Developing genome-editing strategies to ameliorate HbE/β-thalassaemia

Mohsin Badat

New College



Thesis submitted for the degree of Doctor of Philosophy
University of Oxford

MRC Molecular Haematology Unit
Weatherall Institute of Molecular Medicine University of Oxford

Michelmas term 2021

Abstract

The thalassaemias are the most common monogenic disorders worldwide. Both α- and β-thalassaemia are caused by genetic mutations that result in a decrease in globin gene expression and protein synthesis. The more clinically significant of the two disorders, β-thalassaemia is mainly caused by point mutations in the β-globin gene or proximal sequences. Patients develop severe anaemia early in life and are dependent on life-long blood transfusions. The relative excess free α-globin chains in developing erythroblasts causes apoptosis and necrosis, resulting in ineffective erythropoiesis and the resulting severe phenotype. Decades of careful clinical observational studies have shown that individuals with severe β-thalassaemia who co-inherit α-globin gene mutations often have a mild or asymptomatic phenotype. This is especially, but not limited to, individuals with HbE/β-thalassaemia which causes half of all severe β-thalassaemia worldwide. The α- and β-globin gene clusters are among the most finely characterised loci in the human genome, and this understanding of gene regulation in normal and pathological contexts allows the development of tailored therapeutic strategies. In this study, I have to tried to use this knowledge to develop two therapeutic strategies to cure β-thalassaemia. The first aims to tunably reduce α-globin expression by editing its main enhancer, MCS-R2, to restore globin chain balance in patients with severe β-thalassaemia. The second aims to recapitulate the asymptomatic carrier β-thalassaemia trait state by using new technology to repair the HbE mutation *in-situ.*

Initially I aimed to further characterise the role of the main enhancer of α-globin, MCS-R2. This work was enabled through the study of a patient with severe HbH disease who had an extremely rare mutation, leaving her with two α-globin genes and a deletion of MCS-R2 on the same allele. CD34+ cells were obtained and differentiated erythroid cells were studied using various methods, including ATAC-seq and Capture-C to assess gene-expression, chromatin accessibility and architecture in the absence of the main enhancer. Next a variety of techniques were used to profile the sequence of MCS-R2 to identify key regulatory sequences that are important to its function. Several putative transcription factor binding sites were identified, and these were targeted in the development of the first therapeutic strategy. Individual binding sites were mutated using CRISPR/Cas9 with high efficiency to try and tunably reduce α-globin expression. The mutation of binding sites for canonical erythroid transcription factors resulted in the reduction of α-globin expression commensurate with a single α-globin gene deletion, which has been shown previously to be of clinical benefit in individuals with severe β-thalassaemia mutations. Next base editors, a modular variant of the CRISPR/Cas9 system, were used to repair the pathogenic codon 26 mutation in the *HBB* gene that causes HbE/β-thalassaemia. This system was optimised and then used in patient-derived CD34+ cells, resulting in high base-deamination efficiencies and robust production of normal β-globin protein. Serial mouse xenotransplantation experiments were performed, proving that base editing is possible in long-term repopulating haematopoietic stem cells, and the off-target landscape was assessed.

In summary, through the development and utilisation of knowledge of how the main α-globin enhancer controls gene expression, and the adoption of new gene-editing technologies, this work outlines two new therapeutic strategies for curing HbE/β-thalassaemia. These have the potential to deliver significant benefit, both alone or in conjunction with gene-editing therapies currently being trialled in the clinic.

Acknowledgements

This work would not have been possible without many people who have shown a great deal of kindness and help. Firstly, I would like to express my deep gratitude towards my supervisors Doug Higgs, Richard Gibbons and James Davies. I will always be grateful to Doug and Richard for taking a chance and inviting me to work on this exciting project. Doug has provided support, guidance and mentorship throughout the DPhil, and has been and continues to be an inspiration to me in matters scientific and non-scientific. He was also a great companion during the 2018/2019 season; it may not have ended the way we wanted but we'll always have that night in Amsterdam. My thanks go to James for choosing me to be the inaugural student of the Davies lab. He helped guide the project at every step, offering much needed direction, and patiently fielded my all-too-frequent requests for assistance, be it related to the bench or bioinformatics. I shall remember the trip to Cold Spring Harbour with fondness, and will watch the development of the lab from afar with great anticipation and interest. I would not have come to Oxford were it not for the advice and assistance of Irene Roberts, who is an inspiration for aspiring paediatric haematologists everywhere. I thank them all for letting me become a leaf on this tree.

