Rokudai, S. *et al.* (2013) '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*, 110(10), pp. 3895–900. doi: 10.1073/pnas.1300490110.

Rossetto, D., Avvakumov, N. and Côté, J. (2012) 'Histone phosphorylation: A chromatin modification involved in diverse nuclear events', *Epigenetics*. Taylor & Francis, pp. 1098–1108. doi: 10.4161/epi.21975.

Roy, B. *et al.* (2015) 'Nonsense suppression by near-cognate tRNAs employs alternative base pairing at codon positions 1 and 3', *Proceedings of the National Academy of Sciences*, 112(10), pp. 3038–3043. doi: 10.1073/pnas.1424127112.

Rozman, M. *et al.* (2004) '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 and Cancer*, 40(2), pp. 140–145. doi: 10.1002/gcc.20022.

Sabari, B. R. *et al.* (2016) 'Metabolic regulation of gene expression

through histone acylations', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 18(2), pp. 90–101. doi: 10.1038/nrm.2016.140.

Sabari, B. R. R. *et al.* (2015) 'Intracellular Crotonyl-CoA Stimulates Transcription through p300-Catalyzed Histone Crotonylation', *Molecular Cell*. Elsevier Inc., 58(2), pp. 203–215. doi: 10.1016/j.molcel.2015.02.029.

Sainsbury, S., Bernecky, C. and Cramer, P. (2015) 'Structural basis of transcription initiation by RNA polymerase II', *Nature Reviews Molecular Cell Biology*, pp. 129–143. doi: 10.1038/nrm3952.

Sakaguchi, K. *et al.* (1997) 'Phosphorylation of Serine 392 Stabilizes the Tetramer Formation of Tumor Suppressor Protein p53', *Biochemistry*. American Chemical Society, 36(33), pp. 10117–10124. doi: 10.1021/bi970759w.

Saleh-Gohari, N. and Helleday, T. (2004) 'Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells.', *Nucleic acids research*, 32(12), pp. 3683–8. doi: 10.1093/nar/gkh703.

Samuelsson, M. K. *et al.* (1999) 'p57Kip2, a glucocorticoid-induced inhibitor of cell cycle progression in HeLa cells.', *Molecular endocrinology (Baltimore, Md.)*, 13(11), pp. 1811–22. doi: 10.1210/mend.13.11.0379.

Satoh, C. *et al.* (2017) 'Three brothers with a nonsense mutation in KAT6A caused by parental germline mosaicism', *Human Genome Variation*, 4(June), p. 17045. doi: 10.1038/hgv.2017.45.

Sawicka, A. and Seiser, C. (2012) 'Histone H3 phosphorylation - A versatile chromatin modification for different occasions', *Biochimie*, pp. 2193–2201. doi: 10.1016/j.biochi.2012.04.018.

Scheper, G. C. *et al.* (2002) 'Phosphorylation of Eukaryotic Initiation Factor 4E Markedly Reduces Its Affinity for Capped mRNA', *Journal of Biological Chemistry*, 277(5), pp. 3303–3309. doi: 10.1074/jbc.M103607200.

Schoch, K. *et al.* (2017) 'A Recurrent De Novo Variant in NACC1 Causes

a Syndrome Characterized by Infantile Epilepsy, Cataracts, and Profound Developmental Delay', *American Journal of Human Genetics*, 100(2), pp. 343–351. doi: 10.1016/j.ajhg.2016.12.013.

Schoenberg, D. R. and Maquat, L. E. (2012) 'Regulation of cytoplasmic mRNA decay', *Nature Reviews Genetics*, 13(6), pp. 448–448. doi: 10.1038/nrg3254.

Schreiber, N. B. and Spicer, L. J. (2012) 'Effects of fibroblast growth factor 9 (FGF9) on steroidogenesis and gene expression and control of FGF9 mRNA in bovine granulosa cells', *Endocrinology*. The Endocrine Society, 153(9), pp. 4491–4501. doi: 10.1210/en.2012-1003.

Schubert, U. *et al.* (2000) 'Rapid degradation of a large fraction of newly synthesized proteins by proteasomes', *Nature*, 404(6779), pp. 770–774. doi: 10.1038/35008096.

Schuettengruber, B. *et al.* (2007) 'Genome Regulation by Polycomb and Trithorax Proteins', *Cell*. Elsevier, pp. 735–745. doi: 10.1016/j.cell.2007.02.009.

Scott, E. K., Lee, T. and Luo, L. (2001) 'enok encodes a Drosophila putative histone acetyltransferase required for mushroom body neuroblast proliferation', *Current Biology*. Elsevier, 11(2), pp. 99–104. doi: 10.1016/S0960-9822(01)00020-3.

Seifert, A. *et al.* (2015) 'Role of Hox genes in stem cell differentiation.', *World journal of stem cells*. Baishideng Publishing Group Inc, 7(3), pp. 583–95. doi: 10.4252/wjsc.v7.i3.583.

Selvi, R. B. *et al.* (2015) 'Inhibition of p300 lysine acetyltransferase activity by luteolin reduces tumor growth in head and neck squamous cell carcinoma (HNSCC) xenograft mouse model.', *Oncotarget*, 6(41), pp. 43806–43818. doi: 10.18632/oncotarget.6245.

Sen, S. *et al.* (2016) 'Genome-wide positioning of bivalent mononucleosomes', *BMC Medical Genomics*. BMC Medical Genomics, 9(1), p. 60. doi: 10.1186/s12920-016-0221-6.

Serravalle, S. *et al.* (2010) 'A novel specific signature of pediatric MOZ-CBP acute myeloid leukemia', *Leukemia Research*. Elsevier, pp. e292-3. doi: 10.1016/j.leukres.2010.05.026.

Shahbazian, M. D. and Grunstein, M. (2007) 'Functions of Site-Specific Histone Acetylation and Deacetylation', *Annual Review of Biochemistry*, 76(1), pp. 75–100. doi: 10.1146/annurev.biochem.76.052705.162114.

Shashi, V. *et al.* (2016) 'De Novo Truncating Variants in ASXL2 Are Associated with a Unique and Recognizable Clinical Phenotype', *American Journal of Human Genetics*, 99(4), pp. 991–999. doi: 10.1016/j.ajhg.2016.08.017.

Sheikh, B. N., Phipson, B., *et al.* (2015) 'MOZ (MYST3, KAT6A) inhibits senescence via the INK4A-ARF pathway', *Oncogene*, 34(47), pp. 5807–5820. doi: 10.1038/onc.2015.33.

Sheikh, B. N., Lee, S. C. W., *et al.* (2015) 'MOZ regulates B-cell progenitors and, consequently, Moz haploinsufficiency dramatically retards MYC-induced lymphoma development', *Blood*, 125(12), pp. 1910–1921. doi: 10.1182/blood-2014-08-594655.

Sheikh, B. N. *et al.* (2016) 'MOZ (KAT6A) is essential for the maintenance of classically defined adult hematopoietic stem cells', *Blood*, 128(19), pp. 2307–2318. doi: 10.1182/blood-2015-10-676072.

Sheikh, B. N. *et al.* (2017) 'MOZ and BMI1 act synergistically to maintain hematopoietic stem cells', *Experimental Hematology*. Elsevier Ltd, 47, pp. 83-97.e8. doi: 10.1016/j.exphem.2016.10.006.

Shih, Y. Y. *et al.* (2012) 'Calreticulin mediates nerve growth factor-induced neuronal differentiation', *Journal of Molecular Neuroscience*. Humana Press Inc, 47(3), pp. 571–581. doi: 10.1007/s12031-011-9683-3.

Shiota, M. *et al.* (2010) 'Tip60 promotes prostate cancer cell proliferation by translocation of androgen receptor into the nucleus', *Prostate*, 70(5), pp. 540–554. doi: 10.1002/pros.21088.

Shogren-Knaak, M. *et al.* (2006) 'Histone H4-K16 acetylation controls

chromatin structure and protein interactions.', *Science (New York, N.Y.)*, 311(5762), pp. 844–7. doi: 10.1126/science.1124000.

Shukla, S. *et al.* (2018) 'The Kat in the HAT: The Histone Acetyl Transferase Kat6b (MYST4) Is Downregulated in Murine Macrophages in Response to LPS', *Mediators of Inflammation*, 2018(1), pp. 1–11. doi: 10.1155/2018/7852742.

Shulman, E. *et al.* (2014) 'Designer aminoglycosides that selectively inhibit cytoplasmic rather than mitochondrial ribosomes show decreased ototoxicity a strategy for the treatment of genetic diseases', *Journal of Biological Chemistry*, 289(4), pp. 2318–2330. doi: 10.1074/jbc.M113.533588.

Simithy, J. *et al.* (2017) 'Characterization of histone acylations links chromatin modifications with metabolism', *Nature Communications*. Springer US, 8(1), p. 1141. doi: 10.1038/s41467-017-01384-9.

Skokowa, J. *et al.* (2014) 'Cooperativity of RUNX1 and CSF3R mutations in severe congenital neutropenia: A unique pathway in myeloid leukemogenesis', *Blood*, 123(14), pp. 2229–2237. doi: 10.1182/blood-2013-11-538025.

Snead, D. and Eliezer, D. (2014) 'Alpha-Synuclein Function and Dysfunction on Cellular Membranes', *Experimental Neurobiology*. Korean Society for Brain and Neural Science, 23(4), p. 292. doi: 10.5607/en.2014.23.4.292.

