The role of chromatin remodeller ATRX in gene expression



Yuqi Shen

Lady Margaret Hall
MRC Weatherall Institute of Molecular Medicine
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2022

Abstract

ATRX is a chromatin remodeller involved in a wide range of nuclear processes, including transcription. Germline mutations in *ATRX* give rise to alpha-thalassaemia/intellectual disability X-linked (ATR-X) syndrome and downregulate alpha globin expression through a mechanism which is poorly understood. To understand how ATRX regulates gene expression, here I use alpha thalassaemia as an exemplar to study how ATRX loss downregulates alpha globin expression, which would provide insights into understanding the general gene regulation mechanism of ATRX. Specifically, I aim to recapitulate the alpha thalassaemia phenotype observed in ATR-X syndrome patients in an ATRX deficient erythroid cell system, and then to determine the molecular mechanisms that cause perturbed alpha globin expression.

To this end, an ATRX deficient CD34+ haematopoietic stem cell model was generated via CRISPR-Cas9. However, the cells did not show distinct alpha thalassaemia phenotype at the bulk level when differentiated into erythrocytes. Further gene expression analysis in single erythroid colony and single erythroid cell RNA-seq revealed that ATRX deficiency led to alpha thalassaemia in only a subset of cells. Differential gene expression analysis indicated the downregulation of alpha-like globin was likely associated with replicative stress/DNA damage, in particular the activated pATM-RNF168-ubiquitination pathway.

To validate the mechanisms, an erythroid progenitor cell line containing a dTAG-13 inducible ATRX degron system was generated, in which the alpha-like globin gene was tagged with a fluorescent reporter for the enrichment of the affected cell subpopulation. This cell model recapitulated the alpha thalassaemia phenotype and was used to investigate the function and gene regulation mechanism of ATRX. It has been shown that ATRX mainly regulates alpha globin expression at the basophilic erythroblast stage, and ATRX loss-induced alpha globin downregulation is not associated with DNA methylation.

As a prime candidate for causing replicative stress/DNA damage in the absence of ATRX, a putative G4 DNA-forming region ψζ VNTR in the alpha globin cluster was studied to explore its contribution to the gene regulation. It is demonstrated that ATRX binds to this region in a gene expression dependent manner. G4 stabilizer treatment downregulated α-globin gene expression and deletion of this region in erythroid cells abolished the effect of ATRX depletion on alpha globin expression, indicating ATRX regulates alpha globin expression through binding to the G-rich ψζ VNTR.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors Prof Richard Gibbons and Prof Douglas Higgs for all their support, encouragement and guidance throughout my Dphil journey. It has been a privilege to have them as my mentors. I would also like to thank Prof Irene Roberts and Prof Bethan Psaila on my thesis committee for the helpful discussions and support. This endeavour would not have been possible without the generous studentship supported by CAMS Oxford Institute and China Scholarship Council. I also thank my college Lady Margaret Hall for the support in the past four years.

I am grateful to the many people who have helped and supported me through this whole process, including all members of the Gibbons group and Higgs group, the core facility members in the WIMM and other postdocs and students. In particular, thanks go to Chris Fisher and Barbara Xella, who have always been supportive both in my project and when I felt lost and down. I would like to extend my sincere thanks to Caroline Scott and Mohsin Badat for guiding me through many experiments when I initially started in the lab. I am indebted to Julia Truch, Aude-Anaïs Olijnik, Robert Beagrie, Yavor Bozhilov, Anna Rose, Tomas Goncalves and Rebekah Tillotson for their help and support. I also acknowledge the technical advice from Deena Iskander, Dominic Owens, Alastair Smith, Jackie Sloane-Stanley and Joseph Blayney. I feel lucky and thankful to have received much assistance from the wonderful core facilities within the WIMM: Flow Cytometry Facility, Genome Engineering Facility, Single Cell Facility and the CyTOF Facility. Without them, the work in this thesis could not have been completed. I am also grateful to the Oxford Biomedical Data Science training programme for helping me gain coding skills. Thanks should also go to Sean Wen and Liezel Tamon for their help in data analysis as well as their supports as friends. For their guidance and suggestions in my project, I would like to thank a number of other group leaders: David Clynes, Anindita Roy, Philip Hublitz and Adam Mead.

I could not have undertaken this journey without the support and love I have received from many of my friends and my family. Special appreciation to my dearest friend Siyu Liu who has been by my side for the past thirteen years both in work and in life and has always been understanding and supportive. Lastly, my heartfelt gratitude goes to my husband Xu Liu for his unfailing caring and love and being my happy place.

List of Abbreviations

A	Adenine
AAVS1	Adeno-associated virus integration site 1
ADD	ATRX-DNMT3-DNMT3L
ALT	Alternative lengthening of telomere
APC	Allophycocyanin
APC-Cy7	Allophycocyanin-cyanine7
ATAC	Assay for transposase accessible chromatin
ATM	Ataxia-telangiectasia mutated kinase
ATMDS	Alpha thalassaemia myelodysplastic syndrome
ATR-X	Alpha thalassaemia/intellectual disability syndrome X-linked
BFU-E	Burst forming unit-erythroid
bp	Base pair
BSA	Bovine serum albumin
BV	Brilliant violet
C	Cytosine
Cas	CRISPR associated
CD	Cluster of differentiation
cDNA	Complimentary DNA
CDS	Coding sequences
CFU	Colony forming unit
CFU-E	Colony forming unit-erythroid
CGIs	CpG islands
ChIP	Chromatin immunoprecipitation
CO_2	Carbon dioxide
CpG	Cytosine guanine dinucleotide
CRISPR	Clustered regularly interspaced short palindromic repeat
CTD	C-teminal domain
CyTOF	Cytometry by time of flight mass spectrometry
DAXX	Death domain-associated protein
DBA	Diamond-Blackfan anemia

DD	Destabilizing domain
ddPCR	Droplet digital PCR
DDR	DNA damage response
DEG	Differentially expressed gene
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleotide acid
DNMTs	DNA methyltransferases
dNTP	2'-deoxynucleotides 5'-triphosphate
DSB	Double stranded break
dTAG	Degradation tag
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced green fluorescence protein
EGS	Ethylene glycol bis (succinimidyl succinate)
Er	Erbium
EPO	Erythropoietin
EZH2	Enhancer of zeste homolog 2
FACS	Fluorescent activated cell sorting
FACT	Facilitates chromatin transcription
FBS	Foetal blood serum
FC	Fold change
FKBP12	12-kDa FK506-binding protein
Flt3L	Fms related receptor tyrosine kinase 3 ligand
FITC	Fluorescein isothiocyanate
FMO	Fluorescence-minus-one
FP	Fluorescent protein
FSC	Forward scatter
g	Gram
G	Guanine
G4	G-quadruplex
GC	Guanine-cytosine
Gd	Gadolinium
GFP	Green fluorescent protein
GO	Gene ontology

GPA	Glycophorin A
gRNA	Guide RNA
GSEA	Gene set enrichment analysis
H3K4me2	Histone H3 lysine 4 dimethylation
H3K4me3	Histone H3 lysine 4 trimethylation
H3K9me3	Histone H3 lysine 9 trimethylation
H3K27me3	Histone H3 lysine 27 trimethylation
Hb	Haemoglobin
HbA	Haemoglobin A (adult haemoglobin)
HBA	Alpha globin
HBB	Beta globin
HbF	Haemoglobin F (fetal haemoglobin)
HBG	Gamma globin
HbH	Haemoglobin H
HDAC	Histone deacetylases
HDR	Homology directed repair
Het	Heterozygous
HMT	Histone methyltransferases
Hom	Homozygous
HP1	Heterochromatin protein 1
HR	Homologous recombination
HSC	Haematopoietic stem cell
HUDEP	Human umbilical cord blood derived erythroid progenitor
HVG	Highly variable genes
IEF	Isoelectric focusing
IgA	Immunoglobulin A
IgG	Immunoglobulin G
IL3	Interleukin-3
IMDM	Iscove's Modified Dulbecco's Medium
iPSC	Induced pluripotent stem cell
kb	Kilo base
kDa	Kilodalton
KO	Knock-out

L	Litre
LCL	Lymphoblastoid cell line
lncRNA	Long non-coding RNA
M	Molar
mA	Milliamp
MACS	Magnetic activated cell sorting
Mb	Megabase
MCM	Minichromosome maintenance complex
MCS	Multi-species conserved sequences
MECP2	Methyl CpG binding protein 2
MEP	Megakaryocyte–erythroid progenitor
mESC	Murine embryonic stem cell
mg	Milligram
$MgCl_2$	Magnesium chloride
MGI	Mouse Genome Informatics
min	Minute
ml	Millilitre
mM	Millimolar
mNPCs	Murine neuroepithelial progenitors
MRN	MRE11-RAD50-NBS1
mRNA	Messenger RNA
n.s.	Not significant
ncRNA	Non-coding RNA
Nd	Neodymium
NEB	New England Biolabs
ng	Nanogram
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
nM	Nanomolar
NPRL3	NPR3 like, GATOR1 complex subunit
OMIM	Online Mendelian Inheritance in Man
PAM	Protospacer adjacent motif
PBMC	Peripheral blood mononuclear cell

PBS	Phosphate buffered saline
PCA	Principal component analysis
PCH	Pericentromeric heterochromatin
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDS	Pyridostatin
PE	Phycoerythrin
PE-Cy7	Phycoerythrin-cyanine7
PerCP/Cy5.5	Peridinin-chlorophyll-protein complex: cyanine 5.5
PIC	Protease inhibitor cocktail
PML-NB	Promyelocytic leukemia nuclear bodies
Pol II	RNA polymerase II
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
P/S	Penicillin/streptomycin
PTM	Post-translational modification
PVDF	Polyvinylidene difluoride
RBC	Red blood cell
rDNA	Ribosomal DNA
RNA	Ribonucleic acid
RNAP	RNA polymerase
RNP	Ribonuclear protein
RPKM	Reads per kilobase per million mapped
RPL13A	Ribosomal protein L13a
RPS18	Ribosomal protein S18
rRNA	Ribosomal RNA
RS-1	RAD51-stimulatory compound 1
RT	Reverse transcription
RT-qPCR	Real-time quantitative polymerase chain reaction
SCF	Stem cell factor
SD	Standard deviation
SDS	Sodium dodecyl sulfate
seq	Sequencing

sgRNA	Single guide RNA
SNP	Single nucleotide polymorphism
SpCas9	Streptococcus pyogenes Cas9
SSC	Side scatter
ssODN	Single-stranded oligo DNA nucleotides
STR	Short tandem repeat
SWI/SNF	SWItch/sucrose non-fermentable
T	Thymine
TAD	Topologically associated domain
TALENs	Transcription activator like effector nucleases
TBS-T	Tris buffered saline with 0.1% Tween-20
TCEP	Tris (2-carboxyethyl) phosphine hydrochloride
TE	Tris-EDTA
TF	Transcription factor
TFBS	Transcription factor binding site
TPO	Thrombopoietin
TR	Tandem repeat
t-SNE	T-distributed stochastic neighbor embedding
TSO	Template-switching oligo
TSS	Transcription start site
µ	Micro
U	Uracil
UCSC	University of California Santa Cruz
UTR	Untranslated region
VNTR	Variable number of tandem repeat
WIMM	Weatherall Institute of Molecular Medicine
WT	Wild-type
XCI	X chromosome inactivation
Xist	X-inactive specific transcript
ZFNs	Zinc finger nucleases
°C	Degrees Celsius

Table of Contents

Abstract i
Acknowledgements ii
List of Abbreviations iii
List of Figures xiv
List of Tables xvii

1 Introduction 1

1.1 Alpha thalassaemia / intellectual disability X-linked syndrome (ATR-X syndrome) 2

1.2 Functions of ATRX 3

1.2.1 ATRX gene and protein 4

1.2.2 Domains in full-length ATRX protein 4

1.2.3 ATRX localisation and its functions 7

1.2.4 ATRX interactors and functions 15

1.2.5 Roles of ATRX in DNA replication, DNA damage response and DNA repair 23

1.3 Consequences of ATRX mutations 27

1.3.1 ATRX mutations cause impaired development by inducing DNA damage and cell apoptosis 28

1.3.2 ATRX mutations give rise to cognitive and behavioural deficits 30

1.3.3 ATRX mutations trigger cancers via perturbing genome stability and gene expression 32

1.4 Human α-globin genes and α-thalassaemia 34

1.4.1 Human α-globin gene cluster and regulation 34

1.4.2 α-thalassaemia 36

1.5 Potential mechanisms of α-thalassaemia phenotype in ATR-X syndrome 38

1.5.1 Replication stress induced epigenetic instability 40

1.5.2 DNA double-strand break induced transcription silencing....42
1.5.3 ψζ VNTR functions as an enhancer blocker boundary....42
1.6 Human cellular models used for studying erythropoiesis....43
1.7 Genome engineering technologies....46
1.8 Summary and outline of the thesis....48
2 Materials and Methods....51
2.1 Cell culture and related techniques....51
2.1.1 Isolation of primary CD34+ haematopoietic stem/progenitor cells....51
2.1.2 Maintenance of primary CD34+ haematopoietic stem/progenitor cell....52
2.1.3 Erythroid differentiation of HSCs....52
2.1.4 Colony-forming unit (CFU) assays....53
2.1.5 Expansion of HUDEP-2 cells....54
2.1.6 Erythroid differentiation of HUDEP-2 cells....54
2.2 Cell biology techniques....55
2.2.1 Flow cytometry of erythroid cell surface marker staining in live cells....55
2.2.2 Flow cytometry of intracellular staining in fixed cells....56
2.2.3 FACS sorting....58
2.2.4 Mass cytometry (CyTOF)....60
2.2.5 Cellular morphological analysis – Wright's stain of cytospin....62
2.3 Genome engineering....62
2.3.1 Knocking out ATRX with RNP in CD34+ cells....62
2.3.2 Genome engineering in HUDEP-2 cells....65
2.4 Molecular biology techniques....71
2.4.1 DNA extraction, general PCR and Sanger sequencing....71
2.4.2 RNA extraction, cDNA synthesis and qRT-PCR....71
2.4.3 Droplet digital PCR....72
2.4.4 Genotype analysis of single BFU-E colony....73

2.4.5 High-throughput gene expression profiling on Fluidigm Biomark dynamic arrays 74
2.4.6 Targeted bisulfite sequencing for DNA methylation detection....75
2.4.7 Protein extraction and western blot....76
2.4.8 Iso-electric focusing for haemoglobin analysis....77
2.4.9 Bionano optical genome mapping....77
2.5 Next generation sequencing related techniques....78
2.5.1 Smart-seq2 single cell RNA-seq....78
2.5.2 Chromatin Immunoprecipitation followed by next generation sequencing (ChIP-seq)....79
2.6 Computational analysis....81
2.6.1 Analysis of Smart-seq2 single cell RNA-seq data....81
2.6.2 Analysis of ChIP-seq data....84
3 Generation of ATRX deficient haematopoietic stem cells and phenotype assessments of the differentiated erythroid cells....85
3.1 Introduction....85
3.2 Results....87
3.2.1 *In vitro* erythroid differentiation and characterisation of CD34+ HSC cells ...87
3.2.2 Knocking out ATRX in CD34+ HSCs via CRISPR/Cas9....91
3.2.3 Phenotype assessment of ATRX KO erythroid cells at a bulk level....98
3.2.4 Phenotype assessment in single ATRX KO erythroid burst-forming unit (BFU-E)....104
3.3 Discussion....115
4 Gene expression analysis of ATRX deficient erythroid cells at the single cell level....120
4.1 Introduction....120
4.2 Results....122

4.2.1 Knock out efficiency and gene expression analysis in bulk cells used for scRNA-seq 122
4.2.2 Smart-seq2 single cell sequencing data processing and visualisation....125
4.2.3 Differentially expressed gene (DEG) analysis between AAVS1 and ATRX KO cells 127
4.2.4 Determination of α-thalassaemia phenotype in single cells 138
4.2.5 Cell cycle analysis in single cells....140
4.2.6 DEG analysis between affected and unaffected cells 141
4.3 Discussion 146

5 Generation of ATRX degron and HBM-mKate2 reporter cell line and recapitulation of α-thalassaemia phenotype 154

5.1 Introduction 154
5.2 Results 158
5.2.1 Generation and characterisation of ATRX degron cell line 158
5.2.2 Recapitulating α-thalassaemia phenotype in ATRX degron cells 169
5.2.3 Generation and characterisation of HBM-mKate2 reporter cell line 180
5.2.4 FACS-based enrichment of affected cells under ATRX depletion....187
5.2.5 ATRX mainly regulates gene expression at the basophilic erythroblast stage during erythroid differentiation 190
5.3 Discussion 192

6 Investigating the role of ψζ VNTR in ATRX regulated α-globin expression 197

6.1 Introduction 197
6.2 Results 199
6.2.1 ATRX binds to the ψζ VNTR in a gene expression dependent manner....199
6.2.2 G-quadruplex stabilization by G4 ligand disturbs HBM expression....201

6.2.3 HBM expression is no longer perturbed by ATRX loss when ψζ VNTR is knocked out 207

6.3 Discussion 213

7 Conclusions and general discussion 219

7.1 Summary of findings 219

7.2 Wider implications 220

7.2.1 ATRX functions to protect cells from G4-induced replicative stress 221

7.2.2 ATRX loss-induced replicative stress activates ATM-RNF168-ubiquitination DNA damage response pathway 223

7.2.3 Replication stress and DNA damage are associated with perturbed gene expression 225

7.2.4 Working model of ATRX regulation of gene expression 227

7.3 Future directions 229

Reference 231

Appendix 264

List of Figures

Figure 1.1 Schematic diagram of ATRX protein showing its motifs/domains.....5
Figure 1.2 Schematic representation of human α-globin locus on chromosome 16 (16p13.3).....35
Figure 1.3 Downregulated gene expression at human α-globin locus in ATR-X syndrome. 39
Figure 3.1 Schematic of experimental approach for a three-phase HSCs erythroid differentiation protocol.....88
Figure 3.2 Characterisation of wildtype CD34+ HSCs erythroid differentiation. 90
Figure 3.3 Workflow of establishing ATRX knockout erythroid cells from CD34+ HSCs.....91
Figure 3.4 Gender determination of the collected donors. 93
Figure 3.5 Schematic of the locations of sgRNAs within *ATRX.* 94
Figure 3.6 Knockout effect of sg9-1 on ATRX.....97
Figure 3.7 Erythroid phenotype analysis in ATRX knockout cells. 100
Figure 3.8 Globin RNA expression analysis in ATRX knockout cells.....102
Figure 3.9 Globin protein expression analysis in ATRX knockout cells. 103
Figure 3.10 Determination of cell plating density in CFU assay.....105
Figure 3.11 Workflow of phenotype analysis of single ATRX KO erythroid burst-forming units.....106
Figure 3.12 FACS sorting of progenitors for CFU assay.....108
Figure 3.13 Clonogenic capacity of the sorted progenitors. 110
Figure 3.14 Genotype determination of BFU-E colonies and their morphology. 112
Figure 3.15 Gene expression analysis in BFU-E colonies. 114
Figure 4.1 Analysis of bulk cells used for scRNA-seq.....124
Figure 4.2 scRNA-seq data processing and visualisation.....126
Figure 4.3 Differentially expressed gene analysis of scRNA-seq data. 128

Figure 4.4 DNA damage response pathway is upregulated in ATRX KO erythroid cells. 132
Figure 4.5 Expression of genes involved in cell cycle / cell division and ATRX binding signal at them in erythroid cells. 134
Figure 4.6 Expression of erythrocyte related genes and ATRX binding at them in erythroid and non-erythroid cells. 137
Figure 4.7 Globin gene expression in AAVS1 and ATRX KO cells at the single cell level. 139
Figure 4.8 Cell cycle analysis of AAVS1 and ATRX KO cells at the single cell level. 141
Figure 4.9 Single cell re-classification based on their HBM expression. 143
Figure 4.10 Workflow of DEG analysis between affected and unaffected cells. 144
Figure 4.11 GSEA gene ontology enrichment analysis between affected and unaffected cells. 145
Figure 5.1 Schematic for ATRX targeting design to generate ATRX degron cell model in HUDEP-2 cells. 159
Figure 5.2 Workflow for ATRX degron targeting in HUDEP-2 cells. 161
Figure 5.3 Screening of ATRX degron clone. 163
Figure 5.4 Genomic validation of the ATRX degron cell line. 164
Figure 5.5 Validation of ATRX-EGFP-FKBP12 fusion protein expression in the ATRX degron cell line. 165
Figure 5.6 dTAG-13 induced protein degradation in the ATRX degron cell line. 168
Figure 5.7 ATRX degradation outcome induced by different concentrations of dTAG-13 treatment in ATRX degron cell line. 170
Figure 5.8 dTAG-13 treatment has no significant effect on the erythroid differentiation in HUDEP-2 WT and ATRX degron cell line. 173
Figure 5.9 Effect of dTAG-13 treatment on gene expression during erythroid differentiation. 176

Figure 5.10 DNA methylation analysis at *HBM* promoter by targeted bisulfite sequencing. 179
Figure 5.11 Schematic for HBM-mKate2 tagging design. 181
Figure 5.12 Screening of HBM-mKate2 tagged cell clones. 183
Figure 5.13 Genomic validation of the HBM-mKate2 tagged cell line. 184
Figure 5.14 Normal erythroid differentiation of ATRX degron-mKate2 tagged cell line. 185
Figure 5.15 ATRX and mKate2 expression in ATRX degron-mKate2 tagged cell line during erythroid differentiation. 186
Figure 5.16 FACS-based enrichment of affected cells in ATRX degron-mKate2 tagged cell line. 189
Figure 5.17 Determining the critical phase of ATRX regulation during erythroid differentiation in the ATRX degron-mKate2 tagged cell line. 191
Figure 6.1 ATRX binds to the ψζ VNTR in a gene expression dependent manner. 200
Figure 6.2 Determination of the optimal treatment concentration of G4 stabilizers. 203
Figure 6.3 Effect of G4 stabilizer treatment on erythroid differentiation. 204
Figure 6.4 Gene expression analysis in G4 stabilizer treated cells. 206
Figure 6.5 Generation of ψζ VNTR knockout cell line. 210
Figure 6.6 Normal erythroid differentiation of ATRX degron-ψζ VNTR KO cell line. 212
Figure 6.7 HBM expression is no longer affected by ATRX loss in ψζ VNTR KO cells. 213
Figure 7.1 Working model for ATRX regulation of gene expression. 228

List of Tables

Table 1.1 Summary of ATRX functions. 27
Table 2.1 Antibodies of erythroid surface markers used for FACS erythroid differentiation detection. 56
Table 2.2 Antibodies used for FACS intracellular globin staining. 57
Table 2.3 Antibodies used for haematopoietic progenitors sorting. 59
Table 2.4 Antibodies used for mass cytometry. 61
Table 2.5 Guide RNA sequences used for knocking out ATRX in CD34+ cells. 63
Table 2.6 PCR primers used for genotyping ATRX KO in CD34+ cells. 65
Table 2.7 Guide RNA sequences used for generation of ATRX degron HUDEP-2 cell line. 66
Table 2.8 Guide RNA sequences used for generation of HBM mKate2 reporter cell line. 67
Table 2.9 PCR primers used for screening edited HUDEP-2 cell clones. 70
Table 2.10 Taqman assays used for 192.24 IFC chip Biomark gene expression analysis. 75
Table 3.1 Knockout efficiency of sgRNAs targeting ATRX, ranked by KO score. 95
Table 3.2 Knockout efficiency of sg9-1 in three independent knockout experiments in CD34+ cells from three donors. 98
Table 4.1 Significantly differentially expressed genes in scRNA-seq between ATRX KO and AAVS1 cells. 130
Table 4.2 Cell cycle and cell division related DE genes and their functions in cell cycle and cell division. 133
Table 4.3 Erythrocyte related DE genes and their functions in erythropoiesis. 136
Table 4.4 Cell numbers in each group after reclassification based on the level of HBM expression. 143
Appendix Table Dysregulated pathways in the ATRX KO HBM^{low} cells 264

1 Introduction

Many human genetic diseases occur when the function of a single protein is impaired by a genetic defect in its encoding gene or its cis-regulatory elements. Harmful mutations within these DNA regions can lead to disorders by either generating insufficient amount of the functional product or affecting the normal function of its coding product. One example of such a disease is alpha thalassaemia (α-thalassaemia), one of the most common and well-studied human molecular diseases, which most commonly results from an inherited defect within the alpha globin genes or their regulatory elements, thus leading to reduced production of functional alpha globin.

In addition to direct genetic changes in the DNA sequence at the gene locus, aberrant chromatin regulation can also lead to disorders at the epigenetic level which alters the pattern of gene expression. This is often observed when there is a defect in genes encoding proteins that regulate DNA and histone modifications that shape the chromatin architecture. In many cases, these defects result in neurodevelopmental disorders and intellectual disability and are frequently linked to physical and developmental abnormalities. Alpha thalassaemia / intellectual disability X-linked syndrome (ATR-X syndrome) is one such disorder, which is caused by mutations in the chromatin remodelling factor ATRX.

Understanding the mechanisms underlying these conditions can elucidate many of the principles by which gene mutations cause human genetic diseases, which

may contribute further to our understanding of the general molecular mechanisms underlying many other orphan and common diseases.

1.1 Alpha thalassaemia / intellectual disability X-linked syndrome (ATR-X syndrome)

ATR-X syndrome (OMIM #301040) is a severe X-linked intellectual disability syndrome associated with profound developmental delay, facial dysmorphism, genital abnormalities and α-thalassaemia. It is a rare genetic disorder with an estimated prevalence of 1/30,000-1/40,000 neonate boys (Wada et al., 2013). So far, more than 200 cases have been reported. The patients usually have very limited expressive language, generalised hypotonia in their early childhood and miscellaneous abnormalities including facial, skeletal and genital abnormalities as reviewed in (R. Gibbons, 2006). The distinctive facial features include microcephaly, hypertelorism, epicanthic folds, flat nasal bridge, a small triangular upturned nose, and a hypotonic flat midface (R. Gibbons, 2006). In ATR-X syndrome, about 75% of cases have α-thalassaemia, however, this does not require treatment since it is mild. The treatment of ATR-X syndrome is directed toward the specific symptoms that are apparent in each individual, which may require the coordinated efforts of a team of specialists.

ATR-X syndrome was first reported in a group of patients who had an intact alpha globin (α-globin) locus on chromosome 16p but developed α-thalassaemia with variable degrees of haemoglobin H (HbH) which is tetramer of beta globin as a result of reduced synthesis of α-globin and which is most commonly manifested as HbH inclusions in red blood cells stained with the supravital dye brilliant cresyl

blue (Weatherall et al., 1981; A. O. Wilkie et al., 1990; Donnai et al., 1991). The degree of red blood cell hypochromia and Hb H levels are milder than usually found in α-thalassaemia so the manifestation of α-thalassaemia in ATR-X syndrome may be missed without thorough examination and in 25% cases α-thalassaemia is absent (R. J. Gibbons, Brueton, et al., 1995; Lossi et al., 1999; R. J. Gibbons & Higgs, 2000). Family studies and linkage analysis confirmed this disease is X-linked recessively inherited and mapped the candidate gene to Xq13.1-q21.1 (Porteous & Burn, 1990; A. O. M. Wilkie et al., 1991; Cole et al., 1991; Donnai et al., 1991; R. J. Gibbons et al., 1991, 1992; R. J. Gibbons, Brueton, et al., 1995; R. J. Gibbons, Picketts, et al., 1995). Consistent with its X-linked recessive inheritance pattern, ATR-X syndrome predominantly affects males, sparing of most female carriers is associated with extreme skewing of X inactivation (R. J. Gibbons et al., 1992) although very rare female cases have been reported in whom the active X chromosome harboured the *ATRX* mutation (Wada et al., 2005; Badens et al., 2006). The causative gene *ATRX* (known as *XH2* originally) on Xq21.1 was identified via hybridization to cDNA from ATR-X syndrome patients and confirmed in other affected individuals (R. J. Gibbons, Picketts, et al., 1995), and it was speculated to encode a trans-acting regulator of α-globin.

1.2 Functions of ATRX

Various studies have attempted to elucidate the function of ATRX. The evidence deciphering the role of ATRX mainly comes from its gene sequence and

predicted protein structure, the effect of its mutations, its localisation in the nucleus / distribution in the genome and the proteins with which it interacts.

1.2.1 ATRX gene and protein

ATRX is highly conserved between human and mouse, with 85% overall identity (Picketts et al., 1998). Its homolog has also been found in *Drosophila* (*dAdd1* and *dATRX/dXNP*; 66% homology) and *C. elegans* (*XNP-1*; 52% homology) (Villard et al., 1999; N. G. Lee et al., 2007). In human, *ATRX* on Xq21.1 spans around 300 kb and comprises 35 exons, coding for 2,492 amino acids (NCBI reference sequence NG_008838.3 and NP_000480.3) producing a 282 kDa full-length protein. An alternative splicing event involving exon 6 exists, producing a smaller product at 278 kDa (Villard et al., 1997), and this alternative splicing is also conserved in mice (Picketts et al., 1998). In addition, there is a truncated protein isoform (ATRXt, Figure 1.1), with a smaller molecular mass (~200 kDa), which is produced due to the failure to splice intron 11 and the use of a proximal intronic poly(A) signal (Garrick et al., 2004).

1.2.2 Domains in full-length ATRX protein

The full-length ATRX protein contains several domains, but two highly conserved domains are essential for its fundamental function - ADD domain and ATPase / helicase domain (Figure 1.1). The majority of mutations giving rise to ATR-X syndrome are missense mutations and fall within these two regions and mostly result in greatly reduced levels of ATRX protein in ATR-X patients, most probably

due to reduced protein stability (Argentaro et al., 2007; R. J. Gibbons et al., 2008; Mitson et al., 2011).

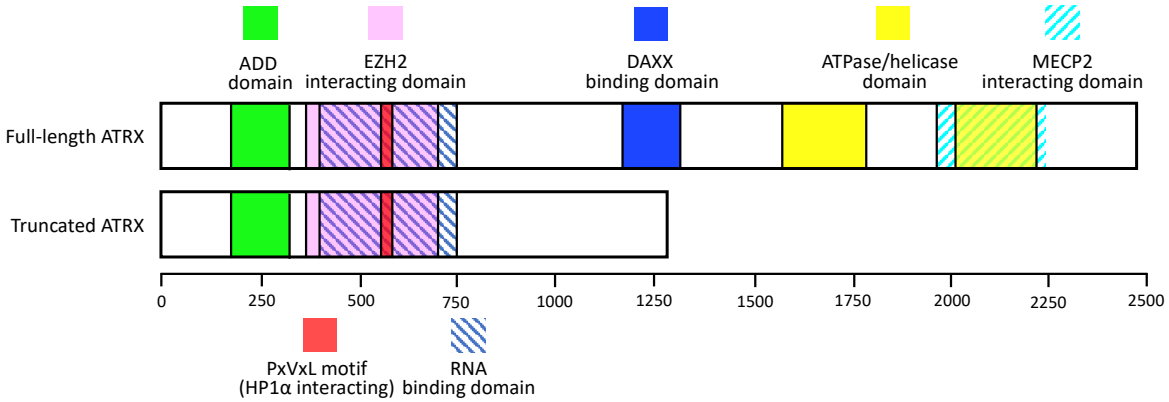


Figure 1.1 Schematic diagram of ATRX protein showing its motifs/domains.

Motifs/domains are annotated with names, along with squares in corresponding colors/patterns.

At the N-terminal is an ATRX-DNMT3-DNMT3L (ADD) domain, which is shared with methyltransferases DNMT3A, DNMT3B, and DNMT3L, all responsible for DNA methylation. This domain is cysteine-rich and contains two different types of zinc-finger motifs at its N-terminal - GATA-like zinc finger and plant homeodomain like domain, allowing access to chromatin, so it is generally thought to determine the genomic localization of ATRX (Bienz, 2006; Argentaro et al., 2007). The C-terminus of the ADD domain is α-helix which can interact with the N-terminal GATA-like zinc finger motif to bring both ends of the ADD domain together to form a globular structure (Argentaro et al., 2007). The ADD domain has been shown as a motif for ATRX binding to histone H3 through lysine 9 trimethylation (H3K9me3) recognition at pericentromeric heterochromatin

(PCH) region, but this binding can be inhibited by modified lysine 4 like dimethylation and trimethylation (H3K4me2, H3K4me3) (Dhayalan et al., 2011; Iwase et al., 2011; Eustermann et al., 2011). ATRX mutants with mutations in its ADD domain would fail in both H3K9me3 interaction and localisation at pericentromeric heterochromatin (Dhayalan et al., 2011; Iwase et al., 2011; Eustermann et al., 2011). Interestingly, this ADD domain has been found in the orthologues of ATRX in vertebrates and plants but is absent in its orthologues in *Drosophila* and *C. elegans* where DNA methylation is absent or rare (Argentaro et al., 2007; Duc et al., 2017).

The C-terminus contains an ATPase/helicase domain that provides DNA-dependent ATPase activity allowing the protein to act as a DNA translocase (Xue et al., 2003). ATRX belongs to the chromatin remodeler SWI2/SNF2 family, members of which are involved in a wide variety of cellular functions including the regulation of transcription, control of the cell cycle, DNA repair and mitotic chromosome segregation by remodeling chromatin, so this domain is considered to play a critical functional role in ATRX (R. J. Gibbons, Picketts, et al., 1995).

Between these two domains, there is one central region which enables ATRX to interact with the chaperone protein death domain-associated protein 6 (DAXX) to form a chromatin remodeling complex involved in the deposition of replication independent histone variant H3.3 (Xue et al., 2003; Tang et al., 2004). More details on ATRX-DAXX interaction are described in section 1.2.4.1.

In addition, there are many other interacting motifs/domains allowing ATRX to interact with other molecules (Figure 1.1), which will be described in section 1.2.4.

1.2.3 ATRX localisation and its functions

Although the primary structure of ATRX indicates its role in chromatin binding and regulation, it is not very clear how ATRX gets involved in this. Looking at the localisation of ATRX would facilitate our understanding on this. Two functional nuclear localisation signals within ATRX protein, between amino acids 685-697 and 1126-1196, enable ATRX to target to the nucleus (Bérubé et al., 2008). In the nucleus, however, ATRX localises in a number of ways.

1.2.3.1 ATRX associates with promyelocytic leukemia nuclear bodies (PML-NB)

The promyelocytic leukaemia nuclear body (PML-NB, also known as nuclear domain 10 in virology) is a dynamic nuclear macromolecular structure presented as discrete nuclear foci in the mammalian cell nuclei. Their number, size and position vary between different cell types, at different cell-cycle phases and in response to cellular stresses such as DNA damage (Bernardi & Pandolfi, 2007). It has been implicated in regulating diverse cellular functions, such as heterochromatin formation, transcriptional regulation, DNA damage response, regulation of apoptosis and intrinsic immunity against double stranded DNA viruses, as reviewed in (Bernardi & Pandolfi, 2007). Its multiple functions result from the diverse proteins localizing to PML-NB, one of which is ATRX.

The localisation of ATRX in PML-NBs was first reported by (Xue et al., 2003) where they showed a partial colocalisation of ATRX, DAXX and PML-NBs in interphase Hela cells by immunofluorescence staining and this has been confirmed by many other subsequent studies. Interestingly, ATRX may localize in PML-NBs by different mechanisms under different circumstances. It has been demonstrated that the recruitment of ATRX to PML-NBs was dependent on its interaction with DAXX, as ATRX failed to localize in PML-NBs when losing its DAXX-interacting domain (Tang et al., 2004) or depleting DAXX in the cells (Morozov et al., 2017). In both studies, PML-NBs and/or DAXX were not at their endogenous levels, which may influence their interactions. Indeed, when using endogenous levels of PML-NB and DAXX, localisation of ATRX to PML-NBs was no longer dependent on DAXX (Bérubé et al., 2008). Instead, a C-terminal region in the full-length ATRX protein was shown responsible for targeting ATRX to PML-NBs, and mutations within this region found in the ATR-X patients disrupted ATRX nuclear speckles and its association with PML-NBs (Bérubé et al., 2008). This suggests that impairment in the ATRX-PML interaction may contribute to the pathogenesis of the ATR-X syndrome.

Furthermore, PML is also involved in the ATRX/DAXX dependent deposition of the histone variant H3.3 (Delbarre et al., 2017). PML can be found in the heterochromatic regions such as telomeres where H3.3 is deposited (Wong et al., 2009), and PML loss impaired H3.3 deposition by ATRX/DAXX at these sites, indicating a PML-dependent role of ATRX/DAXX in H3.3 deposition (Delbarre et al., 2017).

1.2.3.2 ATRX is located at heterochromatin throughout the genome

As a chromatin remodelling factor, ATRX is expected to bind at chromatin regions to regulate chromatin activity. Previous studies have shown that ATRX binds at a wide range of heterochromatic regions, which include pericentromeric heterochromatin, repetitive regions like rDNA, telomeres and retrotransposons, as well as the 3' exons of the zinc finger genes (McDowell et al., 1999; Law et al., 2010; Wong et al., 2010; Sadic et al., 2015; Valle-García et al., 2016). These regions have underlying features in common - they are heterochromatic regions that are frequently transcribed, most are difficult to replicate / transcribe and may form secondary structures such as G-quadruplex (G4) and R-loops (DNA:RNA hybrids).

a. ATRX localizes to pericentromeric heterochromatin

Pericentromeric heterochromatin is constitutive heterochromatin surrounding the centromere of a chromosome, which is highly DNA methylated and enriched in repressive histone modifications such as H3K9me3. The repressive state of PCH is highly organized and maintained by the concerted actions of a group of factors. This includes DNA methyltransferases (DNMTs), histone methyltransferases (HMTs), histone deacetylases (HDACs), chromatin methylation binding factors such as heterochromatin protein 1 (HP1) and methyl CpG binding protein 2 (MECP2), as well as chromatin remodelling factors like ATRX.

PCH localisation of ATRX may involve an elaborate network of interactions including H3K9me3, HP1 and MECP2, among which H3K9me3 probably plays a dominant role. As mentioned in section 1.2.2, the ADD domain of ATRX facilitates ATRX recognizing H3K9me3. H3K9me3 is an epigenetic hallmark of constitutive heterochromatin like PCH, and the state requires HP1 binding to preserve (Maison & Almouzni, 2004; Almouzni & Probst, 2011; Fioriniello et al., 2020). It has been shown that ATRX locates at PCH mainly by binding to H3K9me3 through its ADD domain. Specifically, it has been shown that in mouse embryonic fibroblast cells where H3K9me3 was depleted by removing the H3K9me3 histone methyltransferases Suv39h1 and Suv39h2, neither Atrx nor Hp1α (also known as Cbx5) could be detected at PCH (Iwase et al., 2011). Re-expression of H3K9me3 by recovering Suv39h1 restored the Atrx presence at PCH (Iwase et al., 2011). These suggest that H3K9me3 itself is essential for tethering ATRX to PCH.

The contribution of HP1 in ATRX recruitment at PCH was supported by the observations that mutations in the HP1 interaction motif of ATRX reduced ATRX localisation at PCH, but this effect was less than that of destabilizing the ADD domain (Iwase et al., 2011; Eustermann et al., 2011). This suggested that ATRX targets PCH through a tripartite network of interactions that involves both its ADD domain and its interaction with HP1. Interestingly, the effect of ATRX on HP1 PCH localisation varies in different scenarios. Work by Iwase and colleagues (Iwase et al., 2011) showed that Atrx inhibition by RNAi did not affect the enrichment of Hp1α at PCH in mouse fibroblast cell line NIH3T3 cells, indicating Hp1α protein acts upstream of Atrx and its enrichment at PCH does not require

Atrx. However, another study demonstrated that the absence of ATRX significantly decreased HP1α and HP1γ enrichment at PCH in terminally differentiated neurons derived from mouse embryonic stem cell (mESC) (Marano et al., 2019).

Other studies have also implicated a role of MECP2, an epigenetic reader belonging to the methyl-binding domain family, in the PCH localisation of ATRX particularly in neuronal cells. Interaction between ATRX and MECP2 contributes to both of their PCH localisation reciprocally, as their localisation to PCH were disrupted by losing the expression of the other one (Nan et al., 2007; Marano et al., 2019).

b. ATRX localizes to telomeres, rDNA, retrotransposons and other repetitive sequences

ATRX has been shown to bind G-rich tandem repeats that are predicted to form non-B DNA structures such as G4 (Law et al., 2010). Such repetitive regions within the genome include telomeres and ribosomal DNA repeat sequences.

Telomeres are nucleoprotein structures that locate at the distal ends of chromosomes to protect genome from nucleolytic degradation, unnecessary recombination, repair, and interchromosomal fusion. It has a characteristic tandemly repetitive sequence (TTAGGG)n which occupies several kilo bases (kb) and interacts with specific proteins to maintain genome stability. Studies in mESC showed that ATRX binds at the telomeres during telomere replication at S phase

in conjunction with the incorporation of H3.3, HP1α and PML, but dissociated from the telomeres during mitosis (Wong et al., 2010; Chang et al., 2013). Knockdown of ATRX affected telomere integrity (Wong et al., 2009, 2010). All these evidence indicate that ATRX, together with other factors, binds at telomeres during S phase to maintain their integrity during replication. In recent years, ATRX has been shown to play a role in suppressing the alternative lengthening of telomeres (ALT) pathway since mutations in *ATRX* have been frequently found in many cancers where the cells can maintain their telomeres via ALT. This will be described in more detail in section 1.2.5.

In addition to telomeres, ATRX has also been found binding to other repetitive sequences such as G-rich ribosomal DNA (rDNA) repeats, minisatellite repeats and retrotransposons to promote the local heterochromatin formation (McDowell et al., 1999; Law et al., 2010; Sadic et al., 2015; Truch et al., 2022). ATRX-depleted cells were shown to have reduced copy number and transcription of rDNA as well as rDNA instability, which was associated with the disruption of H3.3 deposition at these regions (Udugama et al., 2018). Knocking out ATRX also impaired the establishment and maintenance of heterochromatin at retrotransposons (Sadic et al., 2015). In addition, mutations in ATRX can alter the pattern of DNA methylation at these regions which may affect local gene expression (R. J. Gibbons et al., 2000). Taken together, these findings demonstrate that ATRX contributes to the stability and DNA methylation of its binding targets in repetitive regions.

c. ATRX localizes to 3' exons of the highly duplicated zinc finger genes

ATRX ChIP-seq data identified the 3' exons of zinc finger genes (ZNFs) as a new class of ATRX targets (Valle-García et al., 2016; Truch et al., 2022). Zinc finger proteins are a large group of putative transcription repressors in human and other primate genomes (Emerson & Thomas, 2009). Their 3' exonic regions encode the zinc finger motifs, ranging from 1 to 40 copies per ZNF gene and very interestingly, they are enriched in both the repressive heterochromatin mark H3K9me3 and paradoxically the transcriptional elongation mark H3K36me3 (Blahnik et al., 2011; Valle-García et al., 2016). This is, however, not an imprinting effect as SNP analysis of RNA-seq data demonstrated that both alleles of certain ZNF genes having H3K9me3 and H3K36me3 were transcribed (Blahnik et al., 2011).

It has been shown that ATRX binds to the 3′ exons of ZNFs where it forms a complex with H3K9 methyltransferase SETDB1, the co-repressor TRIM28, and the transcription factor ZNF274 (Valle-García et al., 2016). ATRX and ZNF274 loss gave rise to reduced level of H3K9me3 at the ZNF 3′ exons and the cells displayed increased levels of DNA damage, indicating that ATRX and the ZNF274/TRIM28/SETDB1 complex bind to the 3′ exons of ZNFs to help maintain their genomic stability through preservation of H3K9me3 mark (Valle-García et al., 2016). As described in the earlier sections, ATRX can recognize and bind H3K9me3 via its ADD domain, but it is unclear if the underlying interaction between ATRX and the ZNF274/TRIM28/SETDB1 complex is mediated by ADD-H3K9me3 recognition. It is possible that the H3K9 methyltransferase SETDB1

establishes basal H3K9me3 levels at ZNF 3′ exons that recruit ATRX through its ADD domain to recognize H3K9me3 and in turn promotes further H3K9me3 enrichment.

1.2.3.3 ATRX binds CpG islands and active regulatory elements in euchromatin

Most previous efforts trying to determine the binding sites of ATRX have been focused on its binding at heterochromatin. Nevertheless, its binding in euchromatin has been presented in several studies (Danussi et al., 2018; Qin et al., 2022; Truch et al., 2022). These binding sites mainly include CpG islands (CGIs) and active regulatory elements.

The human genome contains ~30,000 CGIs which comprise a large number of CpG dinucleotide repeats. They are usually located within or close to gene promoters. CGIs in promoters are often unmethylated, associated with gene transcription and active chromatin modifications, whereas intragenic CGIs lying within gene bodies are methylated (Jeziorska et al., 2017). A previous study showed that ATRX was bound to a series of intragenic CGIs, many of which were methylated and corresponded to the silenced imprinted alleles associated with H3.3 and H3K9me3 modification (Voon et al., 2015). Recently, the binding pattern of ATRX at CGIs has been comprehensively analysed in human lymphoblastoid cell lines derived from normal donors (Truch et al., 2022). It has been shown that ATRX binding sites were present in 29% of annotated genome-wide CpG islands and most of the ATRX bound CpG islands were associated with promoter regions and remained unmethylated (Truch et al., 2022).

Importantly, the unmethylated ATRX-bound promoter CpG islands were enriched in the active chromatin mark H3K4me3 and depleted of heterochromatic mark H3K9me3, whereas the methylated sites were correlated with high levels of H3K9me3 (Truch et al., 2022), indicating ATRX may contribute to the establishment of the chromatin environment at CpG islands.

In addition to CGIs, ATRX exhibits extensive binding at regions of open chromatin associated with active regulatory elements (Danussi et al., 2018; Qin et al., 2022; Truch et al., 2022). Danussi C and colleagues first reported that Atrx displayed a gene regulatory binding pattern in murine neuroepithelial progenitors where Atrx bound transcriptional start sites (TSSs) and enhancer regions, and Atrx deficiency changed genome-wide chromatin accessibility which was associated with altered gene expression (Danussi et al., 2018). Another study implicated the binding of ATRX at the regulatory elements of cell-cycle related genes and revealed that ATRX-deficient cells showed dysregulated cell-cycle-related gene expression and cell-cycle phase transition (Qin et al., 2022). Consistent with this, the work from our lab showed that ATRX was recruited to transcriptionally active elements such as enhancers, promoters and gene bodies, which was defined by the process of transcription (Truch et al., 2022).

1.2.4 ATRX interactors and functions

Interactions between ATRX and its partners impart multiple roles to ATRX. In this section, a number of primary ATRX direct interactors and the function associated with these interactions are described.

1.2.4.1 ATRX interacts with DAXX to deposit H3.3

The most well-known binding partner of ATRX is the transcription cofactor DAXX. ATRX can interact with DAXX through a conserved helical DAXX-binding motif to form a chromatin remodeling complex that deposits replication independent histone variant H3.3 into repetitive heterochromatic regions and putative G-quadruplex forming sequences for transcription-associated heterochromatin reorganization (Xue et al., 2003; Tang et al., 2004; Goldberg et al., 2010; Drané et al., 2010; Lewis et al., 2010; Hoelper et al., 2017; Park et al., 2018).

There are three histones H3 variants in mammalian cells, of which histones H3.1 and H3.2 are closely related and expressed during S-phase when replication occurs, whereas the H3.3 variant, encoded by two genes *H3F3A* and *H3F3B*, is expressed throughout the cell cycle. H3.3 deposition appears to target to particular loci including active rDNA, indicating its deposition may be associated with specific transcription states (Ahmad & Henikoff, 2002). Biochemical investigation has identified the histone chaperone complexes responsible for H3.3 deposition. One of them, the histone chaperone HIRA complex, was implicated in depositing H3.3 at euchromatic regions of the genome, such as transcriptional start sites and gene bodies (Ahmad & Henikoff, 2002; Tagami et al., 2004). And ATRX-DAXX was identified as a H3.3 chaperone complex accounting for H3.3 deposition at telomeres (Goldberg et al., 2010; Lewis et al., 2010) and pericentric heterochromatin (Drané et al., 2010). However, recent work from our lab indicated that ATRX/DAXX/H3.3 complex may also involve in the deposition of H3.3 within euchromatin, given that ATRX binding sites in euchromatin are correlated with levels of histone H3.3 (Truch et al., 2022).

Within the ATRX/DAXX complex, DAXX is the component providing the histone chaperone activity that binds H3.3 variant-specific residues through its conserved N terminus (Drané et al., 2010; Lewis et al., 2010), while ATRX is in charge of recruiting DAXX to telomeric, pericentromeric, and other DNA regions through its ADD domain by recognizing H3K9me3 (Iwase et al., 2011; Eustermann et al., 2011). ATRX/DAXX/H3.3 complex plays an important role in chromatin regulation throughout the genome, and diverse nuclear activities are disturbed when its components are mutated. ATRX loss alone can exert an effect on genome stability by affecting multiple processes such as DNA replication, DNA repair, DNA methylation, chromatin modification and gene expression (R. J. Gibbons et al., 2000; Law et al., 2010; Clynes et al., 2014, 2015; Truch et al., 2022), which will be described in more detail later. Besides, somatic mutations in ATRX, DAXX and H3.3 have been frequently found in a group of cancers where the cells showed enhanced replication stress, altered chromatin modification and gene expression, and the underlying defects in the chromatin architecture may be involved in the pathogenesis of these diseases (Jiao et al., 2011; Schwartzentruber et al., 2012; Chan et al., 2013; Lewis et al., 2013; Helming et al., 2014; Bhargava et al., 2022).

1.2.4.2 ATRX associates with HP1α to maintain heterochromatin

In mammals, heterochromatin protein 1 family contains three paralogs: HP1α (also known as CBX5), HP1β and HP1γ. They have important functions in the maintenance of heterochromatin integrity including PCH and telomeric regions and gene expression regulation (Probst & Almouzni, 2008; Almouzni & Probst, 2011; Chavez et al., 2017; Schoelz & Riddle, 2022). There are two functional

domains in HP1 protein: the N-terminal chromoshadow domain allowing interaction with nuclear proteins and the C-terminal ‘chromatin organisation modifier’ chromo domain which binds to methylated histone H3 including H3K9me3 (Lechner et al., 2005; Zeng et al., 2010).

ATRX associates with HP1α via its conserved pentapeptide motif Pro-Xaa-Val-Xaa-Leu (PxVxL) downstream the ADD domain (Figure 1.1) which can be recognized by the N-terminal chromoshadow domain of HP1α (Lechner et al., 2005). HP1α is able to bind directly to H3K9me3 via its C-terminal chromo domain that locates HP1α at heterochromatin including PCH and telomeres (Zeng et al., 2010). Although ATRX and HP1α were found colocalizing at similar genomic regions and cooperating in heterochromatin regulation (McDowell et al., 1999; Chavez et al., 2017), failure in copurifying these two proteins from cellular extracts indicates that their interaction may be tissue-specific or transient, such as during the recruitment of ATRX to H3K9me3-enriched regions (Lechner et al., 2005; Eustermann et al., 2011; W. A. Scott et al., 2021). It is possible that association of ATRX with HP1α is a secondary effect when both ATRX and HP1α simultaneously binds to H3K9me3-enriched heterochromatins to maintain their stability.

1.2.4.3 ATRX partners with MECP2 to facilitate heterochromatin organization and to silence the imprinted gene

Methyl-CpG binding protein 2 (MECP2) is an epigenetic reader involved in the regulation of locus-specific transcription and genome architecture like PCH organization. ATRX has been demonstrated to interact with MECP2 through its

C-terminal helicase motif (Figure 1.1) (Nan et al., 2007) and the K171 acetylation site within MECP2 protein is essential for their interaction (Pandey et al., 2015). In section 1.2.3.2.a, we have mentioned that ATRX and MECP2 are reciprocally dependent for their expression and PCH localisation (Nan et al., 2007; Marano et al., 2019). Both factors modulate HP1 localisation at PCH to help maintain heterochromatin organization.

Furthermore, ATRX and MECP2 interaction has also been implicated in silencing a subset of imprinted genes during critical stages of postnatal brain development (Kernohan et al., 2010, 2014). It was demonstrated that ATRX cooperates with MECP2 and cohesin, together with the transcription factor/insulator protein CTCF, to silence the developmental expression of specific imprinted loci such as *Igf2/H19* and *Dlk1/Gtl2* (Kernohan et al., 2010). They showed that ATRX was required for the enrichment of the other proteins at imprinted loci and loss of ATRX affected developmental silencing of the gene by changing the histone modification rather than DNA methylation (Kernohan et al., 2010). A later follow-up study addressed the underlying molecular mechanism, where they revealed that MECP2 was required for ATRX binding at these imprinted loci and deficiency of either ATRX or MeCP2 gave rise to decreased CTCF occupancy and altered chromatin looping configurations thus leading to dysregulated gene expression (Kernohan et al., 2014). These works indicate a potential mechanism that expression change in these genes may contribute to the brain defects associated with ATR-X syndrome and Rett syndrome caused by *MECP2* mutations. It is worth noting that a differential regulation of imprinting loci by ATRX was reported in an independent study, in which ATRX/H3.3/H3K9me3 were shown to be

required for allele specific silencing of numerous imprinted genes in mESCs by depositing H3.3 to the methylated alleles to maintain its heterochromatic state and silence the imprinted allele at these loci (Voon et al., 2015); in this case MECP2 is not required, indicating the differences in gene silencing mechanisms at imprinted loci in mESCs versus differentiating cells.

1.2.4.4 ATRX recruits EZH2 to deposit H3K27me3 to repress gene expression

Enhancer of zeste homolog 2 protein (EZH2) is the functional enzymatic subunit of the transcription repressive Polycomb Repressive Complex 2 (PRC2) and it catalyses the addition of methyl groups to histone H3 at lysine 27 (H3K27). The methyltransferase activity of EZH2 promotes H3K27 methylation, facilitates heterochromatin formation and thereby represses gene transcription (Joshi et al., 2008). EZH2 contains a conserved histone lysine methyltransferase domain - SET domain at its C-terminus (Joshi et al., 2008).

A specific interaction between the XNP (previous alias of ATRX) and EZH2 was revealed by the yeast two-hybrid assay analysis coupled with *in vitro* binding assays (Cardoso et al., 1998). ATRX interacts with the SET domain of the EZH2 protein through an interacting domain downstream of the ADD domain (Figure 1.1) (Cardoso et al., 1998). Co-IP assays in another study detected ATRX binding EZH2 in human glioblastoma cells *in vivo* and demonstrated ATRX/EZH2 complex regulated specific gene expression by promoting H3K27me3 in gene promoter region (Han et al., 2020). Moreover, re-expression of the silenced PRC2 gene targets in ATRX-deficient malignant tumours (Oppel et al., 2019)

added more evidence that ATRX could interact with PRC2 subunit EZH2 to silence gene expression.

In addition, Sarma and colleagues reported the formation of a ternary complex ATRX-RNA-PRC2 in mouse embryonic stem cells where ATRX acts in targeting PRC2 to polycomb target genes including the X-inactive specific transcript (*Xist*) RNA to regulate target gene expression (Sarma et al., 2014). In particular, the ATRX-Xist-EZH2 interaction contributed to X chromosome inactivation (XCI) (Sarma et al., 2014). In this case, the interaction between ATRX and EZH2 occurs through the long non-coding RNA (lncRNA) Xist, which is necessary for ATRX recruitment and facilitates the loading of EZH2 to deposit H3K27me3 on the inactivated X chromosome (Sarma et al., 2014). More recently, they identified the RNA binding domain in ATRX (Figure 1.1) and proved that this domain contributed to the localisation of ATRX to chromatin and EZH2 enrichment (Ren et al., 2020). However, this domain contains the complete HP1 interaction motif responsible for heterochromatic localisation and they did not exclude this contribution (Wada et al., 2005; Badens et al., 2006; Muers et al., 2007)

1.2.4.5 ATRX interacts with RNA

In addition to the protein interactors, ATRX has been shown to interact with different RNAs, such as the lncRNA Xist mentioned above (Sarma et al., 2014; Ren et al., 2020). Studies have also unveiled interactions of ATRX with a range of other RNAs. One of these is the lncRNA TERRA (telomeric repeat containing RNA), which interacts with ATRX at the target genes like telomeres to antagonize

each other functionally (Nguyen et al., 2017; H. P. Chu et al., 2017). Another ATRX-interacting lncRNA is ChRO1 (muscle-specific Chromatin reorganization 1), which associates with ATRX/DAXX/H3.3 complex at chromocenters to facilitate transcription-mediated heterochromatin reorganization and cell differentiation through promoting H3.3 deposition (Park et al., 2018). Recently, the RNA binding domain in ATRX was identified and it was found contributing to the chromatin localisation of ATRX via mediating the interactions between ATRX and RNA (Ren et al., 2020), which puts forward another possible mechanism for ATRX to regulate chromatin organization and gene expression.

In addition to the interactors mentioned in this section, there are potentially many other ATRX partners, some of which have been captured in a recent study utilizing proximity-dependent biotinylation proteomic screen approach to identify proximal associating proteins of ATRX (W. A. Scott et al., 2021). Apart from the known interactors, this study has identified new interacting proteins such as CCDC71 associated with chromatin maintenance, FAM207A associated with ribosome biosynthesis and the DNA damage response factor SLF2, which suggests new possible mechanisms underlying ATRX functions (W. A. Scott et al., 2021).

1.2.5 Roles of ATRX in DNA replication, DNA damage response and DNA repair

ATRX was reported playing a role in DNA replication and DNA damage response (DDR) when it was initially shown to be recruited to stalled replication fork sites during S phase, interacting with the MRE11-RAD50-NBS1 (MRN) complex, to facilitate restart of stalled replication forks and recovery from replication stress (Leung et al., 2013; Clynes et al., 2014). ATRX-depleted mESC displayed higher sensitivity to hydroxyurea-induced replication fork stalling and delayed restart of replication (Leung et al., 2013; Clynes et al., 2014). This was consistent with the previous observations in ATRX-depleted myoblasts, where the cells showed delayed S phase, accumulated DNA damage during replication which induced the activation of p53-ATM DDR pathway and increased the incidence of genomic instability (Huh et al., 2012). A follow-up study looking more into the underlying mechanism demonstrated that PARP-1 and ATM (ataxia-telangiectasia mutated kinase) activation can function as a compensation pathway to protect stalled forks from damage, in response to elevated DNA damage caused by ATRX deficiency, insufficient loading of HR proteins and increased degradation of replication forks (Huh et al., 2016). This further supports the role of ATRX in protecting stalled replication forks during cellular proliferation. Moreover, ATRX has also been identified as an interactor of the Fanconi anaemia protein FANCD2 to promote recovery of stalled replication fork in S phase via histone deposition-dependent homologous recombination (HR) DNA repair mechanism, which also requires DAXX for the deposition of histone variants H3.1/H3.3 and MRN to recruit key factors involved in HR (Raghunandan et al., 2020). Recently, a study reported a direct interaction between ATRX and the DNA replication helicase

minichromosome maintenance (MCM) complex, and showed that loss of ATRX promoted the formation of G4 structures at replication forks and thus inducing replication stress (Teng et al., 2021), which provides another mechanism for ATRX regulation of DNA replication.

In fact, the role of ATRX in DNA repair has been implicated in the studies exploring the function of ATRX in DNA replication and DNA repair. For instance, ATRX is required to recruit HR proteins RAD51 to repair damaged DNA through HR dependent DNA repair pathway (Huh et al., 2016). Nevertheless, some direct evidence manifesting the involvement of ATRX in DNA repair stands out.

Work by Juhasz and colleagues showed that ATRX promotes HR-mediated DNA repair of exogenously induced DNA double-strand breaks (DSBs) (Juhász et al., 2018). This occurs downstream of Rad51-dependent homology search step, where ATRX promotes extended DNA repair synthesis and sister chromatid exchanges by cooperating with DAXX to incorporate histone H3.3 and interacting with two DNA repair synthesis proteins - proliferating cell nuclear antigen (PCNA) and Replication factor C 1 (Juhász et al., 2018). A follow -up study from the same group indicated that ATRX-mediated HR repair pathway involves the formation of HR crossover intermediates and this subpathway usually outcompetes RECQ5-dependent Synthesis-Dependent Strand Annealing pathway in the repair of most DSBs in human cells (Elbakry et al., 2021). The process of the subpathway choice is dependent on the interaction of ATRX and RECQ5 with PCNA (Elbakry et al., 2021).

Interestingly, it seems that ATRX functions differently at telomeres with regard to HR-dependent DNA synthesis, since it has been widely shown that ATRX inhibits HR at telomeres and acts as a tumour suppressor. This involves an alternative telomerase-independent ALT pathway. Certain repetitive DNA regions in the genome are particularly prone to damage upon DNA replication stress, such as common fragile sites and telomeres. Due to their inherent repetitive nature, these regions tend to form secondary structure which can block replication forks to induce DNA replication stress and DNA damage. The G-rich tandem (TTAGGG)n hexanucleotide repeat in telomeres are susceptible to form secondary structures like G4 and R-loop during transcription which ATRX can bind and may help resolve (Law et al., 2010; Watson et al., 2013; Clynes et al., 2015; Nguyen et al., 2017). In the absence of ATRX, DNA replication fork stalls at the formed secondary structure, which causes a subsequent DNA damage response and serves as a substrate for MRN-mediated HR repair to maintain the length of telomere through ALT pathway (Law et al., 2010; Clynes et al., 2015; Nguyen et al., 2017). The ability of ATRX to suppress ALT is DAXX-dependent and around 15% of cancers maintain their telomere by ALT pathway where mutations in ATRX and DAXX are frequently found (Heaphy et al., 2011).

Additionally, interaction between ATRX and the macro-histone variant macroH2A1.2 also contributes to protecting the DNA from damage upon fork stalling and replication stress (J. Kim, Sturgill, et al., 2018; J. Kim et al., 2019). This involves a protective epigenetic landscape where macroH2A1.2 are deposited at damaged DNA sites in a histone chaperone FACT (facilitates chromatin transcription) dependent manner to set up a repair-promoting

chromatin environment (J. Kim, Sturgill, et al., 2018). In ATRX mutated ALT cells, macroH2A1.2 is transiently lost at telomeres upon acute replication stress to facilitate DSB formation and induce DNA damage response, which in turn recruits macroH2A1.2 back to trigger HR-associated ALT pathway (J. Kim et al., 2019). This dynamic process provides another mechanism linking ATRX loss and DNA damage to ALT pathway.

To summarise the multiple molecular functions of ATRX described in this section, it can be classified into three main categories: chromatin organization, genome stability maintenance and gene expression regulation (Table 1.1). These functions confer ATRX protein multiple roles in the biological processes including cell proliferation, cell division, cell differentiation, senescence as well as intrinsic immunity.

Chromatin organization	Mainly contributes to heterochromatin maintenance and organization, eg. PCH, telomeres, rDNA and other repetitive regions. This is dependent on interactions with DAXX/H3.3, H3K9me3, HP1, PML, MECP2, EZH2 and RNA.
Genome stability maintenance	Facilitates DNA replication, eg. interacts with MRN/FANCD2 complex, activates PARP-1 and ATM, resolves secondary structures formed at G-rich regions.
	Responds to DNA damage and promotes DNA repair, eg. interacts with macroH2A1.2, promotes HR-mediated DNA repair.

Gene expression regulation	Interacts with EZH2 to deposit repressive histone marks H3K27me3; silences imprinted genes by interacting with MECP2 or depositing H3.3.
	Binds CpG islands and regulatory elements to regulate gene expression with unclear mechanisms.

Table 1.1 Summary of ATRX functions.

1.3 Consequences of ATRX mutations

As stated in the last section, ATRX functions in a wide range of nuclear processes and these processes can be disrupted when *ATRX* is mutated. Therefore, its mutations would be expected to be linked with some disorders. Indeed, germline mutations in *ATRX* give rise to ATR-X syndrome in human, and the patients present symptoms in various systems. Of note, the patient mutations are hypomorphs rather than knockout of gene function. Besides, acquired *ATRX* somatic mutations are frequently found in a group of cancers. Based on the phenotypes of patient carrying naturally occurring *ATRX* mutations and the observations in models carrying artificial *ATRX* mutations, herein the primary consequences of ATRX mutations are summarized in this section.

1.3.1 ATRX mutations cause impaired development by inducing DNA damage and cell apoptosis

ATR-X patients often show global developmental delay in various systems. In their early childhood, there is generalised hypotonia, including muscle hypotonia, delayed ambulation, kyphosis and hypotonia induced distinctive facial traits, suggesting a role of ATRX in skeletal muscle development (R. Gibbons, 2006). This has been supported by the observations in ATRX deficient muscle satellite cells, in which defects in muscle growth and regeneration was detected and this was associated with compromised genome instability at the molecular level (Huh et al., 2016). Targeted loss of ATRX in the limb mesenchyme of mice resulted in shorter digits or brachydactyly, reduced grip strength and altered gait in adulthood, the underlying mechanism of which appeared to be DNA damage-induced apoptosis (Solomon et al., 2013), which further supports the role of ATRX in skeletal development. However, deleting Atrx in mouse cartilage had little effect on the development or growth of cartilage, indicating chondrocytes are not involved in the impaired skeletal development induced by ATRX mutations (Solomon et al., 2009).

Many other tissue-specific Atrx-deficient mouse models reveal the role of ATRX in development in the context of different systems. Conditional ablation of Atrx in mouse forebrain led to neuronal loss during corticogenesis associated with increased cell death upon differentiation, partially triggered by DNA damage and the activation of the p53-dependent apoptotic pathway (Bérubé et al., 2005; Seah et al., 2008), suggesting a role of ATRX in brain development. Conditional inactivation of Atrx in the retina gave rise to loss of neurons, amacrine and

horizontal cells due to the defect in interneuron differentiation and survival postnatally (Medina et al., 2009; Lagali et al., 2016), supporting a role of Atrx in neuron survival. Mice with ATRX deficiency in neuroprogenitors displayed a range of symptoms including reduced growth, shortened life span, cataracts and reduced mineral bone density (Watson et al., 2013). When Atrx was specifically knocked out in Sertoli cells of the mouse testis, the mice developed small testes and discontinuous tubules as well as spermatogenesis defect, which was shown resulting from prolonged G2/M phase in cell cycle and cell apoptosis (Bagheri-fam et al., 2011), implicating a role of ATRX in genital development.

In addition, findings in the ES cell and conventional mouse models where ATRX level has been changed also demonstrate ATRX is essential for cell survival and development. It has been shown that overexpression of ATRX in transgenic mice caused growth retardation and neurodevelopmental abnormalities during embryonic development process and led to a high incidence of prenatal and perinatal lethality, indicating the right dosage of ATRX is critical during early embryonic development (Bérubé et al., 2002). Consistently, ATRX loss gives rise to developmental defects. $Atrx^{null}$ ES cells showed a growth defect and $Atrx^{null}$ male mice cannot survive beyond embryonic day 9.5 (E9.5) due to a specific defect in trophoblast development (Garrick et al., 2006). Interesting, the skewed XCI observed in $Atrx^{null}$ female mice has been shown as a result of cell selection during which cells not expressing Atrx were gradually lost (Garrick et al., 2006; Muers et al., 2007).

Collectively, ATRX mutation-induced developmental deficiency is widely associated with DNA damage and cell apoptosis triggered by ATRX loss at the molecular level. And p53 activation has been shown to act as a mediator downstream to ATRX in this process. Indeed, Atrx deficient macrophages and other Atrx-null primary cells showed higher sensitivity to DNA damaging agents which could be rescued by removing p53 (Conte et al., 2012). And p53 inactivation was able to abrogate abnormalities caused by ATRX loss in neuroblastoma cells, including arrested cell cycle, increased G4 formation, replication stress and DNA damage (Akter et al., 2021). However, this is not the sole mechanism, as ATRX/p53 double knockout mice rescued cell death to some extent but not completely (Seah et al., 2008).

1.3.2 ATRX mutations give rise to cognitive and behavioural deficits

In last section, it has been mentioned that ATRX loss is associated with abnormal brain development, supported by the fact that its loss resulted in brain size reduction and loss of neuron cells (Bérubé et al., 2005; Seah et al., 2008; Medina et al., 2009). Apart from microcephaly, ATR-X patients also suffer from severe intellectual disability, indicating ATRX may play a role in neuron functions such as learning, memory and behaviour. This has been supported by the observations in ATRX deficient models. Conditional deletion of Atrx in the retina led to compromised visual function (Medina et al., 2009), and mice with conditional Atrx loss in neuroprogenitors were affected by cataracts (Watson et al., 2013). Similarly, ophthalmic defects have also been observed in some ATR-X patients, including optic atrophy or refractive errors like myopia (R. Gibbons, 2006). Additionally, deleting exon 2 of *Atrx* to produce truncated and reduced

expression of Atrx in mice leads to compromised hippocampal function and impaired contextual fear memory (Nogami et al., 2011). Consistently, heterozygous female mice with brain-specific Atrx deletion also showed deficits in contextual fear, spatial memory and novel object recognition memory, associated with hippocampal structural defects and fewer presynaptic vesicles (Tamming et al., 2017).

In addition to ATR-X syndrome, *ATRX* mutations has been identified in autism spectrum disorder patients (J. Li et al., 2017; López-Garrido et al., 2022). However, when deleting Atrx in forebrain excitatory neurons postnatally at 2-3 weeks of age, the mice seemed to have normal social behaviours instead of leading to autism-like behaviours (Martin-Kenny et al., 2020), indicating ATRX may contribute to autistic phenotypes in an early stage of development during embryogenesis. Another crucial evidence supporting ATRX's role in the central nervous system (CNS) is that ATRX somatic mutations have been frequently associated with brain tumours, which will be discussed in the next section.

The underlying mechanism of defects in CNS caused by ATRX loss is still unclear and might vary under different conditions. It has been reported that loss of ATRX led to defective mitosis and compromised chromosome congression and segregation in neuroprogenitors (Ritchie et al., 2008). This is probably because of incomplete DNA replication. It has also been demonstrated that there was accumulated replicative DNA damage at telomeres and pericentromeric heterochromatin in Atrx-null embryonic mouse brain that was associated with ATM activation and can be exacerbated by p53 inactivation (Watson et al., 2013).

In addition, the dysregulated synaptic function being responsible for cognitive deficits has been associated with ATRX depletion. The first report indicating ATRX loss affects synaptic function came from a study investigating the role of ATRX in retina, where Atrx-null retina showed defective synaptic organization and function (Medina et al., 2009; Lagali et al., 2016). Later on, some direct evidence appeared. One study indicated the imprinted gene *Xlr3b* was upregulated in the brain of $Atrx^{\Delta E2}$ mouse and this reduced synaptic plasticity (Shioda et al., 2018). Atrx was found to bind to G4 in CpG islands of *Xlr3b* to regulate its expression. Interestingly, treating the $Atrx^{\Delta E2}$ mice with G4-binding small molecule 5-aminolevulinic acid attenuated elevated Xlr3b expression and counteracted the cognitive deficits of mice (Shioda et al., 2018). There is, however, no ortholog of *Xlr3b* in human. Another recent study identified the microRNA miR-137 as one target of ATRX in neurons, whose expression can be suppressed by ATRX through recruiting H3K27me3, and miR-137 has been implicated in regulating synaptic pathways, therefore it is probable that ATRX loss leads to learning and memory deficits by affecting miR-137 mediated synaptic functions (Tamming et al., 2020).

1.3.3 ATRX mutations trigger cancers via perturbing genome stability and gene expression

The first indication that ATRX may be involved in malignancy came from a disorder caused by acquired somatic *ATRX* mutation called acquired alpha thalassaemia myelodysplastic syndrome (ATMDS) (R. J. Gibbons et al., 2003). In recent years, an increasing number of studies reported the association between ATRX mutations and cancer. It has been shown *ATRX* is frequently

mutated in a group of tumours that use the ALT pathway to maintain the length of their telomeres, including glioma, neuroblastoma, pancreatic neuroendocrine tumours and paediatric osteosarcoma (Jiao et al., 2011, 2012; Schwartzentruber et al., 2012; Molenaar et al., 2012; Kannan et al., 2012; X. Y. Liu et al., 2012; X. Chen et al., 2014). As a result, *ATRX* gene status has been regarded as a molecular marker by which to characterise tumours in both the newly published 2021 World Health Organization Classification of Tumors of the Central Nervous System and its 2016 predecessor (Louis et al., 2016; Osborn et al., 2022).

Despite *ATRX* mutations being detected in many cancers, patients with ATR-X syndrome seem not to show increased risk of cancer generally. However, two studies recently reported four ATR-X syndrome patients developed osteosarcoma (Ji et al., 2017; Masliah-Planchon et al., 2018). Two of these four patients are brothers and have the same nonsense germline variant of *ATRX* with the third patient, and the other patient having a missense variant (Ji et al., 2017; Masliah-Planchon et al., 2018). This may indicate *ATRX* germline mutations might be a predisposing factor to develop osteosarcomas, but this remains speculative given the fact that ATR-X syndrome is a very rare disorder and thereby very limited number of patients are reported. Notably, one study followed the development of the pancreatic neuroendocrine tumours in patients with multiple endocrine neoplasia-1 (MEN-1) syndrome, a cancer predisposition condition, and they found mutations in *ATRX* were late events in pancreatic neuroendocrine tumour development (de Wilde et al., 2012), indicating mutations in *ATRX* support cancer development rather than being a prime driver of cancer.

With regards to the molecular mechanism underlying cancer induced by ATRX mutations, the genomic instability and transcriptional change are probably the two major causes. As mentioned in previous sections, ATRX deficiency is associated with increased DNA damage, compromised DNA repair, abnormal cell cycle, impaired chromosome segregation and telomere dysfunction, which are all sources of genomic instability. Moreover, ATRX loss drives alteration in chromatin accessibility, histone composition and transcription, which can result in dysregulated gene expression. These have been supported by a number of cancer research (Schwartzentruber et al., 2012; X. Y. Liu et al., 2012; Chan et al., 2013; Marinoni et al., 2014; Clynes et al., 2015; Koschmann et al., 2016; J. Kim et al., 2016, 2019; Danussi et al., 2018; Oppel et al., 2019; Akter et al., 2021; Qin et al., 2022; Bhargava et al., 2022).

1.4 Human α-globin genes and α-thalassaemia

1.4.1 Human α-globin gene cluster and regulation

The human α-globin gene cluster is located in the subtelomeric region of chromosome 16 (16p13.3, Figure 1.2) and consists of α-like globin genes and their regulatory elements which encode α-like globins which form haemoglobin. The normal haemoglobin (Hb) molecule, whose main function is carrying oxygen from blood to organs and tissues, is a tetramer composed of two α-like and two β-like globin chains which form a hydrophobic pocket around a heme moiety. Haemoglobin production is encoded by the α and β globin gene clusters, which are located on chromosome 16 and chromosome 11 respectively in human.

Different forms of haemoglobin make up of different α-like (ζ or α) and β-like (ε, γ, δ or β) globin chains which are synthesised at different stages of development. The genes are arranged in the order by which they are expressed, presented as 5'-ζ-μ-α2-α1-θ-3' at the α-globin gene cluster (Figure 1.2) and as 5′-ε-Gγ-Aγ-δ-β-3′ at the β-globin gene locus. At the embryonic stage, Hb Gower-I (ζ2ε2), Hb Gower-II (α2ε2) and Hb Portland (ζ2γ2) are produced. Later in the fetal stage, they are replaced by fetal haemoglobin HbF (α2γ2). And finally in the adult, adult haemoglobin HbA (α2β2) together with minor HbA2 (α2δ2) becomes the dominant haemoglobin (Higgs et al., 2012). The globin genes at the human α-globin locus starting from the telomeric side are ζ-globin gene *HBZ*, the minor μ-globin gene *HBM*, then the two α-globin genes *HBA2* and *HBA1* and lastly the minor θ-globin gene *HBQ*; there are also two pseudogenes - one is between ζ and μ called ψζ *HBZP1* and the other is between μ and α2 called ψα1 *HBAP1* (Figure 1.2).

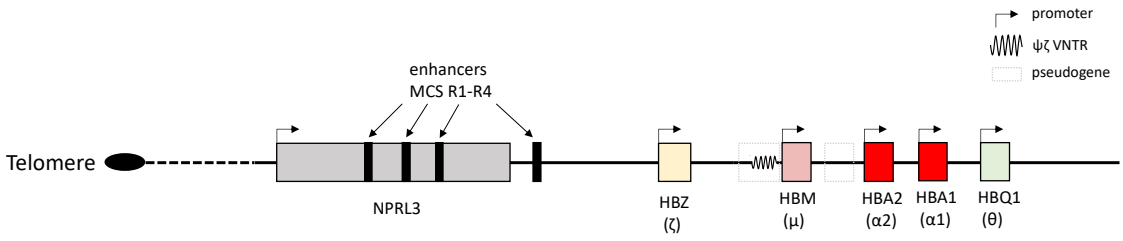


Figure 1.2 Schematic representation of human α-globin locus on chromosome 16 (16p13.3).

The regulatory elements of the α-globin gene cluster have been well-characterised. In the human, there are four distal enhancers regulating faithful α-

globin expression, which are situated at 10 to 48 kilobases upstream of the α-globin gene promoters and described as multi-species conserved sequences (MCS) R1-4 (Figure 1.2). Of these four enhancers, MCS-R2 was reported to be the most effective and strong regulatory element for the transcription of α-globin (Mettananda et al., 2015; Hay et al., 2016). Aside from enhancers, the initiation of α-globin expression also requires an open active chromatin environment and the recruitment of transcription factors. In non-erythroid cells, α-globin genes are primarily silenced by transcription repressive PRC2 complex to establish the H3K27me3 signature (Garrick et al., 2008). When the α-globin transcription is activated during haemopoiesis, H3K27me3 is erased, accompanied by an increasing level of H3K4me3 at the α-globin promoters (Garrick et al., 2008; de Gobbi et al., 2011). Meanwhile, the binding of erythroid-specific transcription factors (eg, GATA-binding factor 1 (GATA1), Kruppel-like factor1 (KLF1)) to enhancers associated with alterations in DNA methylation and histone modification, coupled with the re-organization of 3D genome structure to bring distal enhancers in proximity to the promoters, facilitates the recruitment of RNA polymerase II to drive the initiation of the α-globin gene transcription (de Gobbi et al., 2007; Hughes et al., 2014).

1.4.2 α-thalassaemia

Thalassaemia is one of the most common genetic disorders, with an estimated carrier rate of 1-5% globally (Higgs et al., 2012; Taher et al., 2018). It occurs when there is a reduced production of globin, which mostly results from inherited defects in the two globin gene clusters. α-thalassaemia is caused by down-

regulation of α-globin synthesis, which is characterised by insufficient production of adult (HbA, α2β2) and fetal (HbF, α2γ2) hemoglobin (Higgs et al., 1989). Normally, an individual possesses two pairs of the α-globin genes in total, with two α-globin genes at each chromosome (αα/αα). The severity of the α-thalassaemia depends on the number of α-globin genes affected. For instance, when only one out of the four α-globin genes is affected and downregulated, the carrier is asymptomatic and called 'α-thalassaemia silent carrier'. When two of the genes are damaged, it results in mild to moderate thalassaemia trait. Loss of three α-globin genes would lead to moderate to severe phenotype named HbH disease, in which the excessive β-globin chains form abnormal β4 tetramers (HbH inclusion). In the most severe condition, all the four α-globin genes are impaired which leads to severe α-globin chain deficiency and the formation of γ4 tetramers (Haemoglobin Bart's) in fetal life, known as Hb Bart's hydrops fetalis syndrome. In most cases, a baby with this condition cannot survive (Higgs & Weatherall, 2009).

The genetic cause of α-thalassaemia can be any defect that downregulates α-globin expression, including defects within the coding genes themselves, the regulatory elements and the trans-acting regulators. Among these, an inherited defect within the α-globin gene cluster is the most common reason. A large number of deletions at the α-globin gene cluster that remove one or both adult α-globin genes have been reported to cause α-thalassaemia, which range from 3 to over 100 kb (Higgs & Weatherall, 2009). Deletions around the α-globin promoters and enhancers that leave the α-globin genes intact have also been frequently identified as the genetic causes of α-thalassaemia (Viprakasit et al.,

2003, 2006; Poodt et al., 2006; Coelho et al., 2010; Jiang et al., 2021). In addition, point mutations or nucleotide insertions in the regulatory region that control α-globin expression can cause non-deletional α-thalassaemia (Higgs & Weatherall, 2009; Bozhilov et al., 2021). Occasionally trans-acting mutations that are not physically linked to the α-globin locus can alter the pattern of α-globin expression and give rise to α-thalassaemia. These mutations include those in transcription factors and chromatin-associated factors. ATRX, as a chromatin remodelling factor, is one of these α-globin trans-acting regulators and loss of ATRX can lead to α-thalassaemia in both ATR-X syndrome and ATMDS (R. J. Gibbons, Picketts, et al., 1995; R. J. Gibbons et al., 2003).

1.5 Potential mechanisms of α-thalassaemia phenotype in ATR-X syndrome

Despite considerable progress in defining the properties of the ATRX protein and the consequences of its mutations, knowledge about the function of ATRX in α-globin expression is still limited. Previous work in our lab demonstrated that ATRX binds to G-rich tandem repeats in or near the genes whose expression is perturbed in ATR-X patients (Law et al., 2010). Specifically looking at its targets at the α-globin locus in human erythroid cells by ATRX ChIP-seq, ATRX is enriched at α-globin promoters, enhancers as well as a G-rich ψζ variable number tandem repeat (ψζ VNTR), indicating ATRX may regulate the α-globin expression by binding to these regions (Truch et al., 2022). Ψζ VNTR is a highly polymorphic repetitive region located 1 kb upstream of the *HBM* promotor and it could form G4 structures that could be bound by ATRX *in vitro* (Law et al., 2010;

Y. Li et al., 2016). It has also been shown that ψζ VNTR lies within a block of linkage disequilibrium of SNPs that are correlated with the severity of α thalassaemia (Law et al., 2010). Furthermore, the VNTR size is variable between individuals and the longer the VNTR, the more downregulated is α-globin expression in ATRX patients (r=0.58). And the degree of downregulation of affected genes was found to be related to their proximity to the VNTR - the closer the gene is to the VNTR, the greater the effect (Law et al., 2010) (Figure 1.3). A similar effect was also observed at another ATRX target at the *NME4* locus (Law et al., 2010).

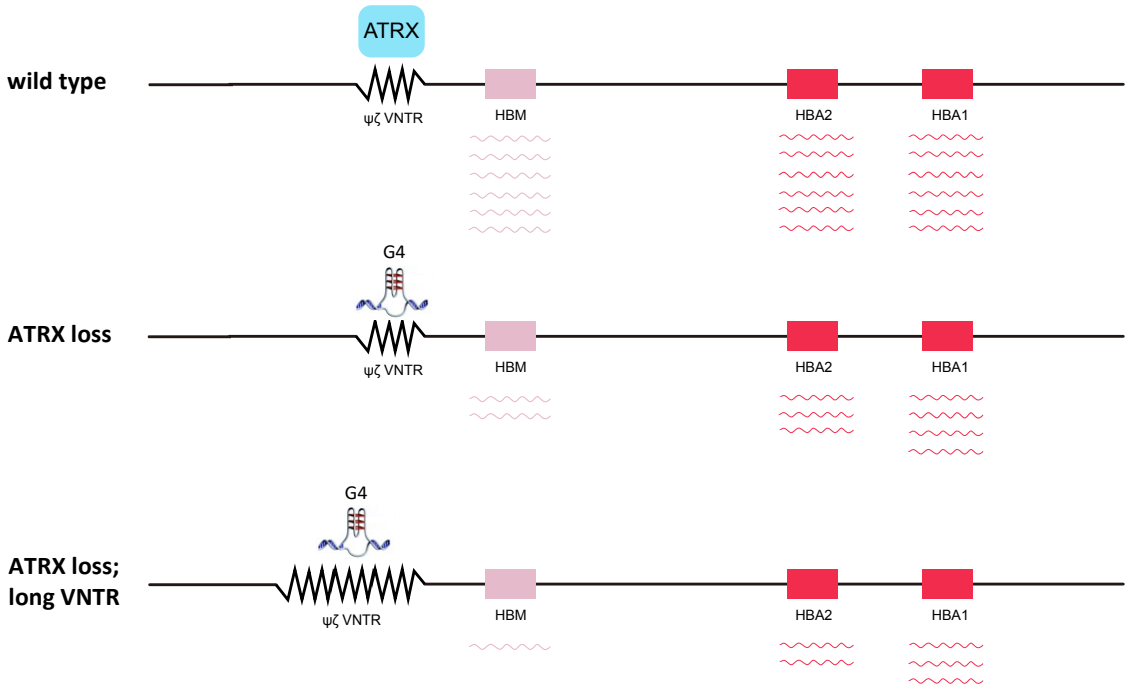


Figure 1.3 Downregulated gene expression at human α-globin locus in ATR-X syndrome.

The wavy coloured lines beneath the genes represent RNA production of the specific genes, and the varying number of the lines indicate the expression level change.

How ATRX regulates gene expression at the α-globin locus and how the size of the VNTR influences this process remain to be determined. Based on our previous work and other publications, we put forward the following three possible mechanisms and they may work in concert.

1.5.1 Replication stress induced epigenetic instability

Precise transmission of parental histone marks to new histones occurring during DNA replication is crucial for the transcriptional stability (Šviković & Sale, 2017). Any impediment on the DNA template, such as DNA damage or secondary structure, or problems associated with regulators of replication may affect this process and result in disturbed transcription.

Previously, work from our lab indicated that ATRX is required for DNA replication through G-rich sequences and in its absence, replication is stalled (Clynes et al., 2014). In addition, loss of ATRX is associated with increased R-loop formation in G-rich regions during transcription which is predicted to promote G4 formation (Nguyen et al., 2017), and loss of ATRX has been associated with an increase in G4 *in vivo* (Y. Wang et al., 2019). As previously discussed, ψζ VNTR at the α-globin locus is able to form G4 structure *in vitro* (Law et al., 2010; Y. Li et al., 2016). In ATR-X patients the epigenetic landscape has been shown to be perturbed with aberrant DNA methylation at repetitive DNA regions (R. J. Gibbons et al., 2000; Schenkel et al., 2017), as well as changes in histone modifications (Kernohan et al., 2010; J. Kim et al., 2019; Truch et al., 2022). One of the histone modifications associated with ATRX and DNA replication stress is macroH2A1, as described in section 1.2.5. MacroH2A1.2 has been shown

deposited at sites suffering replication stress or fork stalling and interacting with ATRX to protect DNA from damage (J. Kim, Sturgill, et al., 2018; J. Kim et al., 2019). In addition, macroH2A can act as an inert chromatin mark. One study concluded that ATRX negatively regulates the chromatin association of macroH2A at *HBA* genes in human erythroleukemic cell line K562 cells which contributes to the downregulation of α-globin in ATR-X patients (Ratnakumar et al., 2012). However, this result needs further testing in a more physiological erythroid cell system where α-globin genes are normally regulated, and particular attention should be paid to macroH2A's enrichment around *HBM*. Additionally, increased deposition of macroH2A at α-globin may be a secondary consequence of replicative stress when ATRX is absent.

All the above-mentioned evidence points to one possibility that replication stress at G-rich region induced by ATRX loss may alter local epigenetic signature and thereby dysregulate gene expression. During DNA replication, DNA becomes single stranded and G-rich regions including ψζ VNTR are particularly likely to form G4 structure in the absence of ATRX, which can lead to stalled DNA polymerase and induce biased incorporation of new histones lacking parental marks like active transcription associated H3K4me3, thus leading to epigenetic instability and changes in gene expression (Sarkies et al., 2010; Schiavone et al., 2014; Papadopoulou et al., 2015). In regard to the α-globin locus, ψζ VNTR is a potential site for replication stress occurring. The longer the ψζ VNTR, the more stable secondary structures are potentially formed, and the more severe replication stress and epigenetic instability. The closer the gene to the replication stress, the more affected its expression.

1.5.2 DNA double-strand break induced transcription silencing

In previous investigations, we showed a direct role of ATRX in DNA repair, and in its absence DSB and DNA damage response are generated (Clynes et al., 2014). Work by others indicated DSB-induced posttranslational modifications on local chromatin could inhibit distal transcription via an ATM kinase dependent manner (Shanbhag et al., 2010). And DSBs could promote a DNA damage signalling-dependent inhibition in gene expression specifically at break-bearing genes, which could be reversed upon DSB repair (J. Kim et al., 2016). Other evidence supporting the role of ATRX in transcription includes ATRX deficiency results in stalling of RNA polymerase II at G-rich intragenic sites, suggesting ATRX could facilitate transcriptional elongation through G-rich region (Levy et al., 2015).

Given the potential occurrence of DSB at ψζ VNTR in the absence of ATRX, it is possible that ATRX may regulate the affected genes by modulating their transcription. This process may act synergistically with the mechanism of replication stress induced epigenetic instability mentioned in the last section.

1.5.3 ψζ VNTR functions as an enhancer blocker boundary

There was one study showing that GATA repeats have enhancer blocker activity which prevents genomic enhancer-promoter communication and supresses gene expression (Kumar et al., 2013). Given the repetitive nature of ψζ VNTR and its location between α-globin enhancers and promoters, it might function as a blocker to cut off enhancer-promoter interaction and thus result in gene

repression. ATRX has been shown to bind to the repetitive ψζ VNTR (Law et al., 2010), with a possible role in helping resolve secondary structures to maintain genome stability. When ATRX is depleted, secondary structures are likely to form at the ψζ VNTR which may interfere with the interaction between enhancers and promoters, thereby leading to gene downregulation.

1.6 Human cellular models used for studying erythropoiesis

Cellular models that can explicitly recapitulate the *in vivo* physiologic and pathologic processes provide a powerful tool for scientific researchers to investigate the underlying mechanisms of particular biological events. They can originate from different species like human, mouse, fish, drosophila or yeast, and sometimes choosing the proper system is critical for reproducing *in vivo* processes. With regard to the erythroid cellular model used for investigating the mechanism by which ATRX regulates α-globin genes, a human cellular model is preferable to a mouse cellular model, as the potentially important element ψζ VNTR and the most-downregulated *HBM* gene are absent in the mouse genome, which makes it less likely to recapitulate the α-thalassaemia phenotype in mouse models. Therefore, only the human cellular models are introduced in this section.

For decades, a number of human cellular models have been used for investigating erythropoiesis. One of the initially used systems is the erythroleukemia cell line K562 cells, which are derived from a female patient with chronic myelogenous leukemia in terminal blast crises (C. B. Lozzio & Lozzio, 1975; B. B. Lozzio & Lozzio, 1979). Whilst K562 cells have provided some

insights into the regulation of globin genes, they exhibit an embryonic-fetal pattern of globin expression (Benz et al., 1980; Rutherford et al., 1981). Given that they are derived from a malignant clone carrying multiple mutations and have an extremely abnormal karyotype, K562 cells do not represent normal erythropoiesis or undergo terminal erythroid differentiation and thereby are not the optimum choice for erythropoiesis study.

There are several better cell models available for investigating the mechanism of disordered erythropoiesis. The first is induced pluripotent stem cells (iPSCs), which benefit from the cell reprogramming technologies that can revert the adult cells back to a pluripotent state and then the cells can be differentiated into the specific cell types of interest. These models maintain the genetic background of the individuals from which the cells were derived, therefore they are essentially human primary cell lines. A number of studies have reported the successful differentiation of human iPSCs along the erythroid pathway with high differentiation efficiency (Dias et al., 2011; Kobari et al., 2012; Olivier et al., 2016; Bozhilov et al., 2021), which is of value for basic studies of erythropoiesis.

The second option is an immortalized erythroid cell system that has been largely used to study erythropoiesis in recent years - human umbilical cord-derived erythroid progenitor (HUDEP) cell line. It was derived from umbilical cord human primary haematopoietic stem cells (HSC) CD34+ cells and was immortalised by lentiviral delivery of a tetracycline-inducible HPV16-E6/E7 transgene (Kurita et al., 2013). Three HUDEP clones were originally generated from three different cord blood samples; the HUDEP-1 and HUDEP-3 clones express the fetal γ-

globin, while the HUDEP-2 clone displays an adult globin expression pattern (Kurita et al., 2013). Due to their extended proliferative capacity and ease of genome editing, HUDEP cells have become one of the favourite erythroid cell models to study erythropoiesis, both in understanding the mechanisms of normal globin gene regulation and in revealing the mechanisms underlying erythropoiesis dysfunction in disease (Moir-Meyer et al., 2018; P et al., 2021; Bao et al., 2021; X. Li et al., 2021; R. King et al., 2021; Lu et al., 2022). In 2017, a similar immortalised erythroid line was generated called BEL-A cell (Bristol Erythroid Line-Adult), created using the same strategy as HUDEP cell generation but derived from adult CD34+ cells rather than cord blood CD34+ cells in generating HUDEP cell (Trakarnsanga et al., 2017). Although BEL-A cells have shown improved performance and more closely recapitulating adult erythropoiesis than HUDEP-2 (Daniels et al., 2020), these cells have not yet been extensively used.

The optimum model with the most physiological characteristics, involves primary cells, either directly from the individual with a particular genetic disorder or genome edited primary cells derived from a normal donor. The primary haematopoietic stem cells CD34+ cells derived from peripheral blood, bone marrow or umbilical cord have been widely utilized to yield mature erythroid cells via *in vitro* culture and differentiation systems (Migliaccio et al., 2002; Dolznig et al., 2005; Giarratana et al., 2005; Ronzoni et al., 2008). In addition, CRISPR/Cas9 system has also been successfully applied to obtain high gene editing efficiency in haematopoietic stem cells (Gundry et al., 2016; Mettananda et al., 2017; Lattanzi et al., 2019). However, their use is limited by restricted

access to donors and difficulties in their long-term *ex vivo* maintenance and expansion.

1.7 Genome engineering technologies

The unprecedented progress in the genome engineering technologies has ushered in an advanced age of functional genomics. These technologies edit the genome sequences typically by introducing DNA double-strand breaks, which are then repaired by the cell's endogenous DNA repair mechanisms, primarily by homology directed recombination (HDR) and non-homologous end joining (NHEJ). Depending upon the repair pathway employed by the cell, the DSB can either result in a gene disruption and consequent lack of expression due to introduction of nucleotide insertions or deletions (indels) via NHEJ, or precise gene editing with assistance of a donor template via HDR. The efficiency of NHEJ and HDR varies across gene loci and cell types.

There exist three main gene editing technologies currently: Zinc Finger Nucleases (ZFNs), Transcription Activator-Like Effector Nucleases (TALENs) and Clustered Regularly Interspaced Short Palindromic Repeats and associated proteins (CRISPR-Cas). They adopt distinct nucleases to cut DNA for creating DSB.

ZFNs is the initial technology that fuses FokI restriction enzyme with the DNA binding zinc finger domains, generating a protein that has both the ability to bind to DNA at a specific locus and then cleave it (Urnov et al., 2005; Geurts et al.,

2009; Rémy et al., 2010). However, limited targeting density, complicated construct design, variable efficiency and expensive cost have restricted its widespread use.

Inspired by zinc finger motifs and transcription activator like effectors (TALEs) derived from pathogenic bacteria found in plants, TALENs were invented, which have revolutionized the field of genome engineering (Boch et al., 2009; Moscou & Bogdanove, 2009; T. Li et al., 2011). TALENs confer an advantage in targeting efficiency and versatility over ZFNs, which makes the technology popular in a variety of areas, particularly in those model organisms that could not previously be modified. However, the construction of large-size DNA segments encoding TALE arrays can be challenging and the large construct may reduce its expression efficiency in a host organism.

In 2012, the first report of RNA-guided nuclease CRISPR-Cas genome editing system was published, where the researchers used Cas9 from *Streptococcus pyogenes* targeting to a specific locus to mediate DSB (Jinek et al., 2012). The CRISPR-Cas system is naturally present and functions in bacteria and archaea as adaptive immunity mechanisms. It integrates fragments of exogenous nucleic acids into CRISPR arrays embedded in microbial genomes and the transcripts of the integrated segments can work as guide RNAs for recognition and inactivation of the cognate targets, thus protecting prokaryotic cells from invading viruses and plasmids (Pourcel et al., 2005; Bolotin et al., 2005; Barrangou et al., 2007). The CRISPR-Cas9 system is comprised of two components - the sgRNA and the Cas9 endonuclease; the sgRNA is designed to bind to specific sequences within

the genome, and the Cas9 protein recognizes the protospacer adjacent motif (PAM) (3'-NGG-5') adjacent to the DNA sequence to introduce a DSB at the target locus. Owing to the versatility and ease of use, the CRISPR-Cas technology rapidly dominated the field of gene editing soon after it was introduced and continues to be the preferred technology. A variety of modifications to the initial Cas9 system have been developed, such as the improvement in specificity by mutating particular sites of the nuclease (Kleinstiver et al., 2016), genome-wide functional genomic screens (Shalem et al., 2014; Wang et al., 2014; Zhou et al., 2014), and the utilization of a catalytically inactive Cas9 (dCas9) to recruit proteins to specific DNA target sites allowing the regulation of gene expression (Qi et al., 2013), as well as approaches like base editing and prime editing (Komor et al., 2016; Rees & Liu, 2018; Anzalone et al., 2019). The CRISPR-Cas9 system in its many forms remains easy to design and use, and therefore accessible to the wider scientific community.

1.8 Summary and outline of the thesis

α-thalassaemia is a hallmark of ATR-X syndrome and results from the reduced expression of α-globin in erythroid cells. The α-globin locus has been studied for decades to understand the principles of gene regulation and a wealth of data has been generated, which makes it an ideal model for understanding how ATRX plays a role in regulating gene expression. Despite the increasing knowledge of ATRX's functions in chromatin organization and gene regulation, the underlying mechanism of ATRX regulating α-globin expression and how ATRX mutations

cause α-thalassaemia remain poorly understood. Therefore, this thesis attempts to establish and characterise an ATRX deficient erythroid cell model and to understand the mechanisms by which α-like globin and α-globin expression is downregulated.

Chapter 3 of the thesis describes the generation and characterisation of ATRX deficient haematopoietic stem cells, and the phenotype assessments of the differentiated erythroid cells both in bulk and single colony. The single colony analysis has allowed me to notice the heterogeneity of α(-like)-globin expression level in cells losing ATRX.

To look into the heterogeneity further, I have performed single cell RNA-seq described in **Chapter 4.** The gene expression pattern was compared between the cells with low α-like globin expression and the cells with high α-like globin expression to indicate the underlying mechanisms that make them distinct.

The nature of stochastic α(-like)-globin changes induced by ATRX loss makes the mechanism investigation more difficult in bulk cultures where only a subset of cells may be affected, but this could be facilitated if the population of cells with low α(-like)-globin expression can be enriched. **Chapter 5** describes the generation and characterisation of a α-like globin reporter cell line with an inducible ATRX degron in HUDEP-2 cells. This model would potentially allow us to investigate the function of ATRX and to validate the gene regulation mechanisms demonstrated by the single cell RNA-seq analysis.

Given the importance of ψζ VNTR in α(-like)-globin expression indicated by the previous work in our lab, I have also investigated its role in ATRX regulated α-globin expression, by either stabilizing its formed secondary structure or deleting it entirely. This is described in **Chapter 6**.

Finally, in **Chapter 7**, I summarise the observations in this thesis and discuss how these findings contribute to our understanding of gene expression regulated by ATRX.

2 Materials and Methods

2.1 Cell culture and related techniques

2.1.1 Isolation of primary CD34+ haematopoietic stem/progenitor cells

Healthy adult peripheral blood CD leucocyte cone samples were obtained from NHS Blood and Transplant (roughly 10 ml) and stored at room temperature until process. Once the cone was collected, the CD34+ cells were separated straightly. Each cone sample was dropped into two 50 ml Corning falcon 50mL conical centrifuge tubes and diluted with 45 ml phosphate buffered saline (PBS, Gibco, Cat No 10010023). Then 10 ml of the mixture was overlaid onto Histopaque-1077 (Sigma, Cat No H8889) in round bottom culture tubes and centrifuged for 25 min at 1800 rpm without brake at room temperature. Harvested peripheral blood mononuclear cells (PBMCs) were washed in PBS and MACS buffer (PBS containing 2 mM ethylenediaminetetraacetic acid (EDTA, ThermoFisher, Cat No AM9260G) and 0.5% bovine serum albumin (BSA, Sigma, Cat No A9576)) and CD34+ cells were purified using positive selection by magnetic activated cell sorting using Human CD34+ Microbead kit (Miltenyi Biotec, Cat No 130-046-702) following the manufacturer's instruction. To increase the purity, eluate containing CD34+ HSPCs was re-purified using a second magnetic column. Both the CD34- faction and the CD34+ cell fraction were cryopreserved in freezing media consisting of 90% foetal blood serum (FBS, Gibco) and 10% dimethyl sulfoxide (DMSO) until required (in aliquots of 100,000 to 300,000 cells per vial/1 ml).

2.1.2 Maintenance of primary CD34+ haematopoietic stem/progenitor cell

$1x10^5$ to $3x10^5$ frozen CD34+ cells were thawed and recovered into 0.5 ml StemSpan SFEM II media (StemCell Technologies, Cat No 9655) supplemented with 100 ng/ml SCF (PeproTech, Cat No 300-07), 100 ng/ml TPO (PeproTech, Cat No 300-18), 100 ng/ml Flt3L (PeproTech, Cat No 300-19) and 1x Penicillin/Streptomycin (P/S, Gibco, Cat No 15070063) for 24 hours at 37°C with 5% CO_2 incubation prior to electroporation. After electroporation, recover the cells into the same maintenance media without P/S for 24 hours until starting differentiation.

2.1.3 Erythroid differentiation of HSCs

CD34+ cells were differentiated along the erythroid lineage using a previously described three-phase liquid culture system (C. Scott et al., 2020), requiring a common base media IMDM (Gibco, Cat No 31980022) containing 3% Solvent/Detergent pooled Human AB plasma (NHSBT), 10 µg/ml human insulin (Merck, Cat No I9278), 3 U/ml heparin (Merck, Cat No H3149), 3 U/ml erythropoietin (EPO, OUH hospital pharmacy), 200 µg/ml iron saturated human holo-transferrin (Bio-Techne, Cat No 2914-HT), 2% (v/v) FBS and 1x P/S. Different combinations of cytokines were supplemented depending on the day of differentiation. In phase I (day 0 to day 7), 1 ng/ml IL3 (PeproTech, Cat No 200-03) and 10 ng/ml SCF were added. In phase II (day 7 to day 11), the IL3 was removed and only 10 ng/ml SCF was supplemented. In phase III (post day 11), only 300 µg/ml holo-transferrin was added. On day 7 and day 11, media was

replaced with Phase II and III media, respectively. The media was changed every two days to maintain cell concentration around 1-2 x 10^5/ml in phase I and phase II, 0.5-1 x 10^6/ml in phase III. Cells were counted and monitored for viability throughout the differentiation process. Samples were harvested at regular intervals during the course of differentiation to confirm the proper differentiation by determining erythroid cell surface marker expression through flow cytometric analysis and cellular morphology through cytospin.

2.1.4 Colony-forming unit (CFU) assays

CFU assay was performed following the technical manual of Human CFU Assays Using MethoCult (STEMCELL Technologies, Cat No H4435). Specifically, cell suspension was diluted in Iscove's MDM with 2% FBS medium (STEMCELL Technologies, Cat No 07700) and then mixed into enriched MethoCult (STEMCELL Technologies, Cat No H4435) to allow progenitor cells to expand and differentiate into myeloid or erythroid colonies at 37°C with 5% CO_2 incubation as per the manufacturer's instructions. After 14-16 days of culture, colonies were inspected and scored to get the number of different type of colonies based on the morphology. Then the single individual BFU-E colony was photographed and picked manually into 0.2 ml PCR tubes containing cold PBS under a light inverted microscope, EVOS XL Imaging System (Manufacturer ThermoFisher Scientific, Cat No AME3300). Then each sample was split into two aliquots - one for genotyping following whole genome amplification (section 2.4.5), one for Fluidigm Biomark gene expression analysis (section 2.4.6). The collected samples in PCR tubes were spun down at 3000 rpm for 10 mins and

the supernatant was carefully discarded. The samples can be directly proceeded to genotyping/gene expression experiments or snapped frozen and stored at -20°C until future use. Cells were counted throughout the differentiation process.

2.1.5 Expansion of HUDEP-2 cells

Wildtype HUDEP-2 cells were obtained from Cell Engineering Division, RIKEN BioResource Center, Tsukuba, Japan(Kurita et al., 2013). During the expansion phase, cells were cultured in StemSpan SFEM medium (STEMCELL Technologies, Cat No 09650) supplemented with 50 ng/ml SCF, 3 U/ml EPO, 0.4 µg/mL dexamethasone (OUH Hospital pharmacy), 2mM L-Glutamine (Gibco, Cat No 25030081), 1x P/S and 2 µg/ml doxycycline (Sigma-Aldrich, Cat No D9891) at 37°C with 5% CO_2 incubation. Cell density was maintained between 0.5 and $1x10^6$ cells/ml. The media is required to change completely every two days to maintain the efficient doxycycline concentration. Cells were kept for a maximum of three weeks.

2.1.6. Erythroid differentiation of HUDEP-2 cells

Differentiation of HUDEP-2 cells used a two-phase liquid culture system, requiring a common base media IMDM containing 3% Human AB plasma, 10 µg/ml human insulin, 3 U/ml heparin, 3 U/ml EPO, 200 µg/ml holo-transferrin, 2% (v/v) FBS and 1x P/S. From day 0 to day 6, 1 ng/ml IL3, 50 ng/ml SCF and 2 µg/ml doxycycline were added to the base media for cell differentiation. After day 6, media was replaced to next phase, where only 300 µg/mL holo-transferrin was

supplemented to the base media. Media was changed every two days. Cell density was maintained at $1x10^6$ cells/ml before day 6 and at $1.5x10^6$ cells/ml thereafter. Cells were counted and monitored for viability throughout the differentiation process. Samples were harvested at regular intervals during the course of differentiation to confirm the proper differentiation by determining erythroid cell surface marker expression through flow cytometric analysis (section 2.2.1) and cellular morphology through cytospin (section 2.2.5).

2.2 Cell biology techniques

2.2.1 Flow cytometry of erythroid cell surface marker staining in live cells

$1x10^5$ cells were washed with PBS and re-suspended in 100 µl FACS buffer (1x PBS supplemented with 2% FBS and 2 mM EDTA) and stained with an erythroid differentiation panel of fluorophore-conjugated antibodies (Table 2.1, use 1 µl of each) and with Hoechst-33258 (ThermoFisher) for live/dead analysis. FACS was carried out on an Attune NxT (ThermoFisher), voltages and compensation were set using Ultra Comp eBeads (ThermoFisher) for antibodies and single stained cells for dyes. Gating was set using fluorescence-minus-one (FMO) controls and unstained controls. Analysis was performed using FlowJo software.

Surface marker	Protein	Fluorophore	Supplier	Cat No
CD235a	GPA	PE	BD Bioscience	555570
CD71	Transferrin	PerCP/Cy5.5	Biolegend	334114
CD49D	α-Integrin	APC	BD Bioscience	561892
CD34	CD34 Antigen	PE/Cy7	Biolegend	343616

CD233	Band3	FITC	IBGRL	9439FI
CD36	Platelet glycoprotein	APC/Cy7	Biolegend	336213
Hoechst 33258	Viability dye	violet	Invitrogen	H3569

Table 2.1 Antibodies of erythroid surface markers used for FACS erythroid differentiation detection.

2.2.2 Flow cytometry of intracellular staining in fixed cells

2.2.2.1 Flow cytometry of globin staining in fixed cells

$2x10^5$ cells were harvested and washed in PBS. Then the cells were resuspended in 900 µl of PBS and fixed on the roller at room temperature for 15 minutes by adding 100 µl (1/10 volume) of 37% formaldehyde to the suspension. After spin down, the cells were resuspended in 2 ml of PBS and can be either directly proceeded to antibody staining or stored in the fridge for up to 1 month. The cells were permeabilized with 1x Permeabilization Buffer (eBioscience, Cat No 88-8824) prior to antibody staining. For antibody staining, the directly conjugated primary antibodies were added to the cells in 1x Permeabilization Buffer and incubated for 30 minutes at room temperature protected from light (Table 2.2). The stained cells were then washed with PBS and analyzed on Attune NxT or BD LSR Fortessa machine. Gating was set using single-stained controls, FMO controls and unstained controls. Analysis was finished using FlowJo software.

Protein	Fluorophore	Volume	Supplier	Cat No
α globin	FITC	1 µl	Santa Cruz Biotechnology	sc-514378
β globin	Alexa Fluor 680	1 µl	Santa Cruz Biotechnology	sc-21757
γ globin	Alexa Fluor 594	5 µl	Santa Cruz Biotechnology	sc-21756

Table 2.2 Antibodies used for FACS intracellular globin staining.

2.2.2.2 Flow cytometry of nuclear ATRX staining in fixed cells

The antibody used for ATRX staining was labelled with fluorescent dye using APEX Antibody Labeling Kits (Invitrogen, Cat No A10478). Anti-ATRX antibody ab97508 (Abcam, rabbit polyclonal) was directly conjugated with Pacific Blue following the manufacturer's guidance and instruction. For ATRX staining, $2x10^5$ cells were harvested and washed in PBS. Then the cells were resuspended in 1 ml of Fixation/Permeabilization working solution provided in the Foxp3 / Transcription Factor Staining Buffer Set (eBioscience, Cat No 00-5523-00) and incubated on the roller at room temperature for 30 minutes. The cells were permeabilized with 1x Permeabilization Buffer prior to antibody staining. Next, the directly conjugated ATRX antibody was added to the cells in 100 ul of 1x Permeabilization Buffer and incubated for 30 minutes at room temperature protected from light. The stained cells were then washed with FACS buffer and analyzed on BD LSR Fortessa machine. Analysis was finished using FlowJo software.

2.2.3 FACS sorting

2.2.3.1 Haematopoietic progenitor stem cell sorting for CFU assays

Cells were washed with FACS buffer and spun down at 300 g for 10 mins. After discarding the supernatant, cell pellets were suspended in 100 µl of human FcR blocking reagent (Miltneyi Biotech, Cat No 130-059-901) diluted in FACS buffer and incubated for 5 mins at room temperature. They were then stained with eleven fluorophore-conjugated antibodies (Table 2.3) for 20 minutes at 4°C and sorted into 1.5 ml eppendorf microcentrifuge tubes with 100 µl IMDM with 2% FBS medium (medium for preparing and washing samples for CFU assays) using a BD LSR Fortessa analyzer or FACS Aria III cell sorter (Becton Dickinson). Gates were set with unstained controls and FMO controls. Data were analyzed on FlowJo software.

Surface marker	Fluorophore	Volume	Supplier	Cat No
CD34	PE/Cy7	2 µl	eBioscience	25-0349-41
CD36	PerCP/Cy5.5	2 µl	Biolegend	336223
CD71	FITC	5 µl	eBioscience	11-0719-41
CD41a	EF450	2 µl	eBioscience	48-0419-41
CD105	PE	2 µl	Biolegend	323205
Lineage cocktail (CD 2/3/56/14/16/19/235a)	APC	3 µl	Thermo Fisher Scientific	22-7776-72
CD38	AlexaFlour 700	2 µl	Biolegend	303523
CD45RA	APC-eFluor 780	2 µl	eBioscience	47-0458-41

CD123	BV605	8 µl	BD Bioscience	564197
CD90	BV711	5 µl	BD Bioscience	740786
CD10	APC	5 µl	eBioscience	17-0106-42
Fixable Viability	BV510	1 µl per 1ml	BD Horizon	564406

Table 2.3 Antibodies used for haematopoietic progenitors sorting.

2.2.3.2 Single nucleated erythroid cell sorting for Smart-seq2

Viability dye Hoechst 33342 (ThermoFisher, Cat No 62249) was added to the culture system of CD34+ differentiated day 13 erythroid cells (control AAVS1 and ATRX KO cells) at a final concentration of 1 µg/ml and incubated with the cell in the 37°C incubator for 5 mins. The cells were then washed and re-suspended in FACS buffer. For cell sorting, one single live nucleated cell was directly sorted into each well of cold 384-well PCR plate (FrameStar 4titude, Cat No 4ti-0384/C) containing cell lysis buffer on Sony SH800S Cell Sorter. The single cell purity mode was used, and the event rate was kept as low as less than 1000/s. After finishing the sorting of a whole plate, the plate was covered with an aluminium PCR film (StarLab), spun down and incubated for 5 minutes at room temperature to allow for protease digestion. Then the plate was snapped frozen on dry ice and stored at -80°C for up to 3 months.

2.2.3.3 Genome-edited cell sorting

HUDEP-2 cells after genome editing through CRISPR-Cas9 were sorted to screen the desired cell clone. The bulk edited cells were sorted based on the

positive signal of fluorescent protein expressed in CRISPR-Cas9 plasmid pX458 on FACS Aria Fusion Cell Sorter (BD Biosciences).

2.2.4 Mass cytometry (CyTOF)

2.2.4.1 Antibody labelling with metal conjugates

Commercially available purified antibodies against β-globin and ATRX were conjugated to specific lanthanide metals using Maxpar X8 Antibody Labeling Kit according to manufacturer instruction (Fluidigm, Cat No PRD002 Rev 14) as described below. Partial reduction of 100 µg of each antibody was performed by incubation in 4mM Tris (2-carboxyethyl) phosphine hydrochloride (TCEP) for 30 min at 37°C. Partially reduced antibodies were then conjugated with lanthanide-loaded polymer for 1h at 37°C. Conjugated antibodies were quantified by measuring the absorbance at 280 nm prior to storage at 4°C in Candor Biosciences Antibody Stabilizer (Cat No 130050) supplemented with 0.05% sodium azide.

2.2.4.2 Mass cytometry nuclear antigen staining

The nuclear antigen staining protocol was used for cytoplasmic haemoglobin and nuclear protein ATRX detection through CyTOF. $3x10^6$ cryopreserved cells were thawed and incubated in their respective original culture media with required cytokines for 1h at 37°C. Staining were then performed following previously described protocol (Palii et al., 2019). Briefly, cells were washed with Maxpar PBS Buffer (Fluidigm, Cat No 201058) and stained with 0.5 µM cisplatin (Fluidigm, Cat No 201064) for 5 min at room temperature to exclude dead cells. Prior to

staining, the cells were washed with Maxpar Cell Staining Buffer (Fluidigm, Cat No 201068) twice, fixed for 30 min at room temperature in 1 ml Maxpar Nuclear Antigen Staining Buffer 1X (Fluidigm, Cat No 201063) and permeabilized by washing twice with 2 ml Barcode Perm Buffer 1X (Fluidigm, Cat No 201057). Cells were then stained with primary antibody cocktail listed in Table 2.4 for 45 min at room temperature. Next, Cells were washed and further stained with secondary antibodies listed in Table 2.4 for 30 min. The cells were then washed with staining buffer and fixed with 1.6% formaldehyde for 10 min at RT. After fixation, the cells were washed and incubated overnight at 4°C in 1 ml of intercalation solution (125 nM Cell-ID Intercalator-Ir (Fluidigm, Cat No 201192A) in Maxpar Fix and Perm Buffer (Fluidigm, Cat No 201067)). Samples were then proceeded to acquisition following Helios sample preparation protocol and acquired on Helios Mass Cytometer (Fluidigm). Helios was tuned and performance was checked before each run according to Fluidigm recommendation.

Protein	**Primary antibody**	**Secondary antibody**	**Volume**	**Supplier**	**Cat No**
α globin	HBA-FITC	144Nd-anti-FITC	2 µl 1st + 1 µl 2nd	Santa Cruz Biotechnology	sc-514378-FITC
β globin	HBB-170Er	-	2 µl	Santa Cruz Biotechnology	sc-21757
γ globin	HBG-PE	154Gd-anti-PE	7 µl 1st + 2 µl 2nd	Santa Cruz Biotechnology	sc-21756-PE
ATRX	ATRX-160Gd	-	1.5 µl	Abcam	218936

Table 2.4 Antibodies used for mass cytometry.

Primary antibodies against HBB and ATRX were directly conjugated with metals. HBA primary antibody was FITC-labelled and the detection was achieved by the combination use of 144Nd-anti-FITC secondary antibody. Similarly, HBG was detected by PE-labelled primary antibody and 154Gd-anti-PE secondary antibody.

2.2.5 Cellular morphological analysis – Wright's stain of cytospin

$1x10^5$ cells were resuspended in 200 µl PBS, spun (400 rpm for 5 mins) onto a glass slide in a Cytospin 4 Centrifuge (ThermoFisher). The slides were air-dried, stained with modified Wright's Stain using Hemateck slide strainer and mounted in DPX (Sigma, Cat No 06522). Cytospins were imaged using an Olympus BX 60 microscope at 200x and 400x magnification.

2.3 Genome engineering

2.3.1 Knocking out ATRX with RNP in CD34+ cells

2.3.1.1 sgRNA design and synthesis

sgRNAs were designed using two online sgRNA design tools - Synthego (https://www.synthego.com/products/bioinformatics/crispr-design-tool) and Benchling (https://www.benchling.com). The target gene was searched and the output sgRNAs recommended by the two tools were compared to confirm the final choices. Selection of sgRNAs is done based on the criteria of (a) have good ON-target score, (b) have the least possible overall calculated OFF-targets, (c) not have off-targets with less than 3 base-pair mismatches. The chosen sgRNAs with 2'-O-methyl 3' phosphorothioate modifications in the first and last three

nucleotides were synthesized at Synthego company. Cas9 protein (Alt-R® S.p. Cas9 Nuclease V3, Cat No 1081058) was bought from IDT company. The sequences of the targeting guides are listed in Table 2.5.

Name	Sequence (5'-3')	Purpose
sgAAVS1	GGGCCACTAGGGACAGGAT	CRISPR sgRNA
sg8	TAGCCGTGACTCAGATGGAA	CRISPR sgRNA
sg9-1	GTGTAAAGACTACACCCTTG	CRISPR sgRNA
sg9-2	GGGTGTAGTCTTTACACGTG	CRISPR sgRNA
sg15	GACGAAAACGTATTGCTGAG	CRISPR sgRNA
sg16	GTGTCCAATAACAACCAAGT	CRISPR sgRNA
ADD-sg8	CAUCCAUUCCAUCUGAGUCA	CRISPR sgRNA
ADD-sg9	UCUACGCAACCUUGGUCGAA	CRISPR sgRNA
ATRX multi-guideRNAs	ATCTATGACTTTATGCTCTT	CRISPR gRNA
	AAATTCCGAGTTTCGAGCGA	
	TACAAAATTACGTCAGCTTA	

Table 2.5 Guide RNA sequences used for knocking out ATRX in CD34+ cells.

2.3.1.2 RNP nucleofection into CD34+ cells

The CD34+ cells were transfected with ribonucleoproteins (RNPs) formed by guide RNA (gRNA) and Cas9 protein. Transfection of RNP complex into CD34+ cells was done using Amaza P3 Primary cell 4D-Nucleofector X Kit (Lonza, Cat No V4XP-3032) and 4D-Nucleofector X Unit device (Lonza). 6 ug Cas9 and 3.2 ug sgRNA were mixed in a 0.2 ml PCR tube and incubated at 25°C for 10 mins

to form RNP. During the incubation, culture medium containing $1x10^5$ cells was spun at 1200 rpm for 5 min. Then the cell pellet was resuspended in 20 µl nucleofection solution and was nucleofected together with the RNP in the cuvette provided using the DZ100 program. Cell suspension was finally transferred to the pre-warmed antibiotic-free SFEM II media and cultured for less than 24 hours either before the initiation of differentiation or before cell sorting.

2.3.1.3 Knockout efficiency assessment

The knockout efficiency at bulk level was examined by analyzing the genotype around the cut site of gRNA. Specifically, $1x10^5$ to $5x10^5$ cells were harvested during differentiation and DNA was extracted from the cells using DNeasy Blood & Tissue Kits (QIAGEN, Cat No 69504). Genotyping PCR against the targeted DNA region was performed using the primers shown in Table 2.6 and the PCR product was subject to Sanger sequencing. The sequencing results of control and test samples were uploaded to TIDE (https://tide.nki.nl) and ICE (https://ice.synthego.com) websites to do *in silico* knockout (KO) analysis.

Name	**Sequence (5'-3')**
ATRX-E8F	CAAAATGTTCTCCCATTCTGTG
ATRX-E8R	TTCGACCAAGGTTGCGTAG
ATRX-E9F	GAGAAATGTGGACTTGGACAGG
ATRX-E9R	CTTTCCCCGCCTGAGTCTTT
ATRX-E15F	TCTTTGCCAACACTTCCTTT
ATRX-E15R	CTTGAGCCCAAGAGTTCG
ATRX-E16F	GGGAGCAGGAGCAAGAGAG

ATRX-E16R	TCCAGGACTTAGCAGGATGTG
ATRX-ADD-9F	GATGATATTAGCCGTGACTC
ATRX-ADD-9R	GCGGAATAAGAGTAGGTTAC
AAVS1-F	CTTGCTTTCTTTGCCTGGAC
AAVS1-R	GATCAGTGAAACGCACCAGA
ATRX-multi-F	CAGAAGTGTTGGCAGGTTCA
ATRX-multi-R	GTCACCCAGAGCCTTTGTTG

Table 2.6 PCR primers used for genotyping ATRX KO in CD34+ cells.

2.3.2 Genome engineering in HUDEP-2 cells

2.3.2.1 Constructs design of tagging ATRX with degron in WT HUDEP-2 cells

All sgRNAs and donor template were designed, prepared and tested by Dr Philip Hublitz and colleagues in the Genome Engineering Core Facility at MRC Weatherall Institute of Molecular Medicine (WIMM). Guide RNAs were cloned into pX458-Ruby (Addgene 48138) vector, which is a modified vector with the GFP tag exchanged for an mRuby cassette and co-expresses Streptococcus pyogenes Cas9 (SpCas9) nuclease and Ruby, along with a single guide RNA (sgRNA) sequence for targeting. Three different gRNAs were tested all targeting C-terminal domain (CTD) of human ATRX gene (sequences shown in Table 2.7). At last, plasmid pX458-Ruby-ATRX-sg2 was used for transfecting HUDEP-2 cells given it achieved highest cutting efficiency tested by Surveyor Assay. The donor was designed to have 800 bp arms at each side, EGFP-FKBP12 F36V in between and directly linking to the absolute CTD of ATRX. All sgRNAs targeting

sites in donor plasmid were done point mutagenesis based on the sequences in the HUDEP-2 cell line without affecting the final encoding amino acids. The donor DNA was synthesised and inserted into pMK-RQ plasmid.

	Sequence (5' -> 3')	Appended sequence at 5' end
ATRX-sg1-F	GGCAATGAGAAGCAAAAATCC	CACC
ATRX-sg1-R	GGATTTTTGCTTCTCATTGCC	AAAC
ATRX-sg2-F	GGACATTGATTTCCCTTGGGA	CACC
ATRX-sg2-R	TCCCAAGGGAAATCAATGTCC	AAAC
ATRX-sg3-F	GGAAATCCAGGACCTTCCCAA	CACC
ATRX-sg3-R	TTGGGAAGGTCCTGGATTTCC	AAAC

Table 2.7 Guide RNA sequences used for generation of ATRX degron HUDEP-2 cell line.

Two Gs are placed before the guide RNA sequences because of the U6 promoter present in the plasmid. Nucleotide sequences "CACC" and "AAAC" were added to the 5' end of the forward and reverse sgRNA respectively for cloning using BbsI restriction enzyme.

2.3.2.2 Constructs design of tagging HBM with mKate2 in ATRX degron cells

General rules and practices of the construct design and clone were same as described in section 2.3.2.1. We aimed to knock in a far-red fluorophore mKate2 into the ATG of the human HBM locus. Three different gRNAs were tested all targeting exon 1 of the human HBM gene (sequences shown in Table 2.8). At

last, plasmid pX458-Ruby-HBM-sg2 was used for transfecting HUDEP-2 cells. Assembly of mKate2 donor ensured inactivation of all sgRNA protospacers by sheer deletion of exon 1 and had 550 bp arms at each side.

	Sequence (5' -> 3')	Appended sequence at 5' end
HBM-sg1-F	GCGGGCCACGAGGCGCAATT	CACC
HBM-sg1-R	AATTGCGCCTCGTGGCCCGC	AAAC
HBM-sg2-F	GGTCCCAGACCTGCGCGATT	CACC
HBM-sg2-R	AATCGCGCAGGTCTGGGACC	AAAC
HBM-sg3-F	GGAGCGCGCCCAAATCGCGC	CACC
HBM-sg3-R	GCGCGATTTGGGCGCGCTCC	AAAC

Table 2.8 Guide RNA sequences used for generation of HBM mKate2 reporter cell line.

Two Gs are placed before the guide RNA sequences because of the U6 promoter present in the plasmid. Nucleotide sequences "CACC" and "AAAC" were added to the 5' end of the forward and reverse sgRNA respectively for cloning using BbsI restriction enzyme.

2.3.2.3 Genome editing design of knocking out ψζ VNTR in ATRX degron cells

Generation of the ψζ VNTR KO cell line was achieved by means of a CRISPR/Cas9 dual guide strategy. ψζ VNTR lies within intron 2 of the pseudo-zeta gene. Given the high homology between *HBZ* and the pseudo-zeta gene,

the 5’ sgRNA was located 1.8 kb upstream of ψζ VNTR within the pseudo-zeta promoter where the sequences have no homology with HBZ and the 3’ sgRNA was at downstream of pseudo-zeta gene, which means the whole pseudo-zeta gene (~2.5 kb) would be deleted. The sequence of the 5’ sgRNA is GCAAACTGGAGCTCATAAGG and the 3’ sgRNA is GGTCGTCCCCACTGTCGTCG.

2.3.2.4 Generation of long single-stranded oligo DNA nucleotides (ssODNs) donor

Initial HBM tagging attempts were done with donor plasmid. However, the Homology directed repair (HDR) was inefficient. Therefore, ssODNs donor was generated for achieving higher efficiency. Briefly, special primers were synthesized from IDT, one of which is 5’ phosphorylated (5’PhoN-N-N----N3’) so this strand can be digested later on, and the other is not phosphorylated and contains 2 phosphorothioate * bonds between the last three 5’ bases. The sequence of the forward primer is 5’-GCTCTATAAGGAGGCCAGGG-3’ and the sequence of the reverse primer is 5’-ACCTGAGCAGCAGCTCCG-3’. The sense ssODN and anti-sense ssODN can be obtained by synthesising one primer with 5’ phosphorylation and the other with phosphorothioate bonds. Double-stranded DNA was synthesized using these two primers by PCR from plasmid template employing Q5 High-Fidelity DNA Polymerase (NEB, Cat No M0491). The PCR product was then purified by AMPure XP Beads (Beckman Coulter, Cat No A63881) and isopropanol precipitated. Then the double-stranded DNA was treated with λ exonuclease (NEB, Cat No M0262) to digest the phosphorylated

strand, purified by AMPure XP Beads (Beckman Coulter, Cat No A63881) and isopropanol precipitated. Final single-stranded donor was quantified using the NanoDrop (Thermo Scientific), and the integrity was checked on 1.5% agarose gel.

2.3.2.5 Nucleofection of HUDEP-2 cells

Prior to nucleofection, cells should be expanded for 5-7 days to have stable decent viability and in an exponential growth phase. The knock-in editing was performed following protocols in (Moir-Meyer et al., 2018). Briefly, 2.5 µg of pX458-Ruby CRISPR-Cas9 plasmid and 2.5 µg of donor were co-transfected into $2x10^6$ HUDEP-2 cells using 2B Amaxa Human CD34 Cell Nucleofector Kit (Lonza, Cat No VPA-1003) on AMAXA Nucleofector 2B (Amaxa) following manufacturer's instruction. The transfected cells were cultured in SFEM expansion media containing 0.75 µM RAD51-stimulatory compound-1 (RS-1) to increase HDR efficiency. Bulk Ruby+ cells were sorted two days later to recovering for 1 week and then single cells were sorted into Terasaki plates (Nunc, Cat No 163118) for clone screening. The sorted bulk Ruby+ cells and the expanded clones derived from single cell were then genotyped using the primers shown in Table 2.9.

For ψζ VNTR knock-out editing, 2.5 µg of pX458-GFP CRISPR-Cas9 plasmid containing 5' sgRNA and 2.5 µg of pX458-Ruby CRISPR-Cas9 plasmid containing 3' sgRNA were co-transfected into $2x10^6$ HUDEP-2 cells. The transfected cells were cultured in SFEM expansion media without RS-1. Bulk

GFP+Ruby+ cells were sorted two days later to recovering for 1 week and then single cells were sorted into Terasaki plates for clone screening. The sorted bulk cells and the expanded clones derived from single cell were then genotyped using the primers listed in Table 2.9.

	Sequence (5' -> 3')	**Length of product (bp)**	**Purpose**
ATRX degron-F	TAGTTAATCAGCTGTGTTCAGCG	Unedited: 417 bp; Edited: 1464 bp	ATRX degron cell screening
ATRX degron-R	TAAAACTAATATGGAAGATTGGC		
HBM-mKate2-F	GCGTGGGTCAATGCCAGAAAC	Unedited: 300 bp; Edited: 1128 bp	HBM-mKate2 cell screening
HBM-mKate2-R	CCTGAGCTCCCGGAGACCC		
VNTR-KO-F1	TCACCTTCTGTGAGGAGTCCG	Unedited: 3119 bp; Edited: 438 bp	VNTR KO cell screening
VNTR-KO-R	AGCCCCGACACGCTTCCAATACG		
VNTR-KO-F2	CAAGACCTACTTCCCGCACTTC	Having 1059 bp product in cells with heterozygous deletion; without product in cells with homozygous deletion.	
VNTR-KO-R	AGCCCCGACACGCTTCCAATACG		

Table 2.9 PCR primers used for screening edited HUDEP-2 cell clones.

2.4 Molecular biology techniques

2.4.1 DNA extraction, general PCR and Sanger sequencing

Genomic DNA from cell pellets was extracted using DNeasy Blood and Tissue Kit (QIAGEN). Genomic DNA from single BFU-E colony was amplified directly from cell lysate using Illustra single-cell GenomiPhi DNA amplification kit following manufacturer's instructions (GE Healthcare, Cat No 25-6600-30).

PCR reactions were performed with Platinum™ Taq DNA Polymerase High Fidelity (Invitrogen, Cat No 11304011). PCR primers used for amplifying *SRY* to confirm the gender of the donors are:
hSRY_F:TCAAGTTAGTACTCTGACGGCCA;
hSRY_R:AGGGCAAGTAGTCAACGTTACTG.

All Sanger sequencing was performed on PCR products purified with Qiaquick Gel Extraction Kit (QIAGEN, Cat No 28704) using the sequencing primers by WIMM sequencing facility.

2.4.2 RNA extraction, cDNA synthesis and qRT-PCR

0.5-2x10^6 cells were collected in 350 ul RLT buffer (from QIAGEN RNeasy Mini Kit), snapped frozen on dry ice and stored at -80°C until use. Total RNA was purified using RNeasy mini kit (QIAGEN, Cat No 74004) according to the manufacturer's instructions. The concentration of RNA was measured using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific). For reverse

transcription, 1 µg of total RNA was used to generate complementary DNA (cDNA) using SuperScript III First Strand Synthesis SuperMix (ThermoFisher, Cat No 11752-250) following the manufacturer's instructions. General quantitative reverse transcription PCR (qRT-PCR) were carried out using inventoried TaqMan assays (Applied Biosystems; TaqMan IDs: HBA2/HBA1-Hs00361191_g1, HBB-Hs00747223_g1, HBG-Hs00361131_g1, HBM-Hs01392876_g1, NME4-Hs00359037_m1, RPL13A-Hs03043885_g1 and RPS18-Hs01375212_g1) and Taqman Universal PCR Master Mix II (ThermoFisher, Cat No 4440040) in technical triplicate on StepOnePlus Real-Time PCR System (Applied Biosystems) according to the manufacturer's protocol. Nascent qPCR was performed using Luna universal qPCR master mix (NEB, M3003) in triplicates. Data were analyzed using the delta delta CT (ΔΔCT) method. RPL13A and RPS18 were used as a reference for normalization.

2.4.3 Droplet digital PCR

The PCR reaction mix (22 µl, containing: 11 µl ddPCR Supermix for Probes, 0.37 µl VIC-labelled HBA Taqman Assays, 0.37 µl FAM-labelled HBB Taqman Assays, 0.37 µl FAM-labelled HBG Taqman Assays, 2 µl 20x diluted cDNA (up to 50 ng initial RNA) and water) was set up at room temperature in semi-skirted 96-well ddPCR plate (Bio-rad, Cat No 12001925). The dilution of the cDNA depends on the expression abundance of the genes of interest. Starting with 1 ug of initial RNA for reverse transcription, 1:4000 cDNA dilution was finally used for globin gene detection. The plate was set up in columns of 8 including no template controls and no probe controls. If less than 8 wells in a column were used, the

rest of the wells should be filled up with 50/50 mix of Supermix and water. Then droplet was generated on Bio-Rad automated droplet generator. Post droplet generation, the oil/reagent emulsion was automatically transferred to a clean 96 well semi-skirted plate which was later sealed with the recommended foil seal and a PX1 PCR Plate Sealer. Afterwards, the samples were amplified on the Bio-Rad C1000 Touch Thermal Cycler with 96-Deep Well Reaction Module (95 °C for 10 min, followed by 40 cycles of 94°C for 30s and 55°C for 60s, with a final step of 98 °C for 10 min). The plate containing the droplet amplicons was subsequently loaded into the QX200 Droplet Reader (Bio-Rad, CA, USA). Standard reagents and consumables supplied by Bio-Rad were used, including cartridges and gaskets, droplet generation oil and droplet reader oil. Around 10,000-15,000 droplets were evaluated and the ration of α-globin to β-like globin was estimated by comparing the VIC positive droplets against FAM positive droplets.

2.4.4 Genotype analysis of single BFU-E colony

The BFU-E colony samples for genotyping were firstly subject to whole genome amplification (WGA) using illustra Single Cell GenomiPhi DNA Amplification Kit (GE Healthcare, Cat No. 29108039), followed by PCR and Sanger sequencing. WGA was performed following instructions of the kit. In brief, cells were lysed by adding 1 µl of lysis buffer to 1 µl of cells in PBS. The samples are then incubated for 10 min at 65°C before the addition of 1 µl of neutralization buffer to stop lysis. A reaction master mix is then added to perform whole genome amplification.

Afterward, the product was purified and used as the template for the following PCR.

2.4.5 High-throughput gene expression profiling on Fluidigm Biomark dynamic arrays

BFU-E colony samples used for Biomark gene expression analysis were collected in 2x Reaction Mix (with Superase-In) and freezed as described above. Before loaded on the chips, cDNA was pre-amplified. Specifically, a 0.2x assay mix which contains each Taqman assay diluted by low EDTA TE buffer (Thermo-Fisher, Cat No 12090-015) was prepared in the clean room to avoid contamination. When performing the pre-amplification, the samples were thawed, 2.5 µl 0.2x assay mix and 1.2 µl RT/Taq mix (from Superscript III 1 Step Kit, ThermoFisher, Cat No 12574026) were added to each sample, and the samples were amplified on Bio-Rad T100 Thermal (50 °C for 15 min, 95°C for 2 min, followed by 20 cycles of 95°C for 15s and 60°C for 4 min). The resulting cDNA was diluted with 4x vol TE buffer and stored at -20°C for up to 3 months. Then the diluted cDNA and Taqman Assays (all from Applied Biosystems, shown in Table 2.10) was loaded on 192.24 IFC chip for gene expression analysis as per protocols from Fluidigm website. The data analysis was finished in the built-in software, in line with the “Real-Time PCR Analysis” guideline from Fluidigm. The colonies having missing detection in the reference genes were excluded from the analysis. Mean of RPL13A, RPS18, ACTB and GAPDH was used as the reference for normalization.

RPL13A-1	Hs03043885_g1	PABPC1	Hs01598422_mH	HBQ	Hs00362218_g1
RPL13A-2	Hs03043887_gH	HBA	Hs00361191_g1	HBZ	Hs00923579_m1
RPS18-1	Hs01375212_g1	HBB	Hs00747223_g1	ATRX-1	Hs00997529_m1
RPS18-2	Hs02387368_g1	HBG	Hs00361131_g1	ATRX-2	Hs00230877_m1
ACTB	Hs01060665_g1	HBM-1	Hs01392875_g1	ATRX-3	Hs00997532_m1
GAPDH-1	Hs02758991_g1	HBM-2	Hs01392876_g1	NME4	Hs00359037_m1
GAPDH-2	Hs00266705_g1	HBD	Hs00426283_m1	NPRL3	Hs00429221_m1
RPL18	Hs00965812_g1	HBE	Hs00362216_m1	DAXX	Hs00985566_g1

Table 2.10 Taqman assays used for 192.24 IFC chip Biomark gene expression analysis.