Sonenberg, N. and Hinnebusch, A. G. (2009) 'Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets', *Cell*, pp. 731–745. doi: 10.1016/j.cell.2009.01.042.

Song, F. *et al.* (2014) 'Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units', *Science*, 344(6182), pp. 376–380. doi: 10.1126/science.1251413.

Sowa, M. E. *et al.* (2009) 'Defining the Human Deubiquitinating Enzyme Interaction Landscape', *Cell*, 138(2), pp. 389–403. doi: 10.1016/j.cell.2009.04.042.

Spender, L. C. *et al.* (2005) 'Transcriptional cross-regulation of RUNX1

by RUNX3 in human B cells', *Oncogene*, 24(11), pp. 1873–1881. doi: 10.1038/sj.onc.1208404.

Stefanis, L. (2012) 'α-Synuclein in Parkinson's disease.', *Cold Spring Harbor perspectives in medicine*. Cold Spring Harbor Laboratory Press, 2(2), p. a009399. doi: 10.1101/cshperspect.a009399.

Steinberg, R. and Gorman, K. (1992) 'Linked spontaneous CG----TA mutations at CpG sites in the gene for protein kinase regulatory subunit.', *Molecular and cellular biology*. American Society for Microbiology (ASM), 12(2), pp. 767–772. doi: 10.1128/MCB.12.2.767.

Stratton, R. F. *et al.* (1992) 'Deletion of the proximal short arm of chromosome 8', *American Journal of Medical Genetics*, 42(1), pp. 15–18. doi: 10.1002/ajmg.1320420105.

Struhl, K. (1998) 'Histone acetylation and transcriptional regulatory mechanisms', *Genes & Development*, 12(5), pp. 599–606. doi: 10.1101/gad.12.5.599.

Stultz, C. M., Levin, A. D. and Edelman, E. R. (2002) 'Phosphorylation-induced conformational changes in a mitogen-activated protein kinase substrate. Implications for tyrosine hydroxylase activation.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 277(49), pp. 47653–61. doi: 10.1074/jbc.M208755200.

Sun, B. *et al.* (2011) 'Regulation of the histone acetyltransferase activity of hMOF via autoacetylation of Lys274', *Cell Research*. Nature Publishing Group, 21(8), pp. 1262–1266. doi: 10.1038/cr.2011.105.

Szenker, E., Ray-Gallet, D. and Almouzni, G. (2011) 'The double face of the histone variant H3.3.', *Cell research*. Nature Publishing Group, 21(3), pp. 421–34. doi: 10.1038/cr.2011.14.

Takechi, S. and Nakayama, T. (1999) 'Sas3 is a histone acetyltransferase and requires a zinc finger motif', *Biochemical and Biophysical Research Communications*, 266(2), pp. 405–410. doi: 10.1006/bbrc.1999.1836.

Takeuchi, J. K. *et al.* (2003) 'Tbx5 and Tbx4 trigger limb initiation through activation of the Wnt/Fgf signaling cascade.', *Development (Cambridge, England)*. The Company of Biologists Ltd, 130(12), pp. 2729–39. doi: 10.1242/dev.00474.

Tan, E.-P. *et al.* (2015) 'Off-target assessment of CRISPR-Cas9 guiding RNAs in human iPS and mouse ES cells', *Genesis*, 53(2), pp. 225–236. doi: 10.1002/dvg.22835.

Tarpey, P. S. *et al.* (2009) 'A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation', *Nature Genetics*. Nature Publishing Group, 41(5), pp. 535–543. doi: 10.1038/ng.367.

Tarsa, L. and Goda, Y. (2002) 'Synaptophysin regulates activity-dependent synapse formation in cultured hippocampal neurons', *Proceedings of the National Academy of Sciences*, 99(2), pp. 1012–1016. doi: 10.1073/pnas.022575999.

Tham, E. *et al.* (2015) 'Dominant Mutations in KAT6A Cause Intellectual Disability with Recognizable Syndromic Features', *The American Journal of Human Genetics*. The American Society of Human Genetics, 96(3), pp. 507–513. doi: 10.1016/j.ajhg.2015.01.016.

Theriault, F. M., Roy, P. and Stifani, S. (2004) 'AML1/Runx1 is important for the development of hindbrain cholinergic branchiovisceral motor neurons and selected cranial sensory neurons', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 101(28), pp. 10343–10348. doi: 10.1073/pnas.0400768101.

Torshabi, M. *et al.* (2011) 'Runx3 expression inhibits proliferation and distinctly alters mRNA expression of Bax in AGS and A549 cancer cells', *Iranian Journal of Pharmaceutical Research*, 10(2), pp. 355–361

Trinh, J. *et al.* (2018) 'A KAT6A variant in a family with autosomal dominantly inherited microcephaly and developmental delay', *Journal of Human Genetics*. Springer US, pp. 2–6. doi:

10.1038/s10038-018-0469-0.

Tsukahara, M. *et al.* (1995) ‘Interstitial deletion of 8p: report of two patients and review of the literature.’, *Clinical genetics*, 48(1), pp. 41–45

Turner-Ivey, B. *et al.* (2014) ‘KAT6A, a Chromatin Modifier from the 8p11-p12 Amplicon is a Candidate Oncogene in Luminal Breast Cancer’, *Neoplasia*. Neoplasia Press, Inc., 16(8), pp. 644–655. doi: 10.1016/j.neo.2014.07.007.

Turner, B. M. and O’Neill, L. P. (1995) ‘Histone acetylation in chromatin and chromosomes.’, *Seminars in cell biology*, 6(4), pp. 229–36

Ullah, M. *et al.* (2008) ‘Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes.’, *Molecular and cellular biology*, 28(22), pp. 6828–6843. doi: 10.1128/MCB.01297-08.

Usuki, F. *et al.* (2006) ‘Specific inhibition of nonsense-mediated mRNA decay components, SMG-1 or Upf1, rescues the phenotype of ullrich disease fibroblasts’, *Molecular Therapy*, 14(3), pp. 351–360. doi: 10.1016/j.ymthe.2006.04.011.

Vanyai, H. K., Thomas, T. and Voss, A. K. (2015) ‘Mesodermal expression of Moz is necessary for cardiac septum development’, *Developmental Biology*. Elsevier, 403(1), pp. 22–29. doi: 10.1016/j.ydbio.2015.04.011.

Veltman, J. A. and Brunner, H. G. (2012) ‘De novo mutations in human genetic disease’, *Nature Reviews Genetics*, pp. 565–575. doi: 10.1038/nrg3241.

Verdone, L. *et al.* (2006) ‘Histone acetylation in gene regulation’, *Briefings in Functional Genomics and Proteomics*, pp. 209–221. doi: 10.1093/bfgp/ell028.

Vermeulen, S. *et al.* (2002) ‘Kallmann syndrome in a patient with congenital spherocytosis and an interstitial 8p11.2 deletion’, *American Journal of Medical Genetics*, 108(4), pp. 315–318. doi: 10.1002/ajmg.10295.

Vlckova, M. *et al.* (2015) ‘A patient showing features of both SBBYSS

and GPS supports the concept of a KAT6B-related disease spectrum, with mutations in mid-exon 18 possibly leading to combined phenotypes', *European Journal of Medical Genetics*. Elsevier Masson SAS, 58(10), pp. 550–555. doi: 10.1016/j.ejmg.2015.09.004.

Voigt, P., Tee, W.-W. and Reinberg, D. (2013) 'A double take on bivalent promoters', *Genes & Development*. Cold Spring Harbor Laboratory Press, 27(12), pp. 1318–1338. doi: 10.1101/gad.219626.113.

Voon, D. C. C., Hor, Y. T. and Ito, Y. (2015) 'The RUNX complex: Reaching beyond haematopoiesis into immunity', *Immunology*, 146(4), pp. 523–536. doi: 10.1111/imm.12535.

Voss, A. K. *et al.* (2009) 'Moz and Retinoic Acid Coordinately Regulate H3K9 Acetylation, Hox Gene Expression, and Segment Identity', *Developmental Cell*. Elsevier Ltd, 17(5), pp. 674–686. doi: 10.1016/j.devcel.2009.10.006.

Voss, A. K. *et al.* (2012) 'MOZ Regulates the Tbx1 Locus, and Moz Mutation Partially Phenocopies DiGeorge Syndrome', *Developmental Cell*. Elsevier Inc., 23(3), pp. 652–663. doi: 10.1016/j.devcel.2012.07.010.

Voss, A. K. and Thomas, T. (2009) 'MYST family histone acetyltransferases take center stage in stem cells and development', *BioEssays*, 31(10), pp. 1050–1061. doi: 10.1002/bies.200900051.

Walton, R. Z. *et al.* (2006) 'Fog1 is required for cardiac looping in zebrafish', *Developmental Biology*, 289(2), pp. 482–493. doi: 10.1016/j.ydbio.2005.10.040.

Wang, H. *et al.* (2013) 'One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering', *Cell*. Elsevier Inc., 153(4), pp. 910–918. doi: 10.1016/j.cell.2013.04.025.

Wang, J. *et al.* (2017) 'WebGestalt 2017: A more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit', *Nucleic Acids Research*. Oxford University Press,

45(W1), pp. W130–W137. doi: 10.1093/nar/gkx356.

Wang, S. *et al.* (2016) 'RaptorX-Property: a web server for protein structure property prediction', *Nucleic acids research*. Oxford University Press, 44(W1), pp. W430–W435. doi: 10.1093/nar/gkw306.