2.4.6 Targeted bisulfite sequencing for DNA methylation detection

High quality genomic DNA was isolated from $5x10^6$ cells using Pure Link Genomic DNA kit (Invitrogen, Cat No K1820-01). Then unmethylated cytosines (C) in DNA was bisulfite-converted to uracils (U) using EpiTect Bisulfite Kit (QIAGEN, Cat No 59104) following the manufacturer's instructions. Primers were designed based on the sequences of the converted strand and covered the promoter region of the HBM gene. Avoid any 'CG' sequence in the primers, if there has to be one CG in the primer, replace C to Y which stands for unspecified pyrimidine nucleoside. The sequences of the primers used are F: GGGTGTTYGGAGGTTTTATAAGGAGGT, R: ACCCTCATCTCCACCTCCCTAAACTC. PCR was performed with the EpiMark

Hot Start Taq DNA Polymerase (NEB, Cat No M0490). Products were then purified and sequenced by Sanger sequencing.

2.4.7 Protein extraction and western blot

At least $2x10^6$ cells were collected and then lysed in 50-200 µl cell lysis buffer based on cell number. The component of cell lysis buffer is RIPA buffer (Sigma, Cat No R0278) supplemented with digitonin (ThermoFisher, Cat No BN2006), Benzonase (Merck, Cat No 71205-3) and protease inhibitor cocktail (Roche, Cat No 11836153001). Incubate for 40 min to 1 hour in cold room on end-to-end rotator. The lysate was then centrifuged at 13,000 rpm for 10 min at 4°C. Supernatant was collected and mixed with equal volume of 2x Laemmli Sample Buffer (Bio-Rad, Cat No 1610737) containing 1/4 volume of 1M DTT (ThermoFisher, Cat No R0861).

Before performing Western blot, the protein samples mixed with loading buffer were denatured at 95°C for 5 min and then chilled on ice at once. Electrophoresis was performed in a NuPAGE Bis-Tris gel (ThermoFisher Scientific) in a Mini-PROTEAN system (Bio-Rad). The gel was then transferred to a PVDF blotting membrane (Merck, Cat No 03010040001) overnight (14-18 hours) with constant current 25 mA and blocked with 5% skimmed milk (Merck, Cat No 70166) dissolved in TBS-T (Tris buffered saline+0.1% Tween-20). Primary antibody staining was performed at room temperature for 1h in 2.5% skimmed milk diluted with TBS-T. After primary antibody incubation, membrane was washed in TBS-T three times before one-hour secondary antibody incubation. After secondary

staining the membrane was washed for another three times and then exposed on X-ray film. Antibodies used in this study include: anti-ATRX (homemade mouse antibody, 39f), anti-p-ATM (Santa Cruz, mouse monoclonal, sc-47739), anti-GFP (Abcam, rabbit polyclonal, ab290), anti-FKBP-12 (Santa Cruz, mouse monoclonal, sc-133067), anti-gamma H2AX (Abcam, rabbit polyclonal, ab2893), anti-alpha tubulin (Abcam, mouse monoclonal, ab7291) which was used as a control, and secondary antibodies Goat anti-Mouse IgG (H+L) Secondary Antibody-HRP (ThermoFisher, Cat No 31430) and Goat anti-Rabbit IgG (H+L) Secondary Antibody-HRP (ThermoFisher, Cat No 31460).

2.4.8 Iso-electric focusing for haemoglobin analysis

Approximately $5x10^5$ cultured erythroblasts were lysed in haemoglobin elution solution and the haemoglobin analysed by iso-electric focusing (IEF) (RESOLVE® Haemoglobin kit, PerkinElmer, USA) using the manufacturer's instructions on a water-cooled horizontal electrophoresis rig (GE Healthcare). Gels were fixed in 10% trichloroacetic acid and stained with the JB-2 staining system (Perkin Elmer, USA) as per manufacturer instructions.

2.4.9 Bionano optical genome mapping

High molecular weight DNA from HUDEP-2 WT and ATRX degron cell line were isolated using the plug lysis protocol, in accordance with the Bionano Prep Cell Culture DNA Isolation Protocol (30026, Rev F). Labelling and staining of the genomic DNA were performed according to the protocol developed by Bionano Genomics: Bionano Prep Labeling - NLRS Protocol (30024, Rev K). Briefly, 750

ng genomic DNA was nicked by Nt.BspQI enzyme, then labelled, repaired and stained. The labelled samples were loaded onto the Bionano Chip and analysed using the Bionano Genomics Saphyr System according to the Saphyr System User Guide. The generated optical maps were assembled and structural variation detection was conducted using Bionano Access software.

2.5 Next generation sequencing related techniques

2.5.1 Smart-seq2 single cell RNA-seq

Smart-seq2 single cell RNA-seq was performed following the protocol described in (Picelli et al., 2014) and (Rodriguez-Meira et al., 2019). Cell lysis buffer containing barcoded Oligo(dT)-tailed oligonucleotides (same as (Rodriguez-Meira et al., 2019)) were dispensed in each well of 384-well plates on a MANTIS Liquid Handling Platform (FORMULATRIX) and a VIAFLO 96/384 (INTEGRA Biosciences). Then single cell was sorted into each well of the plates (section 2.2.3.2) and snapped frozen to be kept in -80°C for up to 3 months. Two whole 384-well plates were sorted for AAVS1 sample, and three whole 384-well plates were sorted for both AAVS1 sample and ATRX KO sample. Optimization experiment (retrotranscription and PCR) was then performed using several pilot single cell samples to determine the required cycles of cDNA amplification for our specific cell type, which turned out to be 21 cycles. Then retrotranscription and PCR steps were performed for the sorted single cells as described in (Rodriguez-Meira et al., 2019). Briefly, proteinase in the lysis buffer was heat inactivated by incubating the plate for 15 min at 72°C in a thermocycler, during which retrotranscription mix was prepared by mixing SMARTScribe retrotranscriptase

(Clontech, Cat No 639537) with RNase inhibitor (TaKara, Cat No 2313A) and template-switching oligo TSO-LNA (sequence: AAGCAGTGGTATCAACGCAGAGTACATrGrG+G, synthesized by IDT with RNase-free HPLC purification method). For each reaction of a single cell, 0.125 µl of RNase Inhibitor was mixed with 0.05 µl of 100uM TSO-LNA and 0.5 µl of SMARTScribe RT enzyme. Once heat inactivation step was finished, the RT mix was dispensed into each well to finish the first-strand synthesis in a thermocycler. Afterwards, the cDNA was amplified by SeqAMP (Clontech, Cat No 638509) using 21 cycles of PCR amplification as determined in the initial optimization step. Next, the products from the 384 wells of each plate were pooled, bead-purified, tagmented and indexed using a commercial Tn5 enzyme (Illumina Nextera XT Kit, Cat No FC-131-1096) to construct sequencing libraries quickly and efficiently from the amplified cDNA. The generated libraries were purified with AMPure beads and the quality were checked using a capillary array on TapeStation System (Agilent). The concentration of resulting libraries was quantified using the NEBNext Library Quant Kit for Illumina (NEB, Cat No E7630L). The libraries were sequenced on the NextSeq platform with 75 bp single-end reads using NextSeq 500/550 High-output 75 cycle kits (Illumina, Cat No 20024906).

2.5.2 Chromatin Immunoprecipitation followed by next generation sequencing (ChIP-seq)

Chromatin immunoprecipitation (ChIP) of ATRX was carried out on $50x10^6$ cells using the ChIP Assay Kit (Millipore, Cat No 17-295) following the optimised protocol described in (Truch et al., 2018). The cells suspended in PBS were fixed with 2mM ethylene glycol bis (succinimidyl succinate) (EGS, Sigma, Cat No

E3257) at room temperature for 45 min on a roller with constant mixing. Then formaldehyde solution was added to achieve 1% final concentration for the second fixation. The cross-linking reaction was quenched by addition of glycine (Sigma-Aldrich) to a final concentration of 125mM, then washed once in PBS and snapped frozen. Samples were stored at -80°C prior to further processing.

Cells were lysed in SDS lysis buffer with Complete Proteinase Inhibitor Cocktail (Roche). Samples were then transferred to AFA Fiber microtubes and fragmented using the Covaris ME220 ultra-sonicator with the following settings (Peak Incident Power (PIP): 75, Duty factor (DF): 14, Cycles/burst (cpb): 1000, Duration: 900s, Temp: 6°C). 1/10 volume of 10% Triton X-100 was added to the sonicated chromatin samples to neutralize the SDS in the lysis buffer. The samples were then spun down at 18400 × g for 10 min at 4 °C and the supernatant was collected. Before addition of the antibody, 5% volume of samples was taken to be used as an input. The chromatin was pre-cleared with 1:1 mix of Protein A and Protein G Dynabeads (Invitrogen) by incubating at 4°C for 30mins and discarding the beads. before immunoprecipitation overnight with the antibodies. The pre-cleared chromatin was incubated with the appropriate amount of antibody coupled to a 100µl 1:1 mix of Protein A and Protein G Dynabeads at 4°C overnight. On next day, chromatin bound beads were sequentially washed in Low Salt Buffer once, High Salt Buffer once, LiCl Buffer once and TE Buffer twice, with the final TE buffer wash at room temperature. Bound chromatin fragments were eluted from beads in elution buffer (1% SDS, 0.1M NaHCO3). Both ChIP sample and input sample were de-crosslinked by incubating samples at 65°C overnight, then proteinase K treated for one hour at

55°C. DNA was precipitated using phenol/chloroform extraction and ethanol precipitation with GlycoBlue (15mg/mL) (Thermo Scientific, Cat No AM9516) to visualise the pellet prior to dissolving in 20 µl of ultrapure water.

The concentration of precipitated DNA was determined by the Qubit dsDNA HS assay (Invitrogen, Cat No Q32854), then indexed using the NEBNext Ultra II DNA library prep kit (NEB, Cat No E7645L) following the manufacturer's instructions. Adapter stocks and primers stocks were diluted 1:10. Size selection was not performed. PCR amplifications was typically performed for 11 cycles. The concentration of resulting libraries was quantified using the NEBNext Library Quant Kit for Illumina (NEB, Cat No E7630L). The libraries were sequenced on the NextSeq platform using NextSeq 500/550 High-output 75 cycle kits (Illumina, Cat No 20024906).

2.6 Computational analysis

2.6.1 Analysis of Smart-seq2 single cell RNA-seq data

2.6.1.1 Data pre-processing

FASTQ files were generated using bcl2fastq with default parameters. Demultiplexed FASTQ files were trimmed for polyA tails using TrimGalore; files from different lanes were merged together using samtools and aligned to the human genome build hg19 using STAR. Counts for each RefSeq gene were obtained with FeatureCounts. Counts were normalized as follows: counts for

each single cell were divided by the total library size for that cell and multiplied by the mean library size of all cells processed. Genes with normalized count values less than 1 were considered non-detected and expression values for these genes were converted to zero. We further normalized counts into the log2 scale. QC filtering was performed using the following parameters: library size > 3000 reads; percentage of reads mapping to the mitochondrial chromosome < 10% and percentage of mapped reads per cell > 70%. After applying these QC filters, 634 cells from AAVS1 sample were retained and 991 cells from ATRX KO sample were retained.

2.6.1.2 Identification of highly variable genes

Similar to a previously described approach (Giustacchini et al., 2017), highly variable genes (HGV) above technical noise were identified by fitting a LOESS model of the log2(mean expression level) and coefficient of variation for each gene. Genes with a coefficient of variation above the fitted model and log2(mean expression) > = 0 were selected as HVG.

2.6.1.3 Dimensionality reduction and data visualisation

A Principal Component Analysis (PCA) reduction was computed using 20 dimensions. T-distributed stochastic neighbour embedding (t-SNE) was performed using 'Rtsne' package, the implementation of the method in R, with "perplexity = 10, max_iter = 500". Both PCA and tSNE scatter graphs were plotted using ''ggplot2'' package in R.

2.6.1.4 Differential gene expression analysis

Log2(normalized counts) matrix was used for analysis and genes expressed in at least five cells were retained. Differentially expressed genes were identified using a non-parametric Wilcoxon test to compare the mean expression level of each gene in both groups and expressing cell frequency was computed using Fisher′s exact test. *P* values were combined using Fisher's method and adjusted p values were derived using Benjamini & Hochberg (false discovery rate) procedure. Significant genes were selected on the basis of adjusted P value < 0.05 and absolute log2(fold change) > 0.5. Violin plots from selected genes were generated using ‘‘ggplot2’’ package in R.

2.6.1.5 Gene Set Enrichment Analysis of Gene Ontology (GO)

Gene set enrichment analysis (GSEA) was performed using the Clusterprofiler R package as previously described (Yu et al., 2012). Differentially expressed gene set was compared with the gene list from org.Hs.eg.db database from the Bioconductor package for GO analysis with settings ont = "BP", pAdjustMethod = "fdr".

2.6.1.6 Cell cycle stage assignment

A Seurat object was created using the original count table for computing cell cycle scores and assigning cell cycle stages. The data was normalized and then FindVariableFeatures and ScaleData were executed. Next, S-phase and G2M-phase cell cycle score were computed for each cell. The scatter plot was

generated based on the annotated cell cycle label of each cell. Cell cycle difference between groups was examined using Fisher's test.

2.6.2 Analysis of ChIP-seq data

Sequencing quality was assessed by FASTQC. ChIP-seq data was analysed using a customised in-house pipeline (Telenius et al., 2018) (script: https://github.com/Hughes-Genome-Group/NGseqBasic/releases). Paired-end sequencing reads were aligned to the human genome build hg19 using bowtie 2. Output BAM files were sorted and indexed using Samtools, then normalised using Deeptools to RPKM (--normalizeUsingRPKM), then converted to a bigwig file which can be visualised in the USCS genome browser.

3 Generation of ATRX deficient haematopoietic stem cells and phenotype assessments of the differentiated erythroid cells

3.1 Introduction

ATR-X syndrome patients exhibit α-thalassaemia. It has been demonstrated that *ATRX* mutations downregulate gene expression at the α-globin locus, with the α-like globin gene *HBM* most affected, followed by *HBA2* and *HBA1*. To unravel the underlying mechanism, in the absence of readily available erythroid cells from affected individuals, our lab has been endeavoring to establish a model that is able to recapitulate the *in vivo* observations and can be used for investigating the mechanism.

Previous researchers in the lab have developed several ATRX deficient models to address this, including both mouse and cellular models. Since the 'full' *Atrx* knockout mouse is embryonic lethal, dying at E9.5 (Garrick et al., 2006), and *HBM* is absent in the mouse genome, an erythroid-specific conditional *Atrx* knockout humanized mouse was developed where a 120 kb segment containing the mouse α-globin locus and its enhancers was replaced with the equivalent human locus. However, α-thalassaemia was not detectable in this mouse model (Higgs and Gibbons personal communication). Additionally, ATR-X patient-derived iPS cells and ATRX depleted erythroid progenitor HUDEP-2 cellular models were established, but neither of these was able to recapitulate the α-thalassaemia phenotype. Therefore, I attempted to knock out ATRX in a more

physiological cellular system, primary haematopoietic stem cells and differentiated the cells towards erythroid lineage to check their phenotypes.

Erythropoiesis defines the process of differentiation and proliferation from haematopoietic stem cells to mature red blood cells (RBCs), which includes the erythroid commitment of HSCs and the erythroid terminal maturation of precursors into circulating enucleated RBCs (Koury et al., 2002). At the molecular level, erythropoiesis is driven by a combination of factors that act in concert to direct globin gene expression. The *in vitro* cellular models of human erythropoiesis enable us to study normal or pathologic erythroid differentiation and to address the pathophysiological mechanisms underlying erythroid disorders. Primary HSC CD34+ cells derived from peripheral blood, bone marrow or umbilical cord have been successfully used to undergo erythroid differentiation and maturation to reproduce the different steps of normal human definitive erythropoiesis via the *in vitro* culture system (Ronzoni et al., 2008; Mettananda et al., 2018). With little access to patient's HSCs, here I have adopted the CD34+ HSCs derived from the normal donors and then knocked out *ATRX* in these cells via CRISPR/Cas9 to emulate the effects of natural *ATRX* mutations in the patients.

CRISPR/Cas9 genome editing of primary HSCs has been widely used both in basic research to elucidate the contribution of genetics to haematopoietic disorders, and in clinical trials to treat genetic disorders including β-thalassaemia and sickle cell disease (Frangoul et al., 2021). Currently, HSCs are primarily edited via the NHEJ pathway to inactivate a genetic factor or regulatory element,

given the relative inefficiency of the HDR process. An essential prerequisite for CRISPR/Cas9 to function is the efficient delivery of the complex into the target cells. The CRISPR/Cas9 complex can be delivered in the forms of plasmid DNA, messenger RNA (mRNA) or ribonucleoprotein RNP (Cas9 protein complexed with sgRNA). Direct delivery of the RNP complex has emerged as a powerful and popular approach for genome editing for several reasons. First, it enables the swiftest genome editing without the need for intracellular transcription and translation, and permits high editing efficiency with lower cytotoxicity (off-target effects, insertional mutagenesis and immune responses etc). Furthermore, RNP delivery offers a robust genome-editing method for primary cells that are often difficult to edit by conventional approaches.

Herein, I have used a CRISPR/Cas9 RNP directed genome editing strategy to create targeted mutations in *ATRX* through the NHEJ repair pathway in human HSCs to mimic the effects of natural *ATRX* mutations, and then examined if this reproduced the α-thalassaemia phenotype observed in the patients.

3.2 Results

3.2.1 *In vitro* erythroid differentiation and characterisation of CD34+ HSC cells

The CD34+ HSC cells were isolated from peripheral blood of healthy donors and were differentiated down the erythroid pathway. Wild type CD34+ cells from three different donors were differentiated to provide a normal reference. The

differentiation protocol used is a previously described three-phase erythroid *ex vivo* culture system (Griffiths et al., 2012; C. Scott et al., 2020), lasting for 21 days (Figure 3.1).

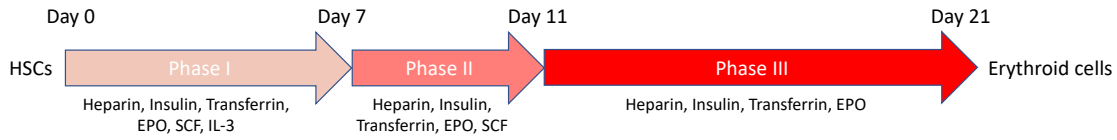


Figure 3.1 Schematic of experimental approach for a three-phase HSCs erythroid differentiation protocol.

The cells mainly focus on expansion in phase I, markedly commit to erythroid differentiation during phase II, and become enucleated in phase III.

During the whole expansion and differentiation process, the number of nucleated cells was checked on automated cell NucleoCounters, and cell samples were collected on specific days to determine the stage of differentiation. The number of the nucleated cells indicated exponential growth by day 13, with > $2x10^4$ fold expansion of erythroid cell numbers during the course of differentiation (Figure 3.2A). The number of the nucleated cells started to drop thereafter due to the process of enucleation (Figure 3.2A).

The differentiation process was tracked and comprehensively assessed by the changes in erythroid cell surface markers, cellular morphology, globin gene expression and haemoglobin analysis. The cell surface marker change was defined by flow cytometric analysis, which showed the expected changes during erythroid differentiation: the expression of glycophorin A (CD235a) and

erythrocyte membrane protein Band 3 increased whereas the expression of CD36, CD49d and transferrin receptor CD71 dropped during differentiation (Figure 3.2B). Cytospin staining showed that the cell morphology changed sequentially from pro-erythroblasts to basophilic erythroblasts, polychromatic erythroblasts, orthochromatic erythroblasts and finally becoming mature red cells after enucleation (Figure 3.2C). The globin gene expression was examined by RT-qPCR, showing the predominant and increasing expression of adult globins HBA and HBB during differentiation. HBM and HBG presented a similar increasing expression pattern, reaching their highest expression levels on day 13 (Figure 3.2D). In addition to the RNA level detection, the haemoglobin expression on protein level was examined by isoelectric focusing (IEF). The result demonstrated that the main haemoglobin expressed during the differentiation was adult haemoglobin HbA (α2β2), and there was subtle fetal haemoglobin HbF (α2γ2) expression after day 13, in line with the qPCR results showing that HBG (γ) attained a higher expression after day 13 (Figure 3.2E).

Taken together, the HSCs underwent expansion and erythroid definitive erythropoiesis to generate mature RBCs in our three-phase *ex vivo* culture system, providing a superior system to study erythropoiesis.

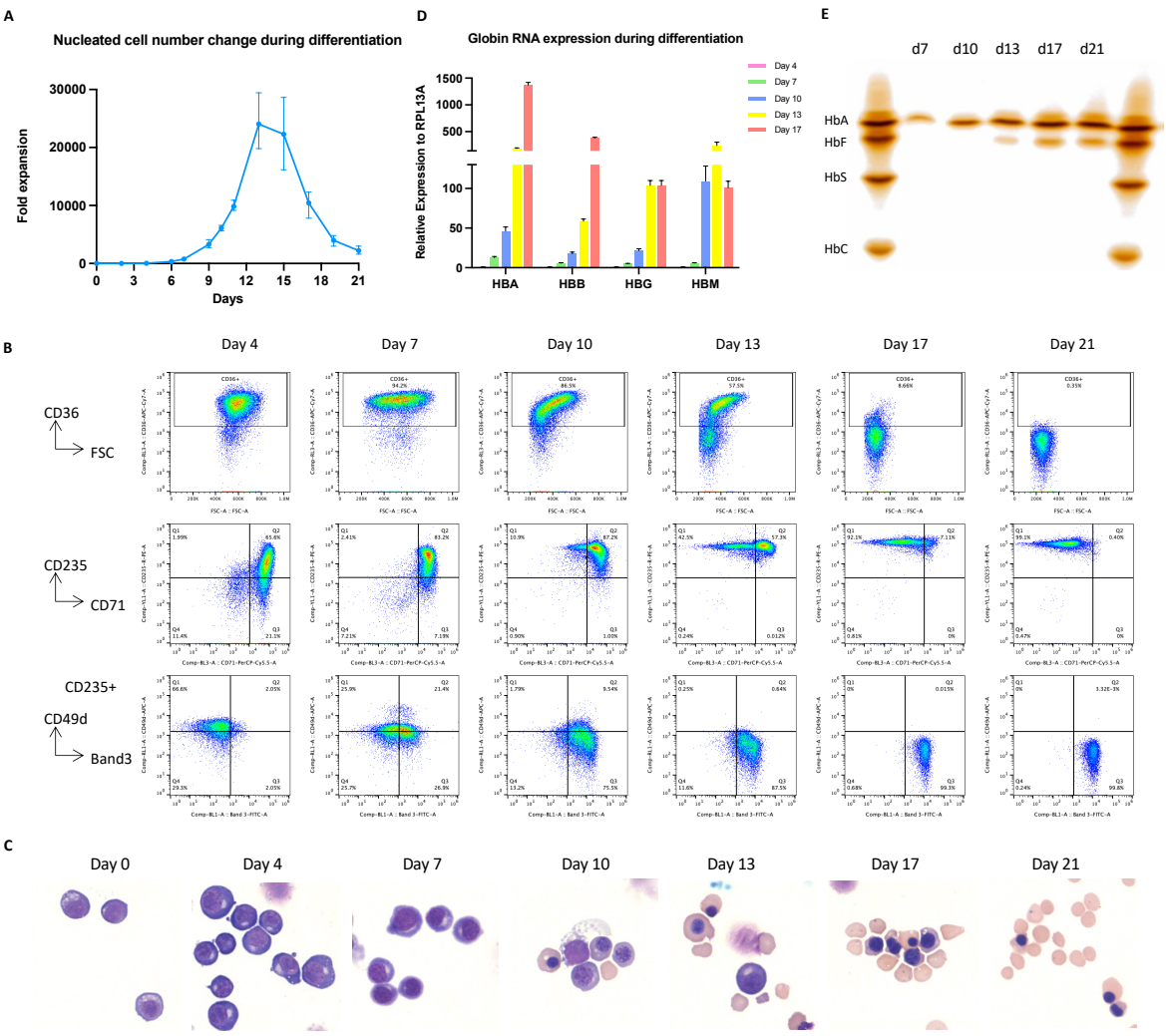


Figure 3.2 Characterisation of wildtype CD34+ HSCs erythroid differentiation.

(A) The nucleated cell number change during differentiation (n=3). Error bars show standard deviations (SD) from three independent experiments.

(B) Flow cytometric analysis of erythroid cell surface markers showing immunophenotype of cultured erythroblasts using a 6-colour antibody panel at day 4, 7, 10, 13, 17 and 21 of differentiation.

(C) Representative cytospins stained with modified Wright's stain (magnification 400x) showing cellular morphology changes during erythroid expansion (day 0-7), differentiation (day 7-10) and enucleation (day 13-21) phases.

(D) Taqman Probe RT-qPCR expression analysis for globin genes (HBA, HBB, HBG, HBM) in cultured cells during differentiation day 4 to day 17, normalized to RPL13A (n=3). (E) Iso-electric focussing (IEF) analysis of haemoglobin in cultured cells during differentiation day 7 to day 21. The left and right lanes were loaded with the haemoglobin AFSC control, which is a pool of normal blood samples with normal haemoglobins A and F and abnormal haemoglobins S and C.

3.2.2 Knocking out ATRX in CD34+ HSCs via CRISPR/Cas9

To recapitulate the α-thalassaemia phenotype in ATR-X syndrome, ATRX was knocked out in isolated CD34+ cells derived from leucocyte cones of healthy donors, and then the knockout cells were differentiated to erythrocytes using the three-phase culture system described in the last section (Figure 3.3).

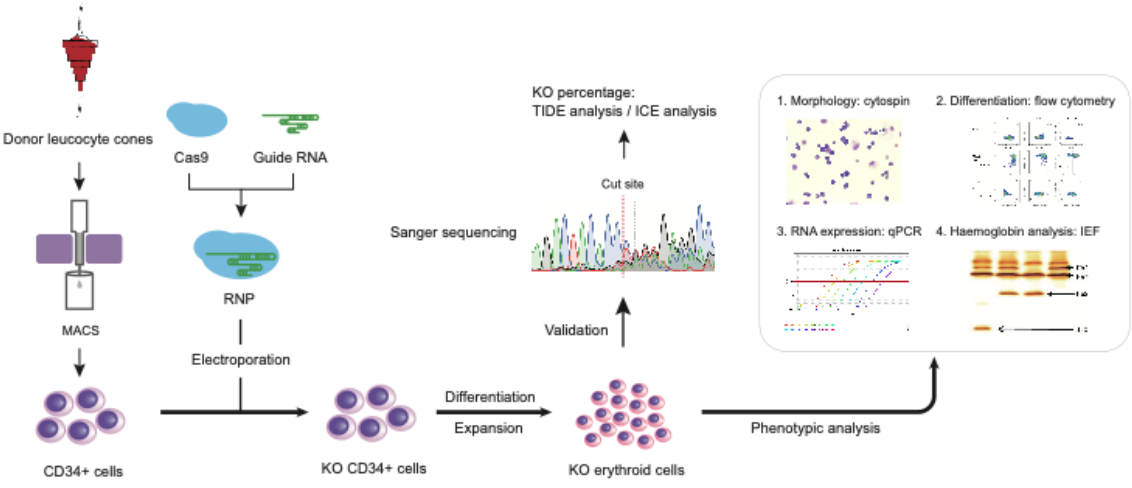


Figure 3.3 Workflow of establishing ATRX knockout erythroid cells from CD34+ HSCs.

CD34+ cells were isolated by magnetic-activated cell sorting (MACS) from healthy donors. ATRX was then knocked out in these cells by RNP complex formed by gRNA

and Cas9. The knockout cells were next differentiated towards erythroid lineage, during which knockout efficiency and erythroid phenotypes were examined.

3.2.2.1 Gender determination of donors

Since the *ATRX* gene is located on chromosome X and ATR-X syndrome is X-linked recessive inherited, it would be preferable to use the CD34+ cells derived from male donors for achieving an efficient ATRX knockout. However, we are blind to the donor's information, therefore the gender of the collected donors was firstly confirmed prior to the knockout experiments.

The genomic DNA of the donor was extracted from the CD34- fraction which was collected during CD34+ cell separation, and then the sex determination gene *SRY* on chromosome Y was amplified by PCR. Donors who have *SRY* would be male. The PCR products were checked by agarose gel electrophoresis which showed that all the collected eight cones were from male donors (Figure 3.4). This benefits the knockout of genes on chromosome X like *ATRX*.

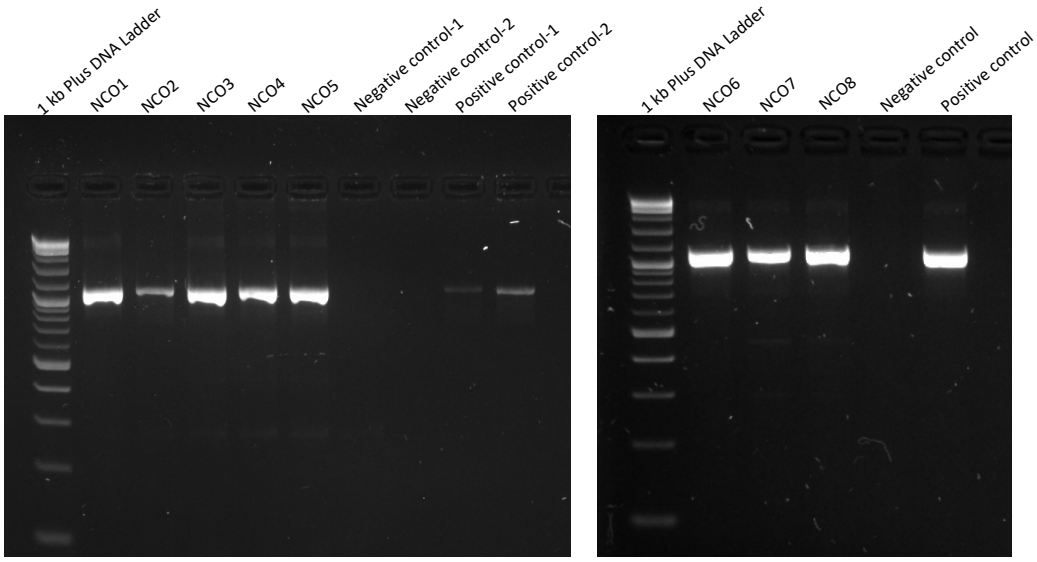


Figure 3.4 Gender determination of the collected donors.

The amplification of the sex determination gene *SRY* in collected eight cones (NCO 1-8) was assessed by DNA agarose electrophoresis. DNA samples from a female and male were used as negative and positive controls respectively. All the donor samples contained *SRY* which means all of them are from male donors.

3.2.2.2 Knockout efficiency of sgRNAs targeting ATRX

After the gender of the donors was confirmed, an ATRX knockout experiment was carried out using the CD34+ cells derived from the male donors. Seven different sgRNAs targeting different regions of *ATRX* were designed (Figure 3.5, sequences are shown in Table 2.5) and individually electroporated into the CD34+ cells to test their efficacy for knockout of *ATRX*.

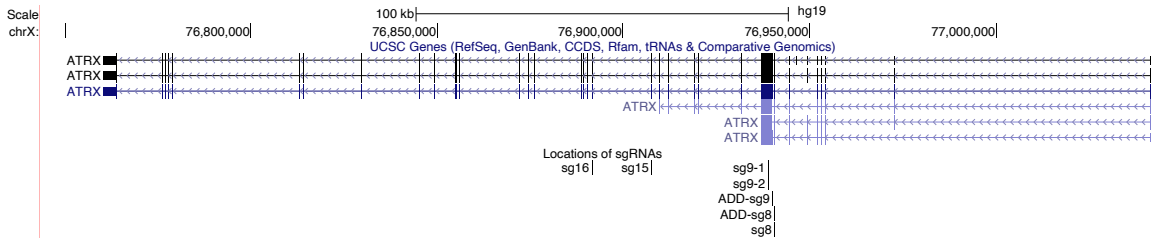


Figure 3.5 Schematic of the locations of sgRNAs within *ATRX*.

Seven sgRNAs were designed, targeting to exon 8, 9, 15 and 16, namely sg8, ADD-sg8, sg9-1, sg9-2, ADD-sg9, sg15 and sg16. Two of them, ADD-sg8 and ADD-sg9, lie within the ADD domain of ATRX.

The cells electroporated with sgRNA were expanded and differentiated down the erythroid pathway, and the cell samples were collected at differentiation day 6 to check the knockout efficiency (Figure 3.3). The corresponding targeting region surrounding the cut site of the sgRNA was amplified by PCR (primers shown in Table 2.6) and the product was then sent for Sanger sequencing. The knockout efficiency was finally assessed by *in silico* knockout analysis tool ICE by comparing the sequencing results of control and knockout samples, as described in section 2.3.1.3. One of the sgRNAs targeting exon 9 of *ATRX,* sg9-1, turned out to be the most efficient, with the highest knockout score of 84 (Table 3.1).

	indel percentage (%)	knockout (KO) score
sg9-1	91	84
ADD-sg8	93	55
ADD-sg9	95	46
sg8	63	33
sg9-2	50	40
sg16	24	20
sg15	11	10

Table 3.1 Knockout efficiency of sgRNAs targeting ATRX, ranked by KO score. Seven individual sgRNAs were used for ATRX knockout experiments and this table shows their KO effect. The indel percentage indicates the proportion of all detected sequences that are different from wild type, which includes indels that may or may not cause a knockout. The knockout score only includes sequences that are frameshift-inducing indels or deletions of 21 nucleotides in length or larger that lead to a putative knockout. The higher the KO score, the higher the percentage of sequences that are likely to result in a functional knockout of the targeted gene.

3.2.2.3 Knockout effect of sg9-1 on ATRX

Given that the knockout efficiency analysis indicated sg9-1 was the most efficient sgRNA for ATRX knockout, sg9-1 was used in subsequent knockout experiments. Meanwhile, a sgRNA targeting the adeno-associated virus integration site 1 (AAVS1) locus was used as a control. The *AAVS1* locus, in intron 1 of *PPP1R12C* (protein phosphatase 1 regulatory subunit 12C), is known as a "safe harbour" because its disruption does not have adverse effects on the cell, and

therefore has been widely used as a control in CRISPR gene knockout experiments (Hayashi et al., 2020).

To confirm that sg9-1 can efficiently exert a functional knockout effect on ATRX, the expression of ATRX at both mRNA and protein level was determined. The ICE KO analysis tool indicated an efficient KO by sg9-1 at the DNA level, with a 92 KO-score (Figure 3.6A). RT-qPCR analysis demonstrated that in ATRX KO cells, ATRX mRNA expression was downregulated to 46.5% of the wildtype level at day 7 and 60.5% at day 13 (Figure 3.6B). Consistently, western blotting and CyTOF analysis of ATRX protein demonstrated a robust ATRX protein loss in ATRX KO cells compared with AAVS1 control cells (Figure 3.6C, D).

To test if the knockout results were reproducible in CD34+ cells derived from different donors, I have repeated the knockout experiments with sg9-1 in other donor derived CD34+ cells. Interestingly, the KO efficiency of sg9-1 was variable between individual experiments and diverse donors, but it could always achieve over 50% KO effect which means it can always lead to at least half of the cells having an ATRX defect (Table 3.2).

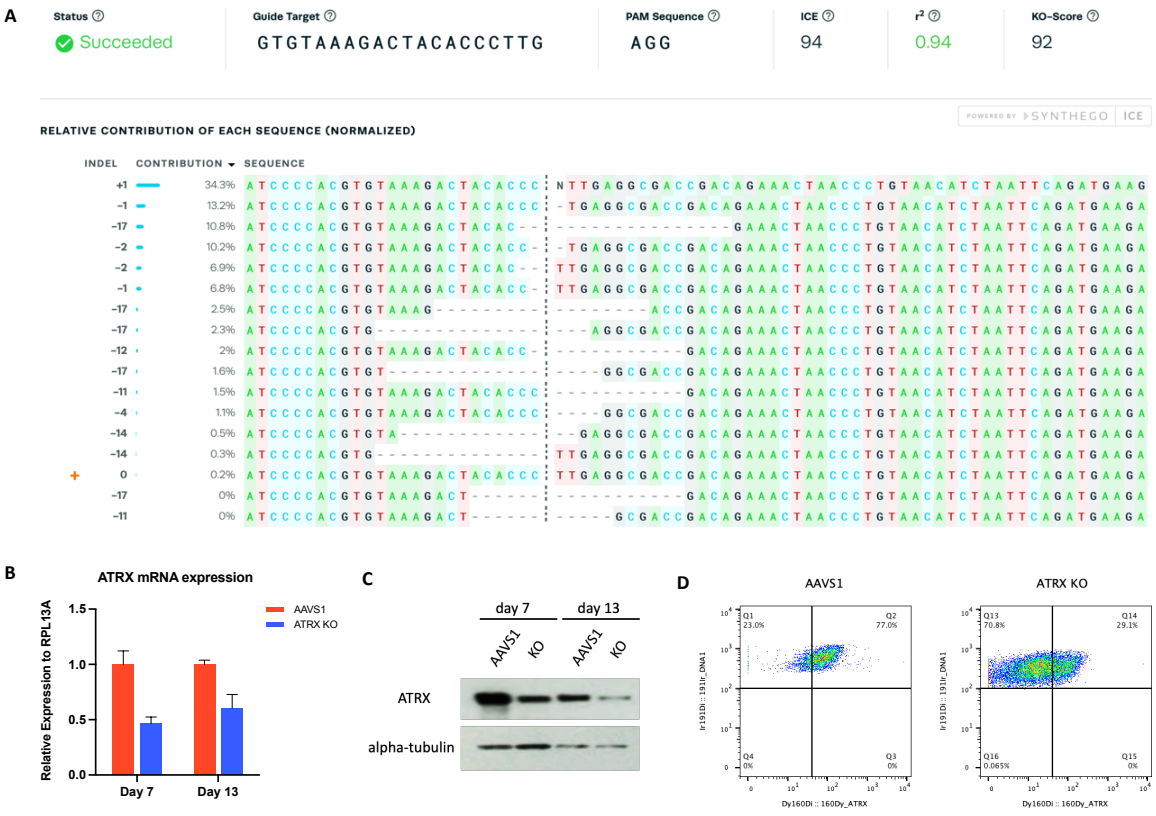


Figure 3.6 Knockout effect of sg9-1 on ATRX.

(A) ICE KO analysis of sg9-1, showing that 92% of the sequences resulted in frameshift indels.

(B) Taqman probe qPCR mRNA expression analysis of ATRX in differentiation day 7 and day 13 cells, normalized to RPL13A (n=3).

(C) Western blotting analysis of ATRX using ab97503 antibody against C-terminal of ATRX protein in differentiation day 7 and day 13 cells.

(D) CyTOF analysis of ATRX expression in day 7 cells showing a large population of cells in ATRX KO sample had reduced ATRX level. X-axis represents 160Dy labelled-ATRX expression and y-axis shows DNA channel 191Ir.

	Donor 1 (NCO 3)		Donor 2 (NCO 4)		Donor 3 (NCO 5)	
	indel percentage (%)	KO score	Indel percentage (%)	KO score	indel percentage (%)	KO score
1st test	87	76	83	52	92	83
2nd test	77	70	88	55	77	71
3rd test	70	61	94	88	94	92

Table 3.2 Knockout efficiency of sg9-1 in three independent knockout experiments in CD34+ cells from three donors.

CD34+ cells derived from three donors NCO 3-5 were electroporated with sg9-1 and the indel percentage and KO score were analysed by ICE KO analysis tool. Three independent experiments were performed for each donor.

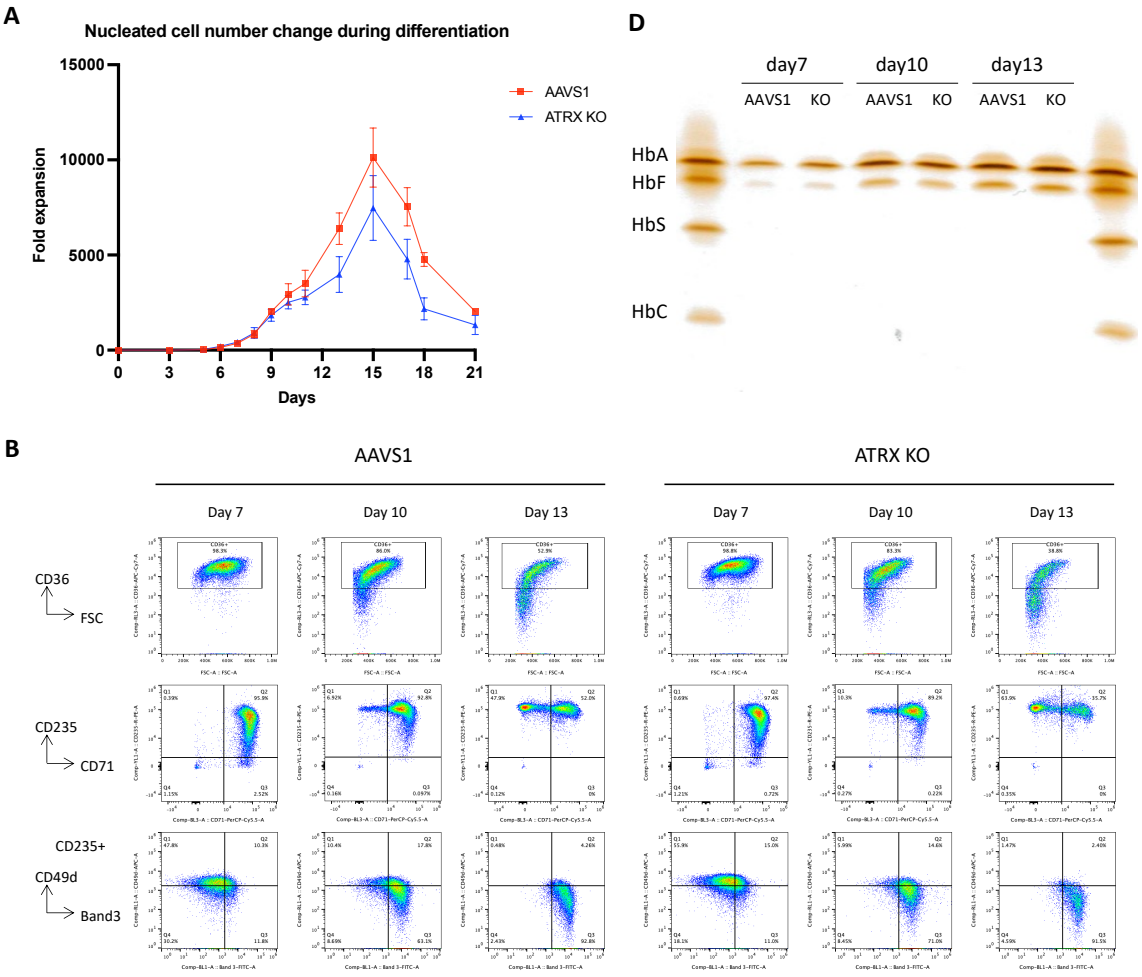
3.2.3 Phenotype assessment of ATRX KO erythroid cells at a bulk level

To assess the erythroid phenotypes of differentiated ATRX KO cells, the cellular morphology was determined by cytospin; cell surface markers were analysed by flowcytometric analysis; haemoglobin expression was examined by IEF; globin gene expression was detected by RT-qPCR and droplet digital PCR (ddPCR) at the RNA level, and by FACS and CyTOF at the protein level.

3.2.3.1 Erythroid phenotype analysis

The most efficient sgRNA sg9-1 was used to knock out ATRX in CD34+ cells derived from three different donors individually, and the cell phenotypes were examined. Phenotypic analysis indicated that knocking out ATRX had no significant effect on the erythroid differentiation process or cellular morphology (Figure 3.7B, C), although the cell growth was slightly slowed (Figure 3.7A). Haemoglobin analysis demonstrated that ATRX KO cells showed similar

haemoglobin expression as control AAVS1 cells, with dominant adult HbA expression and minor fetal HbF expression (Figure 3.7D). In ATR-X patients, about 75% of cases have an unusual mild form of α-thalassaemia manifested by the presence of cells with HbH inclusions rather than reduced red cell indices more commonly seen in classical forms of α-thalassaemia (R. J. Gibbons & Higgs, 2000). Therefore, the presence of HbH inclusions in ATRX KO cells was also checked with extreme care at the very late differentiation stage when the cells were enucleated, but we did not find any trace of HbH inclusions.



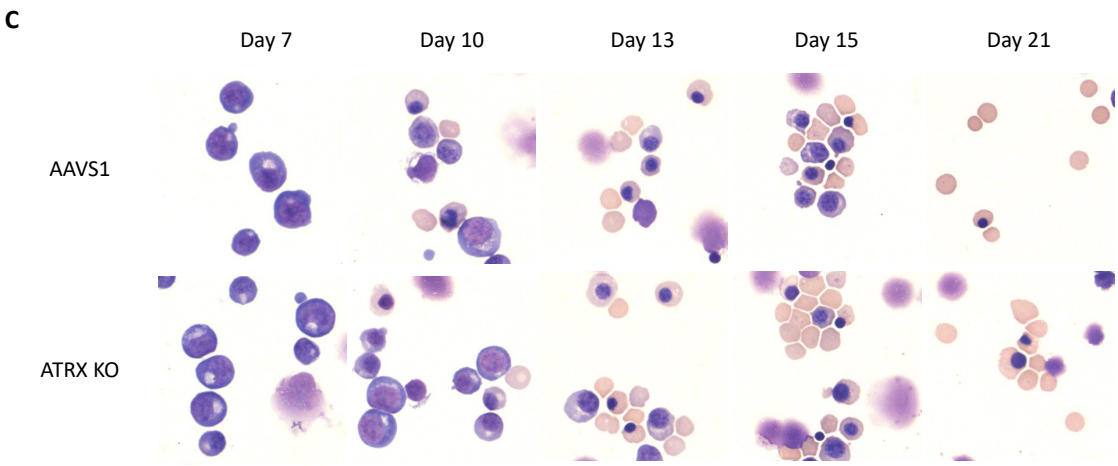


Figure 3.7 Erythroid phenotype analysis in ATRX knockout cells.

(A) The nucleated cell number change during differentiation showing slightly slower cell growth of ATRX KO cells than control cells (n=3). Error bars show standard deviations from three independent experiments.

(B) Representative flow cytometric analysis of erythroid cell surface markers showing ATRX KO cells exhibited similar erythroid differentiation progress as control AAVS1 cells.

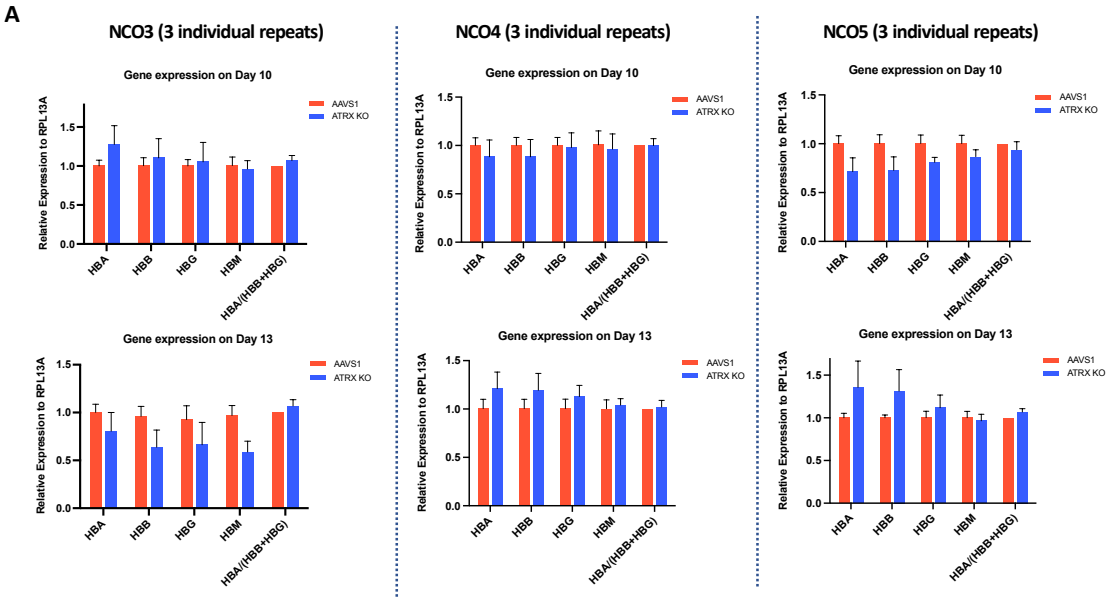
(C) Representative cytospins stained with modified Wright's stain (magnification 400x) showing ATRX KO cells underwent similar cellular morphology change over erythroid differentiation as control cells.

(D) Haemoglobin analysis by IEF in cell samples at different differentiation timepoints. The left and right lanes were loaded with the haemoglobin AFSC control. ATRX KO and AAVS1 samples showed a similar haemoglobin expression pattern.

3.2.3.2 Globin gene expression analysis

To determine the globin gene expression change upon knocking out ATRX, RT-qPCR was performed using cell samples from differentiation day 10 and day 13 to detect mRNA expression of HBA, HBB, HBG, HBM and α to β-like ratio. The

results showed that globin gene expression was variable at different differentiation stages and between different donors, while the α to β-like ratio kept unchanged in all experiments (Figure 3.8A), thereby suggesting that no α-thalassaemia phenotype was detectable in ATRX KO cells. Nevertheless, it is also possible that the gene expression change was too subtle to be detected by the regular qPCR method. Therefore, a more sensitive approach, droplet digital PCR (ddPCR), was performed to detect the expression of HBA, HBB and HBG so the α to β-like ratio can be more accurately measured. To this end, I established a triplex ddPCR system that combines HBA-VIC, HBB-FAM, HBG-FAM probes in a single reaction which would allow us to read out the α to β-like ratio directly by calculating the ratios of the normalized VIC and FAM values. The ddPCR results suggested that the α to β-like ratio did not change significantly in the ATRX deficient cells (Figure 3.8B).



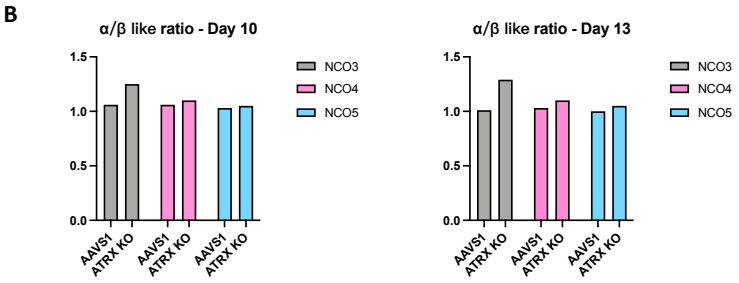


Figure 3.8 Globin RNA expression analysis in ATRX knockout cells.

(A) Taqman Probe RT-qPCR expression analysis for globin genes (HBA, HBB, HBG, HBM) on day 10 and 13 erythroblasts differentiated from CD34+ cells of donors NCO 3-5. Error bars show standard deviations from three independent experiments (n=3).
(B) Droplet digital PCR analysis showing alpha globin to beta-like globin ratio in cultured erythroblasts on days 10 and 13 of differentiation.

To determine if this is the same case at the protein expression level, flow cytometric analysis was performed to detect the expression of the intracellular proteins ATRX, α-globin, β-globin and γ-globin. It is impracticable to detect HBM protein since it has been reported that μ-globin protein is not detectable (Goh et al., 2005) and there is no available antibody for it. The FACS analysis showed the overall expression level of ATRX protein was slightly decreased in the ATRX KO cell group, but without a clear separation between the normal and knockout cells (Figure 3.9A). There was no significant change in the globin expression at the bulk level, but a small number of ATRX KO cells showed reduced α-globin expression (Figure 3.9A). The quantified result indicated the α to β-like ratio was not significantly changed (Figure 3.9B). Given an obvious ATRX protein defect had been shown by Western blotting (Figure 3.6C) but FACS did not show much sensitivity in detecting ATRX downregulation, we next sought to gauge the

protein expression by CyTOF. It turned out that CyTOF showed much better performance in detecting the downregulation of the nuclear protein ATRX than FACS (Figure 3.9C), which allowed us to look at the protein expression in cells with low ATRX expression. Hence, I gated the low ATRX cells from the bulk ATRX KO cells (Figure 3.9C) and then checked the intensity of ATRX and globins in this subset of cells. The results indicated a striking ATRX reduction in the lowATRX cells as expected (Figure 3.9D, E). Similar to the FACS results, expression of α globin and α to β-like ratio remained unaffected (Figure 3.9D, E).

Taken together, the CD34+ cell differentiated ATRX KO erythroid cells do not seem to recapitulate the α-thalassaemia phenotype seen in ATR-X patients.

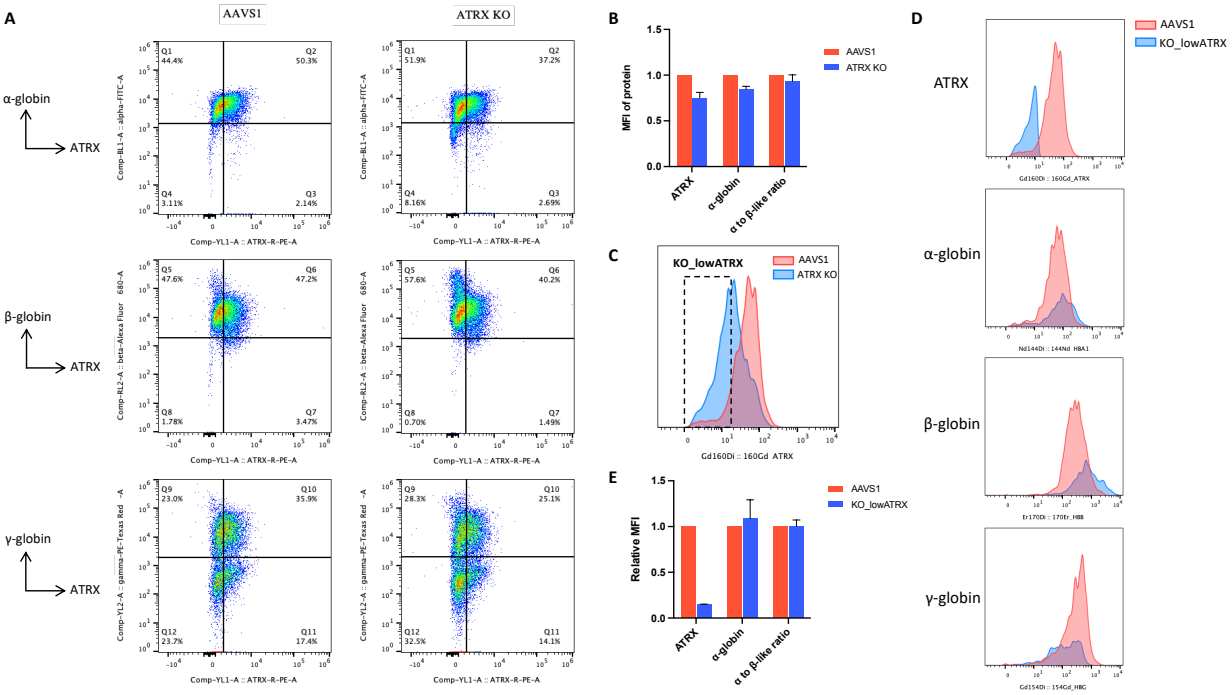


Figure 3.9 Globin protein expression analysis in ATRX knockout cells.

(A) FACS analysis of ATRX and globin protein expression in day 10 AAVS1 control and ATRX KO cells. x-axis shows ATRX expression, manifesting the whole ATRX KO cell

population shifts a bit to the left quadrants. The α-globin/ATRX expression plot shows that there is a tiny population of cells appear in the bottom left quadrant in ATRX KO sample, indicating a ATRX and α-globin double negative feature in them. The quantified results of relative median fluorescence intensity (MFI) ±SD of ATRX, α-globin and α to β-like globin ratio in AAVS1 and ATRX KO samples are shown in (B) (n=3).

(C) Histograms of ATRX expression detected by CyTOF in day 10 AAVS1 and ATRX KO bulk cells. The dash line gates the subpopulation of cells with low ATRX expression.

(D) Histograms of ATRX and globin protein intensity detected by CyTOF in lowATRX subpopulation and AAVS1 controls. The quantified results of relative MFI ±SD of ATRX, α-globin and α to β-like globin ratio are shown in (E) (n=2).

3.2.4 Phenotype assessment in single ATRX KO erythroid burst-forming unit (BFU-E)

One possible reason of not detecting the α-thalassaemia phenotype in ATRX KO cells is that the ATRX defect only affects a small subset of cells, so we might have missed the real effect when looking at this in a bulk population. The earlier FACS analysis showing a tiny population of ATRX KO cells had low α-globin expression also supports this possibility. To address this hypothesis, a Colony-Forming Unit (CFU) assay was performed following the knockout experiment, to allow us to pick out the individual early erythroid progenitor BFU-E colonies and to determine their genotype and phenotype in a single colony.

3.2.4.1 Determination of the cell plating density in CFU assay

As the plating efficiency of CD34+ cells derived from different origins may vary greatly, a pilot CFU assay was firstly performed to determine a suitable seeding density. CD34+ cells were plated into MethoCult semi-solid media at densities of 500 to 2000 cells in 35 mm dishes to promote colony formation. After around 2 weeks, the densities of the formed colonies at each plate were checked, which indicated that the appropriate plating density was 500 cells/plate (Figure 3.10).

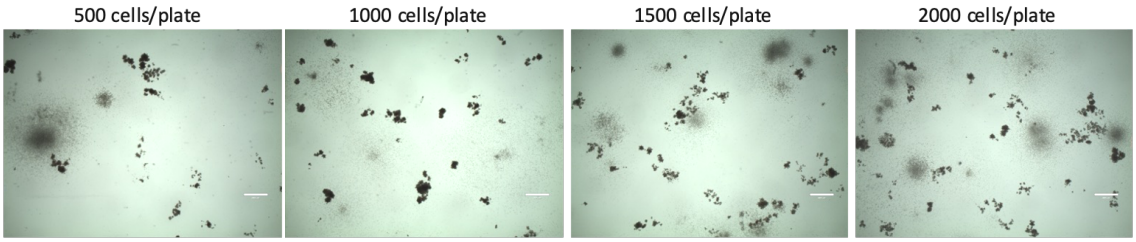


Figure 3.10 Determination of cell plating density in CFU assay.

The growth density of colonies at day 15 of culture after plating CD34+ cells into MethoCult semi-solid media.

3.2.4.2 Haematopoietic progenitors sorting and CFU assay

In the pilot CFU assay, there were too few individual BFU-E colonies generated, which limited the sample size for further genotype-phenotype analysis. To overcome this, a haematopoietic progenitor FACS sorting panel was set up to enrich the progenitors from the bulk cells, including HSC, multi-potent progenitor MPP, early erythroid progenitor EEP/BFU-E and late colony erythroid progenitor LEP/CFU-E. These sorted progenitors were then plated into semi-solid media to

grow into colonies. Two weeks later, the formed BFU-E colonies were collected under microscope and each single colony was divided into two aliquots, one for whole genome amplification to confirm its ATRX KO genotype and the other for Biomark gene expression analysis to determine its phenotype by looking at the globin gene expression (Figure 3.11).

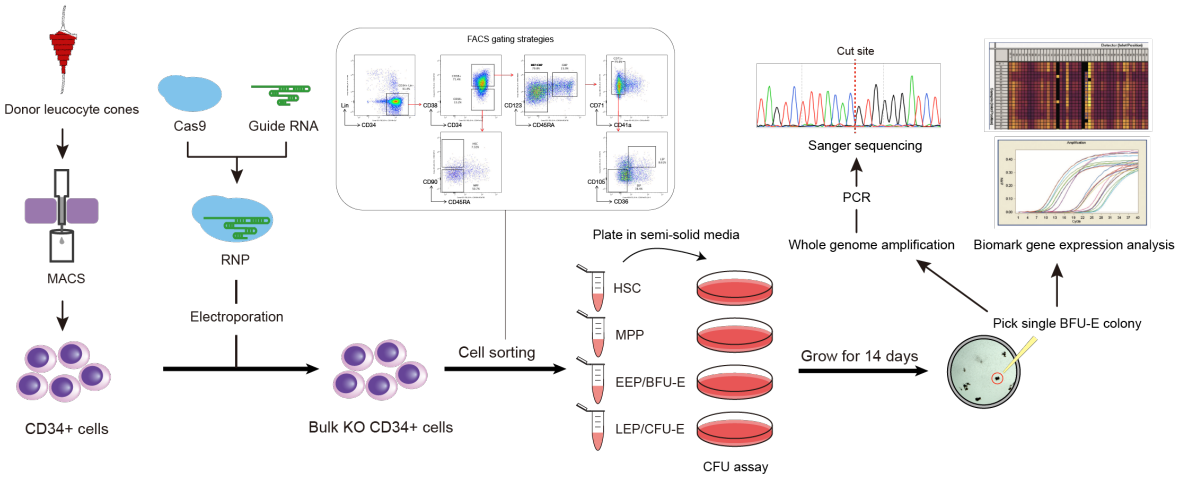


Figure 3.11 Workflow of phenotype analysis of single ATRX KO erythroid burst-forming units.

CD34+ cells isolated from healthy donors first underwent ATRX knockout using RNP complex. The knockout cells were allowed to recover for 24 hours prior to FACS sorting to enrich the four progenitor populations. The sorted cells were plated into the semi-solid media to grow into colonies over 2 weeks prior to genotype and phenotype analysis.

a. FACS sorting of progenitors

The FACS sorting panel used here was previously described for sorting erythroid progenitors (Iskander et al., 2015) (Figure 3.12A, B). HSC, MPP, EEP/BFU-E and LEP/CFU-E were sorted from both CD34+ AAVS1 control and ATRX KO cell samples. The proportion of the progenitors was examined in each sample. EEP/BFU-E was the predominant population in both samples, followed by MPP. Only a tiny fraction of the cells was HSC or LEP/CFU-E. In ATRX KO CD34+ cells, these progenitor frequencies were similar to the controls (Figure 3.12C).

500 cells from each sorted progenitor were plated into the 35mm dish, as determined for optimal seeding density. Given the limited starting cell number of the CD34+ cells (around $1*10^5$ cells for each sample), all the cells from the knockout experiment were used for progenitor sorting. In general, 400-500 HSCs cells were sorted from each sample and all these cells were plated into one 35mm dish. There were around 1500-2000 cells sorted as MPPs, so they were plated into three to four 35mm dishes. The EEP/BFU-E sub-population usually contained 3500-4500 cells, and 500 cells were plated in each dish. The population size of LEP/CFU-E was very small, so it was not plated.

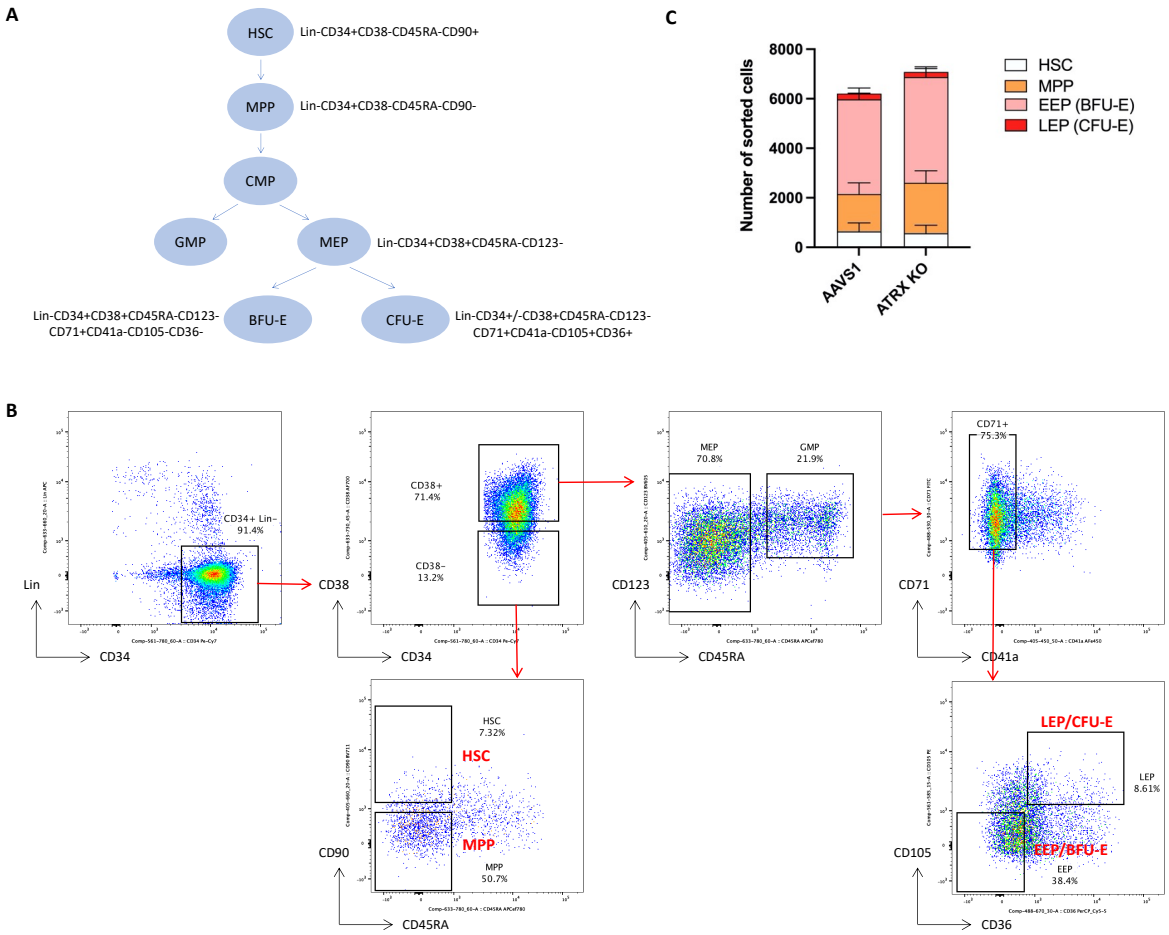


Figure 3.12 FACS sorting of progenitors for CFU assay.

(A) Immunophenotype markers used for distinguishing the different progenitors.

(B) Representative FACS gating strategy for progenitor sorting based on their characteristic markers. HSC, MPP, EEP/BFU-E and LEP/CFU-E are sorted individually. All gating was determined by FMO controls.

(C) Sorted cell number for each progenitor population in AAVS1 control and ATRX KO samples (n=3).

b. Scoring haematopoietic colonies

All types of the formed colonies were identified and counted *in situ* after 7 days and 14-16 days of culture respectively. On day 7, erythroid colony forming units (CFU-E) formed on each plate were first checked and counted. There was, however, no CFU-E colony observed on any plate. On day 14-16, burst-forming unit-erythroid (BFU-E), colony-forming unit-granulocyte (CFU-G), CFU-macrophage (CFU-M), CFU-granulocyte macrophage (CFU-GM) and CFU-granulocyte erythrocyte macrophage megakaryocyte (CFU-GEMM) colonies were identified and counted under microscope. The colonies derived from each sorted progenitor population were scored in both AAVS1 and ATRX KO samples. The scoring results displayed a similar clonogenic capacity of ATRX KO samples as the controls, since no significant difference in colony forming was observed (Figure 3.13A). Meanwhile, the scoring results indicated a high purity of the sorted BFU-E sub-population, as around 90% of the colonies formed by sorted BFU-E progenitors were erythroid colonies (Figure 3.13B).

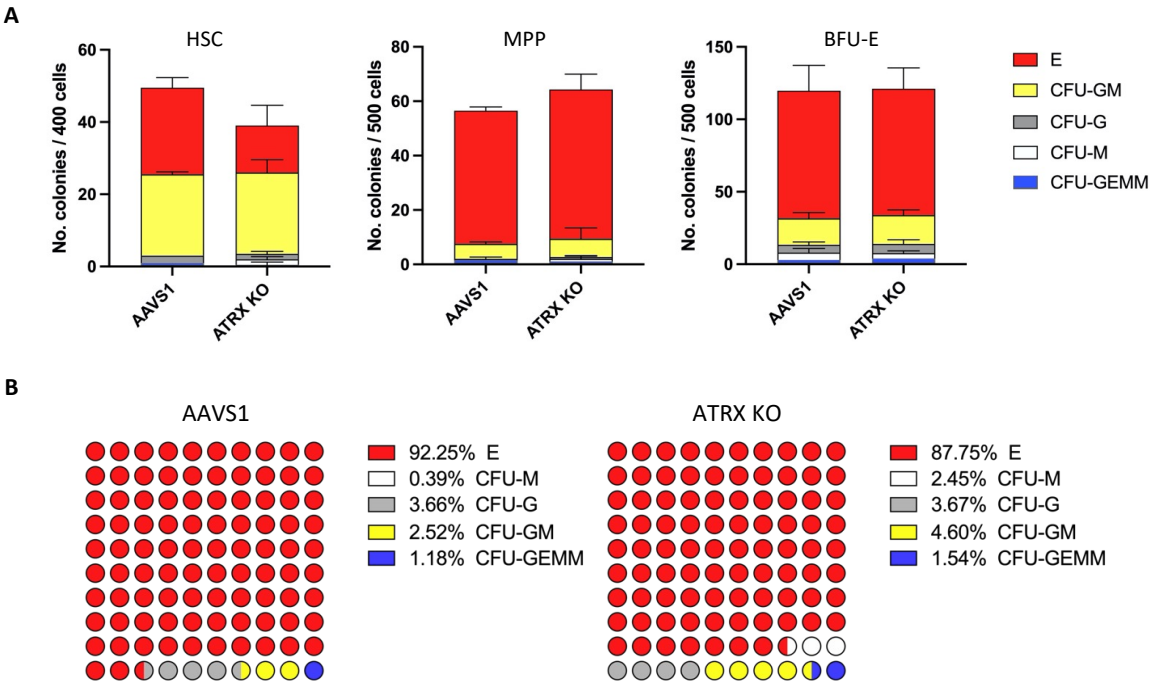


Figure 3.13 Clonogenic capacity of the sorted progenitors.

(A) Colony scoring of the sorted haematopoietic progenitors in AAVS1 control and ATRX knockout samples, showing the frequency of erythroid colonies (E), granulocyte and monocyte-macrophage colonies (CFU-GM), granulocyte colonies (CFU-G), monocyte-macrophage colonies (CFU-M), and mixed granulocyte, erythroid, macrophage, megakaryocyte colonies (CFU-GEMM). The CFU assays were conducted three times independently.

(B) Lineage output of sorted BFU-E progenitors in AAVS1 control and ATRX knockout samples, showing the mean frequency of three experiments.

3.2.4.3 Genotype and phenotype analyses in BFU-E colonies

After finishing the colony counting, the separate and roundish erythroid colonies (BFU-E) were photographed and collected under the microscope. The BFU-E

colonies derived from the ATRX KO cells looked normal and similar to AAVS1 controls in shape and size independent of their genotypes (Figure 3.14A). Each BFU-E colony was then split into two aliquots, one for genotype analysis and the other for gene expression (phenotype) analysis. Genotype analysis was performed in the first place to determine the outcome of gene editing at either *AAVS1* or *ATRX* loci by sequencing the sgRNA targeting region. The phenotype analysis was then conducted in the selected colony samples either being edited at the *AAVS1* locus or carrying an *ATRX* frameshift mutation that is likely leading to a knockout effect.

a. Genotype analysis

In total, 96 BFU-E colonies were collected from AAVS1 sample and 387 BFU-E colonies were collected from ATRX KO sample. These colonies samples were subject to whole genome amplification followed by target amplification at *AAVS1* or *ATRX* loci. The phenotype of each single colony was confirmed by Sanger sequencing.

For the AAVS1 samples, the genotype was only recognized as either “edited” (different from wild type) or “unedited” (wildtype) at *AAVS1* locus. It was shown that 81 out of the sequenced 94 colonies (86.17%) were edited, indicating a high gene editing efficiency. There were two colonies having no PCR product, perhaps due to failed whole genome amplification or having a large deletion beyond the target amplification region.

For the ATRX KO samples, their genotype was classified as “WT” (wild-type), “Frameshift mutation” or “In-frame indel”. Out of the collected 387 colonies, 240 (62.02%) had frameshift mutation, 98 (25.32%) had in-frame indel, 4 (1.03%) were wildtype, and 36 had no amplification (“ND”, not detected) (Figure 3.14B). There were 9 samples having two different genotypes that may be due to a mixture of two colonies.

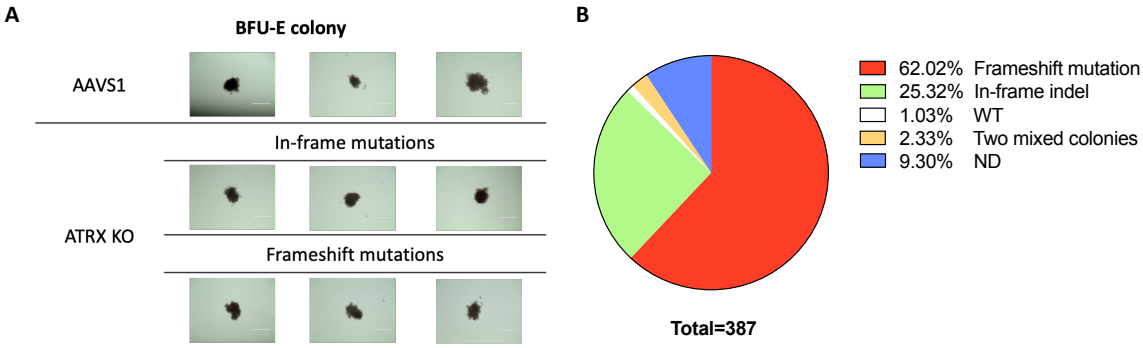


Figure 3.14 Genotype determination of BFU-E colonies and their morphology.

(A) BFU-E colonies derived from AAVS1 control and ATRX KO samples showing similar morphology and size independent of their genotypes.

(B) Genotyping summary of the collected 387 BFU-E colonies derived from ATRX KO samples showing the proportion of each genotype. “ND” represents “not detected”.

b. Gene expression analysis

After the genotypes of the BFU-E colonies were determined, gene expression in the colony samples was analysed using the Fluidigm Biomark high throughput qPCR approach. 50 “edited” AAVS1 colonies and 190 ATRX KO colonies with

“frameshift mutation” were selected for the gene expression analysis to specifically check the expression of ATRX and globin genes.

By comparing the gene expression of ATRX KO samples with control AAVS1 samples, we found that the majority of ATRX KO samples had a low level of ATRX (Figure 3.15B), however only a proportion of these samples showed reduced expression of HBM and HBA (Figure 3.15A), indicating the presence of cellular heterogeneity. Furthermore, the quantified results demonstrated a greater downregulation of HBM than HBA in ATRX KO cells, showing a statistically significant reduction of HBM expression in ATRX KO samples but no significant difference in HBA expression between AAVS1 and ATRX KO groups (Figure 3.15B).

Taken together, from the BFU-E colony results, we conclude that ATRX defect only affects the expression of alpha globin genes in a subset of cells, and HBM is more affected than HBA which is consistent with the observations in ATR-X patients.

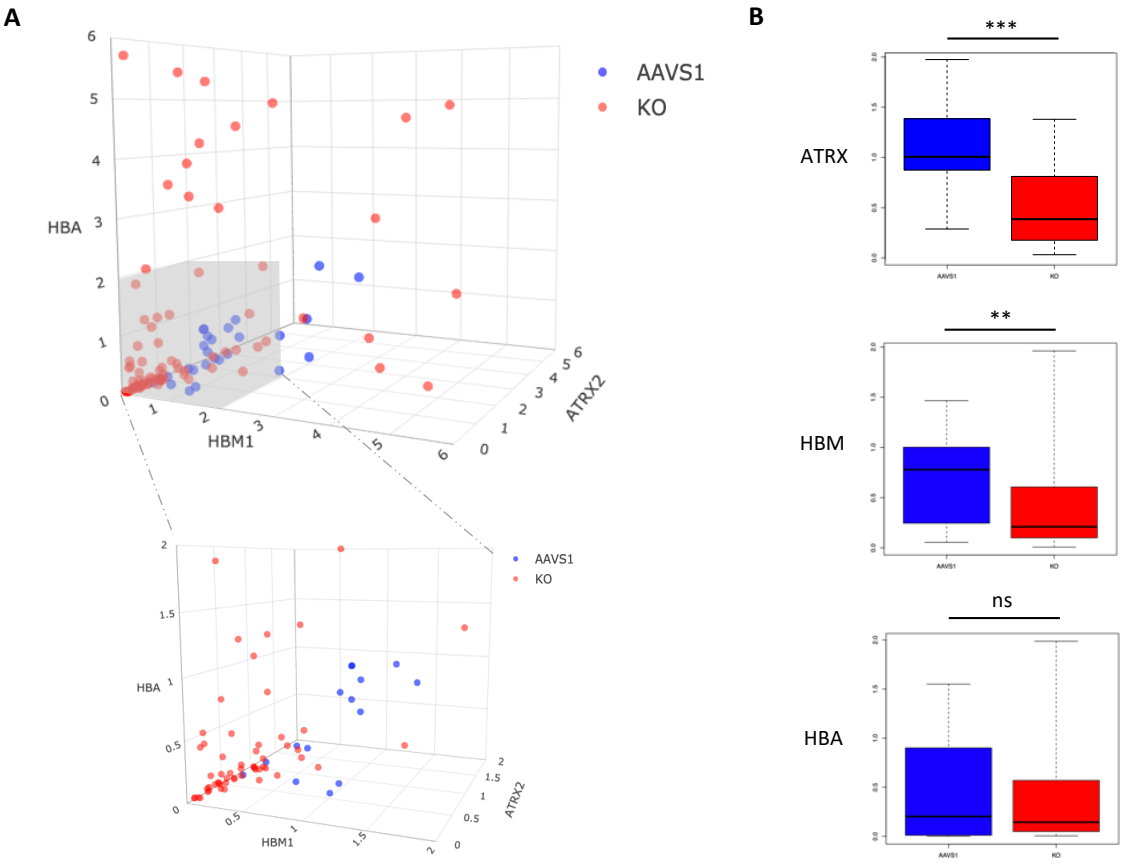


Figure 3.15 Gene expression analysis in BFU-E colonies.

(A) Three-dimensional plot showing expression of ATRX, HBA and HBM in BFU-E colony samples, relative to the mean of RPL13A, RPS18, ACTB and GAPDH reference genes. The colonies missing detection in the reference genes were excluded. Each dot in the plot represents one BFU-E colony, and the edited AAVS1 colonies are shown in blue (n=29) and ATRX KO colonies having frameshift mutation in *ATRX* are shown in red (n=132). The bottom enlarged panel shows the gene expression in the colony densely populated region without outliers. The AAVS1 controls assemble at the internal space whereas the majority of the ATRX KO colonies assemble around the origin showing low expression of all these three genes.

(B) Quantified gene expression in AAVS1 colony group and ATRX KO colony group. Mann-Whitney U test was conducted to measure the statistical difference between the

two groups, *p*-value = 1.493e-05 in ATRX expression, *p*-value = 0.003632 in HBM expression, and *p*-value = 0.9083 in HBA expression. ns: not significant.

3.3 Discussion

ATRX mutations are associated with reduced α-globin expression and lead to α-thalassaemia in ATR-X and ATMDS patients (R. J. Gibbons, Picketts, et al., 1995; R. J. Gibbons et al., 2003), but to date the underlying mechanism remains poorly understood, partially due to the lack of a good model system in which to recapitulate the *in vivo* observations. Although an α-thalassaemia-like phenotype has recently been reported in the atrx deficient zebrafish embryo where the embryonic α-e1 globin expression was reduced (Oppel et al., 2019), these zebrafish are lethal at embryonic stage which makes the mechanism investigation difficult. Considerable efforts have been made to develop a tractable mammalian cellular model in our lab, however neither ATR-X patient-derived iPSC nor ATRX deficient erythroid progenitor HUDEP-2 cell line reproduced the α-thalassaemia phenotype when they were differentiated toward the erythroid lineage. There might be several reasons for this. For instance, the regulation might specifically occur in an early stage of erythropoiesis, but these two cell models are at a later stage and the regulation may therefore have been missed in these systems. Here we use a more physiological stem cell system CD34+ HSC to model this process by knocking out ATRX.

Erythropoiesis is a highly regulated step-wise process coupling cell expansion with differentiation (Barminko et al., 2016). Here, using a three-phase erythroid

differentiation system, we show this can be modelled in an *in vitro* culture system starting from CD34+ haematopoietic stem cells, passing through the expected stages of erythroid differentiation with increasing globin expression and resulting in mature reticulocytes, indicating that this culture system recapitulates the biological process of definitive erythropoiesis and can therefore be used to elucidate mechanisms underlying erythropoiesis dysregulation. It is of note, however, that there is distinct expression of fetal γ-globin (HBG) and fetal haemoglobin HbF after day 13 of differentiation and later stages in our culture system. This is probably because the cells are under mild stress after a long time of *ex vivo* culture thus inducing the expression of γ-globin, which is in agreement with the identity of γ-globin as a cell stress response gene (Schaeffer et al., 2014; Boontanrart et al., 2020). Nonetheless, adult haemoglobin is the predominant haemoglobin, and this culture system has been able to recapitulate the defective erythropoiesis in some inherited disorders including Diamond Blackfan Anaemia and Congenital Dyserythropoietic Anaemia type I, and has been used to uncover the pathogenesis of these diseases (Moniz et al., 2012; O'Brien et al., 2017; C. Scott et al., 2020).

As ATR-X syndrome is rare, with little access to primary patient HSCs, here I have focused on creating a model system using CRISPR/Cas9 gene editing of male CD34+ human peripheral blood progenitor cells. To achieve efficient knockout effect on ATRX, a group of sgRNAs targeting different regions of *ATRX* have been tested. We demonstrated that the sgRNA targeting the largest exon (exon 9) of *ATRX* outside of its functional ADD domain had higher KO efficiency than those targeting the ADD domain directly, and the sgRNA targeting exon 16

which was previously used for producing an ATRX defect in HUDEP-2 cells showed inadequate KO in CD34+ cells, indicating the variability of CRISPR KO efficiency between loci and cell types.

While a competent ATRX knockout is generated by CRISPR/Cas9, here we show the ATRX KO CD34+ differentiated erythroid cells appear to have no detectable α-thalassaemia phenotype. The cells showed similar pattern of differentiation as controls, with appropriate expression of erythroid cell surface markers and haemoglobin. The only noticeable difference between controls and ATRX KO cells was that ATRX KO cells exhibited slower cell growth after day 7 of differentiation when the cells commit to the erythroid lineage and start to actively express globin genes. The cells show high proliferative capacity at this pro-erythroblast stage, along with the active cell cycle-related processes depicted by the transcriptional profile (Ludwig et al., 2019). The slackening of cell growth in ATRX deficient cells highlights a role of ATRX in the cell cycle, which has been supported by a number of previous studies in other cell types showing that ATRX depletion can lead to prolonged cell cycle and defective mitosis by affecting DNA replication and chromosome segregation (Ritchie et al., 2008; Clynes et al., 2014; Huh et al., 2016). Our results suggest this function is also applicable to erythroid cells, although future work is needed to determine whether this is related to DNA replication or chromosome segregation.

The α-thalassaemia phenotype can be reflected by the reduced ratio of α-globin to β-globin at the gene expression level. Given the induction of fetal γ-globin expression at late differentiation stage in our culture system, its contribution to

haemoglobin formation cannot be ignored, and therefore α to β-like (β+γ) globin is adopted to evaluate the α-thalassaemia phenotype in our cells. The RT-qPCR gene expression analysis showed high variability in the globin gene expression pattern between different donors. There are two possible reasons for this. First, the KO efficiency was different between the donors and the individual experiments. Second, the length of the ψζ VNTR are different in these donors, which has been shown to be associated with the degree of the effect (Law et al., 2010). In addition, the gene expression pattern was also variable between day 10 and day 13 in the same donor, indicating this regulation might be a stochastic process during cell differentiation. The α to β-like globin ratio, nonetheless, did not show a significant difference between controls and ATRX KO cells in any of the experiments. This raises the possibilities that the dysregulation only give rise to a subtle effect in the whole population or only a small proportion of the cells are affected, and the effect may therefore be lost in the bulk population. To determine if this is due to a subtle effect, we developed a more sensitive gene expression detection approach ddPCR, which enables a ±10% precision in quantification. However, it showed no change of globin gene expression in ATRX KO cells, which narrows the possibility down to cellular heterogeneity. Therefore, we next examined the genotype and gene expression in individual erythroid BFU-E colonies.

By CFU assay, we demonstrate that ATRX defect in HSCs has no effect on the clonogenic capacity of the progenitors and no effect on the erythroid colony formation. Gene expression analysis in BFU-E colonies reveals ATRX defect gives rise to α-thalassaemia phenotype in only a subset of cells, rather than

causing a subtle change in the whole population, indicating ATRX mediated α-globin expression is a stochastic process. This is consistent with our recent observations in CD34+ cells derived from an individual with ATR-X syndrome (Truch et al., 2022). Moreover, here we show that HBM is more downregulated than HBA by ATRX deficiency, which recapitulates the observation in ATR-X patients, and this may be explained by the proximity of HBM to the ψζ VNTR. However, the Biomark gene expression data of BFU-E colony was somewhat noisy, with large variations between samples, even for the samples with the same genotype. It is therefore obviously of importance to define this with a larger set of samples. To this end, single cell RNA-seq has been performed to depict the gene expression at a larger scale and in a more informative way at the single cell level, which will be the content of the next chapter.

In conclusion, the data presented in this chapter demonstrate that ATRX deficient CD34+ haematopoietic stem cells exhibit normal erythroid differentiation and have no distinct α-thalassaemia phenotype at the bulk level. Further gene expression analysis in single erythroid burst-forming units indicates ATRX deficiency affects gene expression at α-globin locus in only a subset of cells.

4 Gene expression analysis of ATRX deficient erythroid cells at the single cell level

4.1 Introduction

Conventional bulk gene expression analysis measures the total gene expression level from tissues made up of various cell types or from heterogeneous cell samples composed of cells in multiple states. Variability in composition can significantly confound the analysis, such as the identification of differentially expressed genes. Differential gene expression in a subpopulation of cells, in turn, can be weakened by the cellular heterogeneity of a bulk population.

Previous efforts to determine α-thalassaemia in an ATRX deficient cellular model failed because of the cellular heterogeneity. The data presented in the previous chapter demonstrate that the ATRX defect only affects α-globin gene expression in a small subset of cells. Consequently, the downregulation of α-globin expression has probably been below the threshold of detection when checking this in bulk populations in our previous attempts. By targeted gene expression analysis in single BFU-E colonies, we show that the α-thalassaemia phenotype is recapitulated in a subset of ATRX deficient cells. However, exploring the mechanism underlying this phenomenon has been constrained by the limited number of genes can be measured using the Biomark platform. Recent advances in single-cell techniques have allowed a more thorough investigation of cellular heterogeneity and the associated molecular mechanisms.

As the structural functional units of life, cells are typically classified into different types according to their origins and morphology as well as molecular characteristics (Orkin, 2000; Arendt et al., 2016). Changes in cellular activity and identity are reflections of distinct gene expression profiles. Proteomic technologies at the single-cell level can detect the final functional product of gene expression, but they are currently constrained to a pre-selected repertoire of proteins of interest, which hampers their ability to make comprehensive and unbiased analysis of cellular activity (Vistain & Tay, 2021). Instead, single cell transcriptomics provides an alternative means to characterise cellular processes at the molecular level. Single cell RNA-seq (scRNA-seq) enables the transcriptomic profiling of large numbers of single cells in parallel and describes cellular heterogeneity, which has advanced our understanding of health and disease since it was introduced (Svensson et al., 2018; Aldridge & Teichmann, 2020). In the field of haematology, single cell RNA-seq has been extensively applied to identify novel cell subsets, building developmental trajectories, and uncovering distinct pathways (Psaila et al., 2016; Grover et al., 2016; Giustacchini et al., 2017; P. Huang et al., 2020).

Currently, the two most popular scRNA-seq protocols are the plate-based method Smart-seq2 and the droplet-based approach 10X Chromium sequencing. Smart-seq2 requires isolating single cells into each well of a PCR plate containing cell lysis buffer, which is often achieved by FACS sorting, and then is followed by high-fidelity reverse transcription and amplification to increase the cDNA yield for single-cell analysis (Picelli et al., 2014). It is a robust and reliable method, for detecting rarer transcripts in particular, using little or no specialist equipment.

Moreover, the initial cell separation step allows users to enrich rare cell populations of interest for analysis. However, the cell number analysed is limited to the number of the plates being processed (with 96 or 384 wells) which is also associated with greater batch effects. In the droplet-based method, individual cells are encapsulated into droplets containing DNA-barcoded reads for reverse transcription which are broken up later to be subjected to library preparation (Klein et al., 2015; Macosko et al., 2015). This method allows for processing thousands of cells simultaneously, but it requires dedicated equipment like the most commonly used single-cell controller from 10x Genomics and it has higher noise for mRNAs with low expression level (X. Wang et al., 2021).

With the aim of uncovering molecular signatures for the subset of cells affected by an ATRX defect and thereby to elucidate the underlying mechanism, single cell RNA-seq analysis has been performed to analyse the gene expression of ATRX deficient erythroid cells at the single cell level. Given a proportion of erythroid cells are enucleated and lack mRNA material, the Smart-seq2 approach was conducted here to only sort nucleated cells for analysis.

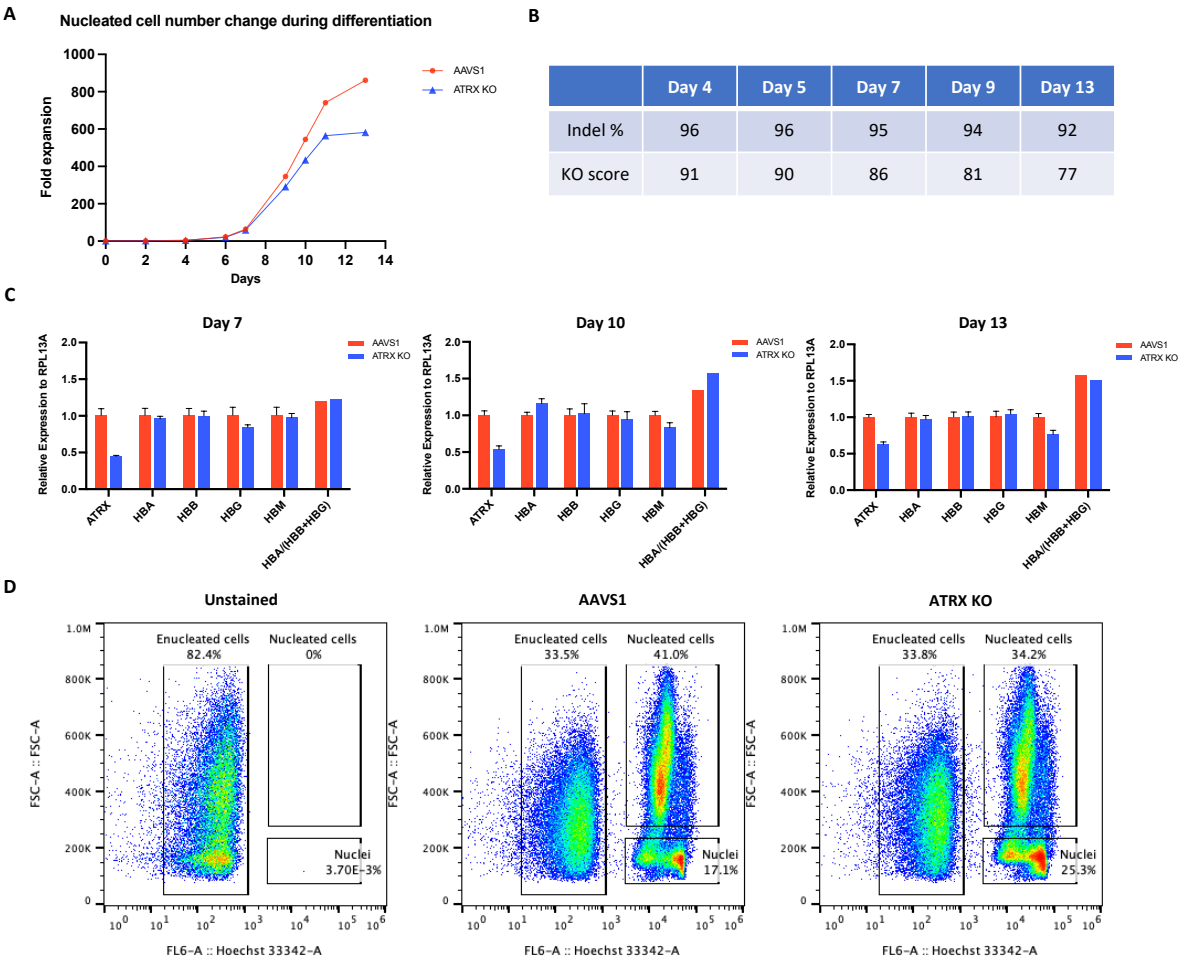
4.2 Results

4.2.1 Knock out efficiency and gene expression analysis in bulk cells used for scRNA-seq

In order to sort single cells for scRNA-seq analysis, CD34+ cells derived from a healthy male donor underwent gene knockout as described in the previous

chapter, using AAVS1 as a control and sg9-1 to knock out ATRX, after which the cells were differentiated down the erythroid pathway. The number of nucleated cells during the differentiation process was recorded and a slower growth of ATRX KO cells was observed (Figure 4.1A). Nonetheless, the ATRX KO cells had a similar differentiation progress to the controls.

To determine the best time point to harvest samples for single cell analysis when the cells are most affected, bulk cell samples at multiple time points were collected during differentiation for assessing knockout efficiency and gene expression analysis. The knockout efficiency was analysed by ICE software and based on the Sanger sequencing of the sg9-1 targeted region. The results showed that the percentage of knocked out cells declined over the course of time (Figure 4.1B). Nonetheless, there were still 77% of the cells having an ATRX KO at day 13 (Figure 4.1B), reflecting the overall high knockout efficiency. Consistent with this, qPCR gene expression analysis revealed a marked reduction of ATRX at the RNA level in ATRX KO samples compared with AAVS1, but the reduction gradually decreased over differentiation, with 55.2% at day 7 and 37.2% at day 13 (Figure 4.1C). RNA expression of the globin genes demonstrated little or no changes in HBA, HBB, HBG and the α to β-like ratio at all time points, but there was an apparent HBM decrease at day 13 (Figure 4.1C), suggesting HBM was most downregulated at day 13 and α-thalassaemia was most likely to be detected in single cells at day 13. Accordingly, single cell analysis was carried out using day 13 cells. Single nucleated cells were sorted into 384-well PCR plates using a cell permeable nuclear stain to exclude enucleated cells and bare nuclei (Figure 4.1D).



	Day 4	Day 5	Day 7	Day 9	Day 13
Indel %	96	96	95	94	92
KO score	91	90	86	81	77

Figure 4.1 Analysis of bulk cells used for scRNA-seq.

(A) The nucleated cell number change in AAVS1 and ATRX KO cells throughout erythroid differentiation until day 13.

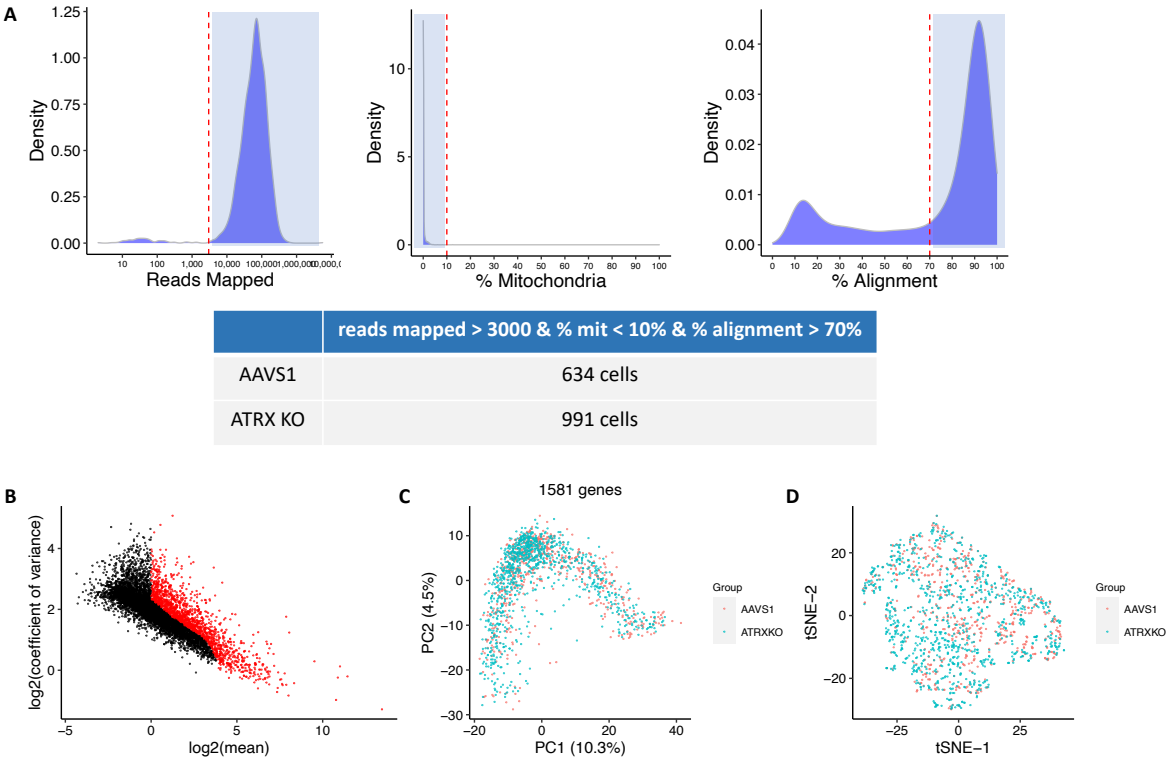
(B) ATRX knockout efficiency analysis by ICE in bulk cells at multiple time points during erythroid differentiation.

(C) Taqman probe RT-qPCR mRNA expression analysis of ATRX and globins in differentiation day 7, day 10 and day 13 cells, normalized to RPL13A. Error bars show SD of three technical repeats.

(D) Single cell sort gating in unstained control, AAVS1 and ATRX KO samples. Nucleated cells and nuclei are both positive for the nuclear stain, but the live nucleated cells are bigger in size than the nuclei, showing higher forward scatter (FSC).

4.2.2 Smart-seq2 single cell sequencing data processing and visualisation

After single cell sequencing data pre-processing and filtering out the poor-quality cells as described in Chapter 2 section 2.6.1.1, 634 cells from the AAVS1 sample and 991 cells from the ATRX KO sample were retained for further analysis (Figure 4.2A). Next, the expression data of these 1625 cells were used to identify genes with high biological variations, namely highly variable genes (HVGs), which contribute to cell-to-cell differences in the cell population. To do so, undetected genes in all cells were firstly excluded and then genes expressed in at least five cells were retained; 6298 genes were finally retained for HVGs detection. Finally, 1581 HVGs were identified by fitting a LOESS model with the mean-variance relationship of log-transformed expression values (Figure 4.2B). Principle component analysis (PCA) and t-distributed stochastic neighbour embedding (t-SNE) were performed using HVGs for dimension reduction and data visualisation. PCA analysis showed that PC1 could only account for 10.3% of the variation in the dataset wherein no distinct difference was observed between AAVS1 and ATRX KO groups (Figure 4.2C). The t-SNE demonstrated no separate clusters or groups but rather a mixed pattern between groups (Figure 4.2D), which would be somewhat expected for the homogenous cultured cells grown and differentiated in the same *in vitro* system.



	reads mapped > 3000 & % mit < 10% & % alignment > 70%
AAVS1	634 cells
ATRX KO	991 cells

Figure 4.2 scRNA-seq data processing and visualisation.

(A) Histograms of quality control metrics and filtering decisions. Top left: histogram of the library size per cell. A threshold is applied at >3000. Top middle: histogram of the percentage of reads mapping to the mitochondrial chromosome. A threshold is applied at <10%. Top right: histogram of the percentage of mapped reads per cell. A threshold is applied at >70%. Bottom table: the number of cells retained in each group after quality control.

(B) Plot of log2(mean) versus log2(coefficient of variance) of the 6298 genes retained after gene filtering, coloured by merging the loess curve. Each dot denotes a gene; black dots represent less variable genes and red dots represent HVGs.

(C) PCA plot based on HVGs. The x- and y-axes denote the top two principal components. Each dot represents one single cell and cells are coloured by groups.

(D) t-SNE visualisation of single cells in AAVS1 and ATRX KO groups using HVGs.

4.2.3 Differentially expressed gene (DEG) analysis between AAVS1 and ATRX KO cells

To determine which genes are expressed differentially in ATRX deficient erythroid cells, DEG analysis was performed between the AAVS1 and ATRX KO groups. Poor-quality cells and genes expressed in less than five cells were filtered out, and then the average expression level of each gene in all the cells from each group was calculated and compared between the two groups. Significantly different genes were selected based on adjusted *p*-value < 0.05 and absolute log2(fold change) > 0.5. After applying these criteria, 71 differentially expressed genes were identified, and these were used for dimension reduction and data visualisation. The t-SNE plot showed that AAVS1 and ATRX KO cells were mixed with no separate clustering (Figure 4.3A). In total, 18 genes were found significantly upregulated in the ATRX KO group and 53 genes were significantly downregulated including the α-like globin gene HBM (Figure 4.3B, Table 4.1). Unfortunately, ATRX gene expression was below the threshold of detection in the majority of the cells in both groups, and therefore its expression was uninformative. The representative DE genes will be discussed in detail later in this section.

To gain insight into the biological function of the differentially expressed genes, gene set enrichment analysis (GSEA) of Gene Ontology (GO) was performed. There was no enriched GO term for the 18 upregulated genes. Importantly, there were 12 GO terms enriched in the 53 downregulated genes, mainly regarding cell cycle, erythrocyte function and cell division (Figure 4.3C). The deep involvement of the downregulated genes in cell cycle and cell division, a total of

6 out of 12 GO terms, may explain why the ATRX KO cells displayed a slower growth during erythroid differentiation than the AAVS1 control cells.

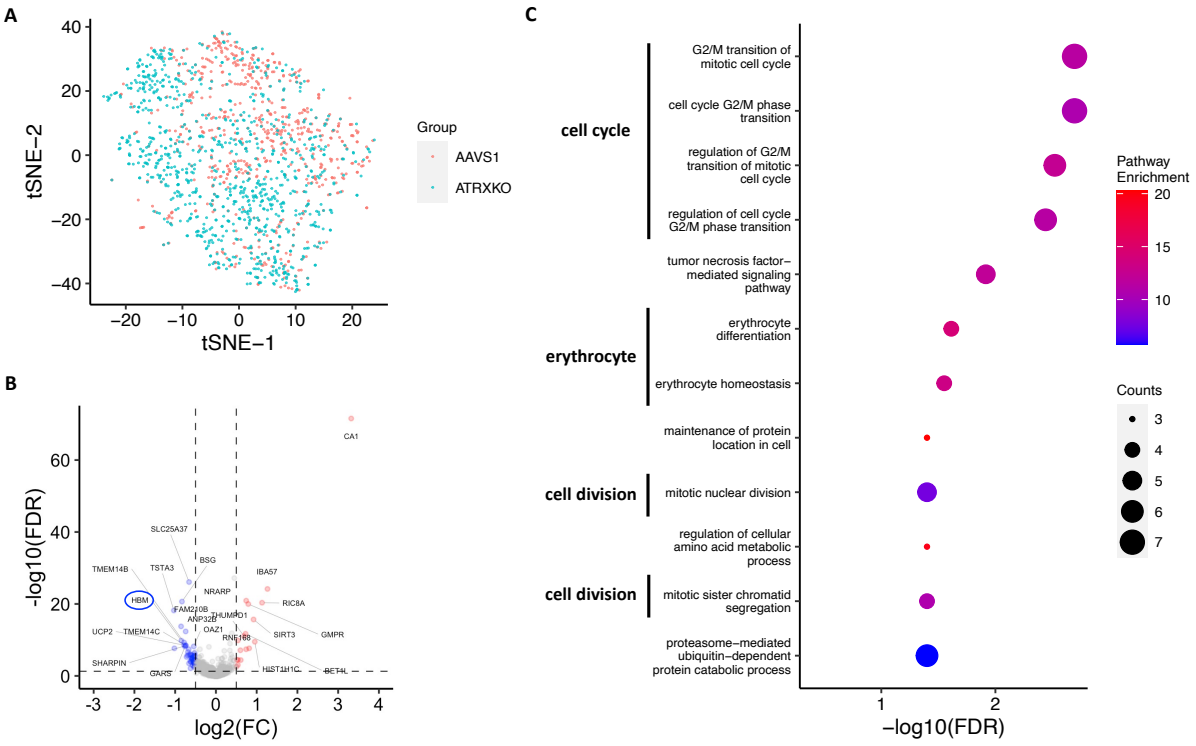


Figure 4.3 Differentially expressed gene analysis of scRNA-seq data.

(A) t-SNE visualisation of single cells in AAVS1 and ATRX KO groups based on identified DE genes.

(B) Volcano plot showing the DE genes in ATRX KO cells versus AAVS1 control in the single-cell transcriptomic analysis. Each dot represents one gene. The horizontal dash line corresponds to a significance value of < 0.05. The two vertical dash lines bound the minimal fold-change (absolute value > 0.5) for the most-differentially-expressed genes. Red dots show the upregulated genes and blue dots represent the downregulated genes. The downregulated HBM is highlighted in blue circle.

(C) Dot plot of Clusterprofiler gene ontology enrichment analysis of the DE genes to identify key biological processes attributed to the different transcriptome of ATRX KO

cells from the AAVS1 control cells. x-axis indicates the *p*-value representing enrichment score; the higher the value, the more significant the enrichment. The size of the dot represents number of DE genes attributed to each GO term; the colour of the dot indicates the odds ratio of the GO term; the larger the odds ratio, the higher the relative abundance of this GO term compared with background.

	Gene name	Mean AAVS1	Mean ATRX KO	log2fc	pval	pval.adj
Upregulated genes in ATRX KO cells	CA1	1.991554187	5.311655872	3.320101685	4.33E-76	2.86E-72
	IBA57	0.877765127	2.14313784	1.265372713	4.12E-28	6.79E-25
	RIC8A	1.215962533	2.347944865	1.131982331	4.88E-24	4.60E-21
	HIST1H1C	1.856554833	2.81050689	0.953952056	8.67E-13	3.18E-10
	SIRT3	0.788235916	1.711587716	0.9233518	3.13E-19	2.07E-16
	GPX1	1.608681099	2.424931616	0.816250517	8.75E-11	2.14E-08
	GMPR	4.121037536	4.914509483	0.793471948	1.26E-23	1.04E-20
	NRARP	0.038172632	0.780944292	0.74277166	9.34E-25	1.23E-21
	CA2	1.012511101	1.75363808	0.741126979	1.84E-10	3.92E-08
	BET1L	0.744439814	1.471643709	0.727203895	4.10E-15	1.93E-12
	THUMPD1	0.545526398	1.235643312	0.690116913	2.08E-14	9.16E-12
	GRSF1	2.014621414	2.624413535	0.609792121	4.11E-07	3.93E-05
	EIF1AY	3.631535373	4.23247381	0.600938436	3.60E-10	7.41E-08
	UIMC1	1.91280149	2.453239942	0.540438452	3.43E-07	3.37E-05
	TUBB4B	1.897542992	2.436671228	0.539128236	2.17E-05	0.000930867
	RNF168	0.278287336	0.816685085	0.538397749	3.43E-13	1.41E-10
	MARCH8	0.816912068	1.331540804	0.514628736	3.70E-07	3.59E-05
	ARRDC2	3.149485973	3.654899811	0.505413838	8.10E-05	0.00273252
Downregulated genes in ATRX KO cells	TSTA3	4.823807892	3.791960676	-1.031847216	8.93E-22	6.55E-19
	H1F0	5.109062135	4.088975498	-1.020086638	9.25E-11	2.18E-08
	FAM210B	4.583982777	3.728525741	-0.855457036	2.87E-17	1.72E-14
	HBM	5.4704496	4.623965791	-0.846483809	4.21E-13	1.63E-10
	BSG	6.715223023	5.881851281	-0.833371742	2.00E-24	2.19E-21
	TMEM14B	2.99866429	2.218043668	-0.780620622	1.52E-12	5.28E-10
	UCP2	6.133449141	5.354008652	-0.779440489	9.47E-12	2.98E-09
	GARS	3.643658568	2.885702755	-0.757955813	2.77E-11	7.32E-09
	TMEM14C	2.427624455	1.681563773	-0.746060682	1.06E-11	3.17E-09
	ANP32B	6.255449094	5.513378453	-0.742070641	7.98E-16	4.39E-13
	SHARPIN	3.764607066	3.029547892	-0.735059174	1.28E-11	3.68E-09
	TUBB	2.379461672	1.658917938	-0.720543734	4.00E-08	5.38E-06
	RAN	2.978835483	2.278089508	-0.700745975	1.31E-08	1.96E-06

	TMSB4X	2.41620044	1.728875728	-0.687324712	1.26E-09	2.30E-07
	AARS	2.745382682	2.080000634	-0.665382048	1.71E-10	3.89E-08
	GNAS	5.472941862	4.81200486	-0.660937002	8.38E-09	1.38E-06
	TUBA1B	2.987840922	2.326928685	-0.660912237	5.99E-06	0.000335021
	SLC25A37	7.531309805	6.872664379	-0.658645425	3.65E-30	8.04E-27
	SELENBP1	3.571342782	2.940917111	-0.630425671	9.00E-07	7.51E-05
	CHID1	2.021604364	1.397391784	-0.624212581	5.06E-09	8.79E-07
	HMGB2	4.823685589	4.200368551	-0.623317038	0.000380483	0.007838525
	RNF187	5.359230929	4.745888288	-0.613342641	2.04E-06	0.000147556
	AUP1	2.353615486	1.743537112	-0.610078374	9.83E-09	1.58E-06
	CALM3	2.456404452	1.855839286	-0.600565166	1.20E-07	1.41E-05
	CYC1	3.117787842	2.528976908	-0.588810935	1.23E-07	1.42E-05
	PTTG1IP	2.4917533	1.90463486	-0.58711844	3.33E-07	3.33E-05
	TANGO2	3.738270085	3.158994767	-0.579275318	1.68E-07	1.79E-05
	EPRS	1.986586359	1.408606763	-0.577979597	1.22E-08	1.87E-06
	KLF1	6.156844746	5.5808046	-0.576040146	6.31E-08	7.86E-06
	KIF22	2.02701997	1.454094992	-0.572924978	2.88E-05	0.001189284
	CENPF	1.631172878	1.0794427	-0.551730178	4.66E-06	0.000271861
	MRPL37	2.217167401	1.666760573	-0.550406828	1.04E-06	8.55E-05
	HK1	3.427313106	2.881151196	-0.546161909	2.69E-06	0.000181007
	OAZ1	6.851524892	6.307439688	-0.544085204	2.29E-11	6.30E-09
	WDR34	1.570342191	1.026539004	-0.543803187	8.08E-07	7.10E-05
	TAX1BP1	3.788059152	3.24472698	-0.543332172	2.06E-08	2.95E-06
	EID1	2.684903448	2.143913404	-0.540990044	6.57E-07	5.86E-05
	HES6	3.806027904	3.267679308	-0.538348597	0.00011073	0.003366308
	SLC3A2	5.149312406	4.615975665	-0.533336741	9.41E-08	1.15E-05
	C7orf50	1.59736971	1.065645925	-0.531723785	1.58E-06	0.000124097
	TMEM56	3.460731579	2.930543962	-0.530187618	2.26E-07	2.37E-05
	PSMD7	3.343792234	2.814231262	-0.529560972	3.37E-06	0.000215856
	C9orf78	6.192212165	5.662918039	-0.529294125	2.45E-09	4.36E-07
	RPL8	4.099411483	3.572378896	-0.527032587	8.33E-07	7.14E-05
	EMC9	1.387050921	0.861549755	-0.525501167	2.75E-08	3.87E-06
	HADHA	2.122357078	1.597961194	-0.524395884	3.19E-07	3.24E-05
	PTGES3	3.896185088	3.371870631	-0.524314457	5.03E-05	0.001895776
	EEF1D	2.918343756	2.400368733	-0.517975023	1.16E-05	0.00056664
	NUSAP1	1.850792842	1.33423228	-0.516560562	5.39E-06	0.000309407
	SKP1	5.097729652	4.58343573	-0.514293921	1.15E-08	1.81E-06
	ARL6IP4	2.25676555	1.743848686	-0.512916864	5.76E-07	5.27E-05
	PSMB1	4.024794202	3.517561823	-0.507232379	2.87E-06	0.000189396
	TPX2	1.39810541	0.895469886	-0.502635524	5.11E-07	4.74E-05

Table 4.1 Significantly differentially expressed genes in scRNA-seq between ATRX KO and AAVS1 cells. The genes are ranked by their log2fc.

4.2.3.1 DNA damage response pathway is upregulated in ATRX KO cells

Among the 18 upregulated genes in ATRX KO cells indicated by the scRNA-seq data, the RING finger protein RNF168 (Figure 4.4A) caught our attention, as it has been shown as a key regulator of the DNA damage response and repair signalling at the damaged site. Upon DNA double-strand break, the MRN complex initially senses the damage, and recruits and activates ATM at DSBs which subsequently phosphorylates the histone variant H2A.X at serine 139 to form γH2AX that marks the damaged site and serves as a platform for MDC1 loading (J. Kelliher et al., 2022). MDC1 subsequently recruits E3 ubiquitin ligase RNF8 to DSBs to ubiquitylate histone H1.2 and L3MBTL2, which facilitates the recruitment of E3 ubiquitin ligase RNF168 to the damaged site, thereby amplifying the DNA damage response and promoting the recruitment of downstream key DDR proteins including 53BP1 and BRCA1 to execute DNA repair (J. Kelliher et al., 2022). In short, the MRN-ATM-γH2AX-RNF168 dependent pathway is essential for DNA damage response and repair. Importantly, ATRX has been shown interacting with the MRN complex to overcome replication stress and facilitate DNA repair (Leung et al., 2013; Clynes et al., 2014).

Given the upregulation of RNF168 in ATRX KO cells (Figure 4.4A), it therefore seems likely that the DNA damage response pathway is upregulated as well in ATRX KO cells. To test this, the level of the phosphorylated ATM (p-ATM) and H2AX (γH2AX) in bulk ATRX KO cells was examined by Western blotting. The results showed that both molecules were upregulated in ATRX KO cells (Figure 4.4B), indicating an upregulation of the DNA damage response pathway.

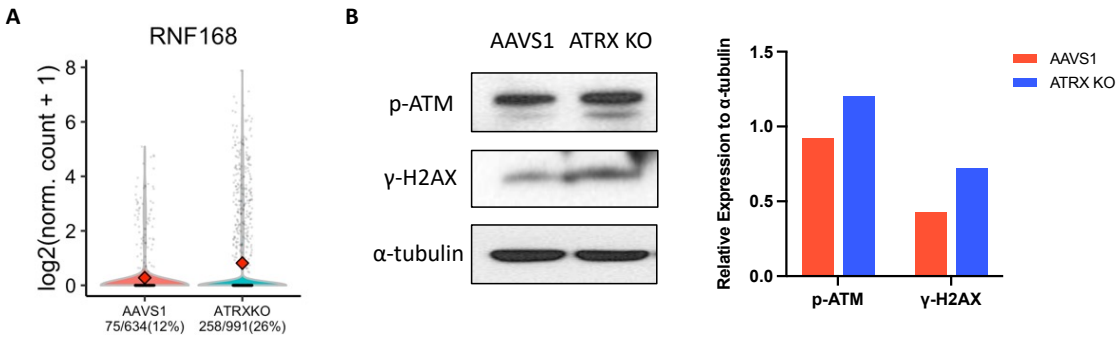


Figure 4.4 DNA damage response pathway is upregulated in ATRX KO erythroid cells.

(A) Violin plot shows RNF168 expression level of each group in the single cells. Each dot represents a single cell. The percentage shown below the group is the number and percentage of cells expressing RNF168 in the whole group. RNF168 is upregulated in ATRX KO group.

(B) Left: representative Western blotting of phosphorylated ATM and H2AX, along with the control protein α-tubulin. Right: quantified result of p-ATM or γ-H2AX protein bands versus α-tubulin.

4.2.3.2 Cell cycle and cell division related pathways are downregulated in ATRX KO cells

The GO analysis of the 53 downregulated genes were enriched in a number of cell cycle and cell division pathways, mainly on G2/M transition, nuclear division and sister chromatid segregation. The major genes contributing to the enrichment of these pathways include SKP1, TUBB, CENPF and RAN, which were all downregulated in ATRX KO cells (Figure 4.5). Their functions are shown in Table 4.2. ATRX has been previously shown binding at active elements in euchromatin and its mutations result in perturbed gene expression associated

with reduced chromatin accessibility, alterations in histone modification and transcription factor binding at these sites (Truch et al., 2022). I therefore checked if ATRX binds to these downregulated genes in erythroid cells by looking at the ATRX ChIP-seq data of normal CD34+ differentiated day 10 erythroid cells (Truch et al., 2022). This demonstrated that ATRX was enriched at the promoters of all these genes (Figure 4.5), suggesting ATRX regulates their expression by binding at their promoters and they are downregulated when ATRX is mutated.

Gene	Protein	Function	Active phase in cell cycle
SKP1	S-phase kinase associated protein 1, an essential component of the **SKP1**-CUL1-F-box (SCF) ubiquitin ligase complex	responsible for the ubiquitination and proteasomal degradation of many key cell cycle regulators including cyclin-dependent kinases (CDKs) to ensure proper cell cycle progression (Thompson et al., 2021)	late G1 to early M
TUBB	beta tubulin	forming microtubules with alpha tubulin which are responsible for the assembly of mitotic spindle (Narvi et al., 2013)	G2/M
CENPF	centromeric protein F	required for kinetochore function and chromosome segregation during mitosis of cell cycle (Y. gang Huang et al., 2022)	G2/M
RAN	a member of the RAS superfamily of small GTPases	RAN-GTP triggers microtubule polymerization at mitotic chromosomes which is required for normal mitotic spindle assembly and chromosome segregation (Militello & Colombo, 2013; Boudhraa et al., 2020)	G2/M

Table 4.2 Cell cycle and cell division related DE genes and their functions in cell cycle and cell division.

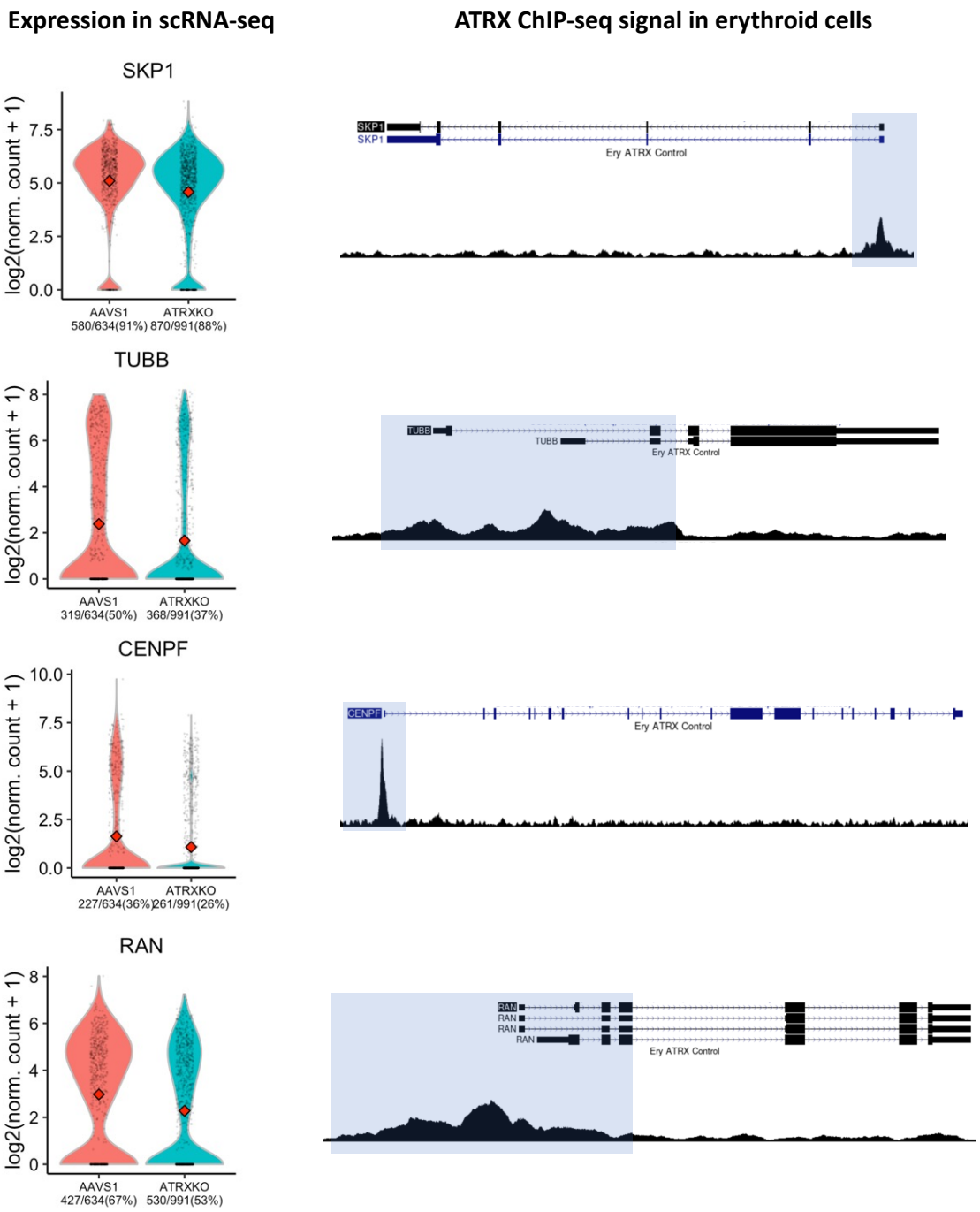


Figure 4.5 Expression of genes involved in cell cycle / cell division and ATRX binding signal at them in erythroid cells.

The left panel shows the gene downregulation in ATRX KO cells at the single cell level, and the percentage shown in each group represents the number and percentage of cells expressing the gene. The right panel shows the ATRX binding (highlighted in blue) at the promoter of the corresponding gene in erythroid cells.

4.2.3.3 Erythrocyte related genes are dysregulated in ATRX KO cells

The biological processes of erythrocyte differentiation and homeostasis are also enriched in the GO analysis of the downregulated gene set, and the contributing genes are FAM210B, TMEM14C, KLF1 and HMGB2 (Figure 4.6). Their functions in erythropoiesis are shown in Table 4.3. Apart from these downregulated genes, I also noticed CA1 and CA2 that have previously been shown associated with erythroid differentiation were upregulated in ATRX KO cells (Table 4.3, Figure 4.6). Given these genes have erythroid specific functions, the ATRX binding at these genes in erythroid and non-erythroid cells was checked in ATRX ChIP-seq data generated from the CD34+ day 10 erythroid cells and human lymphoblastoid cell line (LCL) derived from healthy individuals (Truch et al., 2022). It showed that ATRX had specific or stronger binding at these genes, either at promoter or intragenic region, exclusively in erythroid cells but with little or no binding in non-erythroid LCL cells (Figure 4.6), suggesting ATRX binds to these genes to regulate their transcription in a tissue-specific fashion.

Importantly, considering the functions of these genes in erythroid differentiation (either promoting or suppressing) and their changing pattern in ATRX KO cells (downregulated or upregulated), the erythropoiesis of ATRX KO cells seemed to be suppressed. However, this was not reflected in the tracking of the differentiation progress assessed by the cell surface marker expression (FACS) and cellular morphology changes (cytospin) throughout the differentiation process, since the ATRX KO cells showed similar differentiation and enucleation progression as the AAVS1 controls. One possible reason is that the difference at the molecular level might be indistinguishable at the cellular level.

Gene	Protein	Function	Role in erythropoiesis
FAM210B (downregulated)	an inner mitochondrial membrane protein	facilitating the formation of the mitochondrial iron transport complex that is required for hemoglobinization, proliferation, and enucleation during terminal erythroid maturation (Kondo et al., 2016; Yien et al., 2018)	promoting
TMEM14C (downregulated)	an inner mitochondrial membrane protein	promoting the import of protoporphyrinogen IX into the mitochondrial matrix for heme synthesis and subsequent hemoglobin production (Nilsson et al., 2009; Yien et al., 2014)	promoting
KLF1 (downregulated)	erythroid Kruppel-like transcription factor	acting as a global regulator of erythropoiesis by modulating the expression of a diverse set of genes involved in key erythroid pathways, including hemoglobin synthesis, erythroid membrane and cytoskeleton as well as erythroid turnover and the cell cycle (Tallack & Perkins, 2010; Yien & Bieker, 2013; Gnanapragasam & Bieker, 2017)	promoting
HMGB2 (downregulated)	chromosomal high mobility group protein	regulating erythropoiesis by potentiating the expression of the crucial transcriptional repressor of erythroid differentiation GFI1B (Laurent et al., 2010)	promoting
CA1 (upregulated)	carbonic anhydrase 1	CA1 expression is dropping over erythroid differentiation and thereby is considered as an early erythroid differentiation marker, while CA2 shows the opposite trend (Fraser & Curtis, 1987; Tufts et al., 2003).	suppressing
CA2 (upregulated)	carbonic anhydrase 2		unknown

Table 4.3 Erythrocyte related DE genes and their functions in erythropoiesis.

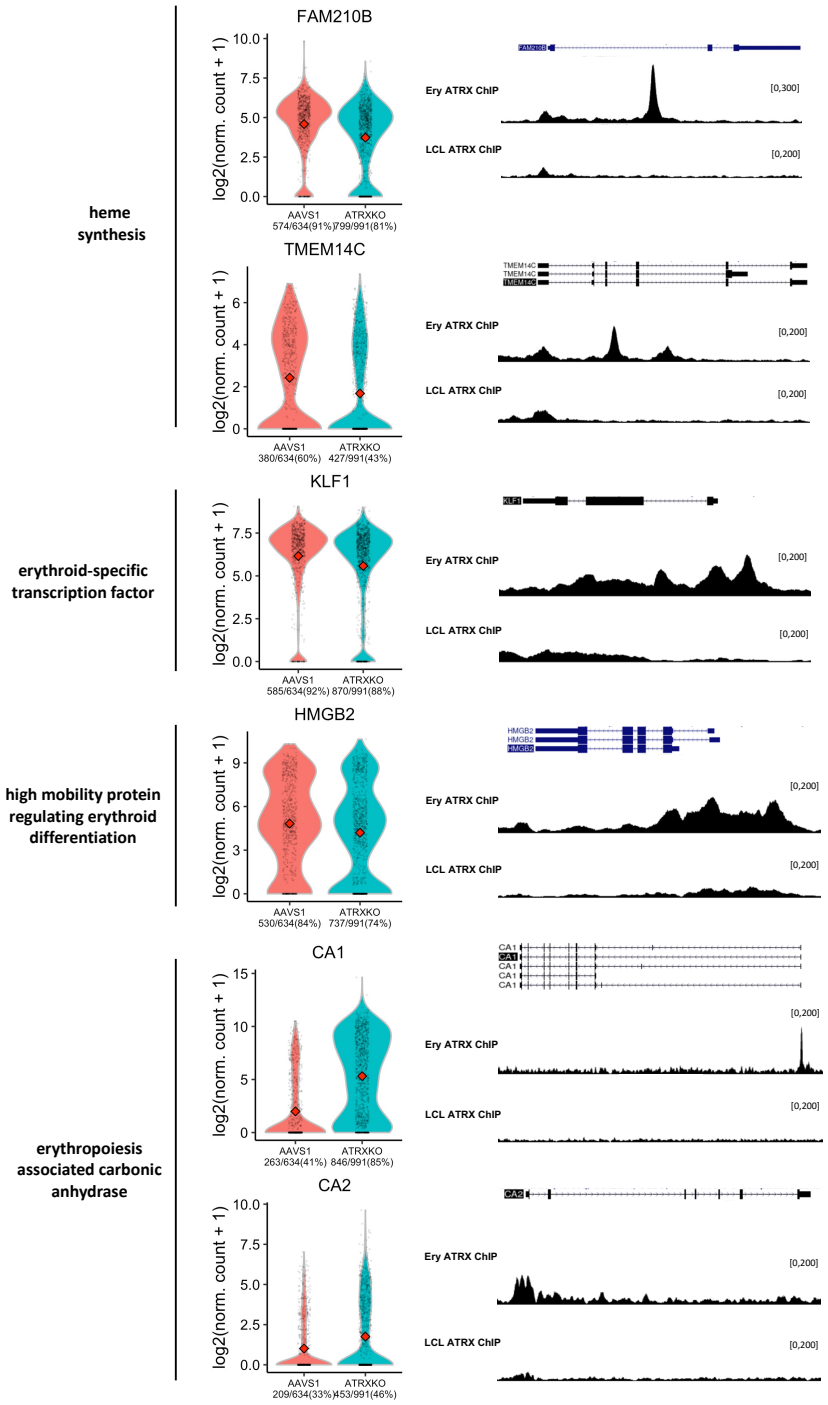


Figure 4.6 Expression of erythrocyte related genes and ATRX binding at them in erythroid and non-erythroid cells.

The left panel shows the main function and the expression of the genes. The right panel shows the ATRX binding at the promoter or intragenic region of the corresponding gene in erythroid (top track “Ery ATRX ChIP”) and non-erythroid cells (bottom track “LCL ATRX ChIP”).

4.2.4 Determination of α-thalassaemia phenotype in single cells

The data shown in the previous chapter demonstrate that ATRX defect gives rise to α-thalassaemia phenotype in only a subset of cells. To determine the α-thalassaemia phenotype in ATRX KO cells at the single cell level, the expression of globin genes including HBA, HBB, HBG and HBM was analysed and compared in single cells between the AAVS1 and ATRX KO groups. The analysis reported no significant difference in the expression of HBA, HBB and HBG between the two groups, but demonstrated a significant reduction of HBM expression in the ATRX KO group (Figure 4.7A). The significant reduction of HBM expression was indicated both by the decreased overall expression level in the ATRX KO group and by a bigger percentage of HBM-negative cells in the ATRX KO group (12%) compared with the AAVS1 controls (6%).

To better observe the correlation between HBA, HBM and HBB expression in single cells, their expression was shown in scatter plots (Figure 4.7B, C). It indicated that there was a subpopulation of cells in the ATRX KO group showing low expression of HBM with normal expression of HBB (Figure 4.7B highlighted with red frame). However, this pattern was less prominent for HBA expression, although there was also a tiny subset of cells with low expression of HBA (Figure 4.7C). Taken together, it can be concluded that, in agreement with the BFU-E data shown in the previous chapter, the scRNA-seq data demonstrate that α-thalassaemia only occurs in a small subset of cells and HBM is more affected than HBA by ATRX loss, which is consistent with the observation in ATR-X patients.

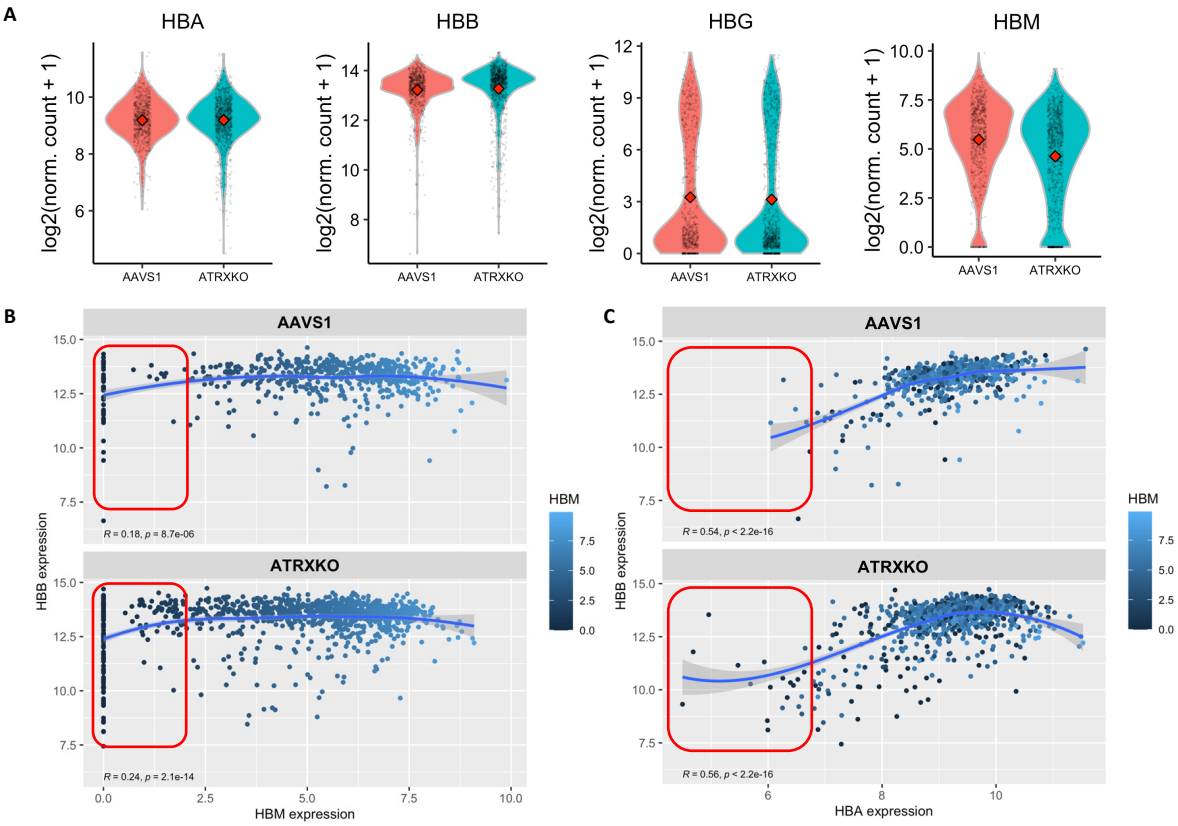


Figure 4.7 Globin gene expression in AAVS1 and ATRX KO cells at the single cell level.

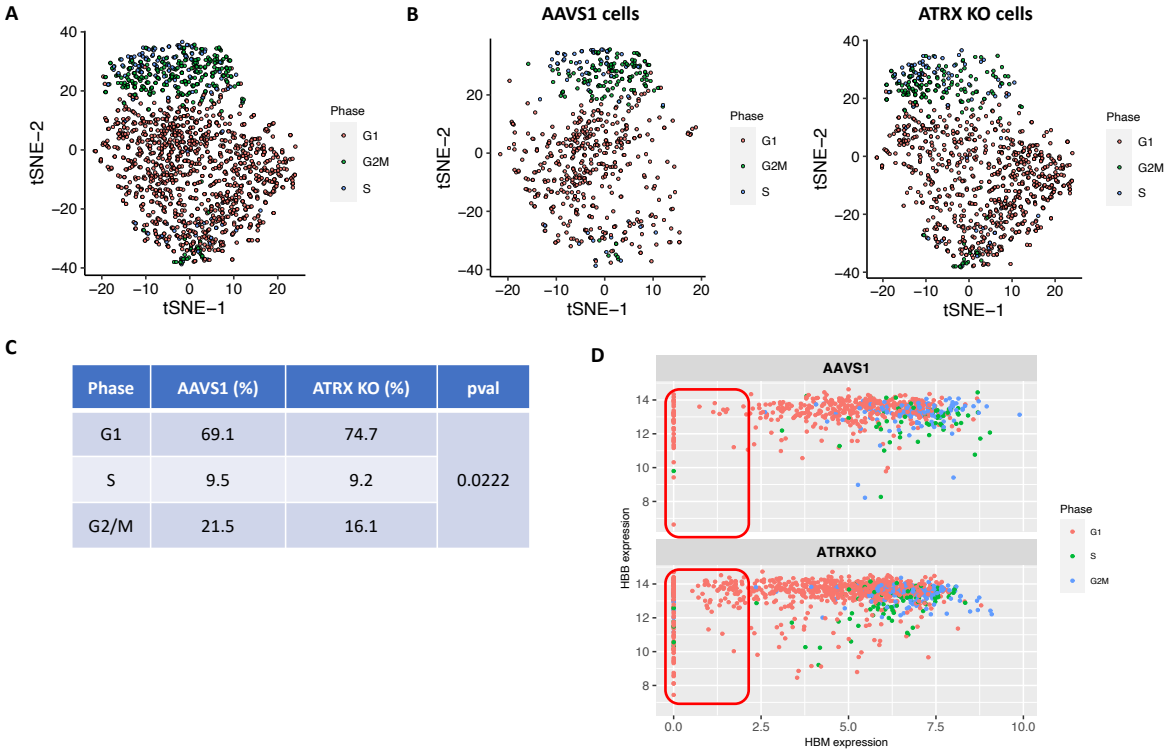
(A) The expression of the globin genes HBA, HBB, HBG and HBM in both groups.

(B) The expression of HBM and HBB in single cells from AAVS1 (top panel) and ATRX KO (bottom panel) group coloured by HBM expression level, showing a subpopulation of cells in ATRX KO group having low expression of HBM.