Wang, Z. *et al.* (2008) 'Combinatorial patterns of histone acetylations and methylations in the human genome', *Nature Genetics*, 40(7), pp. 897–903. doi: 10.1038/ng.154.

Wang, Z. *et al.* (2018) 'Effects of secreted frizzled-related protein 1 on proliferation, migration, invasion, and apoptosis of colorectal cancer cells', *Cancer Cell International*. BioMed Central, 18(1), p. 48. doi: 10.1186/s12935-018-0543-x.

Weischenfeldt, J. *et al.* (2008) 'NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements', *Genes and Development*, 22(10), pp. 1381–1396. doi: 10.1101/gad.468808.

Witt, O., Albig, W. and Doenecke, D. (1996) 'Testis-specific expression of a novel human H3 histone gene', *Experimental Cell Research*, 229(2), pp. 301–306. doi: 10.1006/excr.1996.0375.

Wolpe, S. D. and Cerami, A. (1989) 'Macrophage inflammatory proteins 1 and 2: members of a novel superfamily of cytokines.', *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 3(14), pp. 2565–73

Wong, J. J.-L. *et al.* (2013) 'Orchestrated Intron Retention Regulates Normal Granulocyte Differentiation', *Cell*, 154(3), pp. 583–595. doi: 10.1016/j.cell.2013.06.052.

Woodcock, C. L. and Ghosh, R. P. (2010) 'Chromatin Higher-order Structure and Dynamics', *Cold Spring Harbor Perspectives in Biology*, 2(5), pp. a000596–a000596. doi: 10.1101/cshperspect.a000596.

Wright, C. F. *et al.* (2015) 'Genetic diagnosis of developmental disorders in the DDD study: A scalable analysis of genome-wide research data', *The Lancet*, 385(9975), pp. 1305–1314. doi:

10.1016/S0140-6736(14)61705-0.

Xiong, X. *et al.* (2016) ‘Selective recognition of histone crotonylation by double PHD fingers of MOZ and DPF2’, *Nature Chemical Biology*, 12(12), pp. 1111–1118. doi: 10.1038/nchembio.2218.

Xu, J., Liu, Z. and Ornitz, D. M. (2000) ‘Temporal and spatial gradients of Fgf8 and Fgf17 regulate proliferation and differentiation of midline cerebellar structures’, *Development*, 127(9), pp. 1833–1843. doi: 10.1242/dev.00555.

Yan, K. *et al.* (2016) ‘The Chromatin Regulator BRPF3 Preferentially Activates the HBO1 Acetyltransferase but Is Dispensable for Mouse Development and Survival’, *Journal of Biological Chemistry*, 291(6), pp. 2647–2663. doi: 10.1074/jbc.M115.703041.

Yan, K. *et al.* (2017) ‘Mutations in the Chromatin Regulator Gene BRPF1 Cause Syndromic Intellectual Disability and Deficient Histone Acetylation’, *American Journal of Human Genetics*, 100(1), pp. 91–104. doi: 10.1016/j.ajhg.2016.11.011.

Yan, Y. *et al.* (2000) ‘Crystal structure of yeast Esa1 suggests a unified mechanism for catalysis and substrate binding by histone acetyltransferases’, *Molecular cell*. Sterner and Berger, 6(5), pp. 1195–1205. doi: 10.1016/S1097-2765(00)00116-7.

Yang, C. *et al.* (2012) ‘Autoacetylation of the MYST lysine acetyltransferase MOF protein’, *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 287(42), pp. 34917–34926. doi: 10.1074/jbc.M112.359356.

Yang, X.-J. (2004) ‘The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases.’, *Nucleic acids research*, 32(3), pp. 959–76. doi: 10.1093/nar/gkh252.

Yang, X.-J. and Ullah, M. (2007) ‘MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells.’, *Oncogene*, 26(37), pp. 5408–5419. doi: 10.1038/sj.onc.1210609.

Yang, X. (2015) ‘MOZ and MORF acetyltransferases: Molecular interaction, animal development and human disease’, *BBA* -

Molecular Cell Research. Elsevier B.V., 1853(8), pp. 1818–1826. doi: 10.1016/j.bbamcr.2015.04.014.

Ye, J. *et al.* (2012) 'Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction.', *BMC bioinformatics*, 13(1), p. 134. doi: 10.1186/1471-2105-13-134.

Yilmaz, R. *et al.* (2015) 'A recurrent synonymous KAT6B mutation causes Say-Barber-Biesecker/Young-Simpson syndrome by inducing aberrant splicing', *American Journal of Medical Genetics, Part A*, 167(12), pp. 3006–3010. doi: 10.1002/ajmg.a.37343.

Yoon, Y. *et al.* (2018) 'Streamlined ex vivo and in vivo genome editing in mouse embryos using recombinant adeno-associated viruses', *Nature Communications*. Springer US, 9(1), p. 412. doi: 10.1038/s41467-017-02706-7.

Yu, C. *et al.* (2015) 'Small molecules enhance crispr genome editing in pluripotent stem cells', *Cell Stem Cell*. Elsevier Inc., 16(2), pp. 142–147. doi: 10.1016/j.stem.2015.01.003.

Yu, H. C. *et al.* (2014) 'An individual with blepharophimosis-ptosis-epicanthus inversus syndrome (BPES) and additional features expands the phenotype associated with mutations in KAT6B', *American Journal of Medical Genetics, Part A*, 164(4), pp. 950–957. doi: 10.1002/ajmg.a.36379.

Yu, L. *et al.* (2016) 'Identification of MYST3 as a novel epigenetic activator of ERα frequently amplified in breast cancer', *Oncogene*, (May), pp. 1–9. doi: 10.1038/onc.2016.433.

Zaidi, S. *et al.* (2013) 'De novo mutations in histone-modifying genes in congenital heart disease.', *Nature*, 498(7453), pp. 220–3. doi: 10.1038/nature12141.

Zetoune, A. B. *et al.* (2008) 'Comparison of nonsense-mediated mRNA decay efficiency in various murine tissues', *BMC Genetics*. BioMed Central, 9, p. 83. doi: 10.1186/1471-2156-9-83.

Zetsche, B. *et al.* (2015) 'Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System', *Cell*, 163(3), pp. 759–771. doi: 10.1016/j.cell.2015.09.038.

Zhang, L. H. *et al.* (2015) 'TRIM24 promotes glioma progression and enhances chemoresistance through activation of the PI3K/Akt signaling pathway', *Oncogene*, 34(5), pp. 600–610. doi: 10.1038/onc.2013.593.

Zhang, Q. *et al.* (2016) 'Structural Insights into Histone Crotonyl-Lysine Recognition by the AF9 YEATS Domain', *Structure*. Elsevier Ltd, 24(9), pp. 1606–1612. doi: 10.1016/j.str.2016.05.023.

Zhang, Y. (2003) 'Transcriptional regulation by histone ubiquitination and deubiquitination', *Genes and Development*, pp. 2733–2740. doi: 10.1101/gad.1156403.

Zhao, J., Hyman, L. and Moore, C. (1999) 'Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis.', *Microbiology and molecular biology reviews : MMBR*. American Society for Microbiology (ASM), 63(2), pp. 405–45

Zhou, B. *et al.* (2018) 'Comprehensive, integrated and phased whole-genome analysis of the primary ENCODE cell line K562', *BioRxiv*, pp. 1–49. doi: 10.1101/192344.

Zwaveling-Soonawala, N. *et al.* (2017) 'Variants in KAT6A and pituitary anomalies', *American Journal of Medical Genetics Part A*, 173(9), pp. 2562–2565. doi: 10.1002/ajmg.a.38330.

Zwaveling-Soonawala, N. *et al.* (2018) 'Clues for polygenic inheritance of pituitary stalk interruption syndrome from exome sequencing in 20 patients', *Journal of Clinical Endocrinology and Metabolism*, 103(2), pp. 415–428. doi: 10.1210/jc.2017-01660.