(C) The expression of HBA and HBB in single cells from AAVS1 (top panel) and ATRX KO (bottom panel) group colour by HBM expression level, showing a tiny subset of cells in ATRX KO group having low expression of HBA.

4.2.5 Cell cycle analysis in single cells

Since many cell cycle and cell division regulating genes were downregulated in ATRX KO cells as shown in section 4.2.3.2, cell cycle analysis was performed to assign a cell cycle phase (G1, S, G2/M) for each single cell based on their gene expression profiling. It indicated that the majority of the cells were in G1 phase and very few cells were in S phase (Figure 4.8A). Further examination of this in individual groups showed that the proportion of S cells was similar between the AAVS1 and ATRX KO groups, whereas there were more G1 cells and fewer G2/M cells in the ATRX KO group in contrast with the AAVS1 controls (Figure 4.8B, C). Fisher's test was performed to check the difference between the two groups and the *p*-value of 0.0222 suggested there is significant difference in cell cycle distribution between AAVS1 and ATRX KO groups. To check the cell cycle stage of the minor affected cell population that had downregulated HBM expression in the ATRX KO group, the scatter plot of HBM~HBB expression was coloured by the cell cycle phase, and it showed that the vast majority of the affected cells were at G1 phase (Figure 4.8D).



Phase	AAVS1 (%)	ATRX KO (%)	pval
G1	69.1	74.7	0.0222
S	9.5	9.2	
G2/M	21.5	16.1	

Figure 4.8 Cell cycle analysis of AAVS1 and ATRX KO cells at the single cell level.

(A) t-SNE plot of scRNA-seq data showing the cell cycle phase of each individual cell.

(B) t-SNE plot showing the cell cycle stage in AAVS1 and ATRX KO groups separately.

(C) Percentage of each cell cycle phase in both samples and the p-value between the two groups.

(D) The expression of HBM and HBB in single cells from AAVS1 (top panel) and ATRX KO (bottom panel) group coloured by cell cycle phase, showing the affected cells in the ATRX KO group are mostly in G1 phase.

4.2.6 DEG analysis between affected and unaffected cells

With the results presented above, it is clear that α(-like)-globin genes are only affected by ATRX loss in a subset of cells and the α-like globin HBM is more affected than the α-globin genes. Therefore, it would be more informative to do

DEG analysis between the affected and the unaffected cells, in order to unveil the underlying mechanism of the α-thalassaemia phenotype observed in the affected cells.

4.2.6.1 Classification of affected and unaffected cells

To determine the affected and the unaffected cells, the scRNA-seq data was re-analysed based on the expression level of the most affected globin gene HBM. All cells were re-grouped into HBM^{low}, HBM^{mid} and HBM^{high} subgroups based on their HBM expression level. Cells with HBM expression lower than 2.5 were classified as HBM^{low} cells, higher than 7 were classified as HBM^{high} cells, and others in between were HBM^{mid} cells (Figure 4.9A). With these criteria, the top 17.7% cells were labelled as HBM^{high} cells which can be regarded as the unaffected cells, and the bottom 15.9% cells were labelled as HBM^{low} cells which were perceived as affected cells. The cell number in each group was shown in Table 4.4. tSNE plots were then generated and coloured by sample group (AAVS1, ATRX KO) or HBM group (low, mid, high). It showed that HBM^{low} and HBM^{high} cells were more separated in the tSNE plot than AAVS1 and ATRX KO cells (Figure 4.9B), indicating the cells are more distinct judged by their HBM expression.

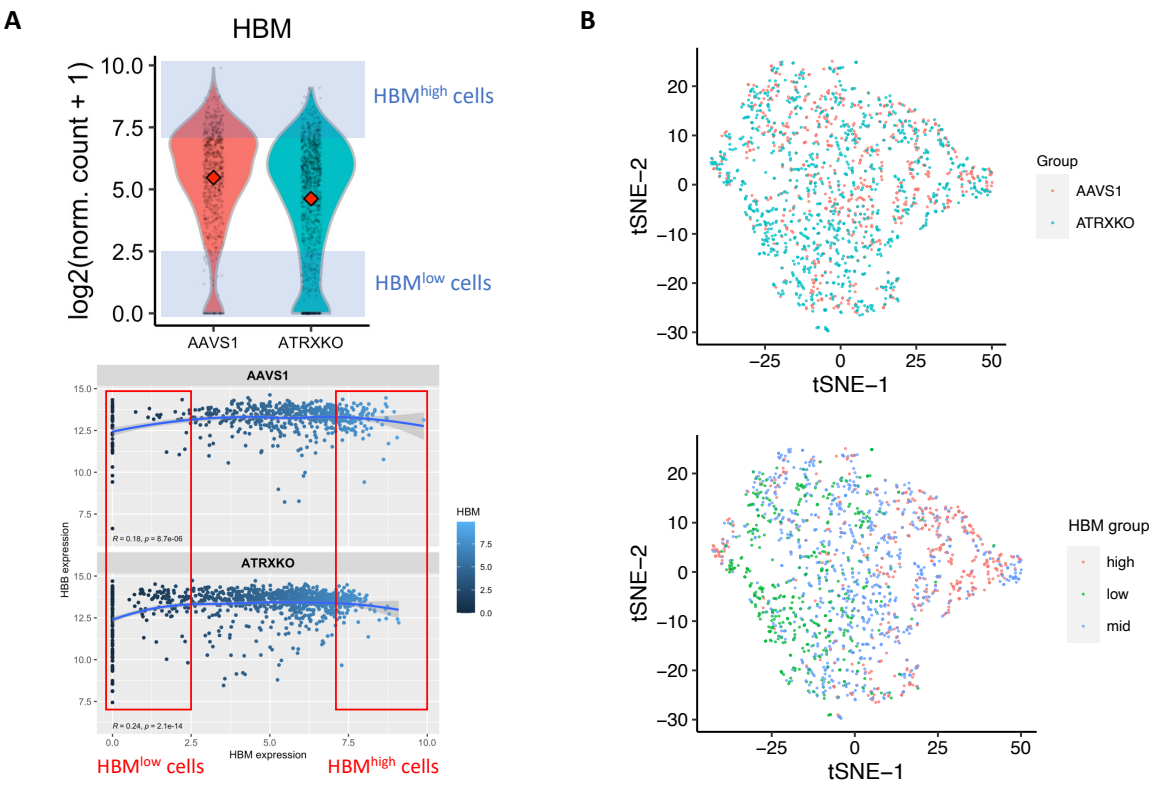


Figure 4.9 Single cell re-classification based on their HBM expression.

(A) Volcano and panel plots of HBM expression overlapping with the classification windows of HBM groups.

(B) t-SNE plots of scRNA-seq data coloured by sample group (top) or HBM group (bottom).

HBM group	AAVS1	ATRX KO	Total cell number (%)
low	59	199	258 (15.9%)
mid	419	661	1080 (66.5%)
high	156	131	287 (17.7%)

Table 4.4 Cell numbers in each group after reclassification based on the level of HBM expression.

4.2.6.2 DEG and GSEA analyses between the affected and the unaffected cells

DEG analysis was performed between HBM^{high} and HBM^{low} cells in ATRX KO and AAVS1 samples separately. GSEA analysis was subsequently conducted using the identified differentially expressed genes to enrich the differential pathways between the unaffected and the affected cells. Finally, the enriched pathways in the ATRX KO sample were compared with that enriched in the AAVS1 sample, and only the pathways exclusively present in the ATRX KO sample were considered to contribute to the phenotype differences caused by ATRX mutations (Figure 4.10). In total, there were 5 pathways upregulated and 56 pathways downregulated exclusively in the ATRX KO affected HBM^{low} cells (Appendix Table).

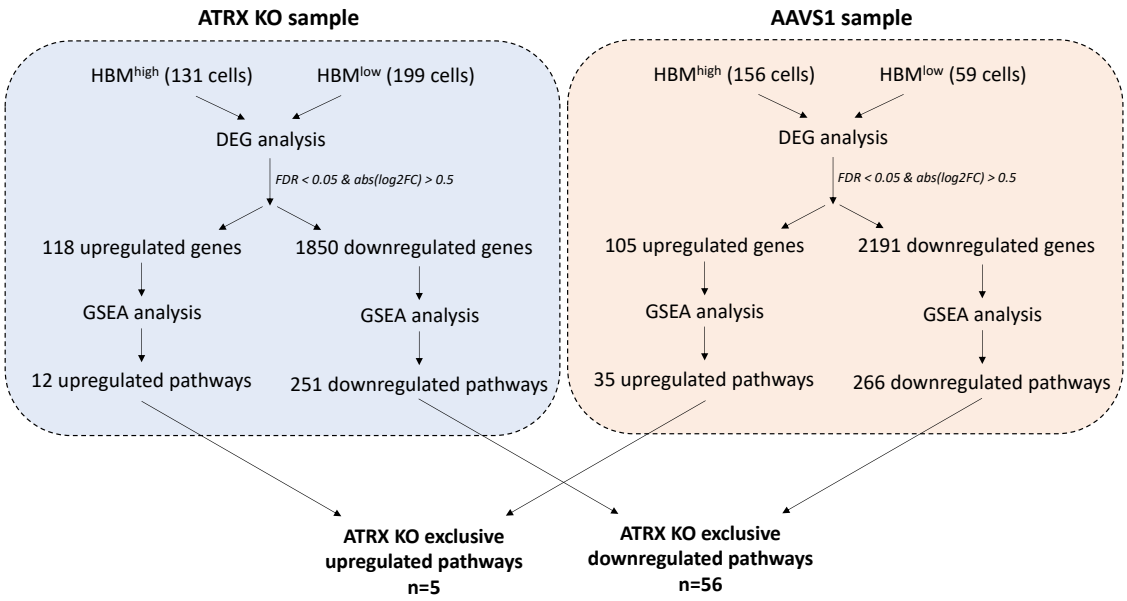


Figure 4.10 Workflow of DEG analysis between affected and unaffected cells.

DEG and GSEA analyses were performed separately in both samples and then ATRX KO sample exclusive differential pathways were determined.

The five upregulated pathways in the affected cells not only include those involved in the elimination of cellular organelles occurring at the terminal differentiation of erythroblasts such as autophagosome organization and mitochondrion disassembly, but also others regulating molecular activities like CDK activity and histone H2A ubiquitination (Figure 4.11A). Interestingly, histone H2A ubiquitination has been shown closely associated with gene repression (Barbour et al., 2020), indicating this upregulated epigenetic regulation of gene silencing might be responsible for the HBM downregulation in the affected cells. Among the 56 downregulated pathways, most are involved in cellular metabolic or catabolic processes and cell cycle-dependent structural assemblies (Figure 4.11B). Aside from that, ATP-dependent chromatin remodelling is enriched (Figure 4.11B), which may be related to disturbed gene expression.

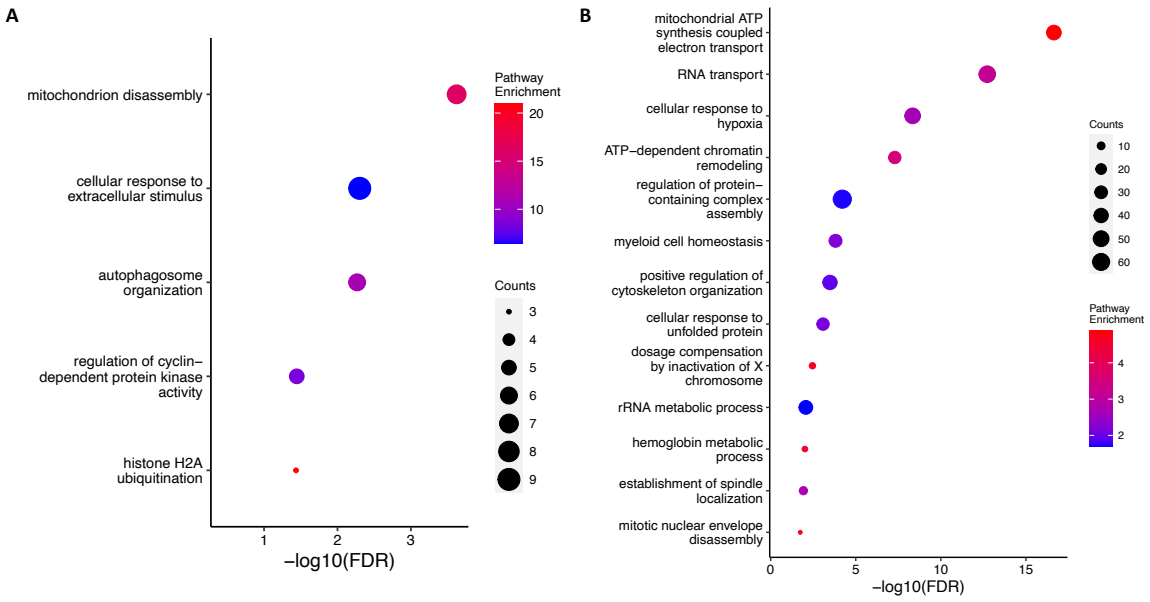


Figure 4.11 GSEA gene ontology enrichment analysis between affected and unaffected cells.

(A) Enrichment of the 5 upregulated pathways in the affected cells.

(B) Selected 13 downregulated pathways in the affected cells.

x-axis indicates the *p*-value; the size of the dot represents number of DE genes attributed to each GO term; and the colour of the dot indicates the abundance of this GO term.

4.3 Discussion

The target gene expression analysis in single BFU-E colonies shown in the previous chapter demonstrate that the ATRX defect gives rise to α-thalassaemia phenotype in only a small subset of cells. To investigate this on a larger scale and in a more informative way, in this chapter, a plate-based scRNA-seq method Smart-seq2 was performed to profile the cellular transcriptome of the control and ATRX KO cells and to identify the differences between them and between cell subpopulations to further reveal the underlying molecular mechanism.

When checking the ATRX knockout efficiency and globin gene expression in bulk cell population at a series of time points during erythroid differentiation, it was shown that the percentage of ATRX knocked out cells declined over the course of time, whereas HBM downregulation was gradually becoming prominent. This seems to be paradoxical as it would be expected that the lower the ATRX level, the more downregulated HBM. However, another critical factor that should be taken into account here is that globin genes including HBM only start to express at the later stages of differentiation, thus leading to the fact that their downregulation would be more noticeable later in differentiation. Therefore, cells used for scRNA-seq were collected at day 13 when the HBM downregulation was evident. The similar population size of enucleated cells in bulk ATRX KO and

controls indicate that ATRX has no obvious effect on the erythrocyte enucleation or maturation, although pathway analysis at the single cell level revealed a molecular difference in genes involved in erythrocyte differentiation and homeostasis between the controls and the ATRX KO cells. Notably, the phenomenon of the decreasing ATRX negative cells during differentiation is in agreement with the previous reported functions of ATRX in cell cycle and cell survival. There are two explanations for this, and they are not mutually exclusive. One is ATRX deficient cells undergo slower cell cycle which leads to a lower proliferation, consistent with a role of ATRX in cell cycle that has been indicated in many studies (Clynes et al., 2014; Huh et al., 2016; Akter et al., 2021; Qin et al., 2022). The other explanation is ATRX KO cells are gradually lost under a cell selection pressure, in line with the previous report showing that there is a strong cellular selection during haematopoiesis in female heterozygous mice against Atrx-deficient cells which results in the increased proportion of Atrx-expressing cells (Muers et al., 2007).

The DE genes enriched in erythroid pathways from the scRNA-seq data include genes encoding mitochondrial membrane proteins, FAM210B and TMEM14C, which are responsible for the normal iron metabolism for heme synthesis during terminal erythropoiesis (Yien et al., 2014, 2018; Kondo et al., 2016), and also genes encoding carbonic anhydrase CA1 and CA2 that regulate carbon dioxide conversion, iron transport and pH level (Brady et al., 1990; J. Chen et al., 2006; Song et al., 2012). Further, pathway analysis in the affected HBM^{low} cells revealed a large number of metabolic related pathways including both mitochondrial and nucleotide metabolism, which could be either due to the cells

being at different differentiation stages, or a direct effect of ATRX regulation. Intriguingly, it has previously been shown that the chromatin factor can regulate cell differentiation to determine cell fate by tuning metabolic activities (Guijas et al., 2018; Rossmann et al., 2021). Therefore, it is possible that ATRX loss leads to changes in cellular metabolism that may exert an effect on erythroid differentiation.

Another class of the dysregulated pathways in ATRX KO cells is cell cycle and cell division related. In accordance with this, the ATRX KO cells showed slower cell growth during differentiation, suggesting they underwent a longer cell cycle. Recent studies in ATRX-mutant cancer cells have correlated ATRX mutations with dysregulation of cell cycle progression associated with perturbed expression of cell-cycle regulatory genes including checkpoint regulators and centromere and spindle assembly regulators (Liang et al., 2020; Qin et al., 2022). But it is unclear as to whether this function of ATRX in regulating the expression of cell-cycle regulatory genes is exclusively present in ATRX-deficient cancer cells as a result of its role as a tumour suppressor or whether this is a comprehensive role of ATRX in all cell types for ensuring proper cell cycle progression. Here I show that ATRX functions as a general cell cycle regulator by regulating the expression of cell cycle and cell division related genes via directly binding to them. Among the genes affected by ATRX loss in erythroid cells, the genes critical during M phase for centromere and spindle assembly were also identified in a previous study working on ATRX-mutant hepatocellular carcinoma cells (Liang et al., 2020), including TUBB that is present in both data. Additionally, ATRX loss also affects the expression of genes involved in other phases of the cell cycle, such

as SKP1 that is active from late G1 to M phase through the cell cycle (Dang et al., 2021; Thompson et al., 2021). These findings suggest that ATRX regulates cell cycle in diverse cell types and thus is likely acting as a general regulator to ensure proper cell cycle progression.

Given these findings, I performed further cell cycle analysis on the single cell data to determine if the ATRX KO erythroid cells displays distinct cell cycle profiles compared to control cells. Notably, the majority of these erythroid cells at this differentiation stage (day 13) are in G1 and very few of them are in S phase, suggesting the cells are less actively divided, which is expected for terminally differentiating cells. This is consistent with the recent finding that cells in terminal differentiation undergo simultaneous differentiation commitment and permanent cell cycle exit in G1 (Zhao et al., 2020; Sainz de la Maza et al., 2022). Interestingly, the cell cycle profile data showed that the ATRX KO cells were more quiescent as there was a significant increase in the G1 cell population size and decrease in the G2/M population size in ATRX KO cells. There are two possible reasons for this. One is that more ATRX KO cells exit from the cell cycle and stay in G1 phase, and the other is ATRX KO cells are halted in G1. The first reason is relatively less likely because the transcriptomic changes in ATRX KO cells indicate these cells are less differentiated than the controls, so they would not be expected to step into terminal differentiation and exit cell cycle earlier than the controls. There is also other evidence that makes the second reason plausible, which is the upregulated DNA damage pathways and the elevated level of phosphorylated (activated) ATM and γH2AX in ATRX KO cells compared to controls. It is known that the DNA damages that is not repaired during G2/M is

repaired in the subsequent G1 stage primarily by 53 BP1 nuclear bodies, which often occurs at chromosomal fragile sites (Lukas et al., 2011). ATRX has been shown frequently colocalizing with 53BP1 nuclear bodies in G1 phase to modulate the repair of these fragile sites (Pladevall-Morera et al., 2019). I have checked the expression of *TP53BP1* gene in my single cell data and it turned out to be upregulated slightly in ATRX KO cells but not a significant change. Moreover, the increase of phosphorylated ATM and its substrate γH2AX in ATRX KO erythroid cells also show consistency with a recent study reporting similar observations in ATRX-mutant glioma cells (Qin et al., 2022). Taken together, it is likely that ATRX KO cells contain DNA damages at fragile sites that need to be repaired at the G1 stage, thus resulting in more G1 cells in ATRX KO sample. Further experiments are needed to confirm this in ATRX KO erythroid cells, for example analysing the cell cycle with cell samples during differentiation, checking for DNA damage and changes in 53BP1 nuclear bodies or ATRX regulated molecules responsible for DNA repair in G1 phase.

The increased DNA damage in ATRX KO erythroid cells and the activation of the ATM-γH2AX pathway in particular, is consistent with previous observations in other ATRX-depleted cells (Huh et al., 2012, 2016; Qin et al., 2022). However, the upregulation of the E3 ligase RNF168 in ATRX KO cells has not been reported before. RNF168 is known as a key molecule during the DNA damage response by catalysing the ubiquitination of histones and other DDR proteins (J. Kelliher et al., 2022). Histone H2A is its primary histone substrate, and various histone H2A variants can be ubiquitinated by RNF168, including H2A, H2AX, H2AZ and macroH2A1/2 (J. L. Kelliher et al., 2020). Among these, macroH2A1,

an inert histone mark that can interact with ATRX and plays a role in maintaining genome integrity and regulating DDR (Ratnakumar et al., 2012; J. Kim, Sturgill, et al., 2018; J. Kim et al., 2019; Sebastian et al., 2020), is a prime candidate to link ATRX loss, DNA damage and gene repression together. Moreover, here it is shown that the histone H2A ubiquitination pathway is upregulated in the affected HBM^{low} cells and the contributing genes for the enrichment of this pathway are UBR2, UBE2B and PCGF5 that are all able to promote H2A ubiquitination and mediate transcriptional silencing (An et al., 2010; Mou et al., 2013; Yao et al., 2018), indicating α-globin gene downregulation might be associated with increased H2A ubiquitination. Collectively, it therefore appears likely that ATRX loss gives rise to DNA damage, and leads to upregulated ATM-RNF168 which promotes the ubiquitination of H2A and other DNA damage response proteins to facilitate the establishment of the DNA repair environment. Meanwhile, the H2A ubiquitination suppresses the local gene expression. This links the DNA damage to disturbed gene expression induced by ATRX loss. Similar models have been demonstrated previously, showing ATM pathway orchestrates DNA damage induced transcriptional repression and DNA repair processes to maintain genome integrity (Shanbhag et al., 2010; Hu et al., 2020). This model directs future work to detect the occurrence of DNA damage at α-globin locus, mostly likely around the variable tandem repeat ψζ VNTR, and to determine the potential repressive epigenetic marks that link DNA damage to transcriptional silencing, for instance, H2A ubiquitination or macroH2A. It is reasoned that the responsible epigenetic mark is able to propagate from the DNA damage site, around ψζ VNTR in the case of α-globin locus, and thereby generating the distance-dependent gene downregulation pattern – the nearest HBM is more affected than

the further HBA2/1 genes. To test this hypothesis, the affected subset of cells need to be enriched to further determine the mechanism in comparison with the unaffected cells, which will be the contents of the next chapter.

There can be no denying that the scRNA-seq here has drawbacks as well. In the first place, ATRX is not efficiently detected in the single cells, which precludes it from working as an indicator of ATRX abundance in each single cell. This brings to the second downside. That is, I have to use HBM level as a marker to classify the affected and unaffected subpopulation, but this may not fully reflect the real biological consequences of ATRX deficiency since the HBM^{low} cell may have intact ATRX in it. Ideally a method is needed for simultaneously determining ATRX genotype and transcriptome analysis in single cells, for example TARGET-seq (Rodriguez-Meira et al., 2020). Nevertheless, given the overall high ATRX knockout efficiency in the whole population and many of the consistent observations obtained from my scRNA-seq data with the previous reports, the data shown here should be fairly reliable to provide clues to mechanism.

In conclusion, in this chapter, single cell RNA-seq was performed to analyse the transcriptomic changes of CD34+ differentiated ATRX KO erythroid cells at the single cell level. It not only reconfirms the stochastic manner of α-globin downregulation caused by the ATRX defect that was observed in BFU-E colony analysis, but also indicates a number of molecular processes that ATRX is involved in, including DNA damage response pathway, cell cycle related pathways as well as erythroid pathways. DEG analysis between the affected HBM^{low} and the unaffected HBM^{high} subpopulations demonstrates that histone

H2A ubiquitination and metabolic activities are distinguishing features, which may be responsible for the gene dysregulation. The cell cycle analysis shows that the affected cells are largely in G1, which is consistent with our most recent single cell data on CD34+ cells from the ATR-X patients (Truch et al., 2022), suggesting our cellular model recapitulates most of the observations in the patient cells. Importantly, here for the first time, pATM-RNF168-ubiquitination is indicated as a potential DDR and gene regulation pathway in response to ATRX loss.

5 Generation of ATRX degron and HBM-mKate2 reporter cell line and recapitulation of α-thalassaemia phenotype

5.1 Introduction

The work presented so far has clearly demonstrated that ATRX loss perturbs gene expression at the α-globin locus in only a subset of cells. Single cell differential gene expression analysis in the affected subpopulation indicates an association between DNA damage induced epigenetic instability and gene silencing, which might be the mechanism of gene downregulation at the α-globin locus. To validate this, the affected cells need to be enriched from the bulk cell population for further investigation. One way to achieve this is to insert and express a fluorescent protein at the α-globin locus which could serve as a reporter of the local gene expression in cells and also allow FACS-purification of live affected cells. This approach involves the utilization of genome engineering and cell clone selection, which takes a long time. Given the limitations on the long-term maintenance of primary HSCs, an immortalized human erythroid progenitor cell line was selected for the generation of the reporter cell line. To avoid the uncertainty of ATRX knockout efficiency by CRISPR-Cas9, a ligand-induced degron system was adopted to rapidly degrade ATRX directly at the protein level.

The expression of a protein of interest can be controlled and thereby perturbed at the DNA, mRNA or protein level. The ligand-induced degron systems are superior to the methodologies that function at the DNA and mRNA levels such as CRISPR knockout and siRNA, in many respects. For one thing, approaches

targeting a pre-translational process at the DNA and mRNA levels may produce harmful, unexpected and irreversible off-target effects on the cell. Also, there is the possibility that in culture the cells adapt to the loss of the target protein and accumulate diverse secondary effects that may obscure the primary defect. Instead, ligand-induced degron technologies introduce rapid depletion of a degron-fused protein via the ubiquitin-proteasome pathway by leveraging the potency of cell-permeable ligands that regulate the association between a degron-fused target protein and an E3 ubiquitin ligase. The depletion can be abrogated and reversed by simply modulating the presence or absence of the ligand in the cell culture system.

The original degron system made use of the destabilizing domain (DD) in the mutant FK506-binding protein FKBP12 that confers instability to the target protein fused to the DD, which becomes unstable in the absence of a DD-binding ligand like Shield-1 and is made stable by the addition of the ligand (Banaszynski et al., 2006). Subsequently, a number of degron systems were developed to induce the degradation of a degron-fused protein via the addition of a ligand, for example the most well-known auxin and degradation tag (dTAG) molecules (Nishimura et al., 2009; Nabet et al., 2018, 2020; S. Li et al., 2019; Yesbolatova et al., 2020). Up to now, auxin and dTAG induced degron systems have been successfully used in numerous studies to induce protein degradation. dTAG exploits the endogenous E3 ligase CRBN or VHL which makes it a simple and clean system (Nabet et al., 2018, 2020). However, the effectiveness of the target protein degradation can be context-specific and protein-specific (Nabet et al., 2018, 2020), and the limited conservation of both CRBN and VHL in vertebrates

makes it not applicable to other eukaryotic species (Kanemaki, 2022). Conversely, the auxin degradation pathway is highly orthologous, but it falls short in the requirement of an exogenous ubiquitin ligase subunit OsTIR1 in addition to the degron fusion, and usually requires high concentrations of ligand that may cause cytotoxicity (Nishimura et al., 2009; Yesbolatova et al., 2020). Therefore, dTAG inducible degron system has become an increasingly popular system applied in human and mouse cells. The CRISPR-Cas9 genome engineering mediated knock-in of $FKBP12^{F36V}$ at an endogenous locus allows specific and potent endogenous protein degradation induced by dTAG degraders, dTAG-13 in particular (Nabet et al., 2018). In the presence of the ligand dTAG-13, it binds to $FKBP12^{F36V}$ to promote the dimerization of $FKBP12^{F36V}$-targeted protein fusion and the CRBN E3 ligase complex, leading to ubiquitination and proteasomal degradation of the targeted protein (Nabet et al., 2018). Herein, this system is used for degrading ATRX protein in the cells.

To facilitate the observation of changes in gene expression induced by ATRX depletion, a fluorescent protein can be inserted at the α-globin locus to serve as a reporter. Such a strategy has been greatly facilitated by the advances in genome engineering technology, and is widely used by researchers to generate fluorescent reporters with high efficiency and scarless integration at targeting locus (Yang et al., 2013; A. J. King et al., 2021). The choice of an appropriate fluorescent protein is critical for the readout of globin gene expression in erythroid cells, since haemoglobin itself emits in the red channel and can potentially quench some fluorescent signal (J. Chu et al., 2014). Taking the excitation/emission spectrum, brightness, photostability and half-time into

consideration (Shaner et al., 2005; J. Chu et al., 2014), the far-red fluorescent protein mKate2 appears to be the most sensible choice and has been selected as the reporter in this project. Although the detection of mKate2 requires specialized infrared microscopy, it can be easily detected by flow cytometry which enables us to do the FACS sorting.

There are multiple genes at the human α-globin locus, among which the α-like globin gene *HBM* is selected as the targeting gene to be tagged with mKate2 for several reasons. Firstly, HBM is more downregulated than the α-globin genes HBA2/1 and therefore more sensitive to ATRX depletion, which makes it a good indicator of gene expression changes in live cells. The second reason involves the consideration of the genome engineering practice. There is only one copy of *HBM* on chromosome 16 which makes the genome engineering easier than editing of the two homologous α-globin genes. Thirdly, editing *HBM* would ensure all four α-globin genes being expressed and would not induce artificial imbalance between α and β globin. However, it remains uncertain whether *HBM* is translated, since its encoding protein μ-globin was not detected previously in erythrocytes from cord blood and adult blood samples by mass spectrometry (Goh et al., 2005). Nevertheless, the ribosome profiling data in human HSCs differentiated erythroid cells demonstrated that HBM was being actively translated (Khajuria et al., 2018). These findings suggest that HBM can be translated but its protein product might not be stable or abundant enough to be detected by mass spectrometry. The recruitment of ribosomes at HBM implies that insertion of mKate2 at this locus is likely to be translated and can potentially be used as a reporter.

In this chapter, I describe the generation and characterisation of a dTAG-13 inducible ATRX degron HBM-mKate2 reporter cell line that can be used to enrich the affected cells for further mechanistic studies.

5.2 Results

5.2.1 Generation and characterisation of ATRX degron cell line

5.2.1.1 ATRX degron engineering design and targeting strategy

To establish the dTAG-13 inducible degron cellular system, a mutant form of FKBP12 (FKBP12^{F36V}) needs to be fused to the targeted protein, ATRX in this case. This can be achieved via genome engineering of the endogenous gene locus of the targeted protein. Specifically, a CRISPR-Cas9 mediated DNA double-strand break is generated at the *ATRX* locus, and a DNA donor template containing FKBP12^{F36V} is supplied for the cell to repair the break via homology directed repair, thus leading to the incorporation of the FKBP12^{F36V} sequences into *ATRX* locus to generate a fusion protein.

ATRX is a large gene consisting of 35 exons and its encoded protein contains various functional motifs and domains. The full-length ATRX has three transcriptional isoforms that differ in one or two internal exons. Essentially, several considerations need to be taken into account for the design of the FKBP12^{F36V} insertion. Firstly, the insertion should not affect the localisation and normal function of ATRX. The 35th exon which contains the C-terminus of ATRX therefore seems to be an ideal tagging site as it allows the normal production of

functional full-length ATRX protein. Secondly, all full-length transcriptional isoforms should be tagged so the overall protein degradation can be efficient and complete, and tagging at 3' end of *ATRX* should enable this. Last but not least, it would be useful to insert a fluorescent protein simultaneously which would allow direct monitoring of both targeting and later protein degradation, thus the enhanced green fluorescence protein (EGFP) is used here. With all that in mind, EGFP-FKBP12^{F36V} was designed to be inserted and fused to the C-terminus of ATRX, following the coding sequences (CDS) within terminal exon 35 (Figure 5.1).

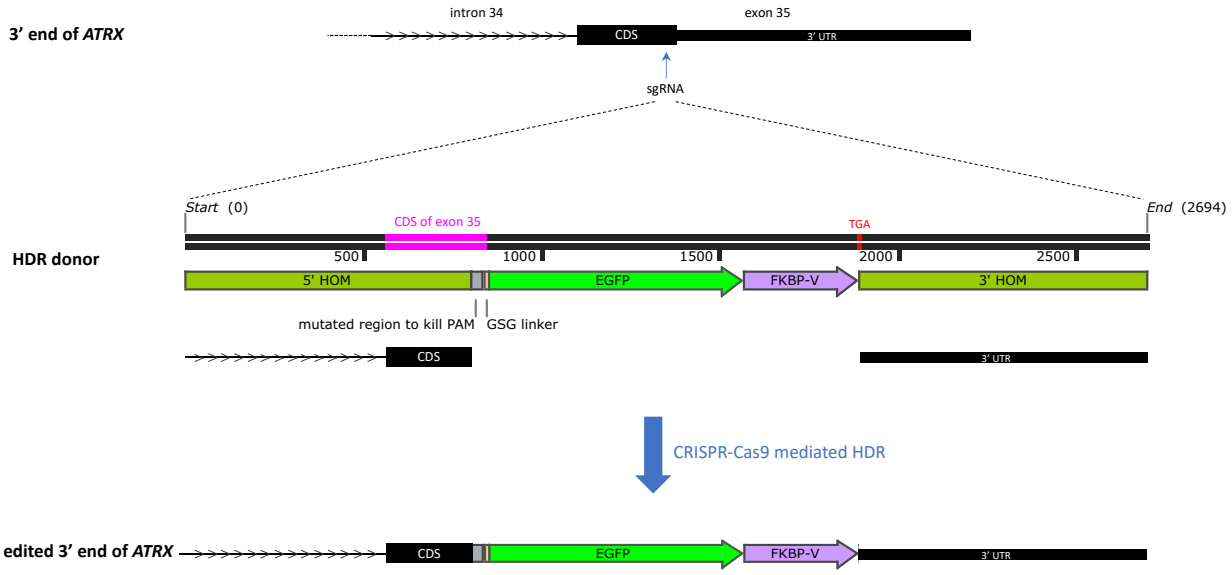


Figure 5.1 Schematic for ATRX targeting design to generate ATRX degron cell model in HUDEP-2 cells.

The top panel shows the 3' end of *ATRX* and the targeting site of the sgRNA. The middle panel denotes the HDR donor design, 2694 bp in length, which contains around 800 bp homology arms at each side of the insertion EGFP-FKBP12^{F36V}, and a "Gly-Ser-Gly" (GSG) oligopeptide linker between ATRX and EGFP-FKBP12^{F36V}. Of note, the PAMs for

all the tested sgRNAs are mutated in the donor DNA to avoid re-targeting but without changing the encoded amino acids. The bottom panel depicts the desired final edited product.

Since *ATRX* is located on chromosome X, it would be preferable to use a male cell line where only one allele needs to be targeted and edited. Therefore, of the two leading erythroid progenitor cell lines, HUDEP-2 cell line (49-53, XY) (Moir-Meyer et al., 2018; Vinjamur & Bauer, 2018) is preferable to the BEL-A cell line (44-48, XX) (Daniels et al., 2020), and was finally selected for ATRX targeting. The targeting strategy is illustrated in Figure 5.2. Briefly, the HUDEP-2 cells were electroporated so that CRISPR plasmid pX458-Ruby and the donor plasmid were introduced into the cells. pX458-Ruby not only encodes sgRNA and Cas9 protein, but also expresses the fluorescent protein Ruby which allows the selection of the transfected cells that have taken up the plasmid and thus likely to generate DSB at the *ATRX* site for further HDR-mediated editing. Hence, the Ruby positive (Ruby+) cells were purified by FACS prior to separation and expansion of single clonal lines. The clonal lines were then screened by a PCR-based strategy. Technical details are included in the Materials and Methods chapter section 2.3.2.

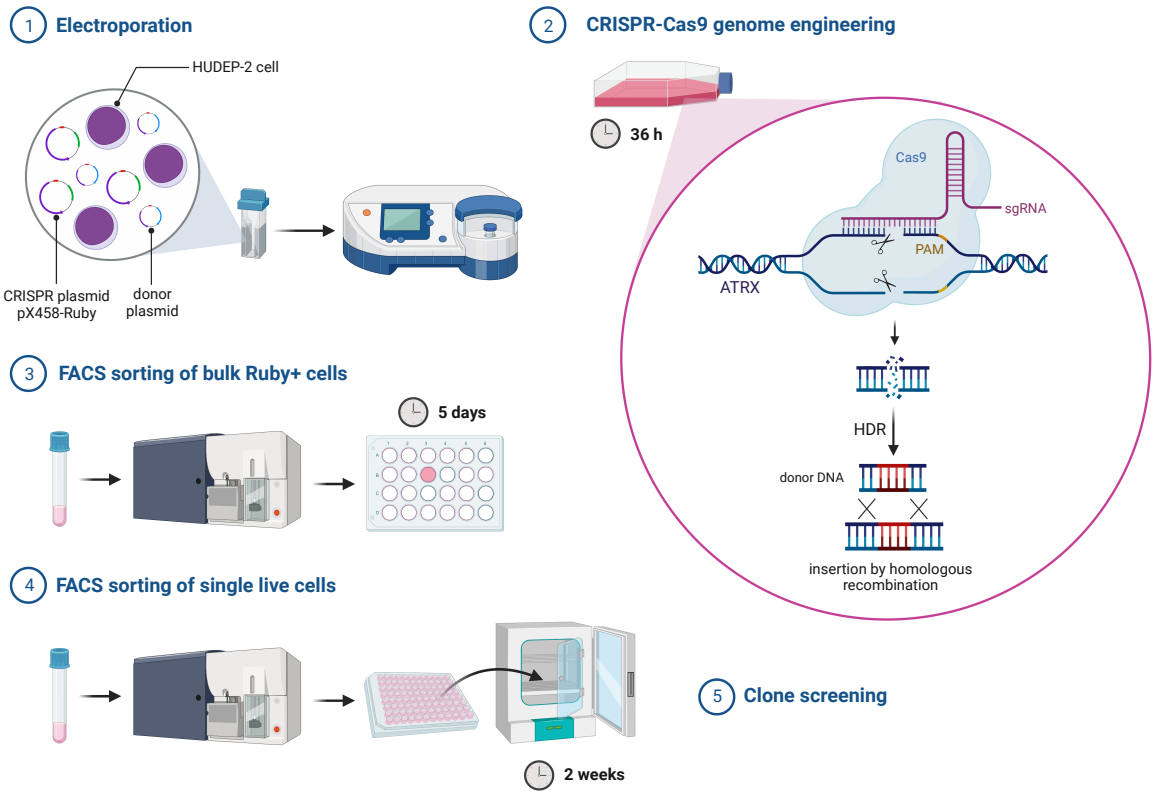


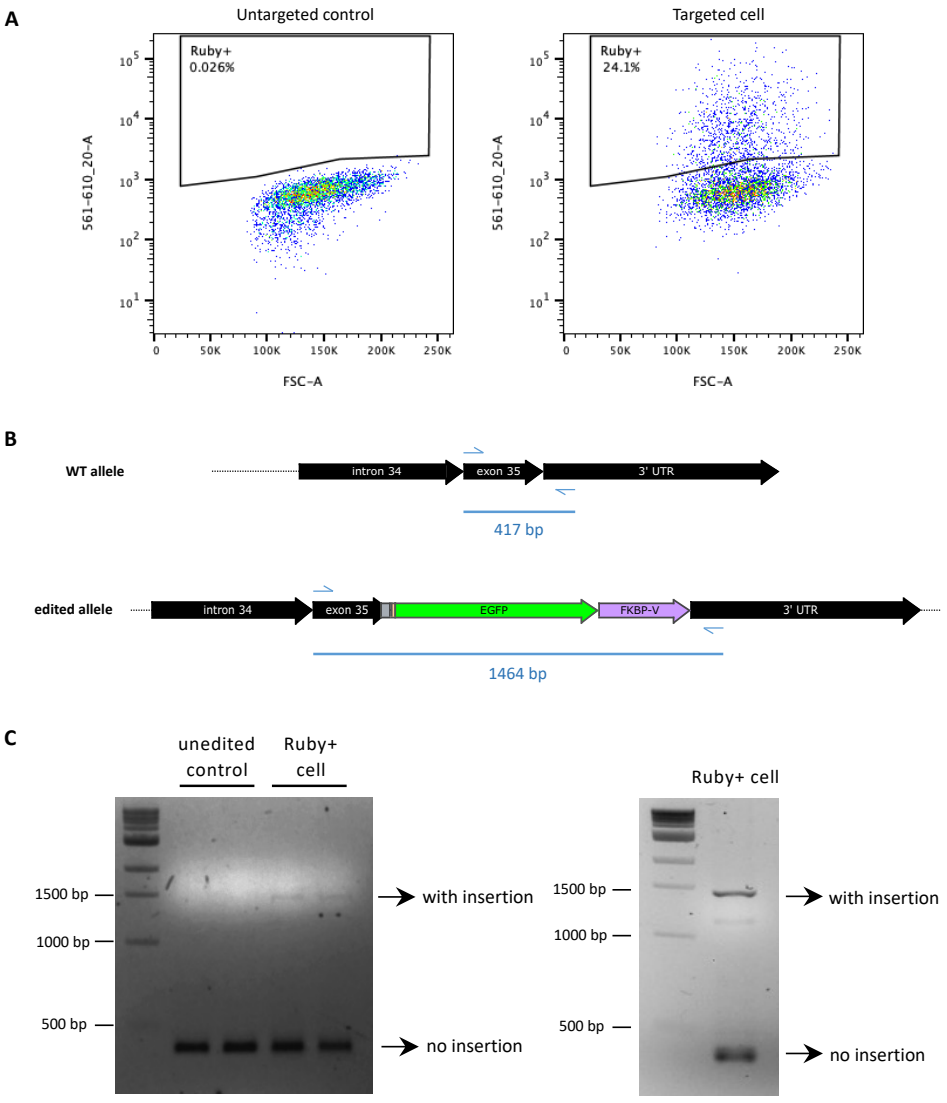
Figure 5.2 Workflow for ATRX degron targeting in HUDEP-2 cells.

The figure was created with BioRender.com.

5.2.1.2 Screening of targeted clone

During the targeting process, bulk Ruby+ cells were initially sorted (step 3 in Figure 5.2) to enrich the cells that had been targeted and expressed the plasmids (Figure 5.3A). Then, the sorted cells were recovered for 5 days prior to the single cell sorting, during which DNA was extracted from the bulk cells to determine the overall editing efficiency by PCR. The PCR primers used for determining editing efficiency in bulk and also for later single clone screening were placed outside of the editing region (denoted by arrows in Figure 5.3B). A PCR reaction was performed with the primers, and the product from the wildtype allele should be

417 bp in length while the edited allele was expected to produce a 1464-bp product (Figure 5.3B). The ATRX degron targeting experiment was performed twice separately. The overall editing efficiency for the first experiment was low (Figure 5.3C left), and no desirable clone was subsequently identified from the sorted 300 single cells. Consequently, the second targeting experiment was performed and the editing efficiency in bulk Ruby+ appeared to be higher than the first time (Figure 5.3C right). Then, 600 single cells were sorted, expanded and PCR screened, among which one positive clone, clone 3-8, was identified with an appropriately sized band (Figure 5.3D).



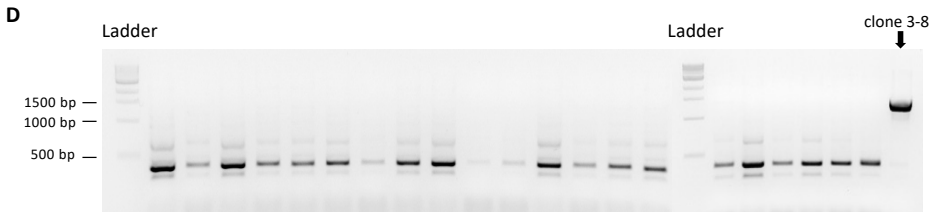


Figure 5.3 Screening of ATRX degron clone.

(A) Bulk Ruby+ FACS sorting plots of untargeted controls and the targeted cells, showing 24.1% of cells are Ruby positive in the targeted cells.

(B) Diagram showing the PCR screening primers and products for WT and edited allele.

(C) PCR products of bulk Ruby+ cells shown in 2% agarose gel, indicating editing efficiency, in the first (left) and the second (right) targeting experiment.

(D) PCR screening results of material from single targeted clones. Clone 3-8 was identified as a positive clone.

The PCR product of this clone was then sequenced and confirmed to have the correct integration (Figure 5.4A). To ensure the *ATRX* locus remained intact after genome engineering in this clone, Bionano optical mapping was performed using high molecular weight genomic DNA, which confirmed the integrity of the locus and also identified the edited insertion at the 3' end of *ATRX* that was absent in the wildtype HUDEP-2 (HUDEP-2 WT) cells (Figure 5.4B). Together, it has been confirmed that the ATRX degron clone has the correct sequence incorporated, and that the rest of the locus remained structurally intact.

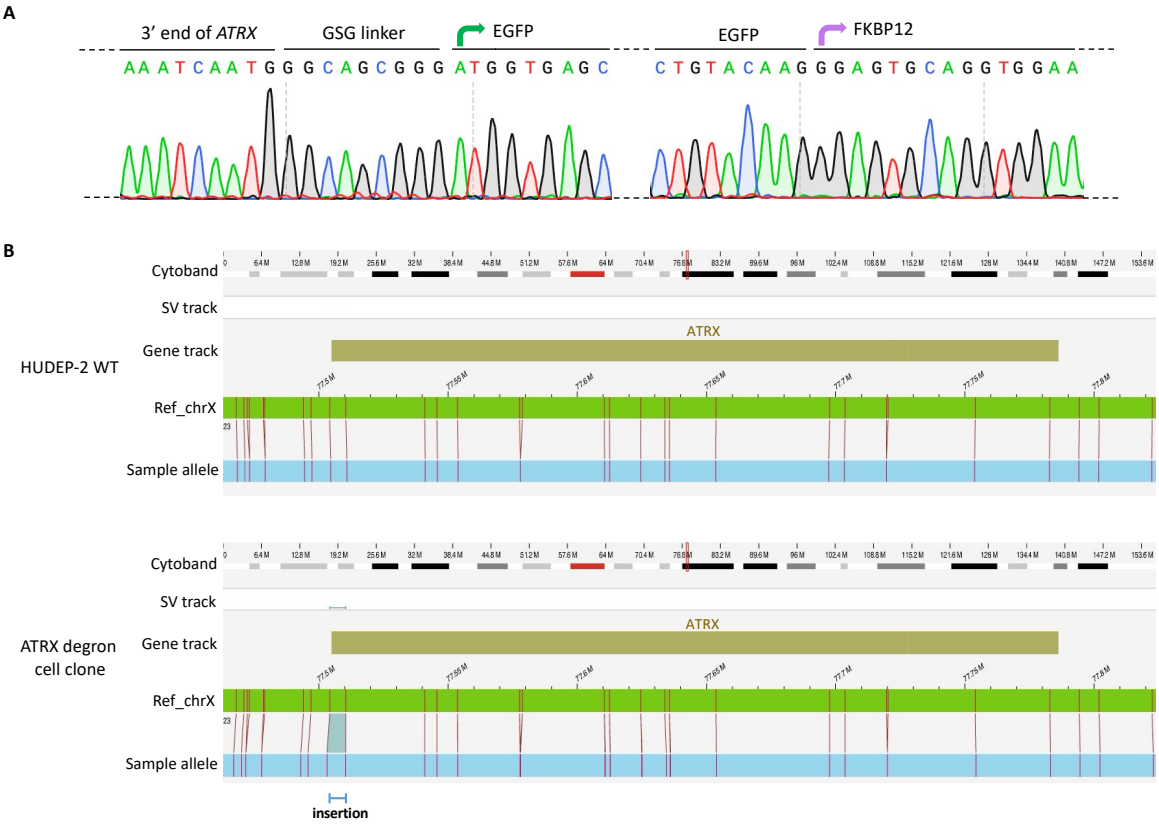


Figure 5.4 Genomic validation of the ATRX degron cell line.

(A) Sanger sequencing result of the targeting *ATRX* locus in the ATRX degron cell clone showing the correct integration.

(B) Schematic of *ATRX* locus visualised with Bionano Saphyr software. The optical mapping with a ~360 kb window (hg38 chrX: 77,462,939-77,823,945) determines normal genomic region in HUDEP-2 WT cells (top panel), and detects the insertion at the 3' end of *ATRX* as a structural variant (the light blue mark in "SV track") and otherwise intact genomic region in the ATRX degron cell line (bottom panel). Ref_chrX is the reference genome assembly provided by the Bionano Saphyr software. Sample allele indicates the allele in HUDEP-2 WT and ATRX degron cell samples.

5.2.1.3 Validation of ATRX-EGFP-FKBP12 fusion protein expression in the ATRX degron cell clone

After confirming the correct integration at the genomic level, the protein expression was next examined to check if the ATRX-EGFP-FKBP12 fusion protein was properly expressed. By performing Western blot with antibodies against different proteins, EGFP and FKBP12 were confirmed to be fused with ATRX in the identified ATRX degron cell clone (Figure 5.5A). Next, the GFP signal in the ATRX degron cell line was checked. It was not conspicuous when checked under fluorescent microscope. Therefore, flow cytometry was performed to analyse the GFP intensity. It demonstrated that ATRX degron cell line had higher GFP intensity than HUDEP-2 WT cell but the difference between them was too subtle to have distinct separation (Figure 5.5B). Given the clear GFP signal detected by Western blot, the possible reasons for the faint GFP signal in ATRX degron cells could be its weak expression or its nuclear localisation that made the detection not sensitive enough.

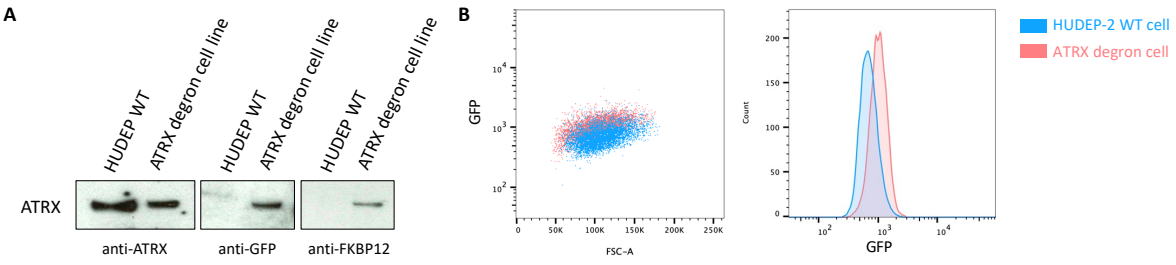


Figure 5.5 Validation of ATRX-EGFP-FKBP12 fusion protein expression in the ATRX degron cell line.

(A) Western blot analysis of ATRX fusion protein in HUDEP-2 WT control and ATRX

degron cell line using anti-ATRX, anti-GFP and anti-FKBP12 antibody respectively. ATRX-EGFP-FKBP12 fusion protein is detected in ATRX degron cell line.
(B) Flow cytometry dot-plot (left) and histogram (right) analysis of GFP signal in HUDEP-2 WT control and ATRX degron cell line.

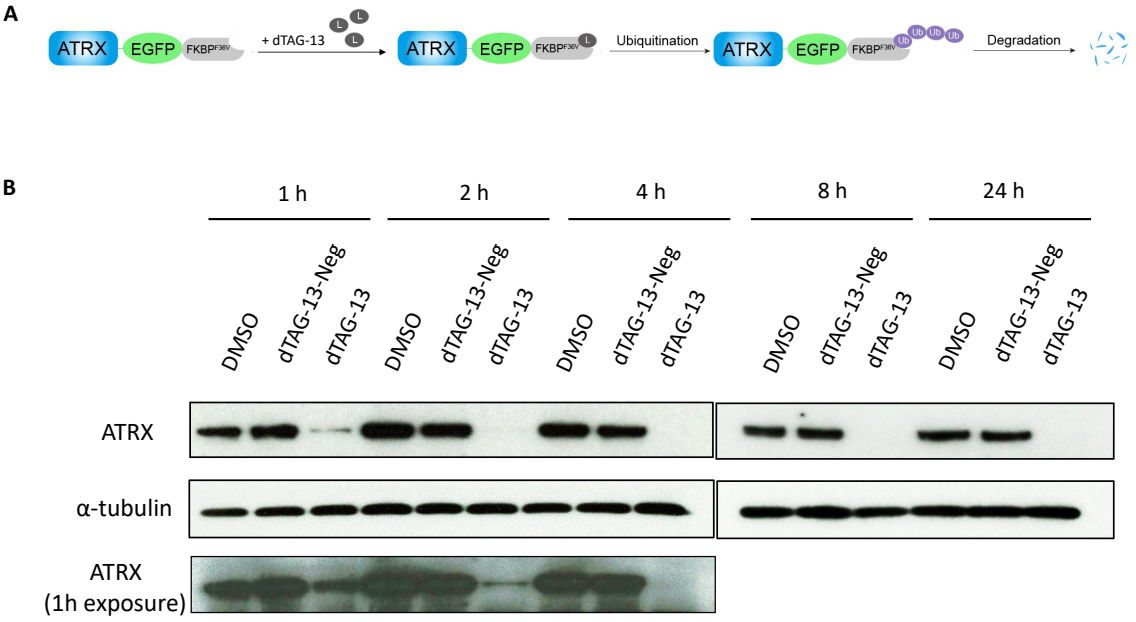
5.2.1.4 dTAG-13 induced protein degradation in ATRX degron cell line

Following the establishment of the ATRX degron cell line, the intention was to induce potent ATRX degradation by using the ligand dTAG-13 in this cellular system (Figure 5.6A). As such, validated ATRX degron cells were treated with 500 nM dTAG-13 as used in previous studies to introduce potent nuclear protein degradation (Nabet et al., 2018, 2020; Weissmiller et al., 2019), and cell samples were harvested at different time points to examine the consequence of dTAG-13 treatment on ATRX stability compared to treatment with mock control DMSO and negative control dTAG-13-Neg (a diastereomer of dTAG-13). Temporal analysis revealed that treatment with dTAG-13 led to rapid degradation of the ATRX fusion chimera in the ATRX degron clones as early as 1 hour post treatment, and ATRX was completely abolished by 4 hours of treatment (Figure 5.6B), establishing a powerful chemical-genetic system to study ATRX function in erythropoiesis.

To detect the protein degradation induced by dTAG-13 via flow cytometry, GFP intensity in HUDEP-2 WT control and ATRX degron cell was evaluated before and after the dTAG-13 treatment respectively. It showed that dTAG-13 treatment had no effect on the HUDEP-2 WT cells (Figure 5.6C). The GFP signal in ATRX degron cells, same as shown above in Figure 5.5B, was higher than the controls,

and its signal was diminished by dTAG-13 treatment to a similar level as that of controls although the difference was not very striking (Figure 5.6C right). Consistently, the intracellular ATRX direct staining detected an overall protein reduction in the cell population upon dTAG-13 treatment, but there was no clear division between positive and negative cells (Figure 5.6D). These results manifest that flow cytometry is not detecting endogenous ATRX-EGFP with much sensitivity, and induced protein loss therefore needs to be determined by Western blot.

Collectively, in the generated ATRX degron cell line, the endogenous ATRX can be rapidly and potently degraded by dTAG-13 as expected, which serves as a compelling model to non-invasively evaluate the consequences of ATRX loss.



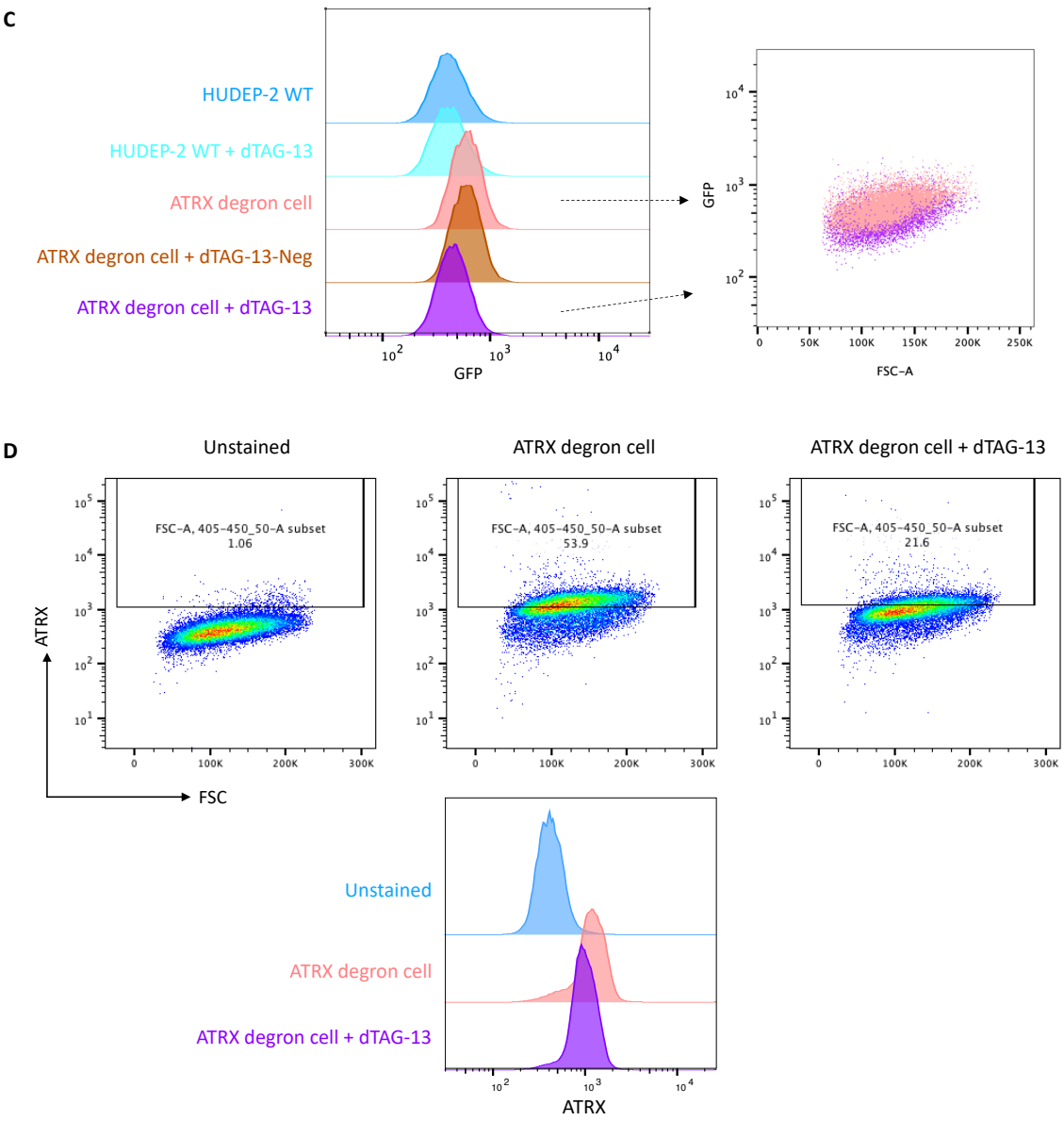


Figure 5.6 dTAG-13 induced protein degradation in the ATRX degron cell line.

(A) Diagram showing dTAG-13 induced ATRX fusion protein degradation.

(B) Immunoblot analysis of ATRX expression in ATRX degron cells treated with 0.1% DMSO, 500 nM dTAG-13-Neg or 500 nM dTAG-13 for the indicated time course. The bottom panel is the long-exposure (1h) blot, showing ATRX is largely depleted by dTAG-13 within 2 hours and absolutely depleted by 4 hours of treatment.

(C) Flow cytometry analysis of GFP expression in HUDEP-2 WT and ATRX degron cell, with or without dTAG-13 treatment. The left offset histograms indicating the reduction

of GFP induced by dTAG-13 in ATRX degron cell. The right dot plot indicating the dTAG-13 treated ATRX degron cells and untreated cells overlap to a great extent.
(D) Flow cytometry analysis of ATRX direct staining in ATRX degron cells, showing protein loss upon dTAG-13 treatment.

5.2.2 Recapitulating α-thalassaemia phenotype in ATRX degron cells

To assess the effect of dTAG-13 induced ATRX degradation on gene expression at the α-globin locus and to determine if this can recapitulate the α-thalassaemia phenotype, the established ATRX degron HUDEP-2 cells were differentiated along the erythroid pathway with and without dTAG-13 treatment and the globin gene expression was evaluated.

5.2.2.1 Determination of the appropriate dose of dTAG-13 for long-term treatment during erythroid differentiation

To mimic the effects of ATRX deficiency in patients caused by the germline mutations in *ATRX*, ATRX protein needs to be depleted through all the phases of erythroid differentiation, thus dTAG-13 treatment was continued throughout the whole differentiation process. The results shown above indicated that 500 nM dTAG-13 is sufficient to induce potent ATRX depletion. To minimise any adverse effect of dTAG-13 treatment on the cells, a series of doses under 500 nM was tested on the cells to determine the lowest dose that can trigger impactful ATRX degradation. It seemed that a dose as low as 25 nM of dTAG-13 was able to induce pronounced ATRX degradation, and the minimal dose leading to absolute

ATRX loss was 100 nM. Therefore, the cells were treated with 100 nM dTAG-13 during their erythroid differentiation in the following experiments.

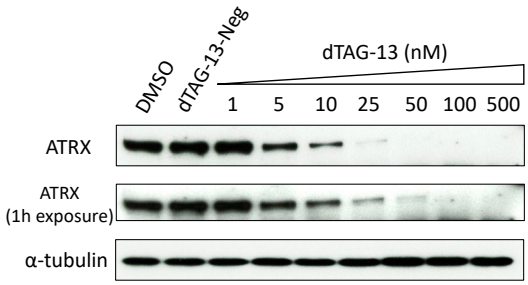


Figure 5.7 ATRX degradation outcome induced by different concentrations of dTAG-13 treatment in ATRX degron cell line.

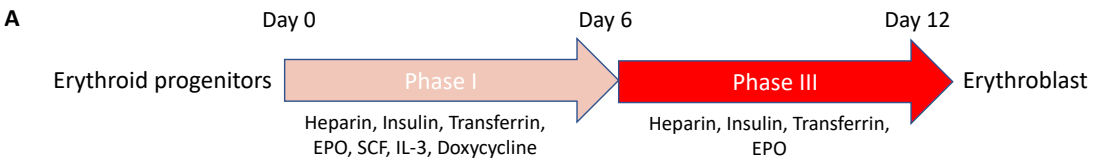
Immunoblot analysis of ATRX expression in ATRX degron cells treated with 0.1% DMSO, 500 nM dTAG-13-Neg, or different concentrations of dTAG-13 for 24 h.

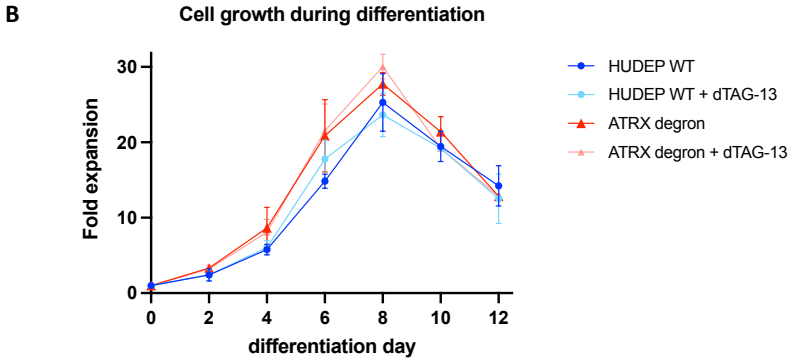
5.2.2.2 dTAG-13 treatment does not affect erythroid differentiation

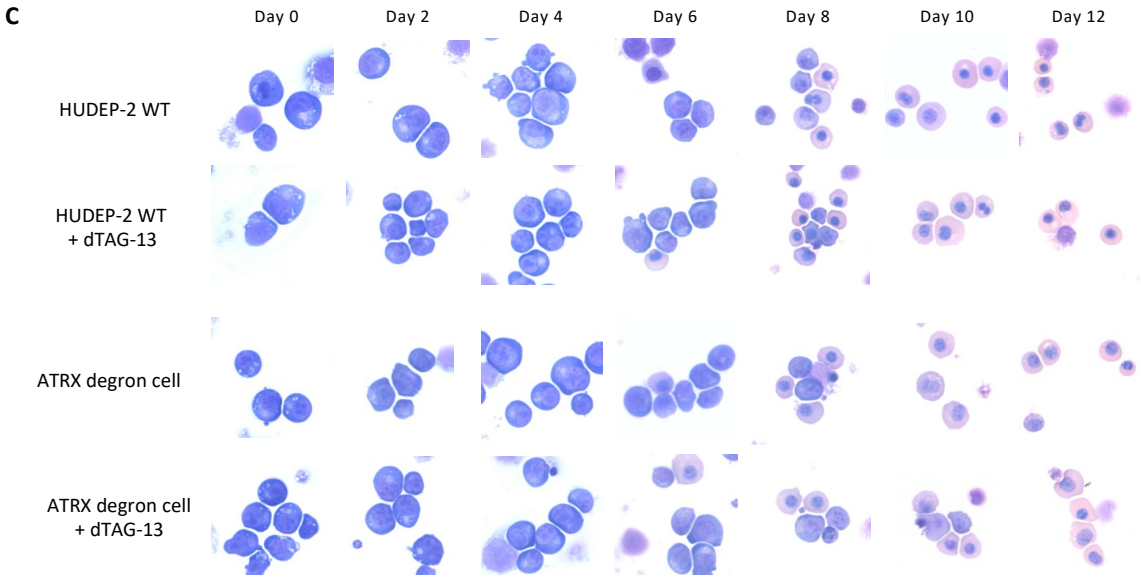
The wildtype HUDEP-2 and ATRX degron cells were differentiated to erythroid cells under normal condition or dTAG-13 treatment, using a two-phase 10 to 12-day differentiation protocol (Figure 5.8A) adjusted from previously described protocols (Kurita et al., 2013; Masuda et al., 2016; Vinjamur & Bauer, 2018). The cell expansion rates were similar in all the four groups, with a 20-30-fold expansion of cell numbers by day 8 and the cell number started to reduce thereafter (Figure 5.8B). This is consistent with a previous report which proposed that the reduction in cell numbers after day 8 was predominantly due to cell apoptosis of orthochromatic erythroblasts that fail to enucleate (Daniels et al., 2020). Importantly, the dTAG-13 treatment had no effect on cell propagation

during erythroid differentiation. The cells exhibited the expected morphological changes as they were differentiated to erythroid cells, from the starting proerythroblast (day 0) through basophilic erythroblast (day 2-6) and polychromatic erythroblast (day 8-10) to the final orthochromatic erythroblast on day 12 (Figure 5.8C). Unlike the HSCs erythroid differentiation described in the Chapter 3, very few of the HUDEP-2 differentiated cells became enucleated in the terminal differentiation stage (Figure 5.8C), in line with previous observations (Vinjamur & Bauer, 2018). Furthermore, the cell surface markers also changed accordingly during the differentiation (Figure 5.8D).

Notably, the edited ATRX degron cell line followed the same normal erythroid differentiation process as the unedited HUDEP-2 WT cells (Figure 5.8C, D), albeit the ATRX degron cells seemed to manifest a slightly slower differentiation, which can be noticed by careful observation of the flow cytometry day 8 plots in which there were fewer CD49d-Band3+ cells in the ATRX degron cell line as compared to HUDEP-2 WT cells (Figure 5.8D). Nevertheless, the difference was trivial and did not affect the conclusion that ATRX degron cells exhibit the same pattern of differentiation as the HUDEP-2 WT cells. Importantly, dTAG-13 treatment had no noticeable effect on the erythroid differentiation of HUDEP-2 WT cells (Figure 5.8B-D), indicating that the cells can be safely treated with dTAG-13 throughout the whole differentiation process. It is interesting to note that dTAG-13 treatment led to a very subtle slowing of erythroid differentiation in the ATRX degron cells as revealed by careful examination on the day 10 flow cytometric analysis of CD49d and Band3 expression (Figure 5.8D).


A
Day 0
Day 6
Day 12
Erythroid progenitors
Phase I
Phase III
Erythroblast
Heparin, Insulin, Transferrin,
EPO, SCF, IL-3, Doxycycline
Heparin, Insulin, Transferrin,
EPO


B
Cell growth during differentiation
Fold expansion
30
20
10
0
0
2
4
6
8
10
12
differentiation day
HUDEP WT
HUDEP WT + dTAG-13
ATRX degron
ATRX degron + dTAG-13


C
Day 0
Day 2
Day 4
Day 6
Day 8
Day 10
Day 12
HUDEP-2 WT
HUDEP-2 WT
+ dTAG-13
ATRX degron cell
ATRX degron cell
+ dTAG-13

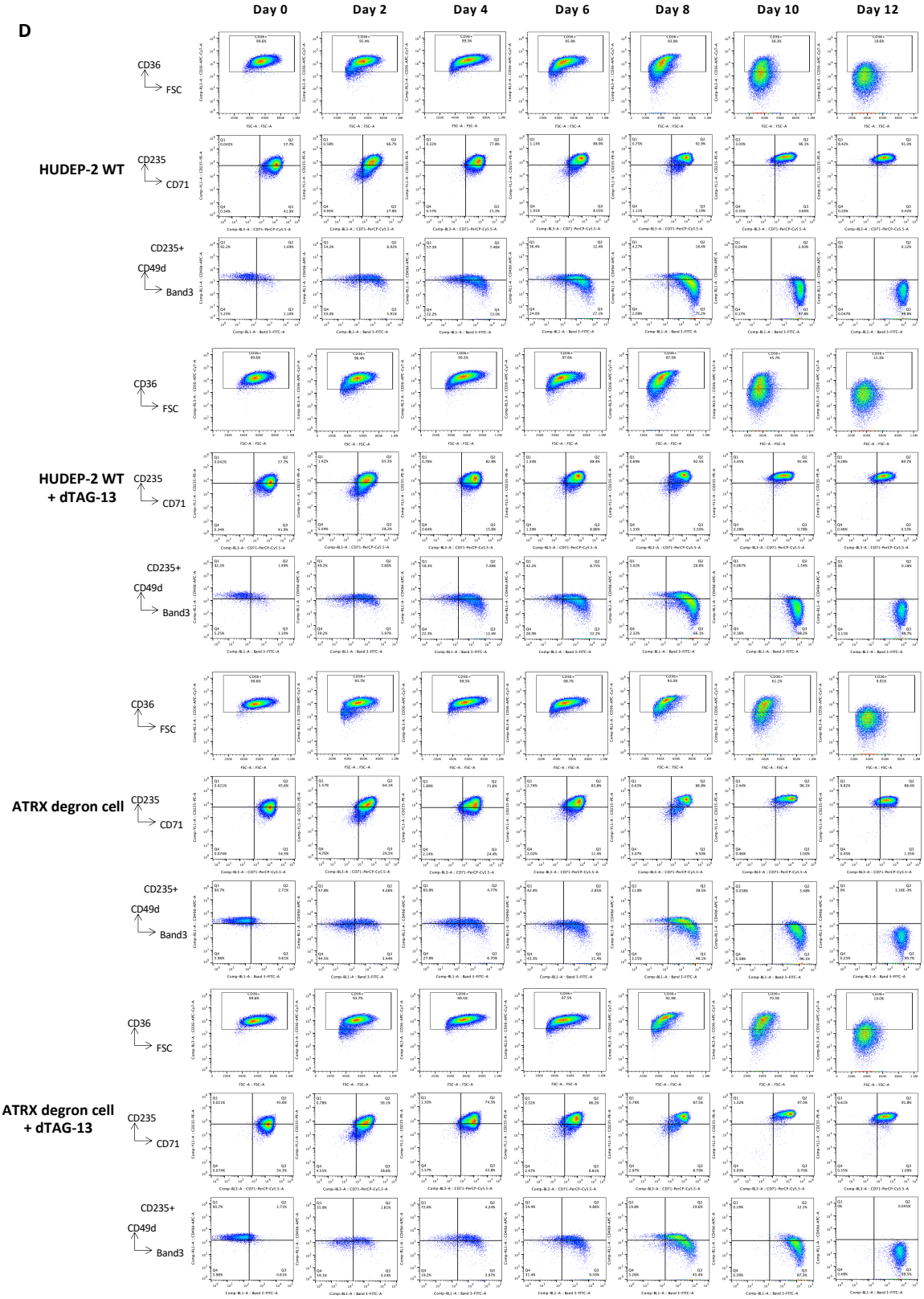


Figure 5.8 dTAG-13 treatment has no significant effect on the erythroid differentiation in HUDEP-2 WT and ATRX degron cell line.

(A) Schematic of the two-phase HUDEP-2 erythroid differentiation protocol. The cells mainly focus on expansion in phase I and commit to erythroid differentiation in phase III. (B) Cumulative fold expansion of cells during differentiation. Live cell numbers were determined by NuceloCounters. Results are mean ± SD from three independent experiments.

(C) Cells stained with a modified Wright's method during differentiation process. HUDEP-2 WT and ATRX degron cells, with or without dTAG-13 treatment, show virtually equivalent differentiation as cells are differentiated to orthochromatic erythroblasts at the final stage.

(D) Flow cytometric analysis of cell surface membrane proteins during differentiation. The expression of CD235a and Band3 are increasing, while CD36, CD71 and CD49d are decreasing during differentiation.

5.2.2.3 Effect of dTAG-13 treatment on gene expression during erythroid differentiation

To assess the effect of dTAG-13 treatment on gene expression, cell samples were collected every other day during erythroid differentiation. ATRX protein expression was firstly examined by Western blot. It showed that ATRX expression in HUDEP-2 WT cells was not affected by dTAG-13 treatment, whereas ATRX was degraded by dTAG-13 in ATRX degron cells (Figure 5.9A). Globin gene expression analysis was performed using RT-qPCR to detect the expression of HBA, HBB, HBG and α to β-like ratio (HBA/(HBB+HBG)). As expected, the expression of all the globin genes was increasing during the erythroid differentiation and dTAG-13 treatment had no significant effect on the globin expression in HUDEP-2 WT cells (comparing red line and orange line in

Figure 5.9B). The edited ATRX degron cells showed similar globin gene expression trend as WT cells, but with relatively lower expression level of globin genes, which was consistent with its slightly slower differentiation (Figure 5.9B). Likewise, the expression of all globin genes was reduced by dTAG-13 treatment in ATRX degron cell (comparing blue line and cyan line in Figure 5.9B), in accordance with its minor delay in differentiation relative to the untreated sample. Therefore, α to β-like ratio was measured to determine the α-thalassaemia phenotype. It demonstrated that dTAG-13 treated degron cell showed slightly lower α to β-like ratio, albeit not reaching statistical significance, compared with control cells (Figure 5.9B), indicating these ATRX deficient cells had mild α-thalassaemia phenotype. Moreover, general HBM to β-like (HBB+HBG) globin ratio and nascent HBM to β-like globin ratio were both significantly decreased in dTAG-13 treated ATRX degron cells on differentiation day 8 and slightly decreased on day 10 (Figure 5.9C). Taken together, by treating ATRX degron cells with dTAG-13 throughout the erythroid differentiation, ATRX protein was efficiently depleted from these cells, and this recapitulated the mild α-thalassaemia phenotype as well as the severe HBM downregulation in ATR-X patients, which provides a tractable system for us to investigate the mechanism of α-thalassaemia in ATR-X syndrome.

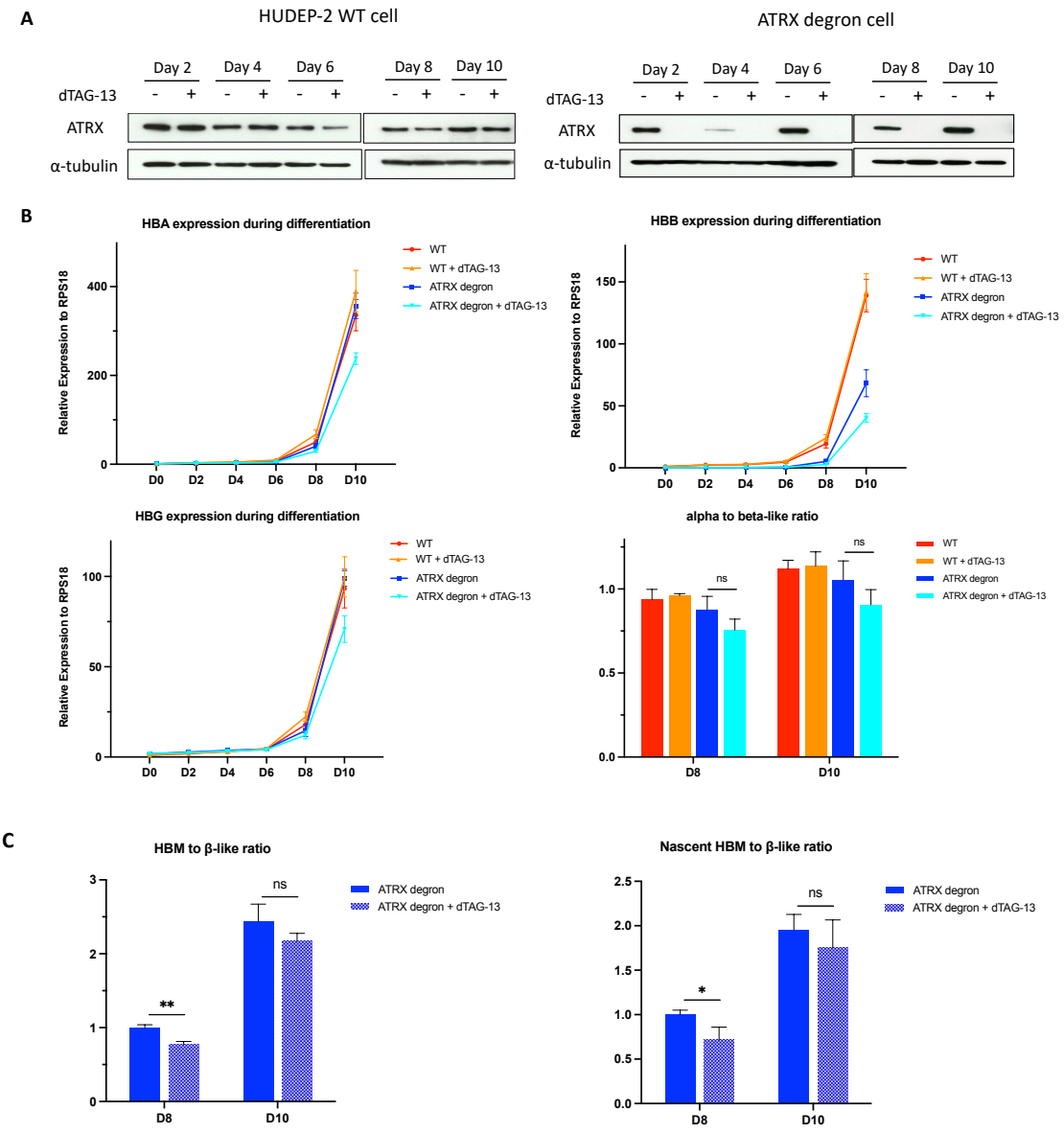


Figure 5.9 Effect of dTAG-13 treatment on gene expression during erythroid differentiation.

(A) ATRX expression in HUDEP-2 WT and ATRX degron cells during erythroid differentiation, showing dTAG-13 induced ATRX depletion only in ATRX degron cells. (B) Taqman Probe RT-qPCR expression analysis of globin genes (HBA, HBB, HBG, α to β-like ratio) in cultured cells during differentiation, normalized to RPS18 (n=3). Error bars show SD from three independent experiments. Multiple unpaired t test was conducted to measure the statistical difference in α to β-like ratio between the control and dTAG-13 treated ATRX cells. ns: not significant.

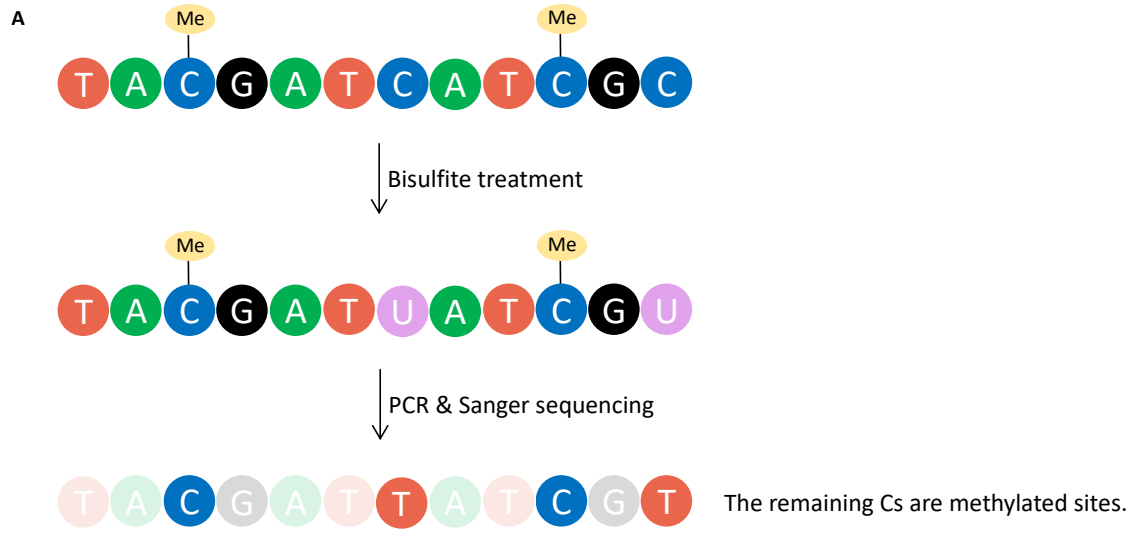
(C) General (left) and nascent (right) HBM expression relative to β-like (HBB+HBG) globin in controls and dTAG-13 treated ATRX degron cells on differentiation day 8 and day 10 (n=3). Data are expressed as mean ± SD with *p <0.05, **p <0.01 according to multiple unpaired t test.

5.2.2.4 DNA methylation is not responsible for HBM downregulation in ATRX deficient cells

Given HBM is downregulated prominently in day 8 dTAG-13 treated cells, these cells could be used as disease model to further investigate the mechanism of gene repression. One mechanism of gene repression could be DNA methylation around TSS. As a 'silencing' epigenetic mark, methylation of cytosine residue at CpG islands and gene promoter is associated with inactive gene expression and involved in maintaining the repression of transcription (Jones, 2012). Of note, hypermethylation at the CpG islands has been observed in ATR-X cases at some genomic loci (Schenkel et al., 2017; Truch et al., 2022). In order to understand whether DNA methylation plays a role in HBM downregulation caused by ATRX loss, targeted bisulfite sequencing was performed at CpG islands and promoter region of *HBM* in undifferentiated ATRX degron cells, untreated and dTAG-13 treated differentiated day 8 degron cells. The unmethylated cytosine (C) in the DNA can be converted to uracil (U) after bisulfite treatment which can subsequently be read as thymine (T) after PCR amplification, whereas the methylated C keeps unchanged post bisulfite treatment and is read as C in subsequent PCR and sequencing results (Figure 5.10A). Thus, by comparing the

sequences of the PCR product with the reference genome sequences, the methylated DNA can be recognised (Clark et al., 2006).

Based on the location of CpG islands and the reported Bisulfite sequencing reads around *HBM* TSS in other cell types, a ~250 bp region overlapping with CpG islands in *HBM* promoter was designed to be targeted in a bisulfite sequencing experiment (Figure 5.10B). In undifferentiated ATRX degron cells, when HBM was silenced, all CpGs at the targeted region were unmethylated (Figure 5.10C). Likewise, the targeted region was generally unmethylated in differentiated day 8 ATRX degron cells in which HBM was actively expressed, and dTAG-13 treatment did not make any difference in it (Figure 5.10C). Based upon the observation that HBM transcription was low in the undifferentiated cells despite the CpGs in the promoter being unmethylated, together with the identical DNA methylation status between untreated and dTAG-13 treated cells, it appears that DNA methylation at *HBM* promoter does not play an important role in its repression.



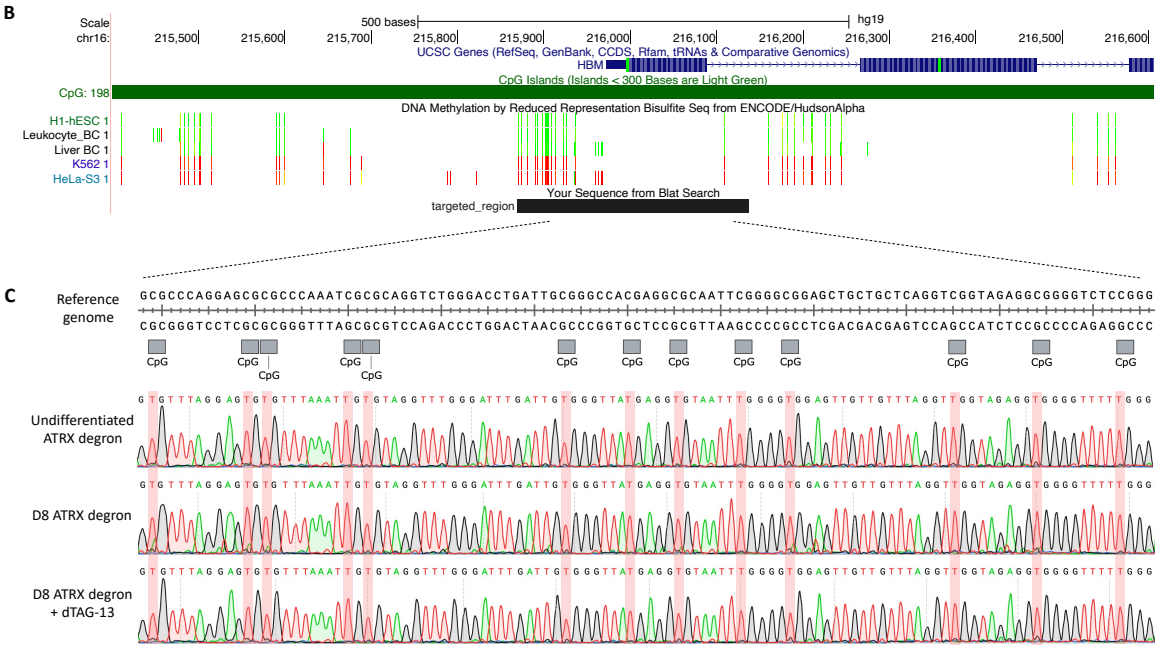


Figure 5.10 DNA methylation analysis at *HBM* promoter by targeted bisulfite sequencing.

(A) Schematic of bisulfite sequencing-based DNA methylation detection.

(B) UCSC tracks showing CpG islands, bisulfite sequencing reads in different cells and designed targeted region around *HBM* promoter. In the bisulfite sequencing reads track, methylation status is represented with a gradient convention: green bar means 0% of molecules sequenced are methylated (completely unmethylated); red bar means 100% of molecules sequenced are methylated (completely methylated). As displayed, the target region is highly methylated in K562 and HeLa cells but unmethylated in other cells.

(C) Top panel shows the reference genome sequences at targeted region, with CpGs denoted as grey bars. Below shows the Sanger sequencing results of undifferentiated ATRX degron cells, untreated and dTAG-13 treated differentiated day 8 degron cells. The Cs within CpGs are highlighted in red and are generally unmethylated as they are all converted to Ts after bisulfite treatment.

5.2.3 Generation and characterisation of HBM-mKate2 reporter cell line

Although HBM is significantly downregulated in dTAG-13 treated ATRX degron cells compared with the control cells on differentiation day 8, the overall reduction by 20~30% at a bulk level may still obscure a larger difference between HBM^{high} and HBM^{low} cells. To overcome this, HBM was tagged with the fluorescent protein mKate2 in the ATRX degron cell line, which would serve as a reporter to reflect the gene expression and also allow us to sort the affected cells having downregulated gene expression by FACS for further mechanism investigation.

5.2.3.1 HBM-mKate2 reporter engineering design and targeting strategy

A CRISPR-Cas9 and HDR mediated knock-in editing approach was adopted to insert mKate2 at the *HBM* locus directly after the start codon ATG, followed by the SV40 polyA to terminate the transcription without HBM being made (Figure 5.11). Given the strong endogenous Kozak sequence of *HBM*, it was kept to ensure proper protein translation. The majority of the first exon and all the remaining *HBM* genomic sequence was maintained in case there is any potential regulatory element or structural regulating mechanism associated with them. A donor DNA template was provided for the HDR-mediated knock in. A donor plasmid was initially used, however, none of the screened clones were positive for the insert due to the extremely low knock-in efficiency. To enhance knock-in efficiency, a single-stranded oligodeoxynucleotides (ssODN) donor was tried and finally used as it has been shown to achieve a high HDR efficiency (M. Liu et al., 2019; Okamoto et al., 2019). Specifically, different combinations of sgRNA (sgRNA-1, sgRNA-2) and ssODN donor (sense ssODN, anti-sense ssODN) were

tried to determine the most efficient combination and thereby increase the probability of identifying a targeted clone.

The targeting strategy was similar to that used for generating the ATRX degron cell line as illustrated in Figure 5.2. Briefly, the ATRX degron cells were electroporated with CRISPR plasmid pX458-Ruby and the ssODN donor. Then, the Ruby+ cells were sorted followed by single cell separation and expansion. The clonal lines were then screened using a PCR-based strategy.

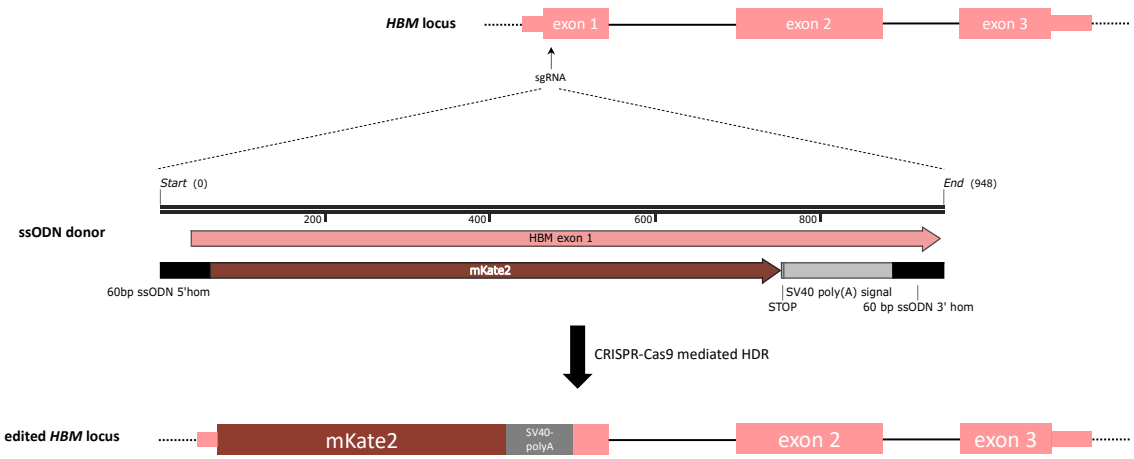
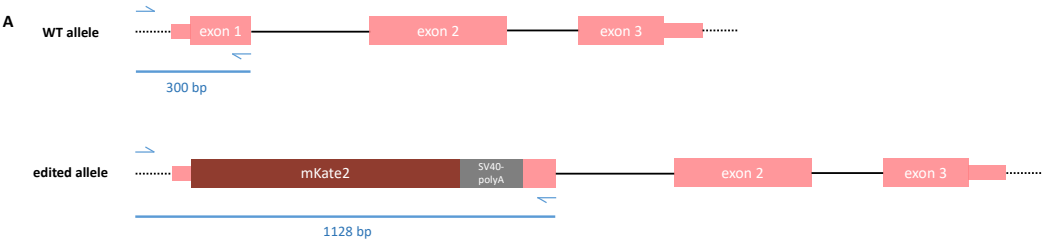


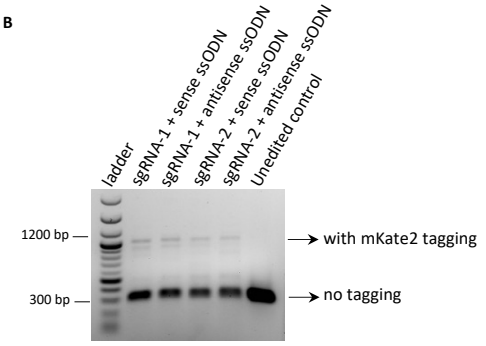
Figure 5.11 Schematic for HBM-mKate2 tagging design.

The top panel shows the endogenous *HBM* and the targeting site of the sgRNA. The middle panel denotes the 948-bp ODN donor design, containing 60 bp homology arms at each side, with mKate2 coding sequence and SV40-polyA in between. Of note, the ODN donor used for targeting was single-stranded, either sense ssODN or anti-sense ssODN. The bottom panel depicts the desired final edited locus.

5.2.3.2 Screening of targeted cell clones

The PCR primers used for determining the editing efficiency in bulk and for clone screening were placed outside of the targeting region and are predicted to generate a 300 bp PCR product from the wildtype allele compared with a 1128 bp product from the edited allele (Figure 5.12A). The sorted bulk Ruby+ cells were initially genotyped to determine the overall editing efficiency by PCR. Among the four combinations of sgRNA and ssODN that were tested, sgRNA-1 in combination with the antisense ssODN achieved the highest editing efficiency as the proportion of the edited allele in this group was highest, at 10.24% (Figure 5.12B). Therefore, 600 single cells were sorted from this group for further clone screening. After the single cell expansion and PCR screening, one positive clone, clone 5-15 was identified (Figure 5.12C).





sgRNA	sgRNA-1	sgRNA-1	sgRNA-2	sgRNA-2
ssODN donor	sense	antisense	sense	antisense
tagging band %	10.03	10.24	8.45	5.63

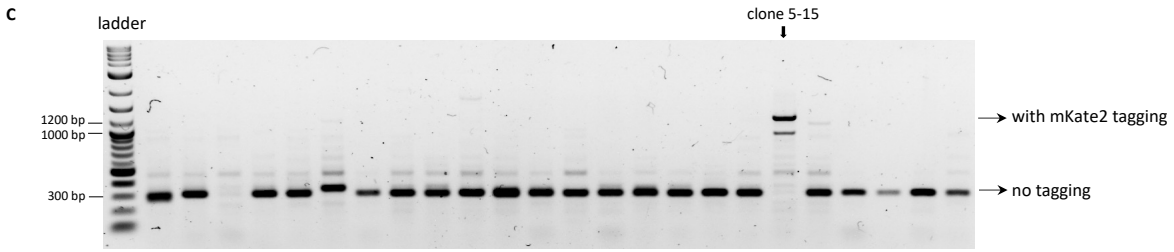


Figure 5.12 Screening of HBM-mKate2 tagged cell clones.

(A) Diagram showing the PCR screening primers and products for WT and edited alleles. (B) PCR products of sorted bulk Ruby+ cells in different groups and the unedited negative control shown in a 2% agarose gel, indicating the editing efficiency. The proportion of the edited band in each group was shown in the right table.
(C) PCR screening results of material from single targeted clones. Clone 5-15 was identified as a positive clone.

Apart from the band with an expected size, there was still one shorter band around 1000 bp generated from this clone. Both bands were sequenced, and the longer bands was confirmed to have the correct integration (Figure 5.13A, B) whereas the shorter band turned out to have a shorter insertion from the CRISPR plasmid scaffold which was not able to produce any functional product (Figure 5.13B). Together, the HBM-mKate2 tagged cell clone was a heterozygote for mKate2 tagging at the *HBM* locus, with one allele expressing fluorescent protein mKate2 reporter and the other allele unable to express mKate2.

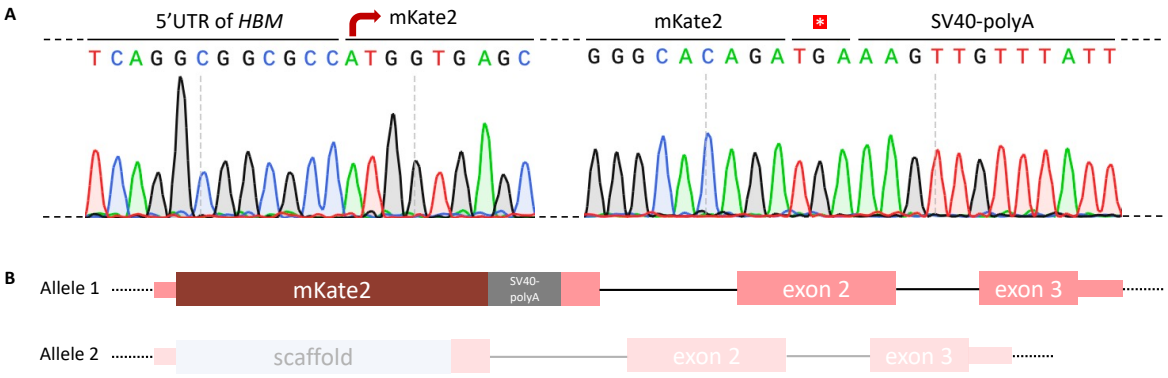


Figure 5.13 Genomic validation of the HBM-mKate2 tagged cell line.

(A) Sanger sequencing result of the longer PCR product in the HBM-mKate2 tagged cell clone showing the correct integration.

(B) Two alleles of the targeted *HBM* locus in the identified HBM-mKate2 tagged cell clone. Allele one has correct mKate2 insertion to produce mKate2 protein that can work as a reporter of gene expression. Allele two has unexpected insertion from the CRISPR pX458-Ruby plasmid and does not produce any functional product.

5.2.3.3 Validation of the mKate2 reporter in erythroid differentiation

To characterise the correctly targetted mKate2 tagged cell clone, the cell line was differentiated to erythroid cells using the two-phase HUDEP-2 differentiation protocol under dTAG-13-Neg treatment or dTAG-13 treatment, together with the ATRX degron cell line which served as a control. The mKate2 tagged cells, both under control conditions (ATRX+) and under dTAG-13 treatment (ATRX-), showed equivalent erythroid differentiation to the ATRX degron cell line, with the expected morphological changes (Figure 5.14A) and erythroid cell surface marker changes during differentiation to erythroblasts (Figure 5.14B).

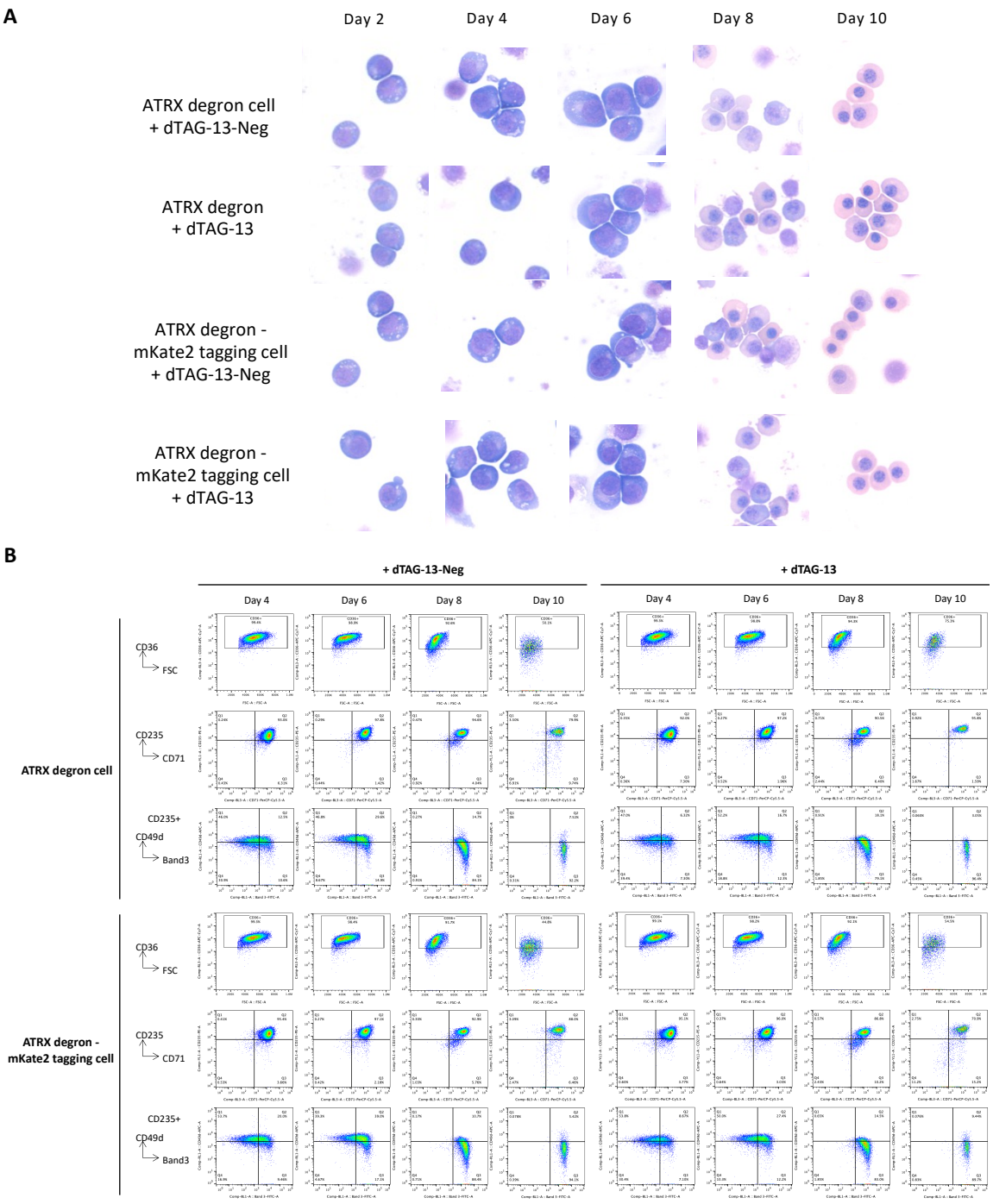


Figure 5.14 Normal erythroid differentiation of ATRX degron-mKate2 tagged cell line.

(A) Cells stained with a modified Wright's method during differentiation. ATRX degron-mKate2 tagged cells display equivalent erythroid differentiation to ATRX degron cells, under both dTAG-13-Neg and dTAG-13 treatment.

(B) Flow cytometric analysis of cell surface markers during differentiation showing similar differentiation between the groups.

To validate whether the mKate2 in the mKate2 tagged cell clone can work as a reporter to reflect the gene expression at *HBM* locus, the mKate2 expression was checked in the ATRX degron-mKate2 tagged cell line during erythroid differentiation. The fluorescence of mKate2 protein in the bulk mKate2 tagged cell population was examined by flow cytometry, comparing treatment with either control dTAG-13-Neg or dTAG-13. Firstly, the depletion of ATRX protein during differentiation in dTAG-13 treated cells was confirmed by Western blot (Figure 5.15A), indicating that the degron system works in the mKate2 tagged cell line. Next, the mKate2 expression was analysed and this demonstrated that the mKate2 fluorescent intensity rose during the erythroid differentiation (Figure 5.15B). Encouragingly, its intensity declined after dTAG-13 treatment (Figure 5.15B), indicating that ATRX loss was associated with mKate2 downregulation. Consistently, at the mRNA level, the expression of mKate2 increased during differentiation and dTAG-13 treatment reduced its expression (Figure 5.15C).

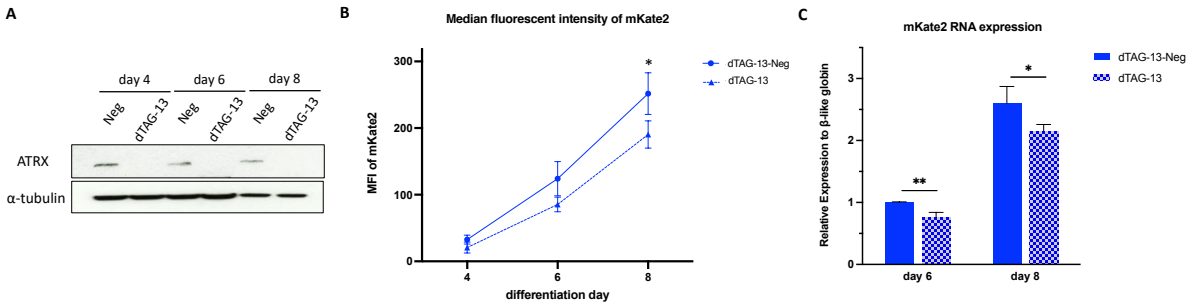


Figure 5.15 ATRX and mKate2 expression in ATRX degron-mKate2 tagged cell line during erythroid differentiation.

(A) Western blot analysis confirming ATRX protein degradation in the dTAG-13 treated mKate2 tagged cell line during erythroid differentiation. “Neg” represents dTAG-13-Neg treated sample.

(B) Median fluorescent intensity changes of mKate2 during differentiation examined by flow cytometry. Error bars show SD from three independent experiments. Multiple unpaired t test was performed to measure the statistical difference between the dTAG-13-Neg and dTAG-13 treated cells. *p <0.05.

(C) mRNA expression of mKate2 on differentiation day 6 and day 8 normalised to β-like globin (HBB+HBG) examined by RT-qPCR using a SYBR method (n=3). Data are expressed as mean ± SD with *p <0.05 and **p <0.01 according to multiple unpaired t test.

Taken together, the established ATRX degron-mKate2 tagged cell line is phenotypically normal when differentiated to erythrocytes. The tagged mKate2 at the *HBM* locus functions as a reporter of gene downregulation when ATRX is depleted from the cell by dTAG-13 treatment.

5.2.4 FACS-based enrichment of affected cells under ATRX depletion

Although the overall intensity of mKate2 in the whole cell population is reduced by ATRX loss in the ATRX degron-mKate2 tagged cell line upon dTAG-13 treatment, the decrease was small and there was still a large overlap between ATRX positive (dTAG-13-Neg treated) and ATRX negative (dTAG-13 treated) cells (Figure 5.16A). Meanwhile, there was no clear separation between high and low mKate2 expressing cells in ATRX deficient cells (Figure 5.16A). In order to

enrich the affected cells to explore the mechanism of perturbed gene expression that was regulated by ATRX, the top 10% of mKate2 expressing cells were defined as unaffected mKate2^{high} cells and the bottom 10% cells were defined as affected mKate2^{low} cells. These two sub-populations were FACS-sorted and the sorted mKate2^{high} cells looked redder than the mKate2^{low} cells, but the cellular morphology and differentiation stage of these two subsets was similar (Figure 5.16B), suggesting the paler mKate2^{low} cells may have an alpha thalassaemia phenotype. When checking the percentage of these two subsets in ATRX positive and negative cells, the affected mKate2^{low} cell subpopulation was increased when ATRX was depleted (Figure 5.16C), suggesting ATRX loss gave rise to mKate2 downregulation and thus an increased proportion of mKate2^{low} cells. This is consistent with the HBM expression pattern in single cells from AAVS1 and ATRX KO samples, as described in the previous chapter. Consequently, the ATRX degron-mKate2 tagged cell model, by sorting the mKate2^{high} and mKate2^{low} subpopulations and comparing their epigenetic profiles, is suitable for determining the underlying mechanism of perturbed gene expression revealed by scRNA-seq.

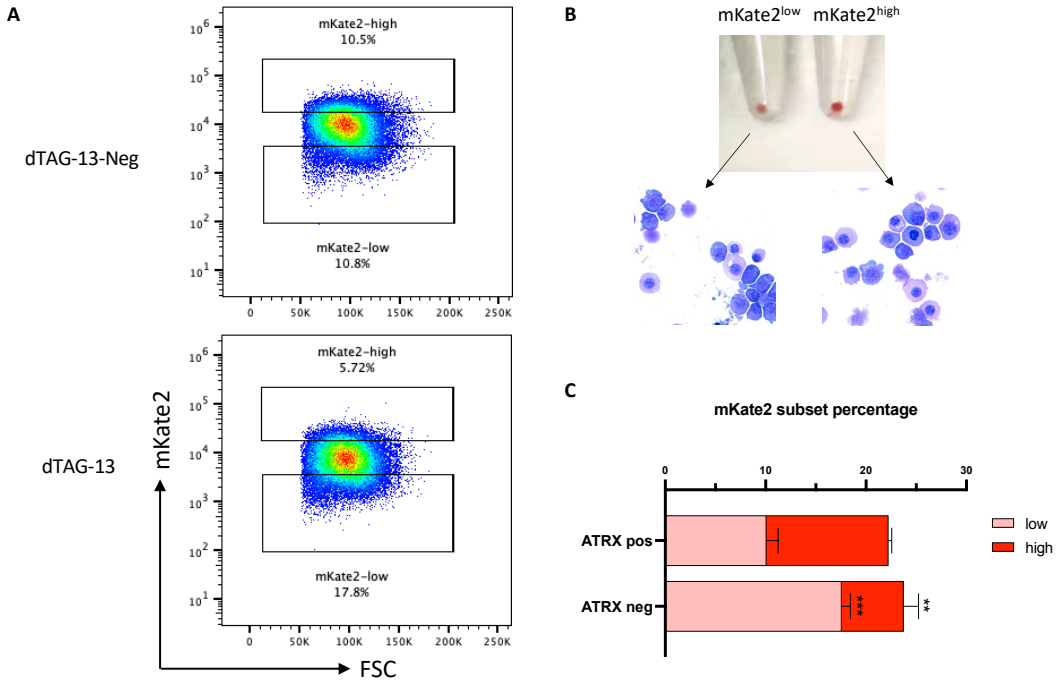


Figure 5.16 FACS-based enrichment of affected cells in ATRX degron-mKate2 tagged cell line.

(A) Flow cytometric analysis of mKate2 intensity in differentiated day 8 cells, under the treatment of dTAG-13-Neg or dTAG-13. The top 10% cells are gated as mKate2^{high} cells and the bottom 10% are gated as mKate2^{low} cells.

(B) Representative images of cell pellet and corresponding morphological Wright's staining of the sorted day 8 mKate2^{low} and mKate2^{high} cells using the sorting gates shown in A.

(C) Percentages of mKate2^{low} and mKate2^{high} cell subpopulations in dTAG-13-Neg treated ATRX positive and dTAG-13 treated ATRX negative cells. Error bars show SD from three independent experiments. Multiple unpaired t test was performed to measure the statistical difference between the ATRX positive and ATRX negative cells. $^{**}p$ <0.01, $^{***}p$ <0.001.

5.2.5 ATRX mainly regulates gene expression at the basophilic erythroblast stage during erythroid differentiation

Since the ATRX degron-mKate2 tagged cell line works as hoped, with dTAG-13 inducible ATRX degradation and a handy readout of gene expression by mKate2 intensity, it has the potential to be a powerful model to dissect the critical phase when ATRX functions during erythropoiesis by depleting ATRX at different phases of erythroid differentiation. The cells were treated with dTAG-13 during differentiation days 0-2, 2-4, 4-6, 6-8, 8-10 to degrade ATRX at these stages, and following the treatment the mKate2 intensity was detected by flow cytometry to assess the effect of ATRX loss on gene expression. Cells treated with dTAG-13-Neg throughout the whole differentiation process were used as a negative control to show the normal mKate2 intensity, and the corresponding dTAG-13 treated ATRX- cells which were depleted for ATRX throughout the entire differentiation process were regarded as the positive control. Therefore, the greater the similarity of the treated cells to the positive control, the more important ATRX is at the corresponding phase. It was shown that depleting ATRX at each stage except days 8-10 had an effect on mKate2 expression, but the effect was subtle on days 0-2 and 6-8 (Figure 5.17A), indicating ATRX plays a major role in gene expression during days 2-6 of differentiation. To further look into this, the cells were treated with dTAG-13 for a single day during days 2-8 of differentiation and the mKate2 intensity was measured. Consistently, cells with ATRX deficiency during days 0-2 and days 6-8 appeared to have minor effect on gene expression (Figure 5.17B) whereas depleting ATRX during days 2-6 led to downregulation of mKate2 to a similar level as the positive control (Figure 5.17B). These data therefore suggest that ATRX functions primarily between days 2 and 6 of

erythroid differentiation, which corresponds to the early basophilic and late basophilic erythroblast stages.

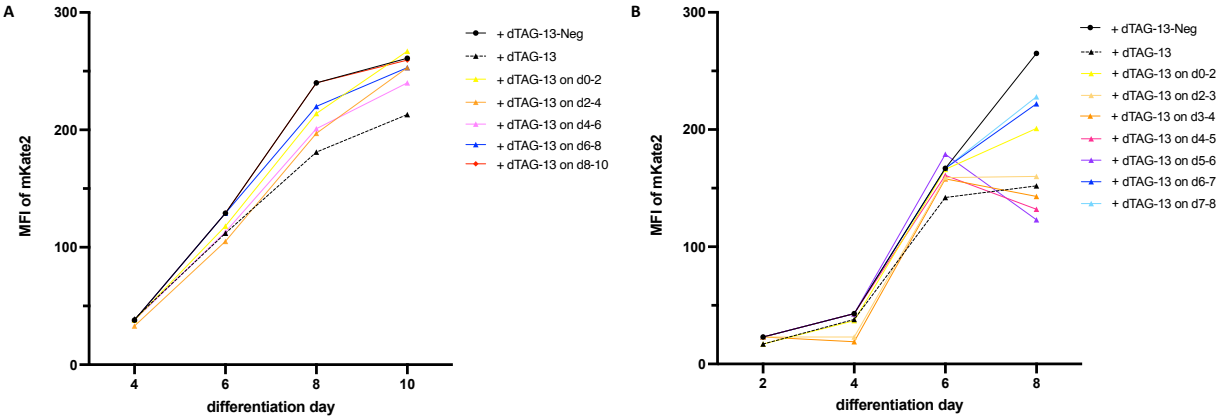


Figure 5.17 Determining the critical phase of ATRX regulation during erythroid differentiation in the ATRX degron-mKate2 tagged cell line.

(A) Depleting ATRX by dTAG-13 treatment at different stages of differentiation, during days 0-2, 2-4, 4-6, 6-8 and 8-10. The black line shows the negative control treated with dTAG-13-Neg; the black dash line shows the positive control treated with dTAG-13 from day 0-10.

(B) Depleting ATRX during days 0-2, 2-3, 3-4, 4-5, 5-6, 6-7 and 7-8 of differentiation. The black line denotes the negative control and the dash line represents the positive control.

5.3 Discussion

Single cell RNA-seq data presented in the previous chapter suggest some potential mechanisms underlying the downregulation of HBM expression in an affected subset of cells depleted of ATRX. To explore the mechanism, the affected cells need to be purified from the ATRX deficient bulk cell population and to be compared with the unaffected cells. To this end, here we established a tractable ATRX degron-mKate2 reporter erythroid progenitor cell line, where ATRX can be rapidly and potently degraded by a degron system and gene expression at *HBM* locus can be reflected by the fluorescent reporter mKate2.

The cellular model used in this project is an immortalized erythroid progenitor cell line HUDEP-2 cells, whose proliferative capacity and karyotype of 49-53, XY render it a suitable system for genome engineering, particularly on chromosome X (Moir-Meyer et al., 2018). Here we fuse the degron FKBP12 F36V together with EGFP at the C-terminus of the ATRX protein, which is not expected to influence the intrinsic function of the ATRX protein. The generated cell line possesses a successful degron system for inducing ATRX loss, but the EGFP intensity is faint by flow cytometry detection although its signal reduces with ATRX degradation. Similarly, flow cytometry detection of ATRX is poor although it is possible to detect a reduced signal after dTAG-13 induced protein degradation. Instead, ATRX degradation needs to be confirmed precisely by Western blot. After the generation of the ATRX degron cell line, we further tagged the *HBM* locus with a mKate2 reporter and obtained a mKate2 tagged cell line. Even though the screened mKate2 tagged cell line is heterozygous, it

demonstrates a normal erythroid progenitor profile and the heterozygous mKate2 works well as a reporter of gene expression.

It is important to note that the HUDEP-2 cell line has shown clonal variability in our lab and by others (Daniels et al., 2020). It is therefore ideal to generate multiple clones with the same genetic modification and compare them to wildtype cells isolated in parallel. However, due to the low HDR efficiency in this project, only a single clone was identified from each editing after screening up to 1200 clones. As a result, the data presented here are from one clone only, with multiple separate differentiations and experiments, which is a potential criticism of the work. Nevertheless, it represents an isogenic model in which ATRX levels can be manipulated. Furthermore, the HBM or mKate2 was downregulated in response to dTAG-13-induced ATRX degradation in both the initially established ATRX degron cell line and the ATRX degron-mKate2 tagged cell line, which is entirely consistent with the observations in an ATR-X patient's sample (Truch et al., 2022) as well as the CD34+ cellular model described in the previous two chapters. Similar to the finding in the CD34+ cellular model, ATRX deficient HUDEP-2 cells showed slightly slower differentiation albeit this difference is very subtle that it is barely noticeable without careful examination on the flow analysis of the markers CD49d and Band3. Hence, HBM/mKate2 RNA expression was normalised to β-like globin genes HBB+HBG to eliminate the effect of slower differentiation. Despite these potential minor issues, the established models provide a compelling system to explore the mechanism of the HBM downregulation caused by ATRX loss.

To see whether DNA methylation played a role in HBM repression, targeted bisulfite sequencing was performed to check the DNA methylation status around *HBM* promoter in the undifferentiated cells not expressing HBM and the differentiated cells expressing HBM. This region is covered by a large CpG island and data from a previous comprehensive study demonstrated it is highly methylated in the tumour cell line K562 and HeLa cells but unmethylated in cells like human ESC and leukocyte (Meissner et al., 2008). This highly variable DNA methylation status between cell types makes it possible for being a mechanism of gene regulation. However, given *HBM* promoter region is always unmethylated in the HUDEP-2 cells irrespective of HBM expression, DNA methylation does not appear to play a gene silencing role here. Having excluded DNA methylation as a mechanism, the role of histone modification in HBM downregulation seems to be plausible. In support of this, histone H2A ubiquitination and DNA damage have been indicated as responsible mechanisms by my scRNA-seq data presented in the previous chapter, which directs the future experiments to check the H2A ubiquitination and deposition of DNA damage markers including macroH2A at *HBM* locus that may link DNA damage or replication stress occurring around the upstream ψζ VNTR to a change in HBM transcription. This can be checked by comparing FACS-enriched mKate2^{low} cells with enriched mKate2^{high} cells. Meanwhile, as a prime candidate for causing DNA damage or replicative stress, the importance of the ψζ VNTR in gene regulation needs to be confirmed, which is investigated in the next chapter of this thesis.

Aside from investigating mechanism, the ATRX degron-mKate2 reporter cell line also enables us to dissect the phase during erythroid differentiation when ATRX

function is critical. By depleting ATRX with dTAG-13 at different stages of differentiation, we found ATRX primarily functions at the basophilic erythroblast stage of erythroid differentiation. Depletion of ATRX during the polychromatophilic and orthochromatophilic erythroblast stages (day 6 to day 10) had little or no effect on the mKate2 expression, indicating ATRX has little role in terminal erythroid differentiation. There might, however, be some potential issue with the dTAG-13 treatment in this experiment. Based on my experience using the dTAG-13-induced degron system, it might take up to 48 hours for ATRX protein to recover to its endogenous level after the washout of dTAG-13. This relative slow reversibility has also been reported by others using the same system to degrade another chromatin remodeller BRD4 (Moyer, 2020). The recovery time can be variable between different target proteins due to their different expression level and also between different models due to the distinct proteasomal dynamics, thus further experiment is needed to determine the reversibility dynamics of ATRX in our degron cell model. Nonetheless, the successive and overlapped treatment intervals between groups can eliminate the effects of the recovery time to some degree. Additionally, it is of note that the HUDEP-2 cell line itself is at the proerythroblast stage (Daniels et al., 2020), thus these models fall short in their ability to reveal any function of ATRX at earlier progenitor stages prior to the proerythroblast stage.

In summary, this chapter reports the development of a robust ATRX degron HUDEP-2 cell line as well as a ATRX degron-mKate2 reporter cell line. These cellular models are characterised and confirmed to have normal erythroid progenitor identity which can be differentiated to erythroid cells and during which

ATRX can be degraded via the degron system. Importantly, these models recapitulate the HBM/mKate2 downregulation caused by ATRX loss, which renders them competent systems to investigate the mechanism of ATRX action on gene expression. Using these models, we indicate that DNA methylation is not responsible for the gene regulation which narrows the potential epigenetic mechanism down to histone modification. The models can be, in the future, be used to test the DNA damage or replication stress mediated transcriptional silencing hypothesis. In addition, using the reporter cell model, we demonstrate ATRX primarily regulates gene expression during the basophilic erythroblast stage during the process of erythroid differentiation.

6 Investigating the role of ψζ VNTR in ATRX regulated α-globin expression

6.1 Introduction

There are millions of tandem repeats in the human genome, which impart a huge source of variation to the genome and contribute to genetic diversity. Based on the length of the repetitive units, tandem repeats can be divided into two classes: short tandem repeats (STRs, or micro-satellites) if the repeating unit is 2-6 bp, and variable number tandem repeats (VNTRs, or mini-satellites) if the repeating unit is >6 bp. These elements have been under-investigated in genomic studies due to the technical challenge and complexity in genotyping them with short read DNA sequencing methods. However, with the advances in long read DNA sequencing approaches, their contribution to functional genomics and human disease have been increasingly recognised. Recently, a number of studies have highlighted the role and importance of VNTRs in the regulation of gene expression and the association between VNTR polymorphism and human diseases (Eslami Rasekh et al., 2021; Marshall et al., 2021; Bakhtiari et al., 2021).

At the human α-globin cluster, there is a VNTR called ψζ VNTR which lies in the second intron of the ψζ pseudogene *HBZP1* and 1 kb upstream of the *HBM* promoter. The ψζ VNTR contains a G-rich repeating unit $(CGCGGGGCGGGG)_n$ that can form G-quadruplex (G4) structure *in vitro* which is a tetra-stranded secondary DNA structure formed by four stretches of guanines (Law et al., 2010; Y. Li et al., 2016). ATRX ChIP-seq analysis has demonstrated its enrichment at

the ψζ VNTR in erythrocytes but not in fibroblast and lymphoblastoid cells (Law et al., 2010; Truch et al., 2022), indicating ATRX binding to ψζ VNTR is likely associated with the exclusively active gene expression at α-globin cluster in erythroid cells. In support, ATRX has previously been shown binding to G-rich sequences in a transcription dependent manner (Nguyen et al., 2017). Analysis of the size of the ψζ VNTR in ATR-X syndrome patients has revealed a positive correlation with the downregulation of α-globin gene expression (Law et al., 2010), suggesting an involvement of ψζ VNTR in gene regulation.

In addition, the intrinsic repetitive sequences of ψζ VNTR makes it a likely candidate to cause replicative stress during DNA replication. Notably, ATRX has been shown to be required for DNA replication, and ablation of ATRX leads to increased stalling of replication and an increase in DNA double strand breaks (Clynes et al., 2014). Moreover, loss of ATRX can cause increased R-loops formation in repetitive sequences such as telomeres (Nguyen et al., 2017). It is therefore speculated that ATRX binds the ψζ VNTR and resolves G4 structures to facilitate DNA replication and gene expression; when ATRX is deficient, G4 structures form at the ψζ VNTR which blocks replication and triggers replicative stress or DNA damage thus perturbing gene expression. In support of this, the single cell RNA-seq data shown in Chapter 4 demonstrate the activated DNA damage response pathway and associated pATM-RNF168-ubiquitination gene regulation pathway in ATRX KO cells. Together, these raise the possibility that the ψζ VNTR works as a transcriptional modulator of α-globin expression regulated by ATRX.

In this chapter, I first describe the binding of ATRX at the ψζ VNTR during erythroid differentiation. I then treat the cells with a G4 stabiliser to determine whether a G4 DNA secondary structure is involved in gene regulation at the α-globin locus. Finally, I generate a ψζ VNTR null cell line by CRISPR/Cas9-mediated knockout of ψζ VNTR to assess the dependence of gene regulation on the ψζ VNTR.

6.2 Results

6.2.1 ATRX binds to the ψζ VNTR in a gene expression dependent manner

Previous ATRX ChIP-seq analysis in diverse cell types has demonstrated the specific ATRX binding at the ψζ VNTR in erythroid cells only (Law et al., 2010; Truch et al., 2022), suggesting the recruitment of ATRX to the ψζ VNTR is probably gene expression dependent. To detect the binding of ATRX at the ψζ VNTR during erythroid differentiation and to determine if there is any correlation between ATRX binding and gene expression, the ATRX degron HUDEP-2 cells were differentiated to erythroid cells under normal condition, during which cell samples were collected at different time points to check the gene expression by qPCR and the ATRX binding by ChIP-seq. Given the ψζ VNTR resides in the pseudogene *HBZP1*, which has extremely high sequences homology with the ζ-globin gene *HBZ*, with only 3-bp difference in their coding sequences, it is difficult to distinguish them in gene expression assays. Furthermore, as a hallmark of pseudogenes, an in-frame stop codon appears in the seventh codon of the pseudogene *HBZP1,* which is likely to induce nonsense-mediated mRNA decay.

As a result, it is not easy to detect the expression of this pseudogene. Therefore, here I used the expression of the nearest α-like globin HBM as a surrogate to reflect the local transcription during erythroid differentiation. It was demonstrated that HBM was barely expressed at early differentiation stage (day 0 to day 2) and was gradually activated after day 4 of erythroid differentiation (Figure 6.1A). Interestingly, ATRX ChIP-seq showed that ATRX started to bind the ψζ VNTR and HBM promoter on differentiation day 4 and remained bound thereafter (Figure 6.1B), concomitant to the active transcription of HBM during differentiation, suggesting ATRX binds to the ψζ VNTR in a gene expression dependent manner.

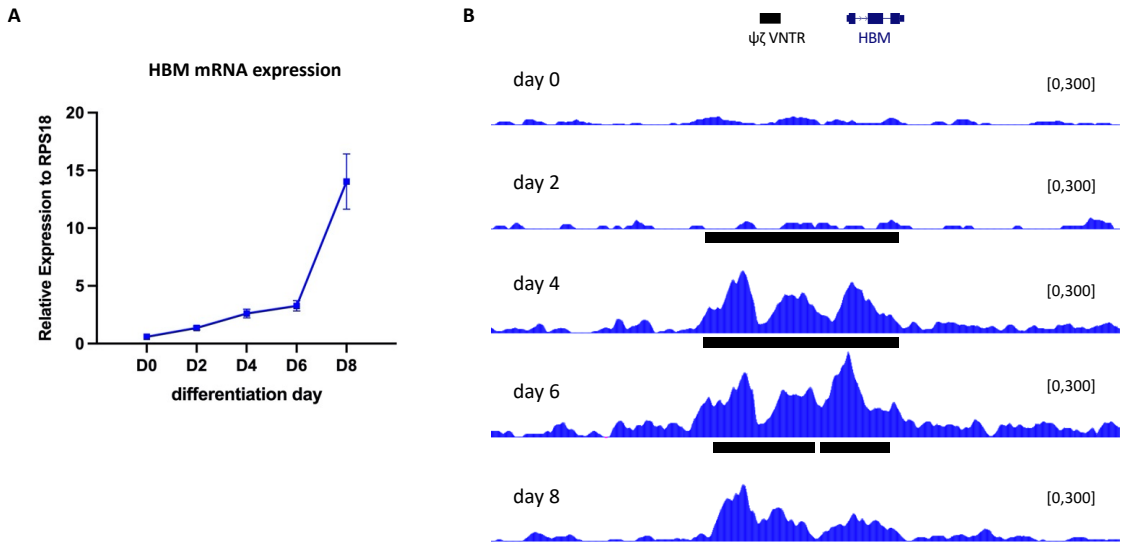


Figure 6.1 ATRX binds to the ψζ VNTR in a gene expression dependent manner.

(A) Taqman probe qPCR mRNA expression analysis of HBM during erythroid differentiation, normalized to RPS18. Error bars show standard deviations from three independent experiments (n=3).

(B) Representative ChIP-seq tracks for ATRX occupancy at the ψζ VNTR-HBM locus in cells at different stages of erythroid differentiation. The black block above the peaks within each track shows the ATRX peak call.

6.2.2 G-quadruplex stabilization by G4 ligand disturbs HBM expression

Since ATRX binding at the ψζ VNTR is correlated with the active HBM expression, it is possible that ATRX is required for local active gene expression, possibly by resolving any potential G4 structure formed at the ψζ VNTR. To assess whether HBM expression is associated with a G4 structure, the HUDEP-2 cells were treated with three different G4 stabilizer compounds during differentiation and the globin gene expression was measured to check if it has been perturbed. These three G4 stabilizers are pyridostatin (PDS), CX-5461 and TMPyP4, which have been commonly used for targeting and stabilizing G4 structures.

6.2.2.1 Determination of the optimal concentration for treatment with G4 stabilizers

Given G4 stabilizers have cytotoxicity, the cytotoxic effects of the three G4 stabilizers were evaluated individually in the undifferentiated ATRX degron cell line under normal conditions to determine the optimal concentration for each G4 stabilizer. The cells were treated with drugs in a series of concentrations for 4 days, and the cell growth, cell viability and cellular morphology were evaluated and compared with untreated cells. As expected, the higher the concentration of the drug and the longer the treatment, the more toxic they were to the cells - with

slower cell growth, reduced cell viability and increased number of abnormal and dead cells (Figure 6.2). Given higher drug concentrations can lead to more reliable stabilization of G4 structures, the optimal concentration for each drug was determined by choosing the highest concentration consistent with cell viability >90%, >5 fold cell expansion and which left the majority of the cells with normal morphology after 4 days of treatment. Finally, 2 μM of PDS, 0.1 μM of CX-5461 and 1 μM TMPyP4 were used for treating cells during erythroid differentiation (Figure 6.2).

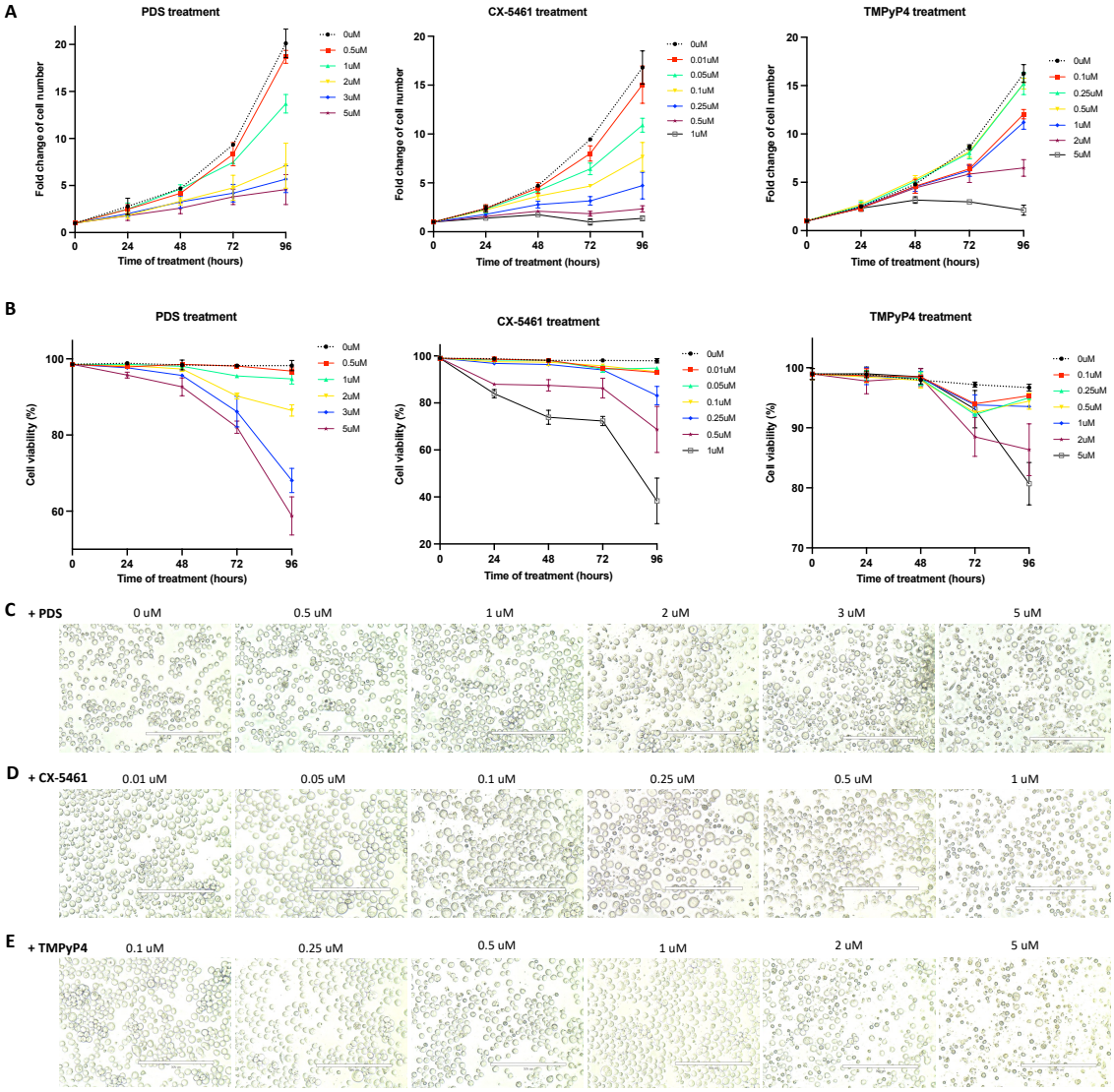


Figure 6.2 (previous page) Determination of the optimal treatment concentration of G4 stabilizers.

(A) The cell expansion reflected by cell number change during the 4 days of treatment. Error bars show standard deviations (SD) from three independent experiments (n=3). The black dash lines represent the untreated (0 μM) control cells.

(B) The cell viability change during the treatment (n=3). The black dash lines represent the untreated control cells.

(C-E) The morphology of cells after 4 days of different concentrations of G4 stabilizer treatment. Scale bar = 200 μm.

6.2.2.2 Treating the ATRX degron cell with G4 stabilizer during erythroid differentiation

After confirming the optimal treatment concentration for each G4 stabilizer, the ATRX degron cells were differentiated to erythroid cells under normal conditions following the HUDEP-2 cell erythroid differentiation protocol described in Chapter 5 section 5.2.2.2, during which the cells were treated with G4 stabilizer from day 6 to day 10 of differentiation when globin genes are actively expressed. The erythroid differentiation process was tracked by flow cytometric analysis of the cell surface markers. Overall, cells treated with G4 stabilizer showed normal and similar erythroid differentiation to untreated cells (Figure 6.3). PDS and TMPyP4 treatment had no significant effect on erythroid differentiation, while CX-5461 treatment seemed to accelerate the differentiation slightly, leading to more $CD49d^{-}Band3^{+}$ cells, albeit this effect was very subtle and not significant (Figure 6.3).

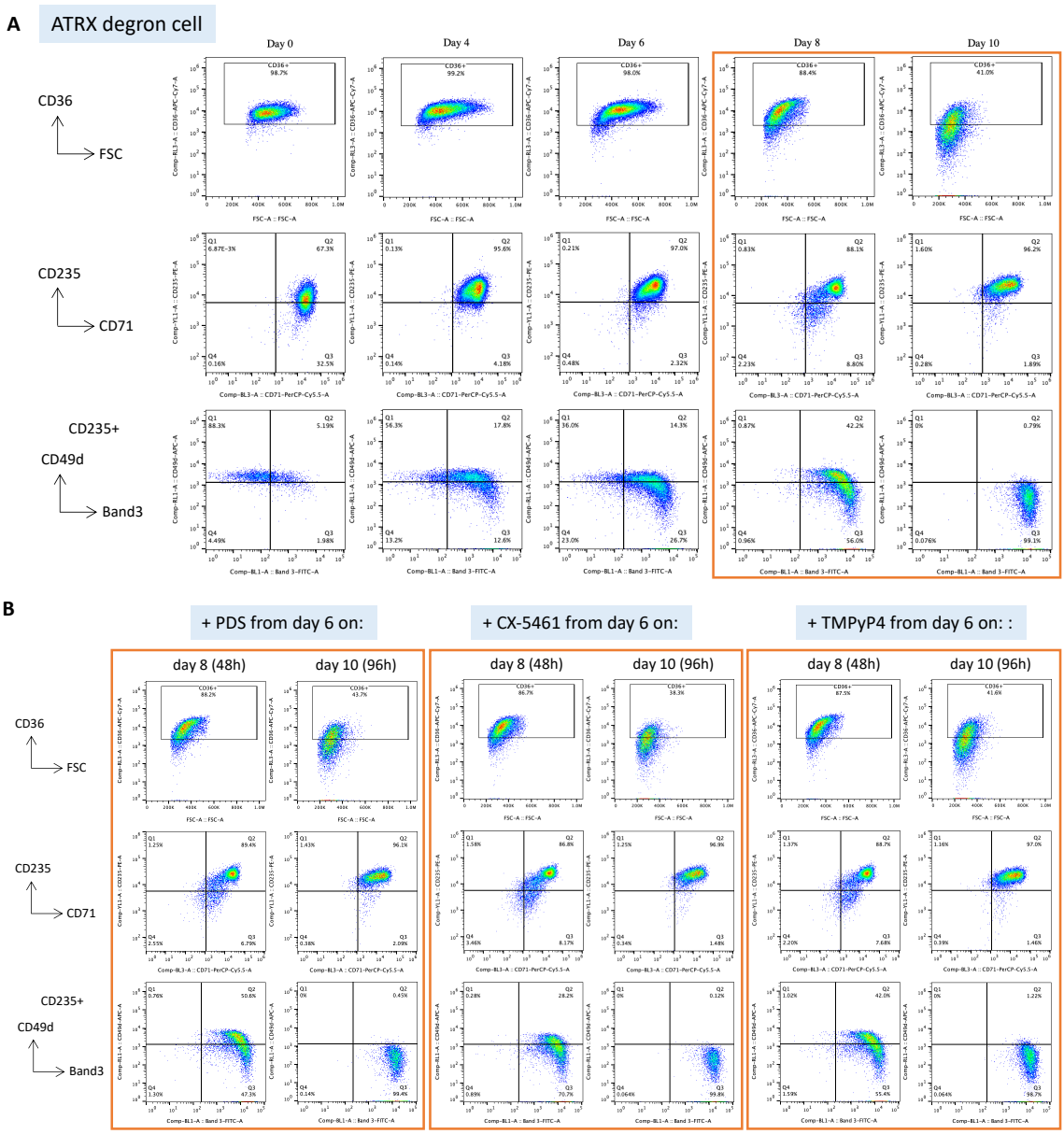


Figure 6.3 Effect of G4 stabilizer treatment on erythroid differentiation.