APPENDICES

Appendix I - KAT6A variants associated with KAT6A GDD

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
3	p.Trp43*	Unknown	Unknown	(Arboleda *et al.*, 2015)	Variant present in the Exome Aggregation Consortium (ExAC) browser, but no clinical information available.	Mainly dominant-negative MOZ protein translation, with low NMD efficacy upon KAT6A PTC mRNA transcripts.
3	p.Gln58*	Unknown	Unknown	(Arboleda *et al.*, 2015)	Variant present in the Exome Aggregation Consortium (ExAC) browser, but no clinical information available.	Mainly dominant-negative MOZ protein translation, with low NMD efficacy upon KAT6A PTC mRNA transcripts.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
3	c.235C>T p.Arg79*	F	• Severe feeding difficulties • Neonatal hypotonia • Poor speech development • Global developmental delay • Ptosis • Broad nasal bridge with a bulbous tip • Thin, small mouth • Dental crowding • Hypothyroidism • Absent pituitary stalk	(Zwaveling-Soonawala *et al.*, 2017)	Diagnosed as having Pituitary Stalk Interruption Syndrome (PSIS) and more specifically Multiple Pituitary Hormone Deficiencies (MPHD), but not KAT6A GDD (although some overlapping features). Hormone treatments have improved motor development to the normal level at age 10 and feeding difficulties have subsided.	Mainly dominant-negative MOZ protein translation, with low NMD efficacy upon KAT6A PTC mRNA transcripts.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Ectopic posterior pituitary lobe			
8	p.Ile368_Thr369ins*	Unknown	Unknown	(Arboleda *et al.*, 2015)	Variant present in the Exome Aggregation Consortium (ExAC) browser, but no clinical information available.	A combination of dominant-negative MOZ protein translation and NMD activity upon KAT6A PTC mRNA transcripts.
8	p.Glu380*	Unknown	Unknown	(Arboleda *et al.*, 2015)	Variant present in the Exome Aggregation Consortium (ExAC) browser, but no clinical information available.	A combination of dominant-negative MOZ protein translation and NMD activity upon KAT6A PTC mRNA transcripts.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
Intron 10-11/exon 11	c.1599-56_1621del (splice site disruption) PREDICTED: p.Pro535Metfs*3	M	• Childhood apraxia of speech (CAS).	(Eising *et al.*, 2018)	No motor impairments, average IQ. No mention of dysmorphic facial features.	A combination of dominant-negative MOZ protein translation and NMD activity upon KAT6A PTC mRNA transcripts.
13	c.1928 A>G p.Asn643Ser	F	• Global developmental delay • Intellectual disability • Hypotonia • Broad nose with a bulbous tip • Widely spaced eyes • Thin upper lip	(Millan *et al.*, 2016)		Exclusively dominant-negative MOZ protein translation due to missense mutation within the MYST domain.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Small ears • Widely space teeth • Flattened midface • Poor speech development			
Intron 16-17/exon17	c.3040-1_3040delGA (splice site disruption) PREDICTED: p.Lys1014Thrfs*4	M	• Poor speech development (completely non-verbal) • Intellectual disability • Epilepsy • Bulbous nose • Obsessive-compulsive disorder	(Millan *et al.*, 2016)	Diagnosed as having Autism Spectrum Disorder. Feeding difficulties resolved.	A combination of dominant-negative MOZ protein translation and NMD activity upon KAT6A PTC mRNA transcripts.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Global developmental delay • Missing olfactory bulb • Feeding difficulties			
17	c.3070C>T p.Arg1024*	M	• Pulmonary hypertension • Global developmental delay • Intellectual disability • Micrognathia • Scaphocephaly • Low set ears • Malocclusion • Hypotonia • Feeding difficulties	(Satoh *et al.*, 2017)	Underwent surgery for undescended testes and an inguinal hernia at 10 months of age.	A combination of dominant-negative MOZ protein translation and NMD activity upon KAT6A PTC mRNA transcripts.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Myopic astigmatism • Poor speech development			
17	c.3070C>T p.Arg1024*	M	• Global developmental delay • Delayed speech development • Intellectual disability • Micrognathia • Scaphocephaly • Low set ears • Malocclusion • Hypotonia • Feeding difficulties • Mild hearing loss	(Satoh *et al.*, 2017)		A combination of dominant-negative MOZ protein translation and NMD activity upon KAT6A PTC mRNA transcripts.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Exotropia			
17	c.3070C>T p.Arg1024*	M	• Respiratory issues • Micrognathia • Scaphocephaly • Low set ears • Hypotonia • Feeding difficulties • Laryngomalacia • Possible global developmental delay	(Satoh *et al.*, 2017)		A combination of dominant-negative MOZ protein translation and NMD activity upon KAT6A PTC mRNA transcripts.
17	c.3070C>T p.Arg1024*	M	• Bilateral cryptorchidism • Esotropia • Feeding difficulties	(Arboleda *et al.*, 2015)		A combination of dominant-negative MOZ protein translation and NMD activity upon KAT6A

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Dysmorphic facial features • Hypotonia • Global developmental delay • Poor speech development (completely non-verbal) • Small lower teeth of abnormal shape • Ocular ptosis • Brachydactyly			PTC mRNA transcripts.
17	c.3070C>T p.Arg1024*	F	• Poor speech development	(Millan *et al.*, 2016)	Feeding difficulties resolved.	A combination of dominant-negative

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Global developmental delay • Hypotonia • Bilateral ptosis • Bulbous nasal tip • Micrognathia • Epicanthal folds • Laryngomalacia • Feeding difficulties			MOZ protein translation and NMD activity upon KAT6A PTC mRNA transcripts.
17	c.3116_3117delCT p.Ser1039*	F	• Global developmental delay • Neonatal hypotonia • Poor speech development	(Tham *et al.*, 2015)	Twin of the sibling below, with the same mutation. Atrial septal defect was repaired using a catheter.	A combination of dominant-negative MOZ protein translation and NMD activity upon KAT6A

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			(completely non-verbal) • Bitemporal narrowing • Broad nasal tip • Low-set, rotated ears • Thin upper lip • Atrial septal defect • Feeding difficulties • Ptosis • Esotropia • Craniosynostosis			PTC mRNA transcripts.
17	c.3116_3117delCT p.Ser1039*	F	• Global developmental delay • Poor speech development	(Tham *et al.*, 2015)	Twin of the sibling above, with the same mutation. Cardiac defect spontaneously	A combination of dominant-negative MOZ protein translation and NMD

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			(completely non-verbal) • Mild hypotonia • Bitemporal narrowing • Broad nasal tip • Low-set, rotated ears • Thin upper lip • Cardiac defect • Feeding difficulties • Ptosis • Esotropia • Craniosynostosis • Seizures		rectified itself. Died suddenly at age 10.	activity upon KAT6A PTC mRNA transcripts.
17	c.3230delA p.Asn1077Metfs*46	F	• Poor speech development	(Millan *et al.*, 2016)		A combination of dominant-negative

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Feeding difficulties • Microcephaly • Global developmental delay • Orbital hypotelorism			MOZ protein translation and NMD activity upon KAT6A PTC mRNA transcripts.
17	c.3287delG p.Cys1096Phefs*27	M	• Poor speech development (completely non-verbal) • Microcephaly • Hypotonia • Global developmental delay • Intellectual disability	(Millan *et al.*, 2016)		A combination of dominant-negative MOZ protein translation and NMD activity upon KAT6A PTC mRNA transcripts.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Recurrent ear infections • Feeding difficulties • Prominent nasal tip • Thin upper lip • Large, rotated ears • Finger and elbow joint hypermobility • 2-3 toe syndactyly • Narrow anterior palate			
17	p.Ser1113*	Unknown	Unknown	(Arboleda *et al.*, 2015)	Variant present in the Exome Aggregation Consortium (ExAC) browser, but no	A combination of dominant-negative MOZ protein translation and NMD activity upon KAT6A

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
					clinical information available.	PTC mRNA transcripts.
18	c.3385C>T p.Arg1129*	M	• Microcephaly • Poor swallowing ability • Atrial septal defect • Brachydactyly • Intermittent esotropia • Global developmental delay • Poor speech development • Dysmorphic facial features • Sleep apnea	(Arboleda *et al.*, 2015)	Atrial septal defect was repaired by surgery. Gastric tube fitted due to recurrent aspiration pneumonias.	Exclusively dominant-negative MOZ protein translation due to the PTC residing within the terminal exon.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Prominent nasal bridge with a full nasal tip • Feeding difficulties • Hypotonia • Dystonia			
18	c.3385C>T p.Arg1129*	F	• Cleft palate • Ventricular septal defect • Esotropia • Bronchomalacia • Recurrent ear, respiratory and urinary tract infections	(Arboleda *et al.*, 2015)	Hospitalised multiple times for respiratory and urinary tract infections. Cleft palate was repaired by surgery, myringotomy tubes fitted to deal with recurrent ear infections. Diagnosed	Exclusively dominant-negative MOZ protein translation due to the PTC residing within the terminal exon.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Poor speech development		with Autism Spectrum Disorder; behavioural therapy improved social interactions and eye contact.	
18	c.3385C>T p.Arg1129*	F	• Microcephaly • Distinct facial features • Hypotonia • Dystonia • Global developmental delay • Poor speech development • Insomnia-like sleep disturbances	(Arboleda *et al.*, 2015)	Tympanic tubes were fitted to deal with the ear infections.	Exclusively dominant-negative MOZ protein translation due to the PTC residing within the terminal exon.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Chronic ear infections • Ocular ptosis • Myopia • Atrial septal defect			
18	c.3462delA p.Gly1155Alafs*21	F	• Poor speech development (completely non-verbal) • Atrial septal defect • Microcephaly • Global developmental delay • Short nose with a broad base • Small ears	(Millan *et al.*, 2016)	Atrial septal defect repaired by surgery.	Exclusively dominant-negative MOZ protein translation due to the PTC residing within the terminal exon.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Hypotonia • Brisk reflexes • Supernumerary left nipple • Coarse facial features			
18	c.3655del p.Leu1219Tyrfs*74	F	• Poor eye contact • Intermittent esotropia • Microcephaly • Thin upper lip • Low set ears • Broad nasal bridge • Global developmental delay • Atrial septal defect • Chronic ear infections	(Elenius *et al.*, 2017)	Atrial septal defect repaired by surgery.	Exclusively dominant-negative MOZ protein translation due to the PTC residing within the terminal exon.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Poor speech development • Severe feeding difficulties			
18	c.3830_3831insTT p.Arg1278Serfs*17	M	• Global developmental delay • Poor speech development • Neonatal hypotonia • Microcephaly • Bitemporal narrowing • Broad nasal tip • Low-set, rotated ears • Thin upper lip • Feeding difficulties	(Tham *et al.*, 2015)	Seizures treated with medication, which resolved the issue.	Exclusively dominant-negative MOZ protein translation due to the PTC residing within the terminal exon.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Neonatal respiratory distress • Seizures • Narrow palate • Low social interaction			
18	c.3850_3851dup p.Gln1285Glyfs*10	M	• Hypotonia • Global developmental delay • Feeding difficulties • Atrial septal defect • Poor speech development • Microcephaly • Bilateral ptosis • Low-set, rotated ears	(Murray *et al.*, 2017)	Patient not officially diagnosed with KAT6A GDD, although displays many clinical features of this syndrome.	Exclusively dominant-negative MOZ protein translation due to the PTC residing within the terminal exon.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Coarse facial features • Difficulty swallowing • Protruding tongue • Esotropia • Myopia • Broad nasal tip • Bitemporal narrowing			
18	c.3879dupA p.Glu1294Argfs*19	F	• Global developmental delay • Poor speech development • Neonatal hypotonia • Microcephaly • Craniosynostosis • Bitemporal narrowing	(Tham *et al.*, 2015)	Atrial septal defect repaired by surgery.	Exclusively dominant-negative MOZ protein translation due to the PTC residing within the terminal exon.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Broad nasal tip • Low-set, rotated ears • Thin upper lip • Atrial septal defect • Feeding difficulties • Clinodactyly • Widely spaced nipples • Social interaction difficulties			
18	c.4108G>T p.Glu1370*	F	• Global developmental delay • Poor speech development • Neonatal hypotonia	(Tham *et al.*, 2015)	Cardiac defect repaired by using a catheter.	Exclusively dominant-negative MOZ protein translation due to the PTC residing within the terminal exon.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Bitemporal narrowing • Broad nasal tip • Thin upper lip • Cardiac defect • Esotropia • Clinodactyly • Short attention span • Narrow palate • Tapered fingers			
18	c.4292dupT p.Leu1431Phefs*8	M	• Global developmental delay • Poor speech development • Neonatal hypotonia • Bitemporal narrowing	(Tham *et al.*, 2015)	Atrial septal defect repaired by surgery.	Exclusively dominant-negative MOZ protein translation due to the PTC residing within the terminal exon.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Broad nasal tip • Low-set rotated ears • Thin upper lip • Atrial septal defect • Widely spaced nipples • Abnormal teeth shape • Micrognathia • Small mouth • Transient neutropenia • Large anterior fontanelle			
18	c.4688_4689del p.Tyr1562*	F	• Severe, intermittent neutropenia	(Gauthier-Vasserot *et al.*, 2017)	Patient not officially diagnosed with KAT6A GDD,	Exclusively dominant-negative MOZ protein translation due to the

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Neonatal respiratory distress • Hypotonia • Feeding difficulties • Motor delays • Unusual withdrawn behaviours • Bitemporal narrowing • Broad nasal tip • Prominent forehead • Low-set ears • Narrow mouth • Brittle nails • Dental abnormalities • Slender extremities		although the authors acknowledge that this individual shares some features of the syndrome.	PTC residing within the terminal exon.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Joint hypermobility			
18	c.5924A>G p.Asn1975Ser	M	• Short stature • Mild global developmental delay • Hypotonia • Microcephaly • Broad nasal tip • Thin upper lip • Ophthalmologic abnormalities • Delayed bone age (noted at age 2.5 years)	(Trinh *et al.*, 2018)	Father of the patient below. Received treatment with growth hormones from age 9-11 years. Currently being treated with medication to control epilepsy.	Exclusively dominant-negative MOZ protein translation due to the PTC residing within the terminal exon.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Unilateral inguinal hernia (developed ~age 36 years) • Carpal tunnel syndrome (developed ~age 36 years) • Latent epilepsy (developed ~38 years)			
18	c.5924A>G p.Asn1975Ser	F	• Global developmental delay • Club feet • Short stature • Microcephaly	(Trinh *et al.*, 2018)	Daughter of the patient above. Her mother did not harbour a KAT6A variant and was phenotypically	Exclusively dominant-negative MOZ protein translation due to the PTC residing within the terminal exon.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Poor speech development • Hypotonia • Obstipation • Strabismus • Limb deformities • Intellectual disability		normal. Underwent several surgical procedures for club feet condition.	
18	p.Gln1995*	Unknown	Unknown	(Arboleda *et al.*, 2015)	Variant present in the Exome Aggregation Consortium (ExAC) browser, but no clinical information available.	Exclusively dominant-negative MOZ protein translation due to the PTC residing within the terminal exon.
MOZ (entire allele deleted)	8p11.21 interstitial deletion	M	• Global developmental delay	(Tham *et al.*, 2015)		Haploinsufficiency of MOZ protein.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Poor speech development • Microcephaly • Bitemporal narrowing • Broad nasal tip • Rotated ears • Thin upper lip • Hypermetropia • Esotropia • Dental crowding • Abnormal teeth shape			
MOZ (entire allele deleted)	8p11.2 interstitial deletion (8p11.1-p21.1)	M	• Global developmental delay • Growth retardation	(Lux *et al.*, 1990)		Haploinsufficiency of MOZ protein.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Bilateral undescended testes • Micropenis • Bilateral strabismus • Spherocytosis			
MOZ (entire allele deleted)	8p11.2 interstitial deletion (8p11-p21)	F	• Spherocytosis • Jaundice caused by gall stones • Short stature • Microcephaly • Micrognathia • Fusion of several vertebrae • Global developmental delay	(Cohen *et al.*, 1991)	Sister of the patient below. Deletion arose *de novo* in the mother. Underwent splenectomy and cholecystectomy.	Haploinsufficiency of MOZ protein.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• 'Bat ears' • No sexual development by the age of 29 • Developed diabetes • Intellectual disability			
MOZ (entire allele deleted)	8p11.2 interstitial deletion (8p11-p21)	M	• Spherocytosis • Short stature • Microcephaly • Micrognathia • Bilateral deafness • Global developmental delay	(Cohen *et al.*, 1991)	Brother of the patient above. Deletion arose *de novo* in the mother.	Haploinsufficiency of MOZ protein.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• No sexual development by the age of 18 • 'Bat ears'			
MOZ (entire allele deleted)	8p11.2 interstitial deletion (8p11.1-p21.1)	F	• Spherocytosis • Hyperbilirubinemia • Seizures • Psychomotor retardation • Unusual facial features • Nystagmus • Intermittent splenomegaly • Micrognathia	(Chilcote *et al.*, 1987)	Underwent splenectomy at the age of 4.	Haploinsufficiency of MOZ protein.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Short stature • Intellectual disability			
MOZ (entire allele deleted)	8p11.2 interstitial deletion (8p11.1-p21.1)	F	• Spherocytosis • Psychomotor retardation • Unusual facial features • Nystagmus • Intermittent splenomegaly • Micrognathia • Short stature • Intellectual disability	(Chilcote *et al.*, 1987)		Haploinsufficiency of MOZ protein.
MOZ (entire allele deleted)	8p11.2 interstitial deletion	M	• Growth retardation	(Kitatani *et al.*, 1988)		Haploinsufficiency of MOZ protein.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
	(8p11.1-p21.1)		• Psychomotor retardation • Microcephaly • Micropenis • Undescended testes • Hypoplastic nails • Spherocytosis • Mild jaundice			
MOZ (entire allele deleted)	8p11.2 interstitial deletion (8p11.2-p21)	M	• Cleft lip and cleft palate • Micropenis • Undescended testes	(Stratton *et al.*, 1992)	Operation to repair cleft lip was performed at 2 months of age. Treated with intramuscular injections of	Haploinsufficiency of MOZ protein.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
					testosterone, which increased penis length. Head was reported to be normal.	
MOZ (entire allele deleted)	8p11.2 interstitial deletion (8p11.21-p11.23)	F	• Global developmental delay • Facial dysmorphism • Low set ears • Thin lips • Micrognathia • Patent ductus arteriosus (CHD) • Depressed nasal tip • High forehead	(Tsukahara *et al.*, 1995)		Haploinsufficiency of MOZ protein.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Psychomotor retardation			
MOZ (entire allele deleted)	8p11.2 interstitial deletion (8p11.21-p12)	M	• Spherocytosis • Jaundice • Growth hormone deficiency • Growth retardation • No sexual development by the age of 20 • Hypogonadism • Micrognathia • Bilateral crumpled ears	(Vermeulen *et al.*, 2002)	Splenectomy performed at the age of 14 years. Treated with growth hormones at the age of 17 years. Diagnosed with Kallman syndrome.	Haploinsufficiency of MOZ protein.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Anosmia after childhood			
MOZ (entire allele deleted)	8p11.2 interstitial deletion (8p11.21-p12)	M	• Spherocytosis • Micropenis • Bilateral cryptorchidism • Right inguinal hernia • Strabismus • Cleft lip and cleft palate • Broad nasal root • Dental anomalies • Abnormal fingers • Patent ductus arteriosus (CHD)	(Melis *et al.*, 2005)	Diagnosed with Kallman syndrome.	Haploinsufficiency of MOZ protein.