(A) Flow cytometric analysis of cell surface membrane proteins during erythroid differentiation showing normal differentiation of the untreated control ATRX degron cells - the expression of CD235a and Band3 are increasing, while CD36, CD71 and CD49d are decreasing. The orange frame highlights the cell differentiation of untreated cells at day 8 and day 10, for the comparison with G4 stabilizer treated cells presented in (B).

(B) Differentiation of cells after 48 hours and 96 hours of G4 stabilizers treatment.

6.2.2.3 Gene expression analysis in G4 stabilizer treated cells

The cell samples were collected at day 8 (after 48 hours of treatment) and day 10 (after 96 hours of treatment) to measure the gene expression as compared with untreated cells. To compare the altered gene expression with that caused by ATRX loss, the ATRX degron cells were differentiated with dTAG-13 treatment throughout the whole erythroid differentiation (day 0 to day 10) in parallel, and the cell samples at day 8 and day 10 were collected for comparison. Aside from the globin genes (HBA, HBB, HBG and HBM), NME4, which is another ATRX target containing a G-rich VNTR, was also included here to check if G4 stabilization would have similar effects on the expression of HBM and NME4.

Gene expression analysis showed that TMPyP4 treatment had no significant effect on the expression of any of these genes in either day 8 or day 10 samples (Figure 6.4). Similarly, PDS treatment had little effect on the expression of most genes. HBM expression appeared to be slightly downregulated by PDS treatment, but the change was not significant when normalised to β-like globin (HBB+HBG) (Figure 6.4). Interestingly, 48 hours of CX-5461 treatment downregulated the expression of putative G4-associated genes, HBM and NME4, but had no effect on the other globin genes, and the degree of their downregulation was equivalent to the effect caused by dTAG-13 induced ATRX loss (Figure 6.4A left). However, CX-5461 downregulated the expression of all the genes after 96 hours of treatment (Figure 6.4B left). Therefore, the α-(like)-globin gene expression was normalised to β-like globin genes (HBB+HBG) to get rid of any effect caused by a shift in erythroid differentiation. Notably, after α to β-like normalisation, it was shown that CX-5461 significantly downregulated HBM expression after both 48h

and 96h of treatment, with a similar effect to dTAG-13 induced ATRX loss, but had no significant effect on HBA expression (Figure 6.4 right). Taken together, G4 stabilization by CX-5461 disturbs gene expression of HBM and NME4, the two genes closely associated with a putative G4 structure, but has no effect on the expression of the other genes.

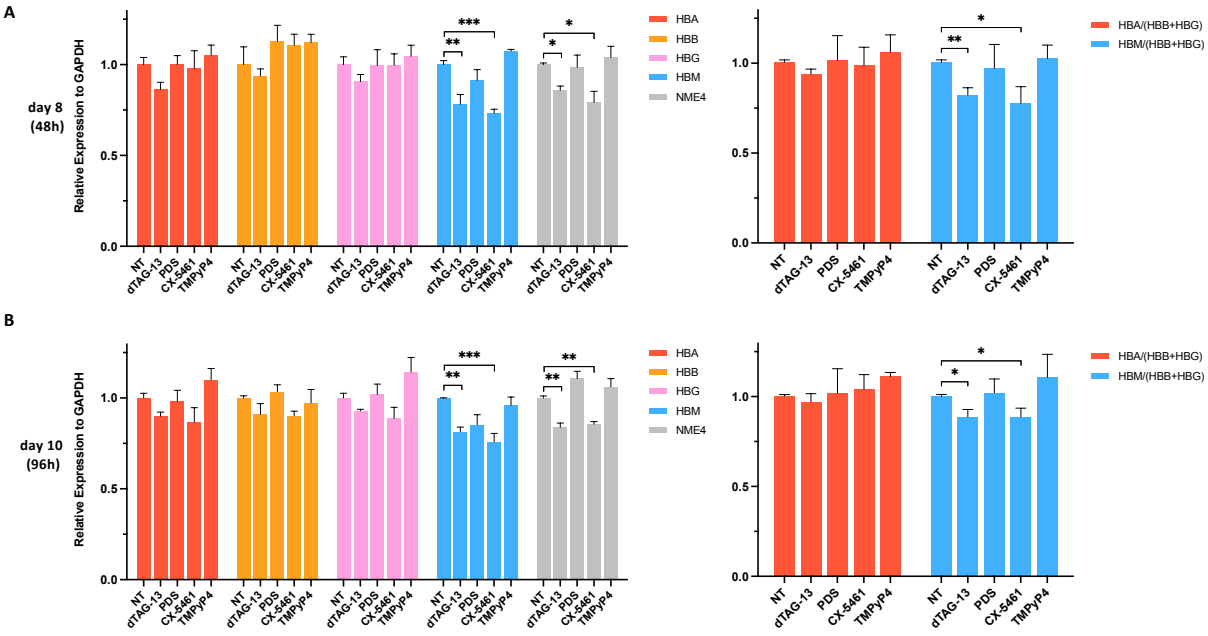


Figure 6.4 Gene expression analysis in G4 stabilizer treated cells.

(A) Taqman Probe RT-qPCR analysis of gene expression (HBA, HBB, HBG, HBM, NME4 and α to β-like ratio) in day 8 cells under different conditions, normalized to GAPDH (n=3). NT: untreated cells; dTAG-13: cells treated with 100 nM dTAG-13 from day 0 to day 8; PDS: cells treated with 2 μM PDS from day 6 to day 8; CX-5461: cells treated with 0.1 μM CX-5461 from day 6 to day 8; TMPyP4: cells treated with 1 μM TMPyP4 from day 6 to day 8. Data are expressed as mean ± SD with *p <0.05, **p <0.01, **p <0.001 according to multiple unpaired t test.

(B) Similar gene expression analysis in day 10 cells. NT: untreated cells; dTAG-13: cells treated with 100 nM dTAG-13 from day 0 to day 10; PDS: cells treated with 2 μM PDS

from day 6 to day 10; CX-5461: cells treated with 0.1 μM CX-5461 from day 6 to day 10; TMPyP4: cells treated with 1 μM TMPyP4 from day 6 to day 10.

6.2.3 HBM expression is no longer perturbed by ATRX loss when ψζ VNTR is knocked out

The data presented so far add further support to our hypothesis that ATRX is required for ensuring normal gene expression at the α-globin locus by resolving G4 structures formed at the ψζ VNTR. To determine whether ATRX regulation of gene expression at the α-globin locus is dependent on ψζ VNTR, ψζ VNTR has been deleted from the genome via CRISPR-Cas9 in ATRX degron cells and the ATRX regulation of the HBM expression is examined in this context.

6.2.3.1 Generation of ψζ VNTR knockout cell line

a. Engineering design and targeting strategy

To knock out the ψζ VNTR (~360 bp in length) in the ATRX degron cell line, a dual-sgRNA CRISPR-Cas9 mediated knockout strategy was designed and adopted to achieve this. In the human α-globin locus, the ψζ VNTR lies within the intron 2 of a pseudo gene called ψζ *HBZP1* which is located upstream of *HBM* and contains three exons in total (Figure 6.5A). Given the extremely high homology of ψζ *HBZP1* with ζ-globin gene *HBZ*, it is difficult to design a specific sgRNA that solely targets ψζ *HBZP1* without affecting *HBZ.* Therefore, the 5' sgRNA was finally designed to locate within the promoter of ψζ *HBZP1* that has reduced homology with *HBZ*; and the 3' sgRNA was designed to locate

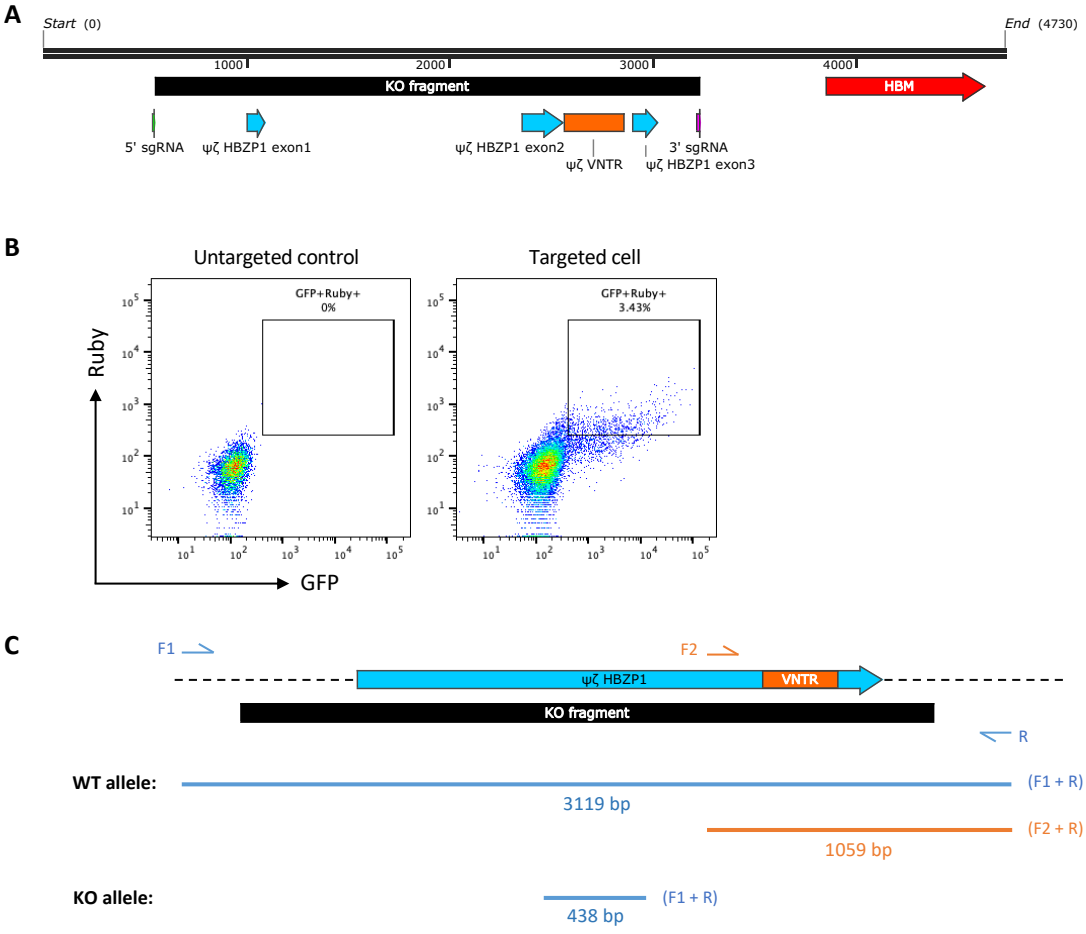
downstream of the ψζ *HBZP1.* As a result, the entire ψζ *HBZP1* (~2.5 kb) would be deleted (Figure 6.5A).

The targeting strategy was similar to the strategy being described in the previous chapter for editing *ATRX* and *HBM* locus. Briefly, the ATRX degron cells were electroporated with two CRISPR plasmids pX458-GFP-5' gRNA and pX458-Ruby-3' gRNA. Next, the bulk GFP+ and Ruby+ double positive cells were purified by FACS prior to separation and expansion of single clonal lines (Figure 6.5B). The clonal lines were then screened by a PCR-based strategy.

b. Screening of targeted cell clone

The bulk GFP+Ruby+ cells were firstly sorted and recovered prior to single cell sorting, during which the overall editing efficiency in the cells was determined by PCR. The PCR primers used for both examining editing efficiency in bulk and single clone screening, are denoted by arrows in Figure 6.5C. The screening includes two sets of primers. The first pair of primers (F1+R) is placed outside of the deleted region, and the product from the WT allele would be 3119 bp in length while the KO allele is expected to produce a 438 bp product (Figure 6.5C). Given the much shorter length of the KO allele compared with the WT allele, the KO allele is expected to have a substantial amplification advantage over the WT allele during PCR in heterozygous clones, resulting in a short PCR product alone. Thus, another pair of primers (F2+R) was designed to distinguish heterozygous and homozygous clones - there will be a 1059 bp PCR product using primers F2+R in the heterozygous deletion clones but without any product in the homozygous clones (Figure 6.5C).

Genotyping in the sorted bulk GFP+Ruby+ cells suggested a marked deletion efficiency (Figure 6.5D). As expected, the WT allele was not efficiently amplified in these samples due to replication bias, while the unedited cells indeed generated an expected 3119 bp PCR product (Figure 6.5D). Single clone screening was performed using both F1+R and F2+R primer sets for each clone. It demonstrated that the majority of the screened clones had a heterozygous deletion but clone 5 had a homozygous deletion (Figure 6.5E). The clone 5 was then sequenced to confirm the correct editing (Figure 6.5F).



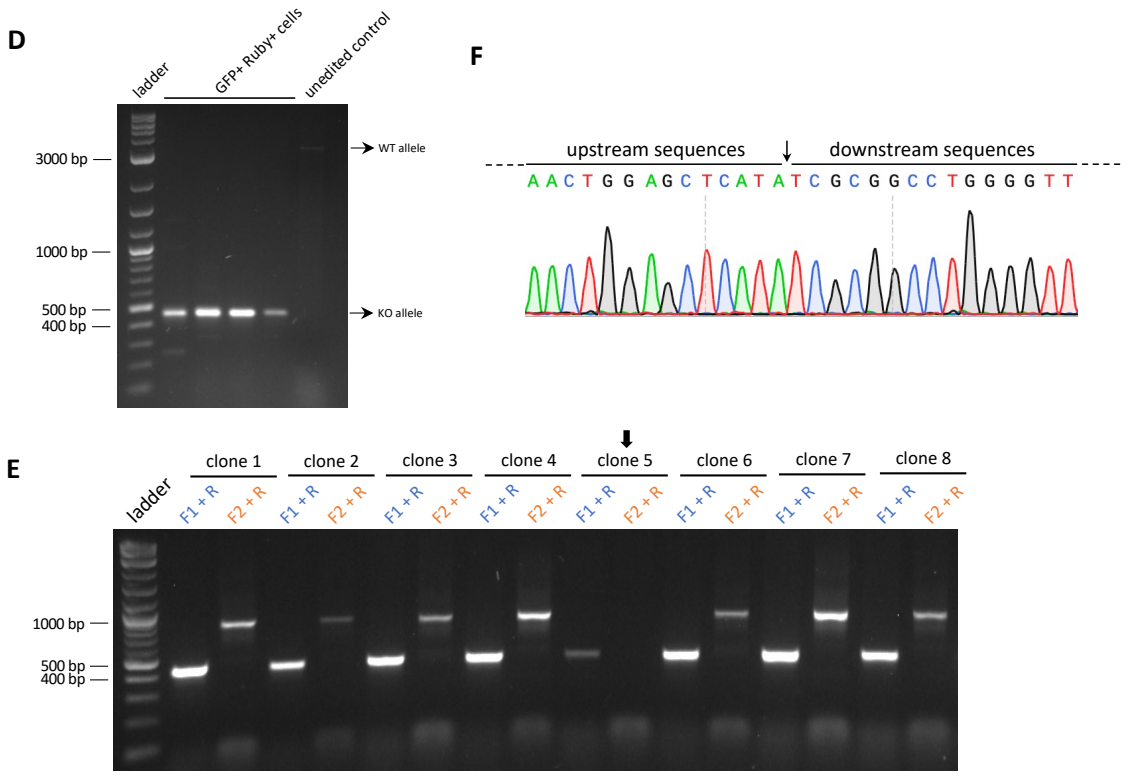


Figure 6.5 Generation of ψζ VNTR knockout cell line.

(A) Schematic for ψζ VNTR knockout design. The orange block represents ψζ VNTR, and the black block denotes the deleted fragment containing the entire ψζ *HBZP1*.

(B) Bulk GFP+Ruby+ FACS sorting plots of untargeted controls and the targeted cells.

(C) Diagram showing the PCR screening primers and products for WT and KO alleles.

(D) PCR products of bulk GFP+Ruby+ cells (from four individual samples) and unedited control using F1+R primers, indicating good editing efficiency.

(E) Representative PCR screening results of single targeted clones. Clone 5 was identified as a homozygous deletion clone and all the others contain heterozygous deletion.

(F) Sanger sequencing result of the ψζ VNTR locus in the ψζ VNTR homozygous knockout clone (clone 5) showing the correct deletion.

6.2.3.2 Characterisation of ψζ VNTR knockout cell line

To characterise the identified ATRX degron - ψζ VNTR homozygous knockout cell line, the cells were differentiated to erythroid cells using the HUDEP-2 differentiation protocol under either dTAG-13-Neg or dTAG-13 treatment, together with the control cell ATRX degron cell line. The ψζ VNTR KO cells showed equivalent erythroid differentiation to ATRX degron cell line under both normal condition (ATRX+) and dTAG-13 treatment (ATRX-), displaying the expected morphological changes (Figure 6.6A) and cell surface marker changes during the differentiation process (Figure 6.6B).

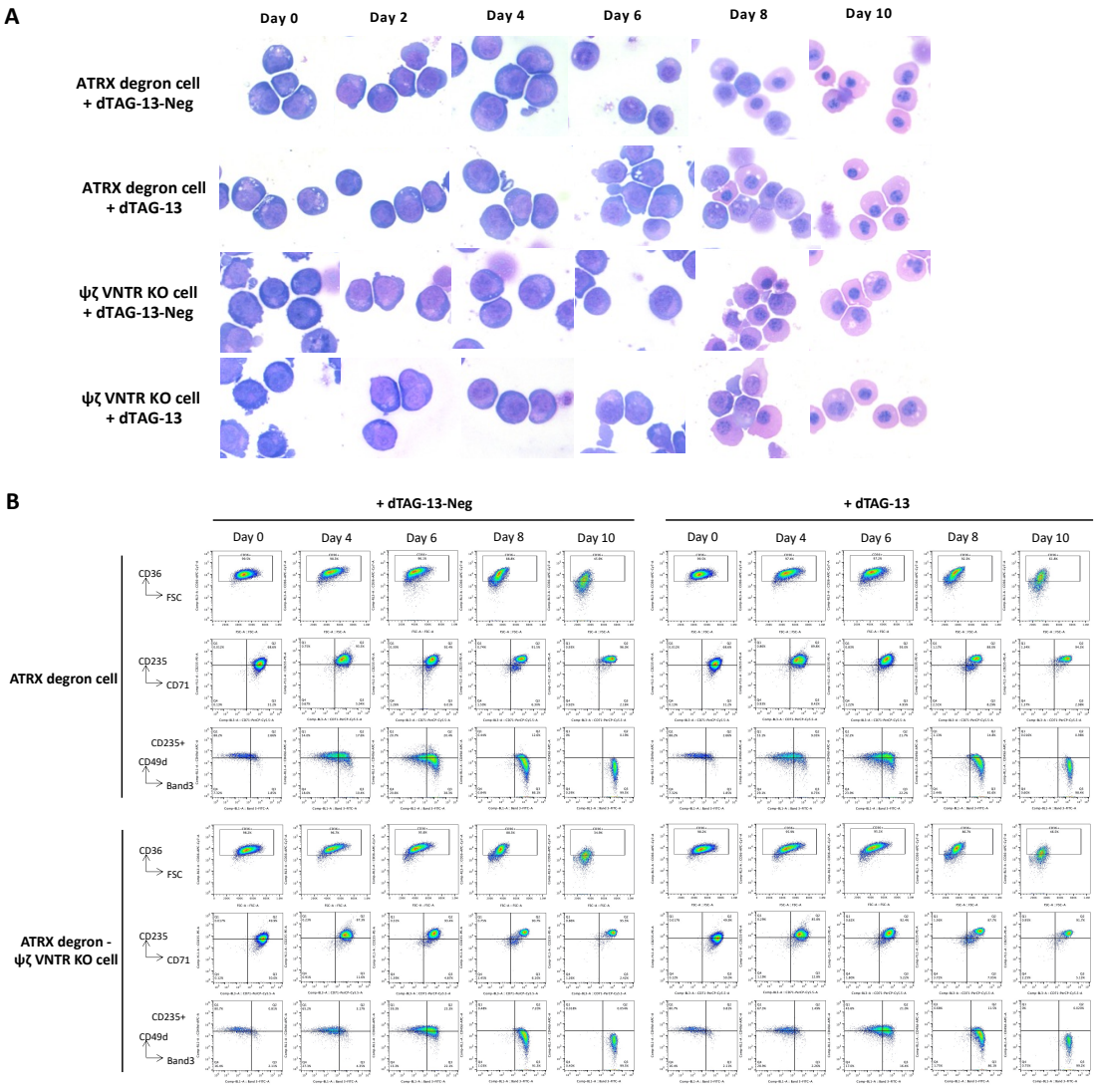


Figure 6.6 (previous page) Normal erythroid differentiation of ATRX degron-ψζ VNTR KO cell line.

(A) Cellular morphology detection by a modified Wright's staining method during differentiation process. ATRX degron-ψζ VNTR KO cells present equivalent erythroid differentiation to ATRX degron cells, both under dTAG-13-Neg and dTAG-13 treatment. (B) Flow cytometric analysis of cell surface markers during differentiation showing similar differentiation process between groups.

6.2.3.3 HBM expression in the ψζ VNTR knockout cell line

To assess the ATRX regulation of HBM expression after ψζ VNTR is deleted, the cell samples with or without ATRX/VNTR were collected at differentiation day 8 to determine the HBM expression and HBM to β-like ratio by qPCR. As shown in the previous chapter, HBM was significantly downregulated in the dTAG-13 treated, ATRX negative cells compared with the normal controls (Figure 6.7). Cells without ψζ VNTR showed a higher expression level of HBM, suggesting a repressive role of ψζ VNTR in HBM expression (Figure 6.7). Interestingly, in ψζ VNTR KO cells, HBM expression was resistant to ATRX loss (Figure 6.7), indicating ATRX regulates HBM expression through ψζ VNTR region.

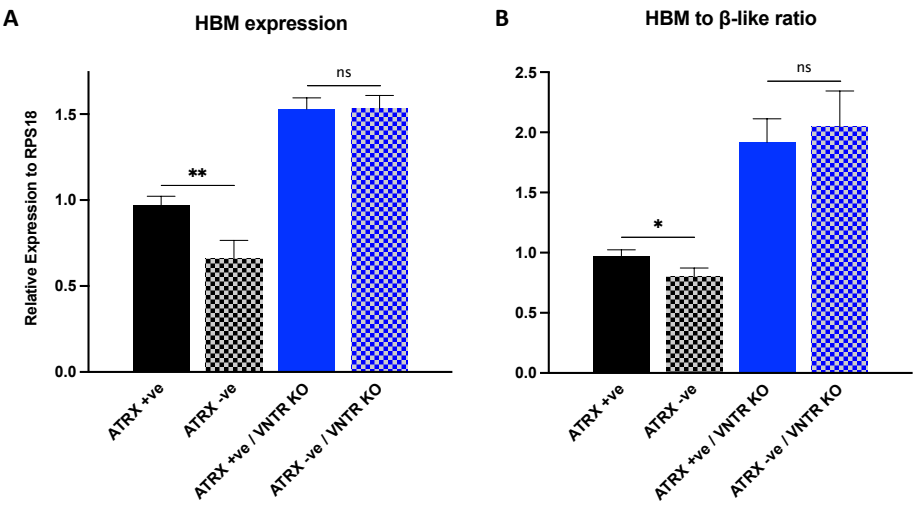


Figure 6.7 HBM expression is no longer affected by ATRX loss in ψζ VNTR KO cells.

Expression of HBM and HBM to β-like (HBB+HBG) globin in cells on differentiation day 8. ATRX +ve is ATRX degron cell treated with dTAG-13-Neg; ATRX -ve is ATRX degron cell treated with dTAG-13; ATRX +ve / VNTR KO is ATRX degron-ψζ VNTR KO cell treated with dTAG-13-Neg; ATRX -ve / VNTR KO is ATRX degron-ψζ VNTR KO cell treated with dTAG-13. Data are expressed as mean ± SD with *p <0.05, **p <0.01 according to multiple unpaired t test (n=3).

6.3 Discussion

It has previously been shown by our lab that ATRX binds the G-rich ψζ VNTR in the human α-globin cluster in erythroid cells, and the length of this repeat is positively correlated with the downregulation of α-(like)-globin expression in ATR-X syndrome patients (Law et al., 2010). This points towards ψζ VNTR as a cis-regulatory element of α-(like)-globin expression regulated by ATRX, but the direct and sufficient evidence is lacking. In this chapter, it is demonstrated that ATRX

regulation of α-(like)-globin expression is associated with binding to the ψζ VNTR, which supports the cis-acting role of ψζ VNTR in α-(like)-globin expression.

Here, the dynamics of ATRX binding at ψζ VNTR during erythroid differentiation correlates ATRX enrichment at ψζ VNTR with active transcription of the α-like globin HBM, which is consistent with the previous finding that recruitment of ATRX to G-rich repetitive sequences is transcription dependent (Nguyen et al., 2017). It is of note that the Illumina short-read sequencing method was used here to show the ATRX binding pattern at the α-globin locus, which made the assembly of highly repetitive genome sequences challenging. As a result, there is a lower number of reads mapping exactly to the ψζ VNTR compared with its flanking sequences in the ATRX ChIP-seq track shown here. Nonetheless, the ATRX peak call using input as a control identified the occupancy of ATRX at the ψζ VNTR, and this has previously been confirmed by interrogating the ATRX ChIP reads for dedicated analysis with repetitive sequences mapping (Law et al., 2010; Truch et al., 2022).

Motivated by the correlation between active gene expression and ATRX binding at the putative G4 sequences at ψζ VNTR, combined with the previous evidence showing ATRX can bind G4 structure (Law et al., 2010), we deduced that G4 structures formed at ψζ VNTR may play a role in gene expression. To test this hypothesis, we treated the cells with three G4-stabilizing ligands (PDS, CX-5461 and TMPyP4) individually to evaluate their effect on gene expression. Surprisingly, they show distinct effects and only CX-5461 affects gene expression at the α-globin locus. This is probably due to unknown *in vivo*

properties of the G4 structure formed by ψζ VNTR as well as the diverse effect of various G4 stabilizers.

In the past years, many efforts have been made to target G4s by various G4 ligands as a therapeutic approach, given the possible involvement of G4s in carcinogenesis and virus biology (Ruggiero & Richter, 2018; Awadasseid et al., 2021). In general, G4 ligands have common physical features that allow the selective interaction and binding with G4s, such as the aromatic surface, positive charges and a particular geometry (Figueiredo et al., 2022). Based on the conformation state and strand direction, there are three types of G4s: parallel, anti-parallel, and hybrid (Ruggiero & Richter, 2018). One previous study showed the ψζ VNTR sequence at the human α-globin locus tended to form a stable parallel G4 structure and even higher-order structures *in vitro* (Law et al., 2010; Y. Li et al., 2016). Different G4 ligands show varying capability in stabilizing G4s in different context. PDS, a G4 DNA-binding small molecule, has been widely used to stabilize G4s and shown the ability to target G4s to modulate the expression of oncogenes (Rodriguez et al., 2012; Moruno-Manchon et al., 2017; L. Y. Liu et al., 2022). However, PDS regulation on gene expression seems to be far more complicated than just acting as a DNA G4-binder. For instance, PDS has been reported to bind RNA and impact gene expression at the level of transcription and translation (He et al., 2021). Similarly, TMPyP4 is also able to stack and stabilize both DNA and RNA G4s, and it has also been reported playing a role in regulating gene expression by stabilizing G4 structure formed in the gene promoter, such as c-MYC (Siddiqui-Jain et al., 2002; S. R. Wang et al., 2016). While one concern of using TMPyP4 as a G4 stabilizer is that it is also

able to stack duplex DNA (Martino et al., 2009), this low selectivity towards G4 structures renders its effect an outcome of multiple mechanisms of action that are difficult to disentangle. Therefore, the insignificant effect of PDS and TMPyP4 treatment on the HBM expression shown here is probably due to the complexity of their gene regulation mechanism.

The third G4 stabilizer used in this study, CX-5461, has previously entered clinical trials for BRCA-deficient tumour treatment by selectively inhibiting ribosomal RNA (rRNA) synthesis that is widely activated and associated with uncontrolled cancer cell proliferation (H. Xu et al., 2017). The two activities of CX-5461, stabilizing G4 structures and inhibiting rRNA synthesis, may not be mutually exclusive. Intriguingly, ribosome biogenesis has been shown to play a regulatory role in erythroid development, as indicated by the erythroid defects observed in congenital Diamond-Blackfan anemia (DBA) or acquired $5q^-$ syndrome caused by mutations in ribosomal genes (Draptchinskaia et al., 1999; Ebert et al., 2008; Gazda et al., 2008), as well as the observation of impaired GATA1 protein production and perturbed erythroid lineage commitment in DBA pathogenesis (Khajuria et al., 2018). This suggests CX-5461 treatment may exert a regulatory function on erythroid differentiation itself at another layer other than stabilizing the G4 structure at ψζ VNTR. While importantly, a recent study has showed that inhibition of ribosome biogenesis by CX-5461 restrained immature erythroblast proliferation at proerythroblasts and early basophilic erythroblasts stages, but had little effect at later stages (le Goff et al., 2021). Therefore, I treated the cells with CX-5461 from differentiation day 6 when most of the cells are already at late basophilic erythroblasts stage, to avoid the secondary effect

on gene expression caused by perturbed ribosome biogenesis. Indeed, we did not observe significant changes in erythroid differentiation when the cells were treated with CX-5461. Meanwhile, we additionally corrected for any effect of CX-5461 on the overall erythroid differentiation process by measuring the ratio of α-(like)-globin to β-like globin, and GAPDH was selected as a reference gene to normalise individual globin gene expression when analysing qPCR data, instead of the ribosomal genes RPS18 and RPL13A being used in previous chapters. By doing so, we could conclude soundly that CX-5461 treatment leads to perturbed expression of HBM likely by stabilizing G4 structure formed at the ψζ VNTR. Further work, of course, is needed to confirm the actual presence of G4 at the ψζ VNTR *in vivo*. Nonetheless, our data indicate that the presence of non-canonical G4 DNA secondary structures supresses gene expression and that ATRX may normally play a role in overcoming their deleterious effect.

To confirm whether the gene regulation role of ATRX is dependent on the ψζ VNTR, we genetically deleted the ψζ VNTR in the α-globin locus and assessed the gene-regulatory influence. Notably, ATRX loss no longer affects HBM expression in the absence of ψζ VNTR, indicating ATRX regulates α-(like)-globin expression directly through ψζ VNTR. This is consistent with the mounting evidence highlighting the importance of VNTRs in gene expression and genetic traits (Örd et al., 2020; Hall et al., 2020; Eslami Rasekh et al., 2021; Bakhtiari et al., 2021). Moreover, given the variable length of ψζ VNTR between individuals, this explains well the variation of α-thalassaemia phenotype between ATR-X patients with the same *ATRX* mutation.

In summary, in this chapter, it is demonstrated that ATRX binds to the ψζ VNTR when the α-(like)-globin genes start to be expressed, and this expression can be perturbed by stabilizing G4 structures. These suggest ATRX may play an important role in preventing or resolving G4 structures formed at ψζ VNTR to ensure proper gene expression. In support of this, when deleting ψζ VNTR in the cell, the expression of α-like globin gene HBM is no longer dependent on ATRX, indicating ATRX regulates gene expression at the α-globin locus through the cis-regulatory element ψζ VNTR.

7 Conclusions and general discussion

7.1 Summary of findings

This thesis aims to explore the role of the chromatin remodeller ATRX in gene expression, using its prime target gene α-globin as an exemplar. From the perspective of gene regulation, α-globin serves as a particularly good example to study, in that the mechanisms of gene regulation at the α-globin gene locus have been well characterised. Additionally, as a gene predominantly expressed in the haematologic system, the experimental samples are easy to collect and obtain. In this thesis, using CRISPR/Cas9 edited haematopoietic stem cells and the immortalized erythroid progenitor cell line HUDEP-2 cells as models, I have managed to recapitulate the α-thalassaemia phenotype observed in the ATR-X patients and to determine the mechanism by which gene expression is perturbed by ATRX loss. The key findings are summarised below.

- ATRX is involved in the DNA damage response, cell cycle regulation and the regulation of gene expression in erythroid cells (**Chapter 4**).
- An ATRX defect leads to α-thalassaemia/downregulation of α-globin in only a subset of cells (BFU-E colony data in **Chapter 3**; scRNA-seq data in **Chapter 4**).
- The downregulation of α-globin genes by ATRX loss is not associated with DNA methylation (**Chapter 5**) but is likely to be associated with DNA damage-induced histone modifications, such as macroH2A1

accumulation or pATM-RNF168 mediated histone ubiquitination (**Chapter 4**).

- The establishment of a dTAG-13 inducible ATRX degron - HBM mKate2 reporter HUDEP-2 cell line recapitulates the α-thalassaemia phenotype and can serve as a tractable system to enrich the affected subpopulation to validate the mechanism affecting gene regulation (**Chapter 5**).
- ATRX primarily functions to regulate α-globin expression at the basophilic erythroblast stage during erythroid differentiation (**Chapter 5**).
- ATRX regulation depends on its binding to the putative G4-containing ψζ VNTR within the α-globin gene cluster (**Chapter 6**). G4 stabilizer CX-5461 downregulates α-globin gene expression and knockout of the ψζ VNTR abolishes the effect of ATRX depletion on α-globin expression (**Chapter 6**).

7.2 Wider implications

The implications of the present study can be extended beyond α-globin gene regulation. ATRX has been found to bind to G4 forming sequences within the genome (Law et al., 2010; Teng et al., 2021; Truch et al., 2022), and ATRX deficiency promotes the formation of G4 structures (Y. Wang et al., 2019), particularly it seems when transcribed (Nguyen et al., 2017). Therefore, the mechanisms involving the regulation of G4-associated, ATRX-dependent genes indicated in this thesis may be applicable widely.

7.2.1 ATRX functions to protect cells from G4-induced replicative stress

7.2.1.1 ATRX interacts with the replisome to prevent G4-induced replicative stress

The DNA replication machinery, also known as the replisome, is a large complex consisting of multiple proteins orchestrating the unwinding and replication of DNA. There have been a number of studies indicating ATRX plays a role in DNA replication by showing ATRX deficiency leads to replication defects (Leung et al., 2013; Clynes et al., 2014, 2015; Huh et al., 2016; Akter et al., 2021). Here I show consistently that DNA replication stress and DNA damage response pathways are upregulated in ATRX deficient erythroid cells (**Chapter 4**). Significantly, a recent study demonstrates that ATRX directly interacts with the DNA replication helicase MCM complex, and loss of ATRX boosts the formation of G4 structures at newly synthesized DNA (Teng et al., 2021). This provides a direct link between ATRX loss, the formation of G4 structures at replication forks and induced replication stress.

Interestingly, direct visualisation of the replication fork and endogenous G4s by single-molecule imaging technology has revealed that G4s can form at newly unwound DNA immediately behind the MCM helicase and prior to nascent DNA synthesis; these replication fork-coupled G4s impede DNA synthesis and are associated with replicative stress (W. T. C. Lee et al., 2021). It therefore seems likely that ATRX functions to resolve the replication fork coupled G4s by interacting with MCM helicase to facilitate DNA synthesis and to prevent replicative stress. In support, Teng et al. have shown that the helicase domain of

ATRX is required to protect cells from G4-induced replicative stress (Teng et al., 2021), further pointing to ATRX helping to resolve replication fork-coupled G4 by its helicase domain. Hence, in the absence of ATRX, G4s accumulate at the replication fork that blocks DNA replication and induces replicative stress and DNA damage, which may have an impact on gene expression.

In addition, MCM complexes have recently been shown as barriers to restrict cohesin-mediated loop extrusion in G1 phase to impede the formation of loops and topologically associating domains, and thereby affecting gene expression (Dequeker et al., 2022). Given the interaction between ATRX and MCM complex (Teng et al., 2021), it is also possible that ATRX regulates gene expression through interacting with MCM to affect loop extrusion. This is, however, less likely the case in regulating α-globin expression although the ψζ VNTR lies between α-globin enhancers and promoters. The reason is that it would be hard to explain the phenomenon of distance-dependent gene downregulation pattern at the α-globin gene locus (HBM>HBA2>HBA1) observed in the ATR-X patients (Law et al., 2010). Moreover, the VNTR at the *NME4* locus that ATRX binds to is within the first intron of *NME4* which is unlikely to affect enhancer-promoter interaction by perturbing loop-extrusion.

In conclusion, the soundest hypothesis to explain the ATRX regulation of α-globin would still be G4-associated replicative stress induced gene dysregulation.

7.2.1.2 ATRX supresses G4s and R-loops forming at G-rich regions during transcription

DNA replication is not the only occasion when G4 forms. Exposure of G-rich single-stranded DNA during transcription also creates opportunities for the formation of G4 structures. It has been shown that ATRX binding increases at G-rich sequences when they are transcribed, and loss of ATRX is associated with increased R-loop formation (Nguyen et al., 2017), indicating ATRX may contribute to resolve the G4s and R-loops formed during transcription. The three-stranded R-loop structure forms when the nascent RNA strand synthesized by RNA polymerase (RNAP) pairs with the complementary template DNA to form an RNA:DNA hybrid, leaving the non-template DNA single-stranded. Notably, the R-loops formed at the G-rich regions can be stabilized by the formation of G4s on the displaced non-template G-rich strand. The persistence of such secondary structures is deleterious and may interfere with replication to induce replication stress (Crossley et al., 2019). In the absence of ATRX, formation of both G4s and R-loops can be increased at the G-rich regions during transcription, which may block DNA replication, result in transcription-replication collisions and eventually causing DNA damage.

7.2.2 ATRX loss-induced replicative stress activates ATM-RNF168-ubiquitination DNA damage response pathway

Studies have indicated that loss of ATRX promotes the formation of G4s (Y. Wang et al., 2019), particularly at the replication forks and thus leading to replicative stress (Teng et al., 2021). The G4 accumulation at the replication fork

would lead to DNA damage and activate the DNA damage response pathway. In mammalian cells, the ATM and ATR are two of the most upstream kinases to be activated in response to DNA damage. DNA damage response signalling pathways regulated by the ATM and ATR kinases are complementary; ATM is mainly involved in double-strand break repair while ATR primarily helps repair single-strand damage (Maréchal & Zou, 2013).

Numerous studies suggest ATM-mediated pathways may play a dominant role over ATR in response to ATRX loss-induced DNA damage. G4-induced replicative stress at replication forks has been shown inhibiting RPA-associated ATR signalling (Teng et al., 2021), indicating ATR pathways do not mediate G4-induced DNA damage response. There are some direct proofs indicating the involvement of ATM in response to ATRX depletion. In the Atrx-null mouse embryonic brain, levels of ATR and p-ATR are unchanged whereas the level of p-ATM is increased (Watson et al., 2013). The ATM related pathways are activated in ATRX deficient cells (Huh et al., 2012, 2016; Qin et al., 2022). In addition, ATRX has been found to promote the recovery from replication stress by interacting with the MRN complex (Leung et al., 2013; Clynes et al., 2014), a sensor of DSB that activates ATM (Maréchal & Zou, 2013). Downstream reaction of the activated ATM pathway often involves a ubiquitination cascade at DSBs that promotes the recruitment of DNA repair proteins such as BRCA1 and 53BP1. RNF168 is one intermediate factor of the ubiquitination cascade and mediates histone H2A ubiquitination (J. Kelliher et al., 2022).

In accordance with these observations, the work presented in this thesis indicates that ATM and RNF168 are upregulated in the absence of ATRX, and histone H2A ubiquitination is upregulated in the affected cells which is potentially responsible for the dysregulated gene expression (**Chapter 4**). The role of ATM-RNF168-ubiqutination pathway in gene expression is further described in the following section.

7.2.3 Replication stress and DNA damage are associated with perturbed gene expression

DNA damage including DSBs can happen following replicative stress, which would disturb the intrinsic stability of chromatin and result in a perturbed epigenetic profile such as alterations in the deposition of histone marks, which is directly linked to gene expression. Here, I discuss two prime histone marks that respond to DNA damage and also play a role in gene regulation, which are likely to act as mediators between DNA damage and perturbed gene expression.

7.2.3.1 Activation of ATM-RNF168 and histone ubiquitination at DNA damage and G4 sites associates with altered gene expression

Increased RNF168 recruitment and histone ubiquitination have been observed at DNA damage and G4 sites, and this DNA damage response-related ubiquitin signalling pathway is frequently altered in cancers in which many genes are expressed aberrantly (Masud et al., 2021; Rong et al., 2022), suggesting this pathway may also involve in gene regulation. Indeed, it has been demonstrated that the activation of ATM-RNF168-histone H2A ubiquitination pathway in

response to DNA damage leads to transcriptional silencing of adjacent genes (Shanbhag et al., 2010).

Consistently, it is shown in this thesis that stabilizing G4 structures at the α-globin and NME4 gene locus leads to downregulated gene expression (**Chapter 6**), similar to the effect of ATRX loss which is accompanied by upregulated ATM-RNF168-histone H2A ubiquitination (**Chapter 4**). This not only provides direct evidence on the connection between an ATRX defect and activated ATM-RNF168-ubiquitination DNA damage response pathway, but also potentially links the two ATRX functions together - genome stability maintenance and gene regulation.

7.2.3.2 Deposition of the histone variant macroH2A1 at sites of replication stress or fork stalling is associated with altered gene expression

The histone variant macroH2A1 has been implicated in diverse nuclear processes. It is involved in ATM-dependent DNA damage response, and both variants, macroH2A1.1 and macroH2A1.2, are recruited to sites of DNA damage (C. Xu et al., 2012; Khurana et al., 2014). Importantly, macroH2A1.2 has been shown to be an epigenetic link between ATRX deficiency and the DNA damage response at sites of recurrent replication stress (J. Kim, Sturgill, et al., 2018; J. Kim et al., 2019). Aside from a role in DNA damage, macroH2A1.2 has been largely associated with gene repression (Ratnakumar et al., 2012; J. M. Kim et al., 2018; J. Kim, Shin, et al., 2018), although it has also been reported to activate gene expression (Dell'Orso et al., 2016). It therefore seems probable that there is an increased deposition of macroH2A1 in response to ATRX loss-induced DNA

damage, and its recruitment at the damaged site may regulate the transcription of local genes.

7.2.4 Working model of ATRX regulation of gene expression

Based upon the findings in this thesis and the discussion above, a working model for ATRX regulation of gene expression at G4 sites is proposed (Figure 7.1). G-rich regions are prone to form G4 structures during replication and transcription. Under normal DNA replication conditions, ATRX interacts with MCM and is recruited to these sites to promote the resolution of G4s alone or together with other factors, allowing for seamless continuation of replication fork progression, which ensures the proper replication and helps maintain the genome stability. In the absence of ATRX, G4s persist and accumulate at replication forks, resulting in persistent fork stalling, replication stress and eventually leading to DNA damage. Under normal transcription conditions, ATRX binds to G-rich regions to prevent the formation of G4s and R-loops, ensuring proper gene expression. In the absence of ATRX, G4 and R-loop persist at these regions which can block the replication fork, leading to transcription-replication collisions, replication stress and DNA damage.

The DNA damage activates the ATM-RNF168 DNA damage response pathway to promote DNA repair, and the recruitment of RNF168 to DSB facilitates ubiquitylation of many substrates including histone H2A, which may mediate transcriptional silencing on chromatin in *cis* to the DSB. Meanwhile, macroH2A may also be recruited to the damaged sites through an ATM dependent pathway

to help DNA repair, which may repress the local gene expression as an inert epigenetic mark. Notably, the epigenetic instability triggered by DNA damage can lead to changes in the local epigenetic landscape which propagate on the chromatin in *cis* to the damaged sites in a distance-dependent manner, therefore leading to the distance-dependent gene dysregulation pattern.

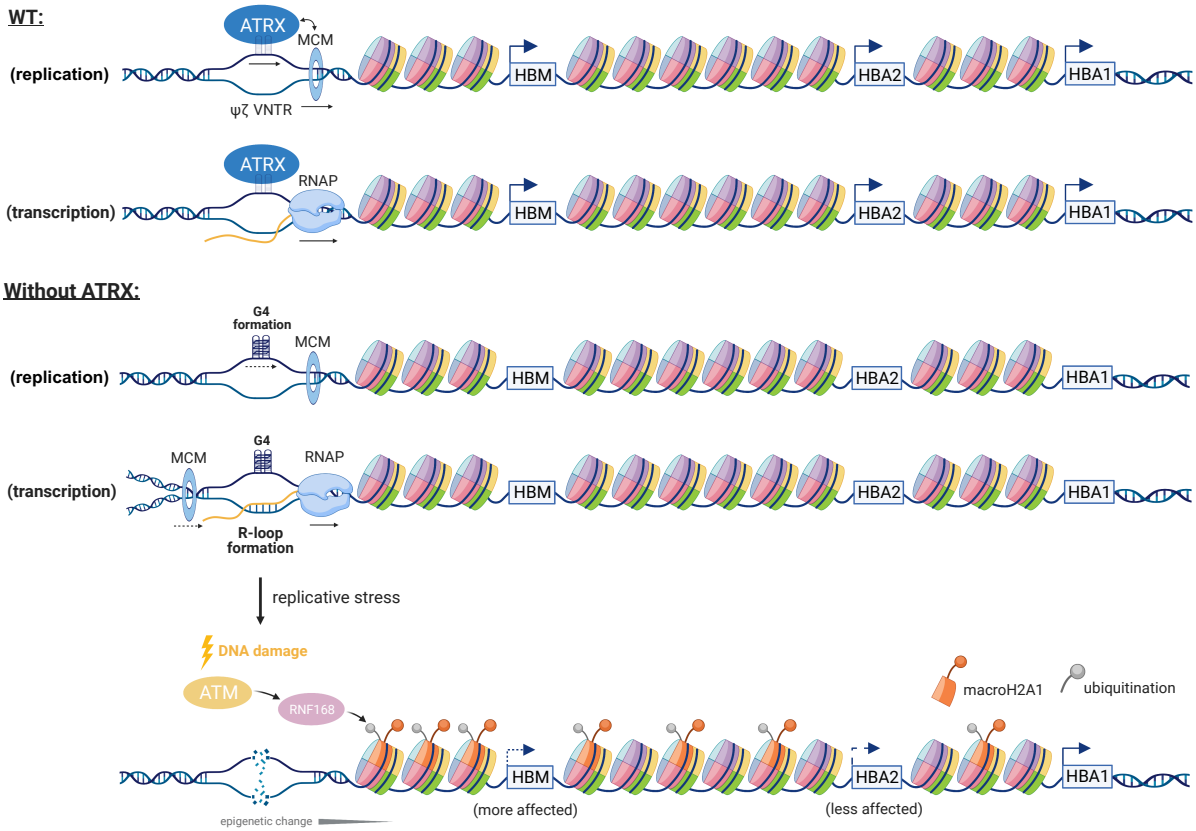


Figure 7.1 Working model for ATRX regulation of gene expression.

In the wild type, ATRX helps resolve G4s to ensure proper DNA replication and transcription. In the absence of ATRX, G4s and R-loops accumulate, leading to replication stress, DNA damage, activation of ATM-RNF168 DNA damage response pathway and perturbed gene expression.

This model can not only nicely explain the puzzling distance-dependent gene dysregulation pattern, but also make the two key findings in this thesis rational. The first is the dependence on the ψζ VNTR for ATRX regulation, given G4 structures are apt to form at this highly G-rich region where ATRX binds. This also explains why there is no α-thalassaemia phenotype observed in all the ATRX deficient mouse models (data from MGI database), as the ψζ VNTR is not present in the mouse genome. The second is that only a subset of ATRX deficient cells are affected. Given the proposed model is based on the formation of G4s during DNA replication and transcription, only those cells suffering from the unresolved G4s/R-loops and replication stress are being affected. Of note, apart from ATRX, other factors can also destabilize G4s to release the cells from this stress, for example the FANCJ DNA helicase and RecQ helicases (Mendoza et al., 2016; W. T. C. Lee et al., 2021). Therefore, the affected cells are those that lack ATRX to resolve G4s and retain G4s that cannot be resolved by other complementary helicases, which would be expected to be only a subset of cells.

7.3 Future directions

Further work needs to be done to strengthen and elaborate the working model. Firstly, there is still no direct evidence showing the formation of G4 structures at the ψζ VNTR sequences *in vivo* in the absence of ATRX. This can be done by identifying G4s in chromatin using the antibody-based G4-ChIP in cells with or without ATRX. Secondly, the presence of replication stress and DNA damage at the ψζ VNTR site needs to be confirmed in the ATRX-depleted cells. This can be

confirmed by checking the recruitment of the DNA damage protein markers like γH2AX and 53BP1 at this site, either through ChIP experiments for these markers or doing immunofluorescence staining of the markers combined with DNA FISH of the α-globin locus. Alternatively, END-seq experiment can be performed to monitor the DNA end resection and DNA double-strand breaks at base-pair resolution *in vivo* (Canela et al., 2016). Finally, the epigenetic mark that links DNA damage to local transcriptional change needs to be explored. The established ATRX degron-HBM mKate2 reporter cell line can contribute to this. The unaffected and affected cell samples can be FACS-sorted separately to perform ChIP-seq experiments for candidate markers (H2A ubiquitination and macroH2A1) and then to compare their differences.

Additionally, the ATRX degron cell line generated in this thesis provides a tractable system to study more general functions of ATRX non-invasively, for example investigating the dynamics of ATRX-DAXX interaction and H3.3 deposition, or the role of ATRX in chromatin conformation as a chromatin remodeller.

Reference

Ahmad, K., & Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Molecular Cell*, *9*(6), 1191–1200.

Akter, J., Katai, Y., Sultana, P., et al. (2021). Loss of p53 suppresses replication stress-induced DNA damage in ATRX-deficient neuroblastoma. *Oncogenesis*, *10*(11).

Aldridge, S., & Teichmann, S. A. (2020). Single cell transcriptomics comes of age. *Nature Communications 2020 11:1*, *11*(1), 1–4.

Almouzni, G., & Probst, A. v. (2011). Heterochromatin maintenance and establishment: Lessons from the mouse pericentromere. *Http://Dx.Doi.Org/10.4161/Nucl.2.5.17707*, *2*(5), 332–338.

An, J. Y., Kim, E. A., Jiang, Y., et al. (2010). UBR2 mediates transcriptional silencing during spermatogenesis via histone ubiquitination. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(5), 1912–1917.

Anzalone, A. v., Randolph, P. B., Davis, J. R., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature 2019 576:7785*, *576*(7785), 149–157.

Arendt, D., Musser, J. M., Baker, C. V. H., et al. (2016). The origin and evolution of cell types. *Nature Reviews Genetics 2016 17:12*, *17*(12), 744–757.

Argentaro, A., Yang, J. C., Chapman, L., et al. (2007). Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(29), 11939–11944.

Awadasseid, A., Ma, X., Wu, Y., & Zhang, W. (2021). G-quadruplex stabilization via small-molecules as a potential anti-cancer strategy. *Biomedicine & Pharmacotherapy*, *139*, 111550.

Badens, C., Martini, N., Courrier, S., et al. (2006). ATRX syndrome in a girl with a heterozygous mutation in the ATRX Zn finger domain and a totally skewed X-inactivation pattern. *American Journal of Medical Genetics, Part A*, *140*(20), 2212–2215.

Bagheri-fam, S., Argentaro, A., Svingen, T., et al. (2011). Defective survival of proliferating Sertoli cells and androgen receptor function in a mouse model of the ATR-X syndrome. *Human Molecular Genetics*, *20*(11), 2213–2224.

Bakhtiari, M., Park, J., Ding, Y. C., et al. (2021). Variable number tandem repeats mediate the expression of proximal genes. *Nature Communications*, *12*(1).

Banaszynski, L. A., Chen, L. chun, Maynard-Smith, L. A., Ooi, A. G. L., & Wandless, T. J. (2006). A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules. *Cell*, *126*(5), 995–1004.

Bao, X., Zhang, X., Wang, L., et al. (2021). Epigenetic inactivation of ERF reactivates γ-globin expression in β-thalassaemia. *American Journal of Human Genetics*, *108*(4), 709–721.

Barbour, H., Daou, S., Hendzel, M., & Affar, E. B. (2020). Polycomb group-mediated histone H2A monoubiquitination in epigenome regulation and nuclear processes. In *Nature Communications* (Vol. 11, Issue 1). Nature Research.

Barminko, J., Reinholt, B., & Baron, M. H. (2016). Development and differentiation of the erythroid lineage in mammals. *Developmental and Comparative Immunology*, *58*, 18–29.

Barrangou, R., Fremaux, C., Deveau, H., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, *315*(5819), 1709–1712.

Benz, E. J., Murnane, M. J., Tonkonow, B. L., et al. (1980). Embryonic-fetal erythroid characteristics of a human leukemic cell line. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(6), 3509–3513.

Bernardi, R., & Pandolfi, P. P. (2007). Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nature Reviews Molecular Cell Biology 2007 8:12*, *8*(12), 1006–1016.

Bérubé, N. G., Healy, J., Medina, C. F., et al. (2008). Patient mutations alter ATRX targeting to PML nuclear bodies. *European Journal of Human Genetics : EJHG*, *16*(2), 192–201.

Bérubé, N. G., Jagla, M., Smeenk, C., et al. (2002). Neurodevelopmental defects resulting from ATRX overexpression in transgenic mice. *Human Molecular Genetics*, *11*(3), 253–261.

Bérubé, N. G., Mangelsdorf, M., Jagla, M., et al. (2005). The chromatin-remodeling protein ATRX is critical for neuronal survival during corticogenesis. *The Journal of Clinical Investigation*, *115*(2), 258–267.

Bhargava, R., Lynskey, M. L., & O'Sullivan, R. J. (2022). New twists to the ALTernative endings at telomeres. *DNA Repair*, *115*.

Bienz, M. (2006). The PHD finger, a nuclear protein-interaction domain. *Trends in Biochemical Sciences*, *31*(1), 35–40.

Blahnik, K. R., Dou, L., Echipare, L., et al. (2011). Characterization of the contradictory chromatin signatures at the 3' exons of zinc finger genes. *PloS One*, *6*(2).

Boch, J., Scholze, H., Schornack, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, *326*(5959), 1509–1512.

Bolotin, A., Quinquis, B., Sorokin, A., & Dusko Ehrlich, S. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, England)*, *151*(Pt 8), 2551–2561.

Boontanrart, M. Y., Schröder, M. S., Stehli, G. M., et al. (2020). ATF4 Regulates MYB to Increase γ-Globin in Response to Loss of β-Globin. *Cell Reports*, *32*(5).

Boudhraa, Z., Carmona, E., Provencher, D., & Mes-Masson, A. M. (2020). Ran GTPase: A Key Player in Tumor Progression and Metastasis. In *Frontiers in Cell and Developmental Biology* (Vol. 8). Frontiers Media S.A.

Bozhilov, Y. K., Downes, D. J., Telenius, J., et al. (2021). A gain-of-function single nucleotide variant creates a new promoter which acts as an orientation-dependent enhancer-blocker. *Nature Communications 2021 12:1*, *12*(1), 1–13.

Brady, H. J. M., Edwards, M., Linch, D. C., et al. (1990). Expression of the human carbonic anhydrase I gene is activated late in fetal erythroid development and regulated by stage-specific trans-acting factors. *British Journal of Haematology*, *76*(1), 135–142.

Canela, A., Sridharan, S., Sciascia, N., et al. (2016). DNA Breaks and End Resection Measured Genome-wide by End Sequencing Molecular Cell Technology DNA Breaks and End Resection Measured Genome-wide by End Sequencing. *Molecular Cell*, *63*, 898–911.

Cardoso, C., Timsit, S., Villard, L., et al. (1998). Specific Interaction between the XNP ATR-X Gene Product and the SET Domain of the Human EZH2 Protein. *Human Molecular Genetics*, *7*(4), 679–684.

Chan, K. M., Fang, D., Gan, H., et al. (2013). The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Genes & Development*, *27*(9), 985.

Chang, F. T. M., McGhie, J. D., Chan, F. L., et al. (2013). PML bodies provide an important platform for the maintenance of telomeric chromatin integrity in embryonic stem cells. *Nucleic Acids Research*, *41*(8), 4447–4458.

Chavez, J., Murillo-Maldonado, J. M., Bahena, V., et al. (2017). dAdd1 and dXNP prevent genome instability by maintaining HP1a localization at Drosophila telomeres. *Chromosoma*, *126*(6), 697–712.

Chen, J., Kremer, C. S., & Bender, T. P. (2006). The carbonic anhydrase I locus contains a c-Myb target promoter and modulates differentiation of murine erythroleukemia cells. *Oncogene*, *25*(19), 2758–2772.

Chen, X., Bahrami, A., Pappo, A., et al. (2014). Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Reports*, *7*(1), 104–112.

Chu, H. P., Cifuentes-Rojas, C., Kesner, B., et al. (2017). TERRA RNA Antagonizes ATRX and Protects Telomeres. *Cell*, *170*(1), 86-101.e16.

Chu, J., Haynes, R. D., Corbel, S. Y., et al. (2014). Non-invasive intravital imaging of cellular differentiation with a bright red-excitable fluorescent protein. *Nature Methods*, *11*(5), 572–578.

Clark, S. J., Statham, A., Stirzaker, C., Molloy, P. L., & Frommer, M. (2006). DNA methylation: Bisulphite modification and analysis. *Nature Protocols 2006 1:5*, *1*(5), 2353–2364.

Clynes, D., Jelinska, C., Xella, B., et al. (2014). ATRX dysfunction induces replication defects in primary mouse cells. *PLoS ONE*, *9*(3).

Clynes, D., Jelinska, C., Xella, B., et al. (2015). Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX. *Nature Communications*, *6*, 7538.

Coelho, A., Picanço, I., Seuanes, F., Seixas, M. T., & Faustino, P. (2010). Novel large deletions in the human alpha-globin gene cluster: Clarifying the HS-40 long-range regulatory role in the native chromosome environment. *Blood Cells, Molecules & Diseases*, *45*(2), 147–153.

Cole, T. R. P., May, A., & Hughes, H. E. (1991). α thalassaemia/mental retardation syndrome (non-deletional type): Report of a family supporting X linked inheritance. *Journal of Medical Genetics*, *28*(11), 734–737.

Conte, D., Huh, M., Goodall, E., et al. (2012). Loss of Atrx Sensitizes Cells to DNA Damaging Agents through p53-Mediated Death Pathways. *PLOS ONE*, *7*(12), e52167.

Crossley, M. P., Bocek, M., & Cimprich, K. A. (2019). R-Loops as Cellular Regulators and Genomic Threats. *Molecular Cell*, *73*(3), 398–411.

Dang, F., Nie, L., & Wei, W. (2021). Ubiquitin signaling in cell cycle control and tumorigenesis. In *Cell Death and Differentiation* (Vol. 28, Issue 2, pp. 427–438). Springer Nature.

Daniels, D. E., Downes, D. J., Ferrer-Vicens, I., et al. (2020). Comparing the two leading erythroid lines BEL-A and HUDEP-2. In *Haematologica* (Vol. 105, Issue 8, pp. E389–E394). Ferrata Storti Foundation.

Danussi, C., Bose, P., Parthasarathy, P. T., et al. (2018). Atrx inactivation drives disease-defining phenotypes in glioma cells of origin through global epigenomic remodeling. *Nature Communications 2018 9:1*, *9*(1), 1–15.

de Gobbi, M., Anguita, E., Hughes, J., et al. (2007). Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. *Blood*, *110*(13), 4503–4510.

de Gobbi, M., Garrick, D., Lynch, M., et al. (2011). Generation of bivalent chromatin domains during cell fate decisions. *Epigenetics and Chromatin*, *4*(1), 1–12.

de Wilde, R., Heaphy, C., Maitra, A., et al. (2012). Loss of ATRX or DAXX expression and concomitant acquisition of the alternative lengthening of telomeres phenotype are late events in a small subset of MEN-1 syndrome pancreatic neuroendocrine tumors. *Modern Pathology,* 25(7),1033-9.

Delbarre, E., Ivanauskiene, K., Spirkoski, J., et al. (2017). PML protein organizes heterochromatin domains where it regulates histone H3.3 deposition by ATRX/DAXX. *Genome Research*, *27*(6), 913–921.

Dell'Orso, S., Wang, A. H., Shih, H. Y., et al. (2016). The Histone Variant MacroH2A1.2 Is Necessary for the Activation of Muscle Enhancers and Recruitment of the Transcription Factor Pbx1. *Cell Reports*, *14*(5), 1156–1168.

Dequeker, B. J. H., Scherr, M. J., Brandão, H. B., et al. (2022). MCM complexes are barriers that restrict cohesin-mediated loop extrusion. *Nature 2022 606:7912*, *606*(7912), 197–203.

Dhayalan, A., Tamas, R., Bock, I., et al. (2011). The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9. *Human Molecular Genetics*, *20*(11), 2195–2203.

Dias, J., Gumenyuk, M., Kang, H., et al. (2011). Generation of red blood cells from human induced pluripotent stem cells. *Stem Cells and Development*, *20*(9), 1639–1647.

Dolznig, H., Kolbus, A., Leberbauer, C., et al. (2005). Expansion and differentiation of immature mouse and human hematopoietic progenitors. *Methods in Molecular Medicine*, *105*, 323–344.

Donnai, D., Clayton-Smith, J., Gibbons, R. J., & Higgs, D. R. (1991). The non-deletion a thalassaemia/mental retardation syndrome: Further support for X linkage. *Journal of Medical Genetics*, *28*(11), 742–745.

Drané, P., Ouararhni, K., Depaux, A., Shuaib, M., & Hamiche, A. (2010). The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes & Development*, *24*(12), 1253–1265.

Draptchinskaia, N., Gustavsson, P., Andersson, B., et al. (1999). The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nature Genetics 1999 21:2*, *21*(2), 169–175.

Duc, C., Benoit, M., Détourné, G., et al. (2017). Arabidopsis ATRX Modulates H3.3 Occupancy and Fine-Tunes Gene Expression. *The Plant Cell*, *29*(7), 1773–1793.

Ebert, B. L., Pretz, J., Bosco, J., et al. (2008). Identification of RPS14 as a 5q-syndrome gene by RNA interference screen. *Nature 2008 451:7176*, *451*(7176), 335–339.

Elbakry, A., Juhász, S., Chan, K. C., & Löbrich, M. (2021). ATRX and RECQ5 define distinct homologous recombination subpathways. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(3).

Emerson, R. O., & Thomas, J. H. (2009). Adaptive evolution in zinc finger transcription factors. *PLoS Genetics*, *5*(1).

Eslami Rasekh, M., Hernández, Y., Drinan, S. D., Fuxman Bass, J. I., & Benson, G. (2021). Genome-wide characterization of human minisatellite VNTRs: population-specific alleles and gene expression differences. *Nucleic Acids Research*, *49*(8), 4308–4324.

Eustermann, S., Yang, J.-C., Law, M. J., et al. (2011). Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. *Nature Structural & Molecular Biology*, *18*(7), 777–782.

Figueiredo, J., Peitinho, D., Paula Cabral Campello, M., et al. (2022). *Screening of Scaffolds for the Design of G-Quadruplex Ligands*. *12*, 2170.

Fioriniello, S., Marano, D., Fiorillo, F., D'esposito, M., & Ragione, F. della. (2020). Epigenetic Factors that Control Pericentric Heterochromatin Organization in Mammals. *Genes 2020, Vol. 11, Page 595*, *11*(6), 595.

Frangoul, H., Altshuler, D., Cappellini, M. D., et al. (2021). CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassaemia. *New England Journal of Medicine*, *384*(3), 252–260.

Fraser, P. J., & Curtis, P. J. (1987). *Specific pattern of gene expression during induction of mouse erythroleukemia cells*.

Garrick, D., de Gobbi, M., Samara, V., et al. (2008). The role of the polycomb complex in silencing alpha-globin gene expression in nonerythroid cells. *Blood*, *112*(9), 3889–3899.

Garrick, D., Samara, V., McDowell, T. L., et al. (2004). A conserved truncated isoform of the ATR-X syndrome protein lacking the SWI/SNF-homology domain. *Gene*, *326*, 23–34.

Garrick, D., Sharpe, J. A., Arkell, R., et al. (2006). Loss of Atrx Affects Trophoblast Development and the Pattern of X-Inactivation in Extraembryonic Tissues. *PLOS Genetics*, *2*(4), e58.

Gazda, H. T., Sheen, M. R., Vlachos, A., et al. (2008). Ribosomal Protein L5 and L11 Mutations Are Associated with Cleft Palate and Abnormal Thumbs in Diamond-Blackfan Anemia Patients. *The American Journal of Human Genetics*, *83*(6), 769–780.

Geurts, A. M., Cost, G. J., Freyvert, Y., et al. (2009). Knockout rats via embryo microinjection of zinc-finger nucleases. *Science*, *325*(5939), 433.

Giarratana, M.-C., Kobari, L., Lapillonne, H., et al. (2005). Ex vivo generation of fully mature human red blood cells from hematopoietic stem cells. *Nature Biotechnology*, *23*(1), 69–74.

Gibbons, R. (2006). Alpha thalassaemia-mental retardation, X linked. *Orphanet Journal of Rare Diseases*, *1*, 15.

Gibbons, R. J., Brueton, L., Buckle, V. J., et al. (1995). Clinical and hematologic aspects of the X-linked α-thalassaemia/mental retardation syndrome (ATR-X). *American Journal of Medical Genetics*, *55*(3), 288–299.

Gibbons, R. J., & Higgs, D. R. (2000). *Molecular – Clinical Spectrum of the ATR-X Syndrome*. *212*, 204–212.

Gibbons, R. J., McDowell, T. L., Raman, S., et al. (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nature Genetics*, *24*(4), 368–371.

Gibbons, R. J., Pellagatti, A., Garrick, D., et al. (2003). Identification of acquired somatic mutations in the gene encoding chromatin-remodeling factor ATRX in the α-thalassaemia myelodysplasia syndrome (ATMDS). *Nature Genetics*, *34*(4), 446–449.

Gibbons, R. J., Picketts, D. J., Villard, L., & Higgs, D. R. (1995). Mutations in a putative global transcriptional regulator cause X-linked mental retardation with α-thalassaemia (ATR-X syndrome). *Cell*, *80*(6), 837–845.

Gibbons, R. J., Suthers, G. K., Wilkie, A. O., Buckle, V. J., & Higgs, D. R. (1992). X-linked alpha-thalassaemia/mental retardation (ATR-X) syndrome: localization to Xq12-q21.31 by X inactivation and linkage analysis. *American Journal of Human Genetics*, *51*(5), 1136–1149.

Gibbons, R. J., Wada, T., Fisher, C. A., et al. (2008). Mutations in the chromatin-associated protein ATRX. *Human Mutation*, *29*(6), 796–802.

Gibbons, R. J., Wilkie, A. O. M., Weatherall, D. J., & Higgs, D. R. (1991). A newly defined X linked mental retardation syndrome associated with α thalassaemia. *Journal of Medical Genetics*, *28*(11), 729–733.

Giustacchini, A., Thongjuea, S., Barkas, N., et al. (2017). Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nature Medicine 2017 23:6*, *23*(6), 692–702.

Gnanapragasam, M. N., & Bieker, J. J. (2017). Orchestration of late events in erythropoiesis by KLF1/EKLF. *Current Opinion in Hematology*, *24*(3), 183–190.

Goh, S. H., Lee, Y. T., Bhanu, N. v., et al. (2005). A newly discovered human α-globin gene. *Blood*, *106*(4), 1466–1472.