MOZ Intron/Exon	Mutation	Gender	Phenotype	Publication	Additional Information	Pathogenesis Hypothesis
			• Hearing loss • Poor growth • Poor speech development • Mild mental retardation • Broad nasal root • Frequent respiratory infections			

Appendix II - Vectors & Plasmids

#	Vector/plasmid name	Gene of interest	Cloned using restriction sites	Oligonucleotides used	Source
1114	pSpCas9n(BB)-2A-GFP (pX461)	-	-	-	Addgene #48140
10068	pX461-*Homo sapiens* KAT6A gRNA E	KAT6A sgRNA E	BbsI (BpiI)	1433/1434	Jonathan Whitchurch
10069	pX461-*Homo sapiens* KAT6A gRNA F	KAT6A sgRNA F	BbsI (BpiI)	1435/1436	Jonathan Whitchurch
10110	pX461-*Homo sapiens* KAT6A sgRNA I	KAT6A sgRNA I	BbsI (BpiI)	1526/1527	Jonathan Whitchurch
10111	pX461-*Homo sapiens* KAT6A sgRNA J	KAT6A sgRNA J	BbsI (BpiI)	1528/1529	Jonathan Whitchurch
8845	YFP-FLAG-full length MOZ WT	KAT6A aa 1-2004 (Full length)	ApaI & XbaI	-	Sachko Matsuda
10117	YFP-FLAG-full length MOZ Arg1129* (c.3385C>T)	KAT6A aa 1-1128	N/A	1574/1575	Jonathan Whitchurch
10118	YFP-FLAG-full length MOZ Arg1024* (c.3070C>T)	KAT6A aa 1-1024	N/A	1572/1573	Jonathan Whitchurch

Appendix III - Oligonucleotides

#	Gene of interest	Region	Purpose	Sequence
1427	Human U6 Promoter	Fwd	Sequencing	ACTATCATATGCTTACCGTAAC
1432	Human KAT6A	Ser1039*	AA substitution via homology-directed repair using CRISPR-Cas9	GAGTCCGAAAGCGCAAACACCACAA TAGCAGTGTAGTCACAGAAACTATTT AGGAGAATTCTGAAGTGTTAGATGAA CCTTTTGAAGATTCTGACTCCGAGAG GCCAATGCCAAGATTAGAACCCACAT TTGAGATCGATGAAGAAGAGGA
1433	Human KAT6A	Guide E Top	CRISPR-Cas9-mediated genome editing	CACCGAGGTTCATCTAACACTTCAG
1434	Human KAT6A	Guide E Bottom	CRISPR-Cas9-mediated genome editing	AAACCTGAAGTGTTAGATGAACCTC
1435	Human KAT6A	Guide F Top	CRISPR-Cas9-mediated genome editing	CACCGTTGAAGATTCTGACTCCGAG

#	Gene of interest	Region	Purpose	Sequence
1436	Human KAT6A	Guide F Bottom	CRISPR-Cas9-mediated genome editing	AAACCTCGGAGTCAGAATCTTCAAC
1468	Human KAT6A	3' of intron 16-17 Fwd	Sequencing	AAAAAGCTTTCTCTATTAAACAGAAACC
1469	Human KAT6A	1091R	Sequencing	AAAAACCTGCGAAGACAAACGACGGAAG
1470	Human KAT6A	1018F	Sequencing	AAAAACTCCACCGAAGGAGGAGAGTCCG
1471	Human KAT6A	1081R	Sequencing	AAAAAGGGAAAAGTTCATTTTCATCCTC
1475	Human KAT6A	Ser1039 WT Fwd	Sequencing	CACAGAAACTATTTCTGAGACCA
1476	Human KAT6A	Ser1039* Fwd	Sequencing	CACAGAAACTATTTAGGAGAATT
1482	Human ANXA1	Fwd	qRT-PCR	AATCAGAAGCCCAAGTCTCCAC
1483	Human ANXA1	Rev	qRT-PCR	TTCACAGTTTGAACATATTCCTGC
1484	Human CALR	Fwd	qRT-PCR	GGGAACCCCCAGTGATTCAG
1485	Human CALR	Rev	qRT-PCR	GTGCCAGACTTGACCTGCC

#	Gene of interest	Region	Purpose	Sequence
1486	Human ETS2	Fwd	qRT-PCR	CTTTGTTAACCCCGTGCAGC
1487	Human ETS2	Rev	qRT-PCR	CCAGGGGTTCTTTGGAATGC
1488	Human CD55	Fwd	qRT-PCR	CTGGCGAGAAGGACTCAGTG
1489	Human CD55	Rev	qRT-PCR	CGCAGCTACGATTGCAGAAC
1506	Human KAT6A	Exon 17 Fwd L1045fs(11)*	qRT-PCR	CACTGAAGTGTTAGGAGAGG
1507	Human KAT6A	Exon 17 Fwd P1048_F1049del	qRT-PCR	CTGAAGTGTTAGATGAAGAAGATT
1508	Human KAT6A	Exon 17 Fwd WT	qRT-PCR	GATGAACCTTTTGAAGATTC
1509	Human KAT6A	Exon 17 & 18 Rev	qRT-PCR	AAGATAGGAGTGTCATCAGCATCA
1510	Human KAT6A	Exon 17 Rev L1045fs(11)*	qRT-PCR	GGCATTGGCCTCTCCTAAC
1511	Human KAT6A	Exon 17 Rev P1048_F1049del	qRT-PCR	CGGAGTCAGAATCTTCTTC
1512	Human KAT6A	Exon 17 Rev WT	qRT-PCR	GAGTCAGAATCTTCAAAAGG
1513	Human KAT6A	Intron 16-17 Fwd	qRT-PCR	GAAAAGATAGTGTTCATAACTG

#	Gene of interest	Region	Purpose	Sequence
1514	Human RHAG	Fwd	qRT-PCR	GGCTGTAGCAGGCATCTTGT
1515	Human RHAG	Rev	qRT-PCR	CATCCACAGAAAGAGAGTCCCAAT
1516	Human GDF15	Fwd	qRT-PCR	ATACTCACGCCAGAAGTGCGG
1517	Human GDF15	Rev	qRT-PCR	CTTGCAAGGCTGAGCTGACG
1520	Human HOXB2	Fwd	qRT-PCR	AGAAATCCGCCAAGAAACCCA
1521	Human HOXB2	Rev	qRT-PCR	CAGGCCATCTGCAGGC
1522	Human HOXB4	Fwd	qRT-PCR	ACTCCGCGTGCAAAGAGC
1523	Human HOXB4	Rev	qRT-PCR	ATTGGGGTTTACCGTGCTCA
1524	Human HOXD13	Fwd	qRT-PCR	CCCGGCTATATCGACATGGTG
1525	Human HOXD13	Rev	qRT-PCR	TAGAGCCACATCCCCTGGAAA
1526	Human KAT6A	Guide I Top	CRISPR-Cas9-mediated genome editing	CACCGTGGATGTATCTGGCTCAAG
1527	Human KAT6A	Guide I Bottom	CRISPR-Cas9-mediated genome editing	AAACCTTGAGCCAGATACATCCAC
1528	Human KAT6A	Guide J Top	CRISPR-Cas9-mediated genome editing	CACCGAAGAAAAAGGGATGGCCCAA

#	Gene of interest	Region	Purpose	Sequence
1529	Human KAT6A	Guide J Bottom	CRISPR-Cas9-mediated genome editing	AAACTTGGGCCATCCCTTTTTCTTC
1530	Human KAT6A	Intron 17-18 Fwd	Sequencing	CCTTCCTGTATAGACACTCC
1531	Human KAT6A	1126F	Sequencing	GTATCTCTTTTGCGAAAACG
1532	Human KAT6A	1177R	Sequencing	CCCGACTCAACTTAAATCC
1533	Human KAT6A	1168R	Sequencing	CGACCAGGTCTTTTCTTCC
1557	Human KAT6A	Exon 17 P1048_F1049del Fwd 2	qRT-PCR	ACCACTGAAGTGTTAGATGAAGAA
1558	Human KAT6A	Exon 17 R1022del Fwd	qRT-PCR	TCTCCACCGAAGGAGAGTC
1559	Human KAT6A	Exon 17 L1044_F1049del Fwd	qRT-PCR	GAGACCACTGAAGTGGAAG
1560	Human KAT6A	Exon 17 E1036fs(1)STOP Rev	qRT-PCR	AAAAGGTTCATCTAACACTCTG

#	Gene of interest	Region	Purpose	Sequence
1561	Human KAT6A	Exon 17 R1022del Rev	qRT-PCR	TTGCGCTTTCGGACTCTCC
1562	Human KAT6A	Exon 17 L1044_F1049del Rev	qRT-PCR	CGGAGTCAGAATCTTCCAC
1563	Human KAT6A	Exon 17 E1036fs(1)STOP Fwd	qRT-PCR	ATAGCAGTGTAGTCACAGAGTG
1570	Human KAT6A	Intron 17-18 Fwd	Sequencing	GTTCTGCCAATTCACTGTGTGAC
1571	Human KAT6A	1208R	Sequencing	CTTCACTCTCCTGGATCTTGGG
1572	Human KAT6A	c.3070C>T (p.Arg1024*) Fwd	Site-directed mutagenesis	CCGAAGGAGGAGAGTCTGAAAGCGC AAACACCA
1573	Human KAT6A	c.3070C>T (p.Arg1024*) Rev	Site-directed mutagenesis	TGGTGTTTGCGCTTTCAGACTCTCCT CCTTCGG
1574	Human KAT6A	c.3385C>T (p.Arg1129*) Fwd	Site-directed mutagenesis	GCCAGTATCTCTTTTGTGAAAACGTG ATGTGAA

#	Gene of interest	Region	Purpose	Sequence
1575	Human KAT6A	c.3385C>T (p.Arg1129*) Rev	Site-directed mutagenesis	TTCACATCACGTTTTCACAAAAGAGA TACTGGC
1578	Human KAT6A	Exons 17 & 18 Fwd	qRT-PCR	CAGATGATGCTGATGACACTCC
1579	Human KAT6A	Exon 18 T1142fs*(17) Rev	qRT-PCR	TGGATTGGTTTGCGGCTCTGT
1580	Human KAT6A	Exon 18 D1141_K1148del Rev	qRT-PCR	TGGGCCATCCCTTTTTCGGC
1581	Human KAT6A	Exon 18 T1142fs*(1) Rev	qRT-PCR	CCTTTTTCTTTTTCAAAGGGTA
1582	Human KAT6A	Exon 17 R1022del Fwd 2	qRT-PCR	TCTCCACCGAAGGAGAGT
1583	Human KAT6A	Exon 18 S1143fs*(4) Rev	qRT-PCR	GGAATTATCGAATGCAATCGG
1584	Human KAT6A	Exon 18 T1142fs*(12) Rev	qRT-PCR	TTTTCTTTTAGGCCATCCCTTG

#	Gene of interest	Region	Purpose	Sequence
1585	Human KAT6A	Exon 18 K1149fs*(5)	qRT-PCR	GATGTATTTTCTTTTTCAAAGGTGT
1613	Human KAT6A	Intron 17-18 Fwd	qRT-PCR	ATATGCTAGTCTTGGAAGAAGG
1592	Human SPARC	Fwd	qRT-PCR	GAAGAGTTTCACAAATCCTTCAAAAT
1593	Human SPARC	Rev	qRT-PCR	AGACTAAGACACATGCAAATCACC
1596	Human RUNX3	Fwd	qRT-PCR	GGCCTCTCCATGCCTTCT
1597	Human RUNX3	Rev	qRT-PCR	AGGAGGGAAGAAACTACAAGGAC
CM01	Human RUNX1	Fwd	qRT-PCR	CTCCCTGAACCACTCCACTG
CM02	Human RUNX1	Rev	qRT-PCR	TGGGGATGGTTGGATCTG
CM03	Human HHEX	Fwd	qRT-PCR	CGGACGGTGAACGACTACA
CM04	Human HHEX	Rev	qRT-PCR	AGAAGGGGCTCCAGAGTAGAG
CM05	Human PROX1	Fwd	qRT-PCR	CCCAGCAGCAGCAACAGA
CM06	Human PROX1	Rev	qRT-PCR	TGACCTTCCGACAGAGTCCT
CM07	Human SERPINE1	Fwd	qRT-PCR	GCTTCCCAGAAATTTGGGTTA
CM08	Human SERPINE1	Rev	qRT-PCR	TTAAAAGTCAAGGAAGATGAAGAGGT
CM11	Human GAPDH	Fwd	qRT-PCR	AGGTGAAGGTCGGAGTCAAC

#	Gene of interest	Region	Purpose	Sequence
CM12	Human GAPDH	Rev	qRT-PCR	GATGACAAGCTTCCCGTTCT
CM13	Human ID1	Fwd	qRT-PCR	CCAGAACCGCAAGGTGAG
CM14	Human ID1	Rev	qRT-PCR	GGTCCCTGATGTAGTCGATGA
F530	Human KAT6A	Fwd	qRT-PCR	ACAGAATCCTTCACCATGGTA
R713	Human KAT6A	Rev	qRT-PCR	CTTTAACACTCAACTCCAATTG
Le 851	Human KAT6A	Fwd	qRT-PCR	GGCAGTGCATCGAGTGTAAA
Le 852	Human KAT6A	Rev	qRT-PCR	CGGTCACATGAATCACAAAAG
CM19	Human HOXB3	Fwd	qRT-PCR	AAGGAGGGGAGATTTGTCG
CM20	Human HOXB3	Rev	qRT-PCR	CTGTTCCAAGCGGCTGAC
CM21	Human HOXB8	Fwd	qRT-PCR	AGCTCTTCCCCTGGATGC
CM22	Human HOXB8	Rev	qRT-PCR	ATAGGGATTAAATAGGAACTCCTTCTC
CM25	Human HDAC6	Fwd	qRT-PCR	AGTTCACCTTCGACCAGGAC
CM26	Human HDAC6	Rev	qRT-PCR	GCCAGAACCTACCCTGCTC
CM27	Human RPL7	Fwd	qRT-PCR	CTGGAACCATGGAGGGTGTA
CM28	Human RPL7	Rev	qRT-PCR	CGCTTTTTCTTAAGGGTTTCTG

#	Gene of interest	Region	Purpose	Sequence
CM29	Human HOXB9	Fwd	qRT-PCR	AAAGAGAGGCCGGATCAAAC
CM30	Human HOXB9	Rev	qRT-PCR	GTGTAGGGACAGCGCTTTTT
CM31	Human VGF	Fwd	qRT-PCR	GTCCCGATCTTCCCCTGT
CM32	Human VGF	Rev	qRT-PCR	CTGAGGGCTTTCATGACCA

Appendix IV – Antibodies

Antibody Name	Monoclonal/ polyclonal	Catalogue Number	Lot Number	Dilution used
Mouse anti-β-actin	Monoclonal	A5441 (Sigma-Aldrich)	11SM4835V	1/500
Mouse anti-MOZ (29D2)	Monoclonal	Generated by the DMH group (Bhattacharya 2013, PhD thesis - unpublished)	N/A	1/500
Chicken anti-mouse IgG-HRP	Polyclonal	sc-2954 (Santa Cruz Biotechnology)	K1113	1/2000

Appendix V - Cell culture media & serum

Medium/Serum Name	Catalogue Number	Lot/Batch Number	Date Used From
Dulbecco's Modified Eagle Medium (DMEM)	BE12-604F (Lonza Biowhittaker)	6MB071	August 2016
Roswell Park Memorial Institute (RPMI 1640)	BE12-167F (Lonza BioWhittaker)	6MB072	August 2016
Roswell Park Memorial Institute (RPMI 1640)	BE12-167F (Lonza BioWhittaker)	6MB135	October 2016
Roswell Park Memorial Institute (RPMI 1640)	BE12-167F (Lonza BioWhittaker)	6MB240	February 2017
Roswell Park Memorial Institute (RPMI 1640)	R0883-500ml (Sigma)	RNBG7483	June 2018
Dulbecco's Modified Eagle Medium (DMEM)	12604F (Lonza Biowhittaker)	0000704255	June 2018
Fetal bovine serum (FBS)	11573397 (Gibco)	42F1452K	May 2015
Fetal bovine serum (FBS, heat inactivated)	10500-064 (Gibco)	08G7280K	June 2018

Appendix VII – DEGs shared between individuals with KAT6A GDD, HEK293 MOZ KO and K562 MOZ siRNA cells

PATIENT X, PATIENT Y & HEK293 MOZ KO	PATIENT X & HEK293 MOZ KO	PATIENT Y & HEK293 MOZ KO	PATIENT X, PATIENT Y & K562 MOZ siRNA
EN1	ACADL	RPRM	PCSK9
HOXB9	EOMES	HPGD	PTGDS
HOXB5	EPHA6	HOXD4	IL16
HOXB8	CCND2	LMX1B	SLC6A15
RASGEF1C	MCOLN3	POU3F1	RNASE1
BCHE	NEBL	TENM1	NRROS
TBX1	NOV	SEMA3G	IL18R1
NPNT	LONRF3	GJB2	AUTS2
HLA-DRB1	HES2	NRXN2	METTL7A
LDLRAD4	NKD1	COL23A1	GPM6B
SLCO4C1	AK7	RORB	AMOT
KIAA1211	LSAMP	RYR2	TNK2
MPPED2	MAP7	LHX8	TCEAL4
FAM181B	SHC2	SLC26A7	MROH6
HLA-DOA	NOTUM	CRTAC1	ALDH3A2
HAGLR	RGS7	GPX3	BACH2
COBL	PLA2G4A	PCDH17	HMGCR
COL14A1	EBF3	RBP1	NPIPB5
TRIM58	FLCN	PPP1R14C	HOXC10
JAKMIP2	LINC01003	PRRG4	GAS6
JAM2	SECTM1	LINGO1	YJEFN3
CD24	NPC2	HES7	ABCA7
CDHR1	GREB1L	TMEM100	SEMA6C
NRN1	CDH4	INHBB	TYK2
FGF13	LRP1	ACVR1C	LENG8
HOXB3	ITPKB	TSLP	BAHCC1
PLXDC2	ATP8A1	TRIM47	CPZ
ZNF704	GLRB	FRMD4B	FHL1
GREB1	CARD6	APLN	IQGAP3
PKNOX2	POLR2A	HOXA2	HIST1H3J
PCDH19	MED12L	GAP43	HIST1H4E
TNFSF9	KCNS3	CRIP2	RPEL1
MAOA	CSF1	ZNF423	
DMKN	NAGLU	COL6A1	