Goldberg, A. D., Banaszynski, L. A., Noh, K.-M., et al. (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell*, *140*(5), 678–691.

Griffiths, R. E., Kupzig, S., Cogan, N., et al. (2012). Maturing reticulocytes internalize plasma membrane in glycophorin A-containing vesicles that fuse with autophagosomes before exocytosis. *Blood*, *119*(26), 6296–6306.

Grover, A., Sanjuan-Pla, A., Thongjuea, S., et al. (2016). Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. *Nature Communications*, *7*.

Guijas, C., Montenegro-Burke, J. R., Warth, B., Spilker, M. E., & Siuzdak, G. (2018). Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nature Biotechnology*, *36*(4), 316–320.

Gundry, M. C., Brunetti, L., Lin, A., et al. (2016). Highly Efficient Genome Editing of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9. *Cell Reports*, *17*(5), 1453–1461.

Hall, A., Moore, A. K., Hernandez, D. G., et al. (2020). A SINE-VNTR- Alu in the LRIG2 Promoter Is Associated with Gene Expression at the Locus. *International Journal of Molecular Sciences*, *21*(22), 1–13.

Han, B., Meng, X., Wu, P., et al. (2020). ATRX/EZH2 complex epigenetically regulates FADD/PARP1 axis, contributing to TMZ resistance in glioma. *Theranostics*, *10*(7), 3351–3365.

Hay, D., Hughes, J. R., Babbs, C., et al. (2016). Genetic dissection of the α-globin super-enhancer in vivo. *Nature Genetics*, *48*(8), 895–903.

Hayashi, H., Kubo, Y., Izumida, M., & Matsuyama, T. (2020). Efficient viral delivery of Cas9 into human safe harbor. *Scientific Reports 2020 10:1*, *10*(1), 1–14.

He, X., Yuan, J., & Wang, Y. (2021). G3BP1 binds to guanine quadruplexes in mRNAs to modulate their stabilities. *Nucleic Acids Research*, *49*(19), 11323–11336.

Heaphy, C. M., de Wilde, R. F., Jiao, Y., et al. (2011). Altered telomeres in tumors with ATRX and DAXX mutations. *Science (New York, N.Y.)*, *333*(6041), 425.

Helming, K. C., Wang, X., & Roberts, C. W. M. (2014). Vulnerabilities of mutant SWI/SNF complexes in cancer. *Cancer Cell*, *26*(3), 309.

Higgs, D. R., Engel, J. D., & Stamatoyannopoulos, G. (2012). Thalassaemia. *Lancet (London, England)*, *379*(9813), 373–383.

Higgs, D. R., Vickers, M. A., Wilkie, A. O. M., et al. (1989). A review of the molecular genetics of the human alpha-globin gene cluster. *Blood*, *73*(5), 1081–1104.

Higgs, D. R., & Weatherall, D. J. (2009). The alpha thalassaemias. *Cellular and Molecular Life Sciences : CMLS*, *66*(7), 1154–1162.

Hoelper, D., Huang, H., Jain, A. Y., Patel, D. J., & Lewis, P. W. (2017). Structural and mechanistic insights into ATRX-dependent and -independent functions of the histone chaperone DAXX. *Nature Communications*, *8*(1).

Hu, K., Li, Y., Wu, W., et al. (2020). ATM-Dependent Recruitment of BRD7 is required for Transcriptional Repression and DNA Repair at DNA Breaks Flanking Transcriptional Active Regions. *Advanced Science (Weinheim, Baden-Wurttemberg, Germany)*, *7*(20).

Huang, P., Zhao, Y., Zhong, J., et al. (2020). Putative regulators for the continuum of erythroid differentiation revealed by single-cell transcriptome of human BM and UCB cells. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(23), 12868–12876.

Huang, Y. gang, Li, D., Wang, L., Su, X. min, & Tang, X. bin. (2022). CENPF/CDK1 signaling pathway enhances the progression of adrenocortical carcinoma by regulating the G2/M-phase cell cycle. *Journal of Translational Medicine*, *20*(1).

Hughes, J. R., Roberts, N., Mcgowan, S., et al. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics*, *46*(2), 205–212.

Huh, M. S., Ivanochko, D., Hashem, L. E., et al. (2016). Stalled replication forks within heterochromatin require ATRX for protection. *Cell Death & Disease 2016 7:5*, *7*(5), e2220–e2220.

Huh, M. S., O'Dea, T. P., Ouazia, D., et al. (2012). Compromised genomic integrity impedes muscle growth after Atrx inactivation. *The Journal of Clinical Investigation*, *122*(12), 4412.

Iskander, D., Psaila, B., Gerrard, G., et al. (2015). Elucidation of the EP defect in Diamond-Blackfan anemia by characterization and prospective isolation of human EPs. *Blood*, *125*(16), 2553–2557.

Iwase, S., Xiang, B., Ghosh, S., et al. (2011). ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. *Nature Structural & Molecular Biology 2011 18:7*, *18*(7), 769–776.

Jeziorska, D. M., Murray, R. J. S., de Gobbi, M., et al. (2017). DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(36), E7526–E7535.

Ji, J., Quindipan, C., Parham, D., et al. (2017). Inherited germline ATRX mutation in two brothers with ATR-X syndrome and osteosarcoma. *American Journal of Medical Genetics. Part A*, *173*(5), 1390–1395.

Jiang, F., Xu, L. L., Tang, X. W., & Li, D. Z. (2021). A novel deletion of the major regulatory element flanking the α-globin gene cluster as a cause of α0-thalassaemia. *International Journal of Laboratory Hematology*, *43*(4), O190–O192.

Jiao, Y., Killela, P. J., Reitman, Z. J., et al. (2012). Frequent ATRX, CIC, FUBP1 and IDH1 mutations refine the classification of malignant gliomas. *Oncotarget*, *3*(7), 709–722.

Jiao, Y., Shi, C., Edil, B. H., et al. (2011). DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science*, *331*(6021), 1199–1203.

Jinek, M., Chylinski, K., Fonfara, I., et al. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, *337*(6096), 816–821.

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics 2012 13:7*, *13*(7), 484–492.

Joshi, P., Carrington, E. A., Wang, L., et al. (2008). Dominant Alleles Identify SET Domain Residues Required for Histone Methyltransferase of Polycomb Repressive Complex 2 *. *Journal of Biological Chemistry*, *283*(41), 27757–27766.

Juhász, S., Elbakry, A., Mathes, A., & Löbrich, M. (2018). ATRX Promotes DNA Repair Synthesis and Sister Chromatid Exchange during Homologous Recombination. *Molecular Cell*, *71*(1), 11-24.e7.

Kanemaki, M. T. (2022). Ligand-induced degrons for studying nuclear functions. *Current Opinion in Cell Biology*, *74*, 29–36.

Kannan, K., Inagaki, A., Silber, J., et al. (2012). Whole-exome sequencing identifies ATRX mutation as a key molecular determinant in lower-grade glioma. *Oncotarget*, *3*(10), 1194–1203.

Kelliher, J., Ghosal, G., & Leung, J. W. C. (2022). New answers to the old RIDDLE: RNF168 and the DNA damage response pathway. *The FEBS Journal*, *289*(9), 2467–2480.

Kelliher, J. L., West, K. L., Gong, Q., & Leung, J. W. C. (2020). Histone H2A variants alpha1-extension helix directs RNF168-mediated ubiquitination. *Nature Communications*, *11*(1).

Kernohan, K. D., Jiang, Y., Tremblay, D. C., et al. (2010). ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain. *Developmental Cell*, *18*(2), 191–202.

Kernohan, K. D., Vernimmen, D., Gloor, G. B., & Bérubé, N. G. (2014). Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome

density, CTCF binding and chromatin looping. *Nucleic Acids Research*, *42*(13), 8356–8368.

Khajuria, R. K., Munschauer, M., Ulirsch, J. C., et al. (2018). Ribosome Levels Selectively Regulate Translation and Lineage Commitment in Human Hematopoiesis. *Cell*, *173*(1), 90-103.e19.

Khurana, S., Kruhlak, M. J., Kim, J., et al. (2014). A macrohistone variant links dynamic chromatin compaction to BRCA1-dependent genome maintenance. *Cell Reports*, *8*(4), 1049–1062.

Kim, J. M., Shin, Y., Lee, S., et al. (2018). MacroH2A1.2 inhibits prostate cancer-induced osteoclastogenesis through cooperation with HP1α and H1.2. *Oncogene 2018 37:43*, *37*(43), 5749–5765.

Kim, J., Shin, Y., Lee, S., et al. (2018). Regulation of Breast Cancer-Induced Osteoclastogenesis by MacroH2A1.2 Involving EZH2-Mediated H3K27me3. *Cell Reports*, *24*(1), 224–237.

Kim, J., Sturgill, D., Sebastian, R., et al. (2018). Replication Stress Shapes a Protective Chromatin Environment across Fragile Genomic Regions. *Molecular Cell*, *69*(1), 36-47.e7.

Kim, J., Sturgill, D., Tran, A. D., Sinclair, D. A., & Oberdoerffer, P. (2016). Controlled DNA double-strand break induction in mice reveals post-damage transcriptome stability. *Nucleic Acids Research*, *44*(7), e64.

Kim, J., Sun, C., Tran, A. D., et al. (2019). The macroH2A1.2 histone variant links ATRX loss to alternative telomere lengthening. *Nature Structural and Molecular Biology*, *26*(3), 213–219.

King, A. J., Songdej, D., Downes, D. J., et al. (2021). Reactivation of a developmentally silenced embryonic globin gene. *Nature Communications 2021 12:1*, *12*(1), 1–15.

King, R., Lin, Z., Balbin-Cuesta, G., et al. (2021). SEC23A rescues SEC23B-deficient congenital dyserythropoietic anemia type II. *Science Advances*, *7*(48).

Klein, A. M., Mazutis, L., Akartuna, I., et al. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, *161*(5), 1187–1201.

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., et al. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, *529*(7587), 490–495.

Kobari, L., Yates, F., Oudrhiri, N., et al. (2012). Human induced pluripotent stem cells can reach complete terminal maturation: in vivo and in vitro evidence in the erythropoietic differentiation model. *Haematologica*, *97*(12), 1795–1803.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420–424.

Kondo, A., Fujiwara, T., Okitsu, Y., et al. (2016). Identification of a novel putative mitochondrial protein FAM210B associated with erythroid differentiation. *International Journal of Hematology*, *103*(4), 387–395.

Koschmann, C., Calinescu, A.-A., Nunez, F. J., et al. (2016). ATRX loss promotes tumor growth and impairs nonhomologous end joining DNA repair in glioma. *Science Translational Medicine*, *8*(328), 328ra28.

Koury, M. J., Sawyer, S. T., & Brandt, S. J. (2002). New insights into erythropoiesis. *Current Opinion in Hematology*, *9*(2), 93–100.

Kumar, R. P., Krishnan, J., Pratap Singh, N., Singh, L., & Mishra, R. K. (2013). GATA simple sequence repeats function as enhancer blocker boundaries. *Nature Communications*, *4*(May), 1844–1847.

Kurita, R., Suda, N., Sudo, K., et al. (2013). Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PloS One*, *8*(3), e59890.

Lagali, P. S., Medina, C. F., Zhao, B. Y. H., et al. (2016). Retinal interneuron survival requires non-cell-autonomous Atrx activity. *Human Molecular Genetics*, *25*(21), 4787–4803.

Lattanzi, A., Meneghini, V., Pavani, G., et al. (2019). Optimization of CRISPR/Cas9 Delivery to Human Hematopoietic Stem and Progenitor Cells for Therapeutic Genomic Rearrangements. *Molecular Therapy*, *27*(1), 137–150.

Laurent, B., Randrianarison-Huetz, V., Marćhal, V., et al. (2010). High-mobility group protein HMGB2 regulates human erythroid differentiation through trans-activation of GFI1B transcription. *Blood*, *115*(3), 687–695.

Law, M. J., Lower, K. M., Voon, H. P. J., et al. (2010). ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell*, *143*(3), 367–378.

le Goff, S., Boussaid, I., Floquet, C., et al. (2021). p53 activation during ribosome biogenesis regulates normal erythroid differentiation. *Blood*, *137*(1), 89–102.

Lechner, M. S., Schultz, D. C., Negorev, D., Maul, G. G., & Rauscher, F. J. (2005). The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. *Biochemical and Biophysical Research Communications*, *331*(4), 929–937.

Lee, N. G., Hong, Y. K., Yu, S. Y., et al. (2007). dXNP, a Drosophila homolog of XNP/ATRX, induces apoptosis via Jun-N-terminal kinase activation. *FEBS Letters*, *581*(14), 2625–2632.

Lee, W. T. C., Yin, Y., Morten, M. J., et al. (2021). Single-molecule imaging reveals replication fork coupled formation of G-quadruplex structures hinders local replication stress signaling. *Nature Communications 2021 12:1*, *12*(1), 1–14.

Leung, J. W.-C., Ghosal, G., Wang, W., et al. (2013). Alpha thalassaemia/mental retardation syndrome X-linked gene product ATRX is required for proper replication restart and cellular resistance to replication stress. *The Journal of Biological Chemistry*, *288*(9), 6342–6350.

Levy, M. A., Kernohan, K. D., Jiang, Y., & Bérubé, N. G. (2015). ATRX promotes gene expression by facilitating transcriptional elongation through guanine-rich coding regions. *Human Molecular Genetics*, *24*(7), 1824–1835.

Lewis, P. W., Elsaesser, S. J., Noh, K. M., Stadler, S. C., & Allis, C. D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(32), 14075–14080.

Lewis, P. W., Müller, M. M., Koletsky, M. S., et al. (2013). Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma. *Science (New York, N.Y.)*, *340*(6134), 857.

Li, J., Wang, L., Guo, H., et al. (2017). Targeted sequencing and functional analysis reveal brain-size-related genes and their networks in autism spectrum disorders. *Molecular Psychiatry 2017 22:9*, *22*(9), 1282–1290.

Li, S., Prasanna, X., Salo, V. T., Vattulainen, I., & Ikonen, E. (2019). An efficient auxin-inducible degron system with low basal degradation in human cells. *Nature Methods 2019 16:9*, *16*(9), 866–869.

Li, T., Huang, S., Jiang, W. Z., et al. (2011). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Research*, *39*(1), 359–372.

Li, X., Chen, M., Liu, B., et al. (2021). Transcriptional silencing of fetal hemoglobin expression by NonO. *Nucleic Acids Research*, *49*(17), 9711–9723.

Li, Y., Syed, J., Suzuki, Y., et al. (2016). Effect of ATRX and G-Quadruplex Formation by the VNTR Sequence on α-Globin Gene Expression. *Chembiochem : A European Journal of Chemical Biology*, *17*(10), 928–935.

Liang, J., Zhao, H., Diplas, B. H., et al. (2020). Genome-wide CRISPR-Cas9 screen reveals selective vulnerability of ATRX-mutant cancers to WEE1 inhibition. *Cancer Research*, *80*(3), 510–523.

Liu, L. Y., Ma, T. Z., Zeng, Y. L., Liu, W., & Mao, Z. W. (2022). Structural Basis of Pyridostatin and Its Derivatives Specifically Binding to G-Quadruplexes. *Journal of the American Chemical Society*, *144*(26), 11878–11887.

Liu, M., Rehman, S., Tang, X., et al. (2019). Methodologies for improving HDR efficiency. *Frontiers in Genetics*, *10*(JAN), 691.

Liu, X. Y., Gerges, N., Korshunov, A., et al. (2012). Frequent ATRX mutations and loss of expression in adult diffuse astrocytic tumors carrying IDH1/IDH2 and TP53 mutations. *Acta Neuropathologica*, *124*(5), 615–625.

López-Garrido, M.-P., Carrascosa-Romero, M.-C., Montero-Hernández, M., Ruiz-Almansa, J., & Sánchez-Sánchez, F. (2022). Brief Report: Evidence of

Autism Spectrum Disorder Caused by a Mutation in ATRX Gene: A Case Report. *Journal of Autism and Developmental Disorders*, 1–10.

Lossi, A. M., Millian, J. M., Villard, L., et al. (1999). Mutation of the XNP/ATR-X gene in a family with severe mental retardation, spastic paraplegia and skewed pattern of X inactivation: Demonstration that the mutation is involved in the inactivalion bias [2]. *American Journal of Human Genetics*, *65*(2), 558–562.

Louis, D. N., Perry, A., Reifenberger, G., et al. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathologica*, *131*(6), 803–820.

Lozzio, B. B., & Lozzio, C. B. (1979). Properties and usefulness of the original K-562 human myelogenous leukemia cell line. *Leukemia Research*, *3*(6), 363–370.

Lozzio, C. B., & Lozzio, B. B. (1975). Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood*, *45*(3), 321–334.

Lu, D., Xu, Z., Peng, Z., et al. (2022). Induction of Fetal Hemoglobin by Introducing Natural Hereditary Persistence of Fetal Hemoglobin Mutations in the γ-Globin Gene Promoters for Genome Editing Therapies for β-Thalassaemia. *Frontiers in Genetics*, *13*.

Ludwig, L. S., Lareau, C. A., Bao, E. L., et al. (2019). Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis. *Cell Reports*, *27*(11), 3228-3240.e7.

Lukas, C., Savic, V., Bekker-Jensen, S., et al. (2011). 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. *Nature Cell Biology*, *13*(3), 243–253.

Macosko, E. Z., Basu, A., Satija, R., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, *161*(5), 1202–1214.

Maison, C., & Almouzni, G. (2004). HP1 and the dynamics of heterochromatin maintenance. *Nature Reviews Molecular Cell Biology 2004 5:4*, *5*(4), 296–305.

Marano, D., Fioriniello, S., Fiorillo, F., et al. (2019). ATRX Contributes to MeCP2-Mediated Pericentric Heterochromatin Organization during Neural Differentiation. *International Journal of Molecular Sciences 2019, Vol. 20, Page 5371*, *20*(21), 5371.

Maréchal, A., & Zou, L. (2013). DNA Damage Sensing by the ATM and ATR Kinases. *Cold Spring Harbor Perspectives in Biology*, *5*(9).

Marinoni, I., Kurrer, A. S., Vassella, E., et al. (2014). Loss of DAXX and ATRX are associated with chromosome instability and reduced survival of patients with pancreatic neuroendocrine tumors. *Gastroenterology*, *146*(2), 453-60.e5.

Marshall, J. N. G., Lopez, A. I., Pfaff, A. L., et al. (2021). Variable number tandem repeats – Their emerging role in sickness and health. *Experimental Biology and Medicine*, *246*(12), 1368.

Martin-Kenny, N., Martin-Kenny, N., Martin-Kenny, N., et al. (2020). Effects of a postnatal Atrx conditional knockout in neurons on autism-like behaviours in male and female mice. *Journal of Neurodevelopmental Disorders*, *12*(1), 1–11.

Martino, L., Pagano, B., Fotticchia, I., Neidle, S., & Giancola, C. (2009). Shedding light on the interaction between TMPyP4 and human telomeric quadruplexes. *The Journal of Physical Chemistry. B*, *113*(44), 14779–14786.

Masliah-Planchon, J., Lévy, D., Héron, D., et al. (2018). Does ATRX germline variation predispose to osteosarcoma? Three additional cases of osteosarcoma in two ATR-X syndrome patients. *European Journal of Human Genetics 2018 26:8*, *26*(8), 1217–1221.

Masud, T., Soong, C., Xu, H., et al. (2021). Ubiquitin-mediated DNA damage response is synthetic lethal with G-quadruplex stabilizer CX-5461. *Scientific Reports 2021 11:1*, *11*(1), 1–20.

Masuda, T., Wang, X., Maeda, M., et al. (2016). Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. *Science (New York, N.Y.)*, *351*(6270), 285–289.

McDowell, T. L., Gibbons, R. J., Sutherland, H., et al. (1999). Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and

the short arms of acrocentric chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(24), 13983–13988.

Medina, C. F., Mazerolle, C., Wang, Y., et al. (2009). Altered visual function and interneuron survival in Atrx knockout mice: inference for the human syndrome. *Human Molecular Genetics*, *18*(5), 966–977.

Meissner, A., Mikkelsen, T. S., Gu, H., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, *454*(7205), 766–770.

Mendoza, O., Bourdoncle, A., Boulé, J. B., Brosh, R. M., & Mergny, J. L. (2016). G-quadruplexes and helicases. *Nucleic Acids Research*, *44*(5), 1989.

Mettananda, S., Clark, K., Fisher, C. A., et al. (2018). Phenotypic and molecular characterization of a serum-free miniature erythroid differentiation system suitable for high-throughput screening and single-cell assays. *Experimental Hematology*, *60*, 10–20.

Mettananda, S., Fisher, C. A., Hay, D., et al. (2017). Editing an α-globin enhancer in primary human hematopoietic stem cells as a treatment for β-Thalassaemia. *Nature Communications*, *8*(1).

Mettananda, S., Gibbons, R. J., & Higgs, D. R. (2015). α-Globin as a molecular target in the treatment of β-thalassaemia. *Blood*, *125*(24), 3694–3701.

Migliaccio, G., di Pietro, R., di Giacomo, V., et al. (2002). In vitro mass production of human erythroid cells from the blood of normal donors and of thalassemic patients. *Blood Cells, Molecules & Diseases*, *28*(2), 169–180.

Militello, R., & Colombo, M. I. (2013). Small GTPases as regulators of cell division. In *Communicative and Integrative Biology* (Vol. 6, Issue 5).

Mitson, M., Kelley, L. A., Sternberg, M. J. E., Higgs, D. R., & Gibbons, R. J. (2011). Functional significance of mutations in the Snf2 domain of ATRX. *Human Molecular Genetics*, *20*(13), 2603–2610.

Moir-Meyer, G., Cheong, P. L., Olijnik, A. A., et al. (2018). Robust crispr/cas9 genome editing of the hudep-2 erythroid precursor line using plasmids and single-stranded oligonucleotide donors. *Methods and Protocols*, *1*(3), 1–17.

Molenaar, J. J., Koster, J., Zwijnenburg, D. A., et al. (2012). Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. *Nature 2012 483:7391*, *483*(7391), 589–593.

Moniz, H., Gastou, M., Leblanc, T., et al. (2012). Primary hematopoietic cells from DBA patients with mutations in RPL11 and RPS19 genes exhibit distinct erythroid phenotype in vitro. *Cell Death & Disease*, *3*(7).

Morozov, V. M., Giovinazzi, S., & Ishov, A. M. (2017). CENP-B protects centromere chromatin integrity by facilitating histone deposition via the H3.3-specific chaperone Daxx. *Epigenetics & Chromatin*, *10*(1).

Moruno-Manchon, J. F., Koellhoffer, E. C., Gopakumar, J., et al. (2017). The G-quadruplex DNA stabilizing drug pyridostatin promotes DNA damage and downregulates transcription of Brca1 in neurons. *Aging*, *9*(9), 1957–1970.

Moscou, M. J., & Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science*, *326*(5959), 1501.

Mou, L., Zhang, Q., Wang, Y., et al. (2013). Identification of Ube2b as a novel target of androgen receptor in mouse sertoli cells. *Biology of Reproduction*, *89*(2).

Moyer, M. R. (2020). *Inducible Protein Depletion of SETD5, a Chromatin Regulator Implicated in Autism Spectrum Disorders* [UNIVERSITY OF CALIFORNIA SAN DIEGO].

Muers, M. R., Sharpe, J. A., Garrick, D., et al. (2007). Defining the Cause of Skewed X-Chromosome Inactivation in X-Linked Mental Retardation by Use of a Mouse Model. *The American Journal of Human Genetics*, *80*(6), 1138–1149.

Nabet, B., Ferguson, F. M., Seong, B. K. A., et al. (2020). Rapid and direct control of target protein levels with VHL-recruiting dTAG molecules. *Nature Communications*, *11*(1).

Nabet, B., Roberts, J. M., Buckley, D. L., et al. (2018). The dTAG system for immediate and target-specific protein degradation. *Nature Chemical Biology*, *14*(5), 431–441.

Nan, X., Hou, J., Maclean, A., et al. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental

retardation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(8), 2709–2714.

Narvi, E., Jaakkola, K., Winsel, S., et al. (2013). Altered TUBB3 expression contributes to the epothilone response of mitotic cells. *British Journal of Cancer 2013 108:1*, *108*(1), 82–90.

Nguyen, D. T., Voon, H. P. J., Xella, B., et al. (2017). The chromatin remodelling factor ATRX suppresses R-loops in transcribed telomeric repeats . *EMBO Reports*, *18*(6), 914–928.

Nilsson, R., Schultz, I. J., Pierce, E. L., et al. (2009). Discovery of Genes Essential for Heme Biosynthesis through Large-Scale Gene Expression Analysis. *Cell Metabolism*, *10*(2), 119–130.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., & Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Methods 2009 6:12*, *6*(12), 917–922.

Nogami, T., Beppu, H., Tokoro, T., et al. (2011). Reduced expression of the ATRX gene, a chromatin-remodeling factor, causes hippocampal dysfunction in mice. *Hippocampus*, *21*(6), 678–687.

O'Brien, K. A., Farrar, J. E., Vlachos, A., et al. (2017). Molecular convergence in ex vivo models of Diamond-Blackfan anemia. *Blood*, *129*(23), 3111–3120.

Okamoto, S., Amaishi, Y., Maki, I., Enoki, T., & Mineno, J. (2019). Highly efficient genome editing for single-base substitutions using optimized ssODNs with Cas9-RNPs. *Scientific Reports*, *9*(1).

Olivier, E. N., Marenah, L., McCahill, A., et al. (2016). High-Efficiency Serum-Free Feeder-Free Erythroid Differentiation of Human Pluripotent Stem Cells Using Small Molecules. *Stem Cells Translational Medicine*, *5*(10), 1394–1405.

Oppel, F., Tao, T., Shi, H., et al. (2019). Loss of atrx cooperates with p53-deficiency to promote the development of sarcomas and other malignancies. *PLoS Genetics*, *15*(4).

Örd, T., Puurand, T., Örd, D., et al. (2020). A human-specific VNTR in the TRIB3 promoter causes gene expression variation between individuals. *PLoS Genetics*, *16*(8).

Orkin, S. H. (2000). Diversification of haematopoietic stem cells to specific lineages. *Nature Reviews Genetics 2000 1:1*, *1*(1), 57–64.

Osborn, A. G., Louis, D. N., Poussaint, T. Y., Linscott, L. L., & Salzman, K. L. (2022). The 2021 World Health Organization Classification of Tumors of the Central Nervous System: What Neuroradiologists Need to Know. *American Journal of Neuroradiology*.

P, H., XQD, W., F, F., et al. (2021). 3'HS1 CTCF binding site in human β-globin locus regulates fetal hemoglobin expression. *ELife*, *10*.

Palii, C. G., Cheng, Q., Gillespie, M. A., et al. (2019). Single-Cell Proteomics Reveal that Quantitative Changes in Co-expressed Lineage-Specific Transcription Factors Determine Cell Fate. *Cell Stem Cell*, *24*(5), 812-820.e5.

Pandey, S., Simmons, G. E., Malyarchuk, S., Calhoun, T. N., & Pruitt, K. (2015). A novel MeCP2 acetylation site regulates interaction with ATRX and HDAC1. *Genes & Cancer*, *6*(9–10), 408–421.

Papadopoulou, C., Guilbaud, G., Schiavone, D., & Sale, J. E. (2015). Nucleotide Pool Depletion Induces G-Quadruplex-Dependent Perturbation of Gene Expression. *Cell Reports*, *13*(11), 2491–2503.

Park, J., Lee, H., Han, N., et al. (2018). Long non-coding RNA ChRO1 facilitates ATRX/DAXX-dependent H3.3 deposition for transcription-associated heterochromatin reorganization. *Nucleic Acids Research*, *46*(22), 11759–11775.

Picelli, S., Faridani, O. R., Björklund, Å. K., et al. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols*, *9*(1), 171–181.

Picketts, D. J., Tastan, A. O., Higgs, D. R., & Gibbons, R. J. (1998). Comparison of the human and murine ATRX gene identifies highly conserved, functionally important domains. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*, *9*(5), 400–403.

Pladevall-Morera, D., Munk, S., Ingham, A., et al. (2019). Proteomic characterization of chromosomal common fragile site (CFS)-associated proteins uncovers ATRX as a regulator of CFS stability. *Nucleic Acids Research*, *47*(15), 8004–8018.

Poodt, J., Martens, H. A. M., Walsh, I. B. B., Felix-Schollaart, B., & Hermans, M. H. A. (2006). A newly identified deletion of 970 bp at the alpha-globin locus that removes the promoter region of the alpha1 gene. *Hemoglobin*, *30*(4), 471–477.

Porteous, M. E. M., & Burn, J. (1990). Unknown syndrome. A possible new X linked retardation syndrome: Dysmorphic facies, microcephaly, hypotonia, and small genitalia. *Journal of Medical Genetics*, *27*(5), 339–340.

Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading, England)*, *151*(Pt 3), 653–663.

Probst, A. v., & Almouzni, G. (2008). Pericentric heterochromatin: dynamic organization during early development in mammals. *Differentiation*, *76*(1), 15–23.

Psaila, B., Barkas, N., Iskander, D., et al. (2016). Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. *Genome Biology*, *17*(1).

Qi, L. S., Larson, M. H., Gilbert, L. A., et al. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173–1183.

Qin, T., Mullan, B., Ravindran, R., et al. (2022). ATRX loss in glioma results in dysregulation of cell-cycle phase transition and ATM inhibitor radio-sensitization. *Cell Reports*, *38*(2).

Raghunandan, M., Yeo, J. E., Walter, R., et al. (2020). Functional cross talk between the Fanconi anemia and ATRX/DAXX histone chaperone pathways promotes replication fork recovery. *Human Molecular Genetics*, *29*(7), 1083–1095.

Ratnakumar, K., Duarte, L. F., LeRoy, G., et al. (2012). ATRX-mediated chromatin association of histone variant macroH2A1 regulates α-globin expression. *Genes and Development*, *26*(5), 433–438.

Rees, H. A., & Liu, D. R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. *Nature Reviews. Genetics*, *19*(12), 770–788.

Rémy, S., Tesson, L., Ménoret, S., et al. (2010). Zinc-finger nucleases: A powerful tool for genetic engineering of animals. *Transgenic Research*, *19*(3), 363–371.

Ren, W., Medeiros, N., Warneford-Thomson, R., et al. (2020). Disruption of ATRX-RNA interactions uncovers roles in ATRX localization and PRC2 function. *Nature Communications*, *11*(1).

Ritchie, K., Seah, C., Moulin, J., et al. (2008). Loss of ATRX leads to chromosome cohesion and congression defects. *The Journal of Cell Biology*, *180*(2), 315.

Rodriguez, R., Miller, K. M., Forment, J. v., et al. (2012). Small-molecule–induced DNA damage identifies alternative DNA structures in human genes. *Nature Chemical Biology 2012 8:3*, *8*(3), 301–310.

Rodriguez-Meira, A., Buck, G., Clark, S. A., et al. (2019). Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *Molecular Cell*, *73*(6), 1292-1305.e8.

Rodriguez-Meira, A., O'Sullivan, J., Rahman, H., & Mead, A. J. (2020). TARGET-Seq: A Protocol for High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *STAR Protocols*, *1*(3), 100125.

Rong, G., Pan, Z., Ding, M., & Wang, L. (2022). RNF168 suppresses the cancer stem cell-like traits of nonsmall cell lung cancer cells by mediating RhoC ubiquitination. *Environmental Toxicology*, *37*(3), 603–611.

Ronzoni, L., Bonara, P., Rusconi, D., et al. (2008). Erythroid differentiation and maturation from peripheral CD34+ cells in liquid culture: cellular and molecular characterization. *Blood Cells, Molecules & Diseases*, *40*(2), 148–155.

Rossmann, M. P., Hoi, K., Chan, V., et al. (2021). Cell-specific transcriptional control of mitochondrial metabolism by TIF1g drives erythropoiesis. *Science*, *372*(6543), 716–721.

Ruggiero, E., & Richter, S. N. (2018). G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Research*, *46*(7), 3270.

Rutherford, T., Clegg, J. B., Higgs, D. R., et al. (1981). Embryonic erythroid differentiation in the human leukemic cell line K562. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(1), 348–352.

Sadic, D., Schmidt, K., Groh, S., et al. (2015). Atrx promotes heterochromatin formation at retrotransposons. *EMBO Reports*, *16*(7), 836–850.

Sainz de la Maza, D., Hof-Michel, S., Phillimore, L., Bökel, C., & Amoyel, M. (2022). Cell-cycle exit and stem cell differentiation are coupled through regulation of mitochondrial activity in the Drosophila testis. *Cell Reports*, *39*(6), 110774.

Sarkies, P., Reams, C., Simpson, L. J., & Sale, J. E. (2010). Epigenetic instability due to defective replication of structured DNA. *Molecular Cell*, *40*(5), 703–713.

Sarma, K., Cifuentes-Rojas, C., Ergun, A., et al. (2014). ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. *Cell*, *159*(4), 869–883.

Schaeffer, E. K., West, R. J., Conine, S. J., & Lowrey, C. H. (2014). Multiple physical stresses induce γ-globin gene expression and fetal hemoglobin production in erythroid cells. *Blood Cells, Molecules & Diseases*, *52*(4), 214–224.

Schenkel, L. C., Kernohan, K. D., McBride, A., et al. (2017). Identification of epigenetic signature associated with alpha thalassaemia/mental retardation X-linked syndrome. *Epigenetics and Chromatin*, *10*(1), 1–11.

Schiavone, D., Guilbaud, G., Murat, P., et al. (2014). Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *The EMBO Journal*, *33*(21), 2507–2520.

Schoelz, J. M., & Riddle, N. C. (2022). Functions of HP1 proteins in transcriptional regulation. *Epigenetics & Chromatin 2022 15:1*, *15*(1), 1–15.

Schwartzentruber, J., Korshunov, A., Liu, X. Y., et al. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature 2012 482:7384*, *482*(7384), 226–231.

Scott, C., Downes, D. J., Brown, J. M., et al. (2020). Recapitulation of erythropoiesis in congenital dyserythropoietic anemia type I (CDA-I) identifies defects in differentiation and nucleolar abnormalities. *Haematologica*, *106*(11), 2960–2970.

Scott, W. A., Dhanji, E. Z., Dyakov, B. J. A., et al. (2021). ATRX proximal protein associations boast roles beyond histone deposition. *PLoS Genetics*, *17*(11).

Seah, C., Levy, M. A., Jiang, Y., et al. (2008). Neuronal death resulting from targeted disruption of the Snf2 protein ATRX is mediated by p53. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(47), 12570–12580.

Sebastian, R., Hosogane, E. K., Sun, E. G., et al. (2020). Epigenetic Regulation of DNA Repair Pathway Choice by MacroH2A1 Splice Variants Ensures Genome Stability. *Molecular Cell*, *79*(5), 836-845.e7.

Shalem, O., Sanjana, N. E., Hartenian, E., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (New York, N.Y.)*, *343*(6166), 84–87.

Shanbhag, N. M., Rafalska-Metcalf, I. U., Balane-Bolivar, C., Janicki, S. M., & Greenberg, R. A. (2010). ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. *Cell*, *141*(6), 970–981.

Shaner, N. C., Steinbach, P. A., & Tsien, R. Y. (2005). A guide to choosing fluorescent proteins. *Nature Methods*, *2*(12), 905–909.

Shioda, N., Yabuki, Y., Yamaguchi, K., et al. (2018). Targeting G-quadruplex DNA as cognitive function therapy for ATR-X syndrome. *Nature Medicine*, *24*(6), 802–813.

Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., & Hurley, L. H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(18), 11593–11598.

Solomon, L. A., Li, J. R., Bérubé, N. G., & Beier, F. (2009). Loss of ATRX in chondrocytes has minimal effects on skeletal development. *PloS One*, *4*(9).

Solomon, L. A., Russell, B. A., Watson, L. A., Beier, F., & Bérubé, N. G. (2013). Targeted loss of the ATR-X syndrome protein in the limb mesenchyme of mice causes brachydactyly. *Human Molecular Genetics*, *22*(24), 5015–5025.

Song, S. H., Kim, A., Dale, R., & Dean, A. (2012). Ldb1 regulates carbonic anhydrase 1 during erythroid differentiation. *Biochimica et Biophysica Acta*, *1819*(8), 885–891.

Svensson, V., Vento-Tormo, R., & Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nature Protocols 2018 13:4*, *13*(4), 599–604.

Šviković, S., & Sale, J. E. (2017). The Effects of Replication Stress on S Phase Histone Management and Epigenetic Memory. *Journal of Molecular Biology*, *429*(13), 2011–2029.

Tagami, H., Ray-Gallet, D., Almouzni, G., & Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell*, *116*(1), 51–61.

Taher, A. T., Weatherall, D. J., & Cappellini, M. D. (2018). Thalassaemia. *Lancet (London, England)*, *391*(10116), 155–167.

Tallack, M. R., & Perkins, A. C. (2010). KLF1 directly coordinates almost all aspects of terminal erythroid differentiation. *IUBMB Life*, *62*(12), 886–890.

Tamming, R. J., Dumeaux, V., Jiang, Y., et al. (2020). Atrx Deletion in Neurons Leads to Sexually Dimorphic Dysregulation of miR-137 and Spatial Learning and Memory Deficits. *Cell Reports*, *31*(13).

Tamming, R. J., Siu, J. R., Jiang, Y., et al. (2017). Mosaic expression of Atrx in the mouse central nervous system causes memory deficits. *Disease Models & Mechanisms*, *10*(2), 119.

Tang, J., Wu, S., Liu, H., et al. (2004). A novel transcription regulatory complex containing death domain-associated protein and the ATR-X syndrome protein. *The Journal of Biological Chemistry*, *279*(19), 20369–20377.

Telenius, J., Consortium, T. W., & Hughes, J. R. (2018). NGseqBasic - a single-command UNIX tool for ATAC-seq, DNaseI-seq, Cut-and-Run, and ChIP-seq data mapping, high-resolution visualisation, and quality control. *BioRxiv*, *16*, 393413.

Teng, Y. C., Sundaresan, A., O'Hara, R., et al. (2021). ATRX promotes heterochromatin formation to protect cells from G-quadruplex DNA-mediated stress. *Nature Communications 2021 12:1*, *12*(1), 1–14.

Thompson, L. L., Rutherford, K. A., Lepage, C. C., & McManus, K. J. (2021). The SCF complex is essential to maintain genome and chromosome stability. In *International Journal of Molecular Sciences* (Vol. 22, Issue 16). MDPI AG.

Trakarnsanga, K., Griffiths, R. E., Wilson, M. C., et al. (2017). An immortalized adult human erythroid line facilitates sustainable and scalable generation of functional red cells. *Nature Communications 2017 8:1*, *8*(1), 1–7.

Truch, J., Downes, D. J., Scott, C., et al. (2022). The chromatin remodeller ATRX facilitates diverse nuclear processes, in a stochastic manner, in both heterochromatin and euchromatin. *Nature Communications 2022 13:1*, *13*(1), 1–16.

Truch, J., Telenius, J., Higgs, D. R., & Gibbons, R. J. (2018). How to Tackle Challenging ChIP-Seq, with Long-Range Cross-Linking, Using ATRX as an Example. In *Methods in Molecular Biology* (Vol. 1832, pp. 105–130). Humana Press Inc.

Tufts, B. L., Esbaugh, A., & Lund, S. G. (2003). Comparative physiology and molecular evolution of carbonic anhydrase in the erythrocytes of early vertebrates. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *136*(2), 259–269.

Udugama, M., Sanij, E., Voon, H. P. J., et al. (2018). Ribosomal DNA copy loss and repeat instability in ATRX-mutated cancers. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(18), 4737–4742.

Urnov, F. D., Miller, J. C., Lee, Y. L., et al. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature 2005 435:7042*, *435*(7042), 646–651.

Valle-García, D., Qadeer, Z. A., McHugh, D. S., et al. (2016). ATRX binds to atypical chromatin domains at the 3' exons of zinc finger genes to preserve H3K9me3 enrichment. *Epigenetics*, *11*(6), 398–414.

Villard, L., Fontès, M., & Ewbank, J. J. (1999). Characterization of xnp-1, a Caenorhabditis elegans gene similar to the human XNP/ATR-X gene. *Gene*, *236*(1), 13–19.

Villard, L., Lossi, A. M., Cardoso, C., et al. (1997). Determination of the genomic structure of the XNP/ATRX gene encoding a potential zinc finger helicase. *Genomics*, *43*(2), 149–155.

Vinjamur, D. S., & Bauer, D. E. (2018). Growing and genetically manipulating human umbilical cord blood-derived erythroid progenitor (HUDEP) cell lines. *Methods in Molecular Biology*, *1698*, 275–284.

Viprakasit, V., Harteveld, C. L., Ayyub, H., et al. (2006). A novel deletion causing alpha thalassaemia clarifies the importance of the major human alpha globin regulatory element. *Blood*, *107*(9), 3811–3812.

Viprakasit, V., Kidd, A. M. J., Ayyub, H., et al. (2003). De novo deletion within the telomeric region flanking the human alpha globin locus as a cause of alpha thalassaemia. *British Journal of Haematology*, *120*(5), 867–875.

Vistain, L. F., & Tay, S. (2021). Single-Cell Proteomics. *Trends in Biochemical Sciences*, *46*(8), 661–672.

Voon, H. P. J., Hughes, J. R., Rode, C., et al. (2015). ATRX Plays a Key Role in Maintaining Silencing at Interstitial Heterochromatic Loci and Imprinted Genes. *Cell Reports*, *11*(3), 405–418.

Wada, T., Ban, H., Matsufuji, M., et al. (2013). Neuroradiologic Features in X-linked α-Thalassaemia/Mental Retardation Syndrome. *American Journal of Neuroradiology*, *34*(10), 2034–2038.

Wada, T., Sugie, H., Fukushima, Y., & Saitoh, S. (2005). Non-skewed X-inactivation may cause mental retardation in a female carrier of X-linked α-thalassaemia/mental retardation syndrome (ATR-X): X-inactivation study of nine female carriers of ATR-X. *American Journal of Medical Genetics*, *138 A*(1), 18–20.

Wang, S. R., Min, Y. Q., Wang, J. Q., et al. (2016). A highly conserved G-rich consensus sequence in hepatitis C virus core gene represents a new anti–hepatitis C target. *Science Advances*, *2*(4).

Wang, T., Wei, J. J., Sabatini, D. M., & Lander, E. S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science (New York, N.Y.)*, *343*(6166), 80–84.

Wang, X., He, Y., Zhang, Q., Ren, X., & Zhang, Z. (2021). Direct Comparative Analyses of 10X Genomics Chromium and Smart-seq2. *Genomics, Proteomics & Bioinformatics*, *19*(2), 253–266.

Wang, Y., Yang, J., Wild, A. T., et al. (2019). G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. *Nature Communications*, *10*(1).

Watson, L. A., Solomon, L. A., Li, J. R., et al. (2013). Atrx deficiency induces telomere dysfunction, endocrine defects, and reduced life span. *The Journal of Clinical Investigation*, *123*(5), 2049–2063.

Weatherall, D. J., Higgs, D. R., Bunch, C., et al. (1981). Hemoglobin H Disease and Mental Retardation. *New England Journal of Medicine*, *305*(11), 607–612.

Weissmiller, A. M., Wang, J., Lorey, S. L., et al. (2019). Inhibition of MYC by the SMARCB1 tumor suppressor. *Nature Communications*, *10*(1).

Wilkie, A. O. M., Pembrey, M. E., Gibbons, R. J., et al. (1991). The non-deletion type of α thalassaemia/mental retardation: A recognizable dysmorphic syndrome with X linked inheritance. *Journal of Medical Genetics*, *28*(10), 724.

Wilkie, A. O., Zeitlin, H. C., Lindenbaum, R. H., et al. (1990). Clinical features and molecular analysis of the alpha thalassaemia/mental retardation syndromes. II. Cases without detectable abnormality of the alpha globin complex. *American Journal of Human Genetics*, *46*(6), 1127–1140.

Wong, L. H., McGhie, J. D., Sim, M., et al. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Research*, *20*(3), 351–360.

Wong, L. H., Ren, H., Williams, E., et al. (2009). Histone H3.3 incorporation provides a unique and functionally essential telomeric chromatin in embryonic stem cells. *Genome Research*, *19*(3), 404.

Xu, C., Xu, Y., Gursoy-Yuzugullu, O., & Price, B. D. (2012). The histone variant macroH2A1.1 is recruited to DSBs through a mechanism involving PARP1. *FEBS Letters*, *586*(21), 3920–3925.

Xu, H., di Antonio, M., McKinney, S., et al. (2017). CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nature Communications 2017 8:1*, *8*(1), 1–18.

Xue, Y., Gibbons, R., Yan, Z., et al. (2003). The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(19), 10635–10640.

Yang, H., Wang, H., Shivalila, C. S., et al. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell*, *154*(6), 1370.

Yao, M., Zhou, X., Zhou, J., et al. (2018). PCGF5 is required for neural differentiation of embryonic stem cells. *Nature Communications*, *9*(1).

Yesbolatova, A., Saito, Y., Kitamoto, N., et al. (2020). The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nature Communications 2020 11:1*, *11*(1), 1–13.

Yien, Y. Y., & Bieker, J. J. (2013). EKLF/KLF1, a Tissue-Restricted Integrator of Transcriptional Control, Chromatin Remodeling, and Lineage Determination. *Molecular and Cellular Biology*, *33*(1), 4–13.

Yien, Y. Y., Robledo, R. F., Schultz, I. J., et al. (2014). TMEM14C is required for erythroid mitochondrial heme metabolism. *Journal of Clinical Investigation*, *124*(10), 4294–4304.

Yien, Y. Y., Shi, J., Chen, C., et al. (2018). FAM210B is an erythropoietin target and regulates erythroid heme synthesis by controlling mitochondrial iron import and ferrochelatase activity. *Journal of Biological Chemistry*, *293*(51), 19797–19811.

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : A Journal of Integrative Biology*, *16*(5), 284–287.

Zeng, W., Ball, A. R., & Yokomori, K. (2010). HP1: heterochromatin binding proteins working the genome. *Epigenetics*, *5*(4), 287–292.

Zhao, M. L., Rabiee, A., Kovary, K. M., et al. (2020). Molecular Competition in G1 Controls When Cells Simultaneously Commit to Terminally Differentiate and Exit the Cell Cycle. *Cell Reports*, *31*(11), 107769.

Zhou, Y., Zhu, S., Cai, C., et al. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature*, *509*(7501), 487–491.

Appendix

Appendix Table Dysregulated pathways in the ATRX KO HBM^{low} cells

	ID	Description	geneID	p.adjust
Up-regulated	GO:0061726	mitochondrion disassembly	BNIP3L/MAP1LC3B/GABARAPL1/ATG14/WDR45/RB1CC1/PINK1	0.00023843
	GO:0031668	cellular response to extracellular stimulus	PIM1/MAP1LC3B/HMOX1/GABARAPL1/ATG14/WDR45/GLUL/CPEB4/FOS	0.00498183
	GO:1905037	autophagosome organization	MAP1LC3B/GABARAPL1/ATG14/WDR45/RB1CC1/PINK1	0.00540514
	GO:1904029	regulation of cyclin-dependent protein kinase activity	CNPPD1/SERTAD1/HEXIM1/HSP90AB1/CCNL1	0.03584214
	GO:0033522	histone H2A ubiquitination	UBR2/PCGF5/UBE2B	0.03667291
Down-regulated	GO:0042775	mitochondrial ATP synthesis coupled electron transport	NDUFS8/CDK1/UQCRFS1/UQCRC1/CYC1/NDUFS7/CYCS/NDUFA6/UQCRH/NDUFA12/CCNB1/COX6C/NDUFA9/NDUFV1/NDUFA13/UQCR10/NDUFB11/NDUFAB1/COX7A2L/SNCA/COX8A/NDUFS3/NDUFA8/COX5A/NDUFA11/NDUFB7/NDUFA4/COX4I1/NDUFB8/NDUFA10/NDUFV3/NDUFB5/PARK7/NDUFB9/NDUFB6/NDUFS2/DLD/COX5B/UQCRC2/COX6A1/SDHA/NDUFB10	2.27E-17
	GO:0051168	nuclear export	ANKLE1/RAN/BARD1/SLBP/ALYREF/NUP37/NUTF2/SRSF7/GTSE1/CSE1L/ENY2/RANGAP1/U2AF1/NUP50/THOC1/RAE1/SRSF1/ANP32B/NUP107/DDX39A/NUP93/PTPN11/TXN/DHX9/THOC7/SP100/SARNP/NUP85/EIF4E/HNRNPA2B1/FIP1L1/NUP62/EIF4A3/XPOT/THOC6/NXT1/SRSF6/CPSF6/TPR/POLR2D/PHAX/XPO6/KHDRBS1/PARK7/ALKBH5/SRSF2/SRSF4/THOC2/WDR33/SEC13/SRRM1/SSB/POLDIP3/IWS1/RNPS1/CALR/EIF6/RBM8A/HSPA9/STRADB/SRSF3	5.62E-16
	GO:0050657	nucleic acid transport	RAN/FLOT1/SLBP/ALYREF/QKI/NUP37/NUTF2/SRSF7/ENY2/U2AF1/CKAP5/NUP50/THOC1/RAE1/SRSF1/NUP107/DDX39A/NUP93/DHX9/THOC7/SARNP/NUP85/EIF4E/HNRNPA2B1/FIP1L1/NUP62/EIF4A3/EIF5A2/XPOT/THOC6/NXT1/G3	1.88E-13

			BP2/SRSF6/CPSF6/HNRNPU/TPR/POLR2D/PHAX/KHDRBS1/ALKBH5/SRSF2/SRSF4/THOC2/WDR33/MRPL18/SEC13/SFPQ/SRRM1/SSB/POLDIP3/IWS1/YBX1/RNPS1/EIF6/RBM8A/SRSF3	
	GO:0050658	RNA transport	RAN/FLOT1/SLBP/ALYREF/QKI/NUP37/NUTF2/SRSF7/ENY2/U2AF1/CKAP5/NUP50/THOC1/RAE1/SRSF1/NUP107/DDX39A/NUP93/DHX9/THOC7/SARNP/NUP85/EIF4E/HNRNPA2B1/FIP1L1/NUP62/EIF4A3/EIF5A2/XPOT/THOC6/NXT1/G3BP2/SRSF6/CPSF6/HNRNPU/TPR/POLR2D/PHAX/KHDRBS1/ALKBH5/SRSF2/SRSF4/THOC2/WDR33/MRPL18/SEC13/SFPQ/SRRM1/SSB/POLDIP3/IWS1/YBX1/RNPS1/EIF6/RBM8A/SRSF3	1.88E-13
	GO:0071456	cellular response to hypoxia	PSMB3/AQP1/CCNA2/EIF4EBP1/SLC29A1/HIGD1A/ACAA2/RWDD3/CCNB1/PSMC5/PSMB2/PSMD14/HP1BP3/OPA1/BAD/PSMC2/PSMA2/PSME1/CHCHD2/SUV39H1/PSMD1/PSMD9/PHB2/P4HB/PSMA4/PSMB1/VHL/RGCC/PSMA5/IRAK1/PSME2/PSMD11/GNB1/PSMC3/PSMD7/RBX1/PSMB4/PSMB6/SIRT2/RBPJ/TMBIM6/PSMA1/PSMA6/PSMD8/PSMA7/PSMD6/NFE2L2/PSMD2/PSMC4	4.55E-09
	GO:0043044	ATP-dependent chromatin remodeling	CENPN/RBBP7/PTMA/CENPW/OIP5/ANP32E/CENPM/CENPK/HJURP/PBRM1/CENPA/MIS18BP1/MBD2/MIS18A/CENPL/HDAC2/HDAC1/SMARCE1/CENPH/ACTB/ACTL6A/SMARCA5/SMARCA4/SMARCC2/NASP/DMAP1/HNRNPC/ZNHIT1	5.10E-08
	GO:0030177	positive regulation of Wnt signaling pathway	ASPM/PSMB3/DEPDC1B/FAM53B/MBD2/PIN1/PSMC5/PSMB2/PSMD14/CSNK1G2/PSMC2/PSMA2/PSME1/PSMD1/PSMD9/PSMA4/CAPRIN2/PSMB1/ATP6AP2/SMARCA4/CSNK2A1/PPM1B/PSMA5/PSME2/PSMD11/PSMC3/VPS35/GSKIP/PSMD7/PSMB4/PSMB6/USP8/RBPJ/CSNK1D/GID8/PSMA1/PSMA6/PSMD8/PSMA7/PSMD6/PSMD2/PSMC4	1.70E-07
	GO:0046501	protoporphyrinogen IX metabolic process	UROD/ALAD/UROS/CPOX/PPOX/FECH/HMBS/ALAS2/EIF2AK1	3.66E-07

	GO:0006779	porphyrin-containing compound biosynthetic process	UROD/ALAD/SPTA1/UROS/CPOX/TSPO/SLC25A38/PPOX/TMEM14C/SLC11A2/FECH/HMBS/ALAS2	2.34E-06
	GO:0033014	tetrapyrrole biosynthetic process	UROD/ALAD/SPTA1/UROS/CPOX/TSPO/SLC25A38/PPOX/TMEM14C/SLC11A2/FECH/HMBS/ALAS2	2.34E-06
	GO:0051098	regulation of binding	CALM3/HMGB2/SRI/HMGB1/STMN1/RAN/SPTA1/NUCKS1/TMSB4X/SUMO3/FLOT1/HJURP/MAD2L2/CDCA5/AURKB/AURKA/PPP1CA/GATA1/MBD2/CDT1/PIN1/GMNN/SUMO1/ITGA4/NEK2/HDAC2/ADD2/MAPRE1/ADD1/PLK1/ACTB/TXN/DHX9/SP100/KDM1A/EIF4A3/ROCK1/PARP1/FKBP1A/RPL11/CARM1/CAPRIN2/ITGB1BP1/LRPAP1/PARK7/ID1/EIF2S1/SNAPIN/RUVBL2/TRIB3/CEBPG/MARK3/EDF1/SIRT2/EIF3D/S100A10/TMBIM6/DPH3/B2M/CSNK2B/CSNK1A1/ZBTB7A	9.15E-06
	GO:0048524	positive regulation of viral process	TOP2A/PPIA/NUCKS1/CFL1/KPNA2/POLR2K/RAD23A/VAPB/DHX9/POLR2J/TRIM21/SRPK1/POLR2D/PPIE/P4HB/SMARCA4/SNW1/PABPC1/STAU1/NELFCD/NELFE/VPS4A/TRIM11/HSPA8/VAPA/TSG101	2.56E-05
	GO:0043254	regulation of protein-containing complex assembly	TRIOBP/FAF1/HMGB1/STMN1/RPA3/SPTA1/TMSB4X/RPS3/CD36/HJURP/ARF6/EIF4EBP1/RDX/CDK5RAP2/SPTB/RAC1/PFN1/BRK1/CDT1/ARPC2/ARL2/RPA2/TFRC/HAX1/PSMC5/SUMO1/RPA1/CKAP5/SNX9/HSP90AA1/ADD2/ARPC1A/MAPRE1/ADD1/SNCA/PSMC2/CREB1/ARPC5/PTPN11/ACTR3/RNF4/LCMT1/PARP1/RHOA/ABI2/MKKS/CAPZA2/ASAP1/IQGAP2/ARPC5L/PARK7/DDB2/SAR1A/SAR1B/BCL2L11/CDC42EP4/DBNL/PSMC3/ARPC3/TMOD1/VPS35/RBX1/DRG1/SIRT2/NR1H2/HSPA8/WASF2/PSMC4/LGALS3	6.20E-05
	GO:0050852	T cell receptor signaling pathway	PSMB3/RPS3/KCNN4/PSMC5/PSMB2/PSMD14/EZR/PSMC2/PSMA2/INPP5D/PSME1/PHPT1/PSMD1/PSMD9/STOML2/FYN/PSMA4/PSMB1/KH	6.43E-05

			DRBS1/PSMA5/PSME2/PSMD11/PSMC3/MAP3K7/PSMD7/PSMB4/PSMB6/RBCK1/CRKL/PSMA1/PSMA6/STK11/PSMD8/PSMA7/PSMD6/ERMAP/PSMD2/PSMC4/LGALS3	
	GO:0002262	myeloid cell homeostasis	AHSP/KLF13/HMGB2/NCAPG2/HMGB1/IKZF1/RHAG/TRIM10/GATA1/SLC25A38/SOX6/TMEM14C/SLC11A2/KLF1/ADD1/INPP5D/FAM210B/SLC4A1/MAPK14/BCL2L11/BAP1/MAEA/CEBPG/ALAS2/EPB42/SOD1/RPS14/HSPA9/B2M/ZBTB7A/PRDX1	0.0001543
	GO:0051495	positive regulation of cytoskeleton organization	CD47/RPS3/SPAG5/RANBP1/ARF6/CFL1/CDK5RAP2/RAC1/PFN1/BRK1/POC1A/ARPC2/ARL2/HAX1/CKAP5/SNX9/ARPC1A/PLK4/MAPRE1/WDR1/ARPC5/ACTR3/NUP62/PDXP/TPM1/RHOA/ABI2/IQGAP2/DSTN/ARPC5L/ITGB1BP1/ID1/RGCC/CDC42EP4/ARPC3/DYNC1H1/DRG1/S100A10/NSFL1C/WASF2	0.00032777
	GO:0038093	Fc receptor signaling pathway	PSMB3/CD47/RAC1/BRK1/ARPC2/PSMC5/PSMB2/PSMD14/HSP90AA1/ARPC1A/PSMC2/ACTB/PSMA2/ARPC5/PSME1/ACTR3/RAP1A/PSMD1/PSMD9/FYN/PSMA4/PSMB1/PSMA5/PSME2/PSMD11/PSMC3/MAP3K7/ARPC3/PSMD7/PSMB4/PSMB6/ACTG1/PSMA1/PSMA6/CALM1/PSMD8/PSMA7/PSMD6/WASF2/PSMD2/PSMC4	0.00043646
	GO:0019430	removal of superoxide radicals	PRDX2/CD36/DHFR/GCH1/PARK7/SIRT3/SOD1/NFE2L2/CCS/PRDX1	0.00048989
	GO:0034620	cellular response to unfolded protein	WIPI1/ASNS/TM7SF3/TLN1/KLHDC3/ADD1/SSR1/HDGF/VAPB/HSPD1/CXXC1/YIF1A/ATF6/RHBDD1/EIF2S1/GOSR2/BCL2L11/COPS5/DNAJB11/CALR/SERP1/TMBIM6/HERPUD1/GFPT1/TPP1/HSPA8/HSPA9/NFE2L2	0.00083905
	GO:0006900	vesicle budding from membrane	CHMP4A/DNM2/RILP/PICALM/SNX3/AP2M1/VAPB/TRAPPC4/TFG/PPP6R1/SCFD1/CD59/KLHL12/SAR1A/SAR1B/GOSR2/SEC13/ARFGAP3/CSNK1D/S100A10/VPS4A/PEF1/VAPA	0.0009101
	GO:0034605	cellular response to heat	RBBP7/RPA3/HTRA2/NUP37/RPA2/SUMO1/RPA1/HDAC2/HSP90AA1/NUP50/RAE1/NUP107/	0.00095773

			NUP93/HSPD1/NUP85/NUP62/TPR/POLR2D/EIF2S1/SEC13/IRAK1/DNAJB1/PTGES3/HSPA8/BAG5	
	GO:0006165	nucleoside diphosphate phosphorylation	DTYMK/CMPK2/NUP37/HK1/NUP50/RAE1/NME4/DDIT4/NUP107/GAPDH/NUP93/AK1/NUP85/NUP62/NCOR1/PFKL/SLC4A1/PGM2L1/PFKM/TPR/CMPK1/SEC13/EIF6/ALDOA/TPI1/ZBTB7A	0.00103696
	GO:0071385	cellular response to glucocorticoid stimulus	PCK2/AQP1/ZFP36L2/GSTP1/EIF4EBP1/CBX3/UBE2L3/DDIT4/EIF4E/FECH/ATP2B1/HNRNPU/REST/BCL2L11/USP8	0.0015317
	GO:0015671	oxygen transport	HBM/HBD/HBA2/HBA1/HBG1/HBQ1/HBG2	0.0015317
	GO:0015949	nucleobase-containing small molecule interconversion	RRM1/DTYMK/DUT/RRM2/NME4/AK1/TXNRD1/CMPK1/GSR	0.00335052
	GO:0009048	dosage compensation by inactivation of X chromosome	EXOSC10/CDYL/BRCA1/RLIM/HNRNPU/SMCHD1/EIF1	0.00351405
	GO:0110053	regulation of actin filament organization	TRIOBP/STMN1/SPTA1/TMSB4X/CD47/ARF6/CFL1/RDX/SPTB/RAC1/PFN1/BRK1/ARPC2/HAX1/SNX9/ADD2/ARPC1A/WDR1/ADD1/F11R/ARPC5/ACTR3/PDXP/ROCK1/TPM1/RHOA/ABI2/MKKS/CAPZA2/IQGAP2/DSTN/ARPC5L/ITGB1BP1/ID1/RGCC/CDC42EP4/DBNL/ARPC3/TMOD1/ACTG1/S100A10/WASF2	0.00397176
	GO:0071616	acyl-CoA biosynthetic process	ACAT1/TECR/PDHA1/ELOVL1/PPT2/MPC2/SLC25A1/ACSL3/ELOVL5/MPC1/PPCS/PPT1/DLD	0.00505135
	GO:0030970	retrograde protein transport, ER to cytosol	AUP1/RHBDD1/UBE2G2/UBAC2/HERPUD1/SEC61B/SEL1L/NPLOC4/BCAP31	0.00564159
	GO:0033144	negative regulation of intracellular steroid hormone receptor signaling pathway	LBH/HDAC1/BRCA1/NCOR1/STRN3/RHOA/PHB2/SMARCA4/CALR/ZBTB7A	0.005882
	GO:0006364	rRNA processing	EXOSC8/MRPS11/METTL5/EXOSC10/ERI1/RRP7A/KAT2B/RPS6/SUV39H1/EIF4A3/IMP3/POP7/FRG1/ISG20/RPS16/EXOSC1/RPL11/RPL14/DKC1/RPP30/RPL35/WBP11/RPS7/NOL11/EXOSC9/TRMT112/EXOSC3/RPS2/RPL5/WDR18/EIF6/RPL7A/RPS14/TSR3/RPL7	0.00724226
	GO:0016072	rRNA metabolic process	EXOSC8/MRPS11/METTL5/EXOSC10/ERI1/RRP7A/KAT2B/RPS6/SUV39H1/EIF4A3/IMP3/POP7/F	0.00858922

			RG1/ISG20/RPS16/EXOSC1/RPL11/RPL14/DKC1/RPP30/RPL35/WBP11/RPS7/NOL11/EXOSC9/TRMT112/EXOSC3/RPS2/RPL5/WDR18/EIF6/RPL7A/SLFN14/RPS14/TSR3/RPL7	
	GO:0020027	hemoglobin metabolic process	AHSP/CAT/ADD1/ALAS2/EPB42/EIF2AK1	0.00970376
	GO:0048308	organelle inheritance	VRK1/CDK1/STK25/YWHAZ/MAP2K2/PDCD10	0.00970376
	GO:0048313	Golgi inheritance	VRK1/CDK1/STK25/YWHAZ/MAP2K2/PDCD10	0.00970376
	GO:0051293	establishment of spindle localization	NUSAP1/NDC80/MAD2L1/MYH9/CDK5RAP2/CENPA/DYNLT1/PLK1/BCCIP/DYNC1H1/NSFL1C	0.01193846
	GO:0060315	negative regulation of ryanodine-sensitive calcium-release channel activity	CALM3/SRI/GSTO1/CALM1/CALM2	0.01238381
	GO:0048199	vesicle targeting, to, from or within Golgi	WIPI1/TRAPPC4/TFG/PPP6R1/SCFD1/CD59/KLHL12/SAR1A/SAR1B/GOSR2/SEC13/ARFGAP3/GOSR1/CSNK1D/PEF1	0.01395792
	GO:0008154	actin polymerization or depolymerization	TRIOBP/SPTA1/TMSB4X/DIAPH3/ARF6/CFL1/RDX/SPTB/RAC1/PFN1/BRK1/ARPC2/HAX1/SNX9/ADD2/ARPC1A/WDR1/ADD1/ARPC5/ACTR3/PDXP/RHOA/ABI2/MKKS/CAPZA2/IQGAP2/DSTN/ARPC5L/CDC42EP4/DBNL/ARPC3/TMOD1/WASF2	0.01520258
	GO:2000171	negative regulation of dendrite development	BCL11A/ARF6/HDAC2/YWHAH/ASAP1/CARM1/ID1/UBE3A	0.01698373
	GO:0007077	mitotic nuclear envelope disassembly	VRK1/CDK1/CCNB2/CCNB1/PLK1	0.01827341
	GO:0034063	stress granule assembly	RPS23/LSM14A/CIRBP/G3BP2/G3BP1/EIF2S1/DYNC1H1	0.02055154
	GO:0150117	positive regulation of cell-substrate junction organization	ARF6/RAC1/LIMS1/IQSEC1/ROCK1/ITGB1BP1/MAP4K4/S100A10	0.02070356
	GO:0006081	cellular aldehyde metabolic process	IDH2/DHRS13/HAGH/GLO1/GRHPR/ALDH9A1/PDXP/ESD/PARK7/REST/PDXK/TPI1/RDH11	0.02156267
	GO:0032535	regulation of cellular component size	TRIOBP/SPTA1/AQP1/TMSB4X/ARF6/CFL1/RDX/SPTB/RAC1/PPP1CA/PFN1/XK/BRK1/E2F4/ARPC2/KCNN4/HAX1/DNM2/SNX9/RILP/HP1BP3/ADD2/ARPC1A/WDR1/PICALM/ADD1/EZR/CREB1/AP2M1/SNF8/ARPC5/KEL/ACTR3/KDM1A/PD	0.02412209

			XP/RHOA/ABI2/MKKS/CAPZA2/IQGAP2/RTN4/DSTN/ARPC5L/GDI1/CDK4/CDC42EP4/DBNL/ARPC3/TMOD1/WASF2	
	GO:0098732	macromolecule deacylation	PPT2/RBM14/HDAC2/HDAC1/ABHD12/C6orf89/CTBP1/MIER1/RCOR1/LYPLAL1/PPT1/REST/SFPQ/HDAC3/SUDS3/SIRT3/SIRT2/MORF4L1/ZNHIT1	0.02613391
	GO:0090322	regulation of superoxide metabolic process	CD36/DHFR/GSTP1/GCH1/PARK7/SIRT3/AATF/SOD1/NFE2L2	0.0279458
	GO:0071549	cellular response to dexamethasone stimulus	PCK2/AQP1/EIF4EBP1/CBX3/DDIT4/EIF4E/FECH/USP8	0.03041267
	GO:0010592	positive regulation of lamellipodium assembly	RAC1/BRK1/ARPC2/DNM2/ACTR3/AKIRIN1/WASF2	0.03180424
	GO:1900026	positive regulation of substrate adhesion-dependent cell spreading	TRIOBP/RAC1/ARPC2/DNM2/LIMS1/P4HB/CALR/CRKL/S100A10	0.03274474
	GO:0006575	cellular modified amino acid metabolic process	DHFR/GSTP1/LPCAT3/HAGH/GLO1/MTHFD2/GGH/ACADM/GCH1/ABHD12/PDZD11/SLC22A4/CHAC2/ALDH9A1/CRAT/PLOD3/GCLM/SLC6A8/PARK7/OSBPL8/PANK2/CTSB/SHMT2/GART/GSR/SOD1/NFE2L2/SERINC1	0.03312154
	GO:0071479	cellular response to ionizing radiation	RHNO1/NUCKS1/RAD51AP1/EEF1D/ECT2/XRCC5/DNM2/RAD1/RAD51/KDM1A/TANK/XRCC6/MAPK14	0.03865874
	GO:0046902	regulation of mitochondrial membrane permeability	TFDP1/YWHAQ/ACAA2/DYNLT1/TP53BP2/BAD/TFDP2/SLC25A5/CNP/YWHAB/YWHAH/BCL2L11/YWHAZ/BLOC1S2	0.04455349
	GO:0048488	synaptic vesicle endocytosis	CALM3/ARF6/DNM2/PICALM/SNCA/ACTB/ROCK1/VAMP2/ITSN2/SH3GL1/ACTG1/ITSN1	0.045941
	GO:0140238	presynaptic endocytosis	CALM3/ARF6/DNM2/PICALM/SNCA/ACTB/ROCK1/VAMP2/ITSN2/SH3GL1/ACTG1/ITSN1	0.045941
	GO:0007131	reciprocal meiotic recombination	TOP2A/MLH1/RAD51/KLHDC3/TRIP13/RAD51C/MND1/MUS81/RAD21/RAD50/FANCD2	0.04840894