MLXIPL	EYA2	MSX2	**PATIENT X & K562 MOZ siRNA**
LAMC3	HIST1H1C	HCN2	APOBR
ARL4C	BAG3	CCNJL	CPT1B
SYK	C7ORF43	HLX	LAT
CD70	MMP2	PTX3	ABCA2
RUNX3	SERPING1	GALNT14	RASSF5
ALPL	ZC3H4	TNFRSF21	TNKS1BP1
ALDH1A3	CASC10	CYBA	ZNF180
NPY1R	SH3BGR	IGFBP5	FADS1
BCL2	BAG1	SRGAP1	SERPINE1
VEGFB	OTUD1	IFIT1	C1ORF54
CACNA1G	MYCBP2	TBX2	HLA-B
RGCC	GPAA1	ALMS1	MET
LMO2	FOXO4	FAT1	DUSP5
EPDR1	LZTS1	TPD52L1	UBE2C
FASN	EIF4EBP1	RBM15	
GAS7	NFE2L3	UCP2	
IRS2	MIR210HG	NF1	**PATIENT Y & K562 MOZ siRNA**
CLCN4	LAPTM4B	NFAT5	
EGR1	NOTCH3	SLC6A17	
RGS16	NDUFA13	CDA	DLK1
UAP1L1	PCK2	CAT	CILP
INSIG1	MTHFD2	ADAMTS1	VANGL2
DSG2	FAM86HP	BIRC6	PAPLN
CDON	E2F7	EGFL7	KCNN4
HOXA3	CENPM	KCNQ5	ARC
SCAF1	GYPC	GATA2	ARHGAP42
SIK1	SORL1	ZNF490	NR4A1
COL2A1	UBE2T	LAMA3	ID3
NCOR2	ESPNL	ARHGEF16	UCHL1
CA8	SPC25	SLC1A1	USP53
HOXC4	TROAP	PCDH9	ID1
PPARGC1B	CXCL12	FBXL16	HHEX
BMP2	KIF22	IGF2BP1	GOLGA8B
KIAA1671	DENND2A	NR4A3	
NACAD	LMOD1	ARHGAP20	**PATIENT X, PATIENT Y, HEK293 MOZ KO & K562 MOZ siRNA**
MAF	PNPLA4	LLGL2	
HSD17B14	DDIT4	HOXA13	
CDKN2B	CHAC1	ELFN2	
ZFP36L2	NRGN	HOXC11	
SLIT3	TK1	HOTAIR	HOXB6
CFD	SPC24		LTBP4
RIPK4	FAM198A		ZDHHC8
C11ORF96	MGAT3		NEAT1

TIAM1	SLCO5A1	CDH2
ABI3BP	EPPK1	ZBED9
HCFC1	RPSAP58	HIST1H4C
SYP	MPZL3	
MBP	SLC2A4	
SETD1B	WNT10B	
ZC3H13	CDH3	
HOXB2		
ZNF628		
NAALAD2		
SOX4		
ZCCHC14		
SETD1A		
SKI		
PPDPF		
PRUNE2		
TGFB1		
INSC		
PTPRD		
EP400		
B4GALT6		
SLC35F2		
REPS2		
PDZRN3		
FBRS		
TLE3		
GLI3		
FTH1		
RCAN2		
IFIH1		
RNPEPL1		
YLPM1		
GALNT16		
IRF2BP2		
CREBBP		
VSTM4		
ZMIZ1		
BCL9L		
SPEN		
BCORL1		
MAP4K5		
GLDC		
KREMEN1		
ZYX		
NUDT8		
FOS		

HOXA5
SLTM
TMEM86A
TMEM171
CTSF
CREB3L1
SLC8A1
TRIM14
AGA
TRIB1
DCLK2
MOCOS
HCLS1
FKBP10
LINC00641
ILK
CPS1
FAM131B
ZNF185
IFITM1
CENPV
FUT1
ATP2B4
AKAP6
TUBA1B
MALAT1
CFI
RASGRP1
HSD17B6
NUP210
PCDH10
PDK1
RAP1GAP2
SCN9A
FAM105A
HID1
APOBEC3B
GALNT3
ADAM19
SYTL5
ANKRD18B
ZNF460
HHIP
MIF
EEF1A2
H19

SLC27A2
FOXE1
CTTNBP2
GAL
CGA
SH3GL2
EPHA5
STRA6

Appendix VIII - Professional Internship for PhD Students (Reflection)

Characterisation of MOZ expression in oral cancer tissues & novel small molecule inhibitors of its activity

A short reflection about my 3-month visit to the Transcription & Disease Laboratory of Professor Tapas K. Kundu at the Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR) in Bangalore, India. Required by the BBSRC DTP.

Background

Oral cancer has an unusually high prevalence in India due to the consumption of smokeless tobacco products and as this cancer is typically diagnosed at a late stage of development, the prognosis is often poor, posing a significant public health burden for the country. A high proportion of individuals who develop oral cancer are from the lower socioeconomic group of society, who often cannot afford the health costs necessary to treat this disease. Of interest is the KAT6A gene, which encodes a histone acetyltransferase that functions to switch on and off particular genes. Interestingly, increased KAT6A expression has been linked to the development and progression of several tumour types; it is an important biomarker in lung, breast and other solid tumours - these cancers are major health burdens both in India and the UK. The enzymatic function of KAT6A in histone acetylation make it an attractive target for targeted small molecule therapies, although at the current time of writing there is no effective inhibitor of this enzyme in the scientific literature.

I set out to determine whether the expression of the KAT6A protein (MOZ) is upregulated in oral cancer tumours, with a specific focus on the Indian population. These cancer tissues are not readily available to us in

our laboratory in the UK, neither is the expertise in working with them. Furthermore, if its expression is increased in this cancer type, it could be a potential target for small molecule inhibitors with therapeutic use in cancer treatment. To this end, another objective was to perform *in vitro* screens of purified KAT6A protein with novel compounds synthesised in the host laboratory and derived from garcinol, a known histone acetyltransferase inhibitor from the fruit rind of *Garcinia indica*. If successful, an inhibitor may be able to help improve the poor prognosis many people receive when they develop a cancer type in which KAT6A is the main gene driving its development and proliferation. This is part of our main, long-term objective and was the rationale for my visit to JNCASR.

Results: expression of KAT6A in oral cancer tumours

To assess the relative expression levels of the KAT6A protein in oral cancer tissues from Indian patients, tissue microarrays were generated from tissue biopsies and 5 µm sections were taken for immunohistochemical analysis. Antibodies specific for particular epigenetic marks of interest, i.e. histone H3K9ac, H3K14ac and H3K18ac, as well as the KAT6A protein were used to stain both normal and tumour tissues. It was found that KAT6A is overexpressed in some, but not all of the oral tumour samples tested, relative to normal tissue from the same patients. In addition, there was an overall increased presence of histone H3K9ac, H3K14ac and H3K18ac marks, all of which are associated with KAT6A activity. It would appear that these results indicate a correlated upregulation in KAT6A-associated epigenetic marks with increased KAT6A expression in oral cancer tumours.

Results: KAT6A inhibitor screening

To identify potential compounds from a small collection derived from the known non-specific histone acetyltransferase inhibitor garcinol, filter-binding assays were performed using 10 µM of each of 12 compounds. To acquire active KAT6A enzyme for these reactions, *Escherichia coli*

BL21 cells were transformed and induced to express the MYST domain of KAT6A (amino acids 510-810, the catalytic domain), with an N-terminal GST affinity purification tag. After optimisation of the recombinant protein expression conditions, purification of the GST-KAT6A protein was done and its ability to catalyse the acetylation of core histones was checked. The activity of the protein was normalised for each reaction and results of this small screen revealed two such compounds which reduced the catalytic activity of KAT6A to ~50%, namely SCJ-03 and SCTK-14.

In the host laboratory, the chemical structure of SCTK-14 had already been determined, so this compound was taken forward to test its specificity of inhibition. Filter-binding assays were repeated using SCTK-14 with normalised counts of several other histone acetyltransferases: p300, TIP60 and PCAF, as well as KAT6A. From these experiments, it appears that the compound SCTK-14 is most effective at inhibiting the activity of KAT6A (~50%); it reduced the activity of p300 by approximately 20%, while enhancing the activity of TIP60 by 20%. There was no impact upon PCAF. In conclusion, these results indicate that the compound SCTK-14 reduces the catalytic activity of KAT6A by 50% *in vitro,* and is a most potent inhibitor towards this histone acetyltransferase, compared to three similar proteins.

Personal experience

The experience was, at times challenging, particularly in the early stages of my visit. I found some of the administrative processes were laborious and acquiring particular reagents was not as easy as in the UK, due to delivery schedules. In addition, it took time for me to adapt to the cultural environment and the cuisine! That said, after about four weeks, I became used to the system arrangements and ways of working and things began to fall into place. My lab colleagues and host supervisor were welcoming and helped me with the necessary administrative arrangements. They also offered constructive criticism in lab meetings and presentations. I believe that my visit has allowed me to facilitate a positive working

relationship with these individuals; it has also laid the foundations for future visits from both institutions and I expect that the preliminary work from this study will form the basis for a continued collaboration.

I enjoy working in research and academia is the career sector in which I wish to remain after my PhD. Having spent some time working in a completely different laboratory environment in another country, will I believe add skills and qualities to my repertoire that will aid me in my career progression. I have gained new knowledge in the field of cancer biology and learnt laboratory techniques such as immunohistochemistry, as well as protein purification and the characterisation of novel compound activity towards proteins of interest, all of which are distinct from my own research project. I am very happy that I was given the opportunity to spend three months doing research abroad and would recommend other PhD students seize similar opportunities, as they are well worth the experience, both from a scientific and cultural perspective.