I feel honoured to have shared my time in Oxford with the members of the Higgs, Gibbons, Davies, Buckle and Hughes labs. It has been a pleasure to work with you, and I apologise for any of my shortcomings. Whilst I cannot mention as many as I would like to here, in particular I must thank: Chris Fisher for teaching me how to use a pipette and other scientific techniques when I arrived, and for being the foundation of all my subsequent work; Damien Downes for helping with computational analysis and being the source of all erythroid NGS assays; Robert Beagrie for being a bioinformatic maven and teaching us throughout lockdown. Damien and Rob have borne my pleas for help at all hours with great patience and generosity. Peng Hua has helped an immense amount with all things mice, stem cells and sequencing. Caroline Scott showed me how to isolate CD34s from blood, without which none of this would have been possible and has a wealth of knowledge about erythroid cells. Yavor Bozhilov and Anthony Cheong taught me much about cell culture, bacterial cloning and cars, and along with Aude-Anaïs Olijnik were companions in plying the dark arts of the HUDEP. Sachith Mettananda more than laid the groundwork for this project, and provided advice and assistance with patient samples. Jelena Telenius and Ron Schwessinger set up some of the data analysis pipelines that were vital to this work, and without Andi Roy and Siobhan Rice the mice work would not have been possible. Lastly, Ed Tunnacliffe provided much good conversation, support and has taught me a lot about Dictyostelium. I must thank Philip Hublitz for making genome engineering comprehensible and enjoyable, as well as the members of the CBRG, sequencing and flow cytometry facilities.

A special thanks to my friends and family. Hassan al-Khoee, Arash Sedighi and Ashraf Hoque all made doing a PhD seem like a romantic endeavour. Thank you for persevering with me through my idiosyncrasies, and providing oceans of support, mirth and slices. Bilal (the original CRISPR) and Ayesha Badat are wonderful siblings, who I hope to see once in a while now that this thesis has been written. I must dedicate this work to those who left this world just before it was completed; my grandmothers Aisha Badat, Fatima Seedat, father-in-law Mohamed Outaleb and my teacher Muhammad Mlamali Adam. My son Moussa was born during my final year. A source of joy, wonder and happiness, he has made me a better man and will always be part of this work. This journey began many moons ago with the selfless efforts and dreams of my grandparents and the love of my parents, who inspired me to do everything I have done. I will never be able to pay back the debt of gratitude that I owe them, but I hope this work goes a small way to making them proud. Finally, I must thank my wife Nora, who orchestrated our move there and back again, and has supported me - often at her own expense - with love, care and attention. I will always be indebted to her for the forbearance, encouragement, companionship and care she has shown. This work belongs to her as much as it does to me. 'What can be the reward for good, except good?'

Abbreviations

AAV	Adeno-associated Virus
ABE	Adenine base editor
ActRIIA	Activin A receptor type IIa
AHSP	Alpha haemoglobin stabilising protein
AML	Acute myeloid leukaemia
ATAC	Assay for transposase-accessible chromatin
ATF4	Activating transcription factor
ATRX	Alpha thalassaemia/mental retardation, X-linked
ATR-16	Alpha thalassaemia/mental retardation, chromosome 16 linked
ATMDS	Alpha thalassaemia/myelodysplastic syndrome
BCL11a	B cell leukaemia/lymphoma 11a
BE	Base editor
BFU-E	Burst forming unit-erythroid
BSA	Bovine serum albumin
BHFS	Bart's hydrops fetalis syndrome
CBE	Cytidine base editor
CD	Cluster differentiation
CFU-E	Colony forming unit-erythroid
CHD4	Chromodomain helicase DNA binding protein 4
ChIP	Chromatin immune precipitation
CNV	Copy number variation
CpG	Cytosine-guanine dinucleotide
CRISPR/Cas	Clustered regularly interspersed short palindromic repeats/CRISPR associated protein
CrRNA	CRISPR RNA
CTCF	CCCTC-binding factor
Ctrl	Control
DNA	Deoxyribonucleic acid
DHS	DNase I hypersensitive sites
DSB	Double stranded break
DsODN	Double stranded oligodeoxynucleotide
EPO	Erythropoietin
EZH2	Enhancer of zeste homolog 2
FLI1	Friend leukaemia integration 1
Flt3L	Fms like tyrosine kinase 3 ligand
GDF11	Growth differentiation factor 11
GH	Growth hormone
HbF	Fetal haemoglobin
HDAC	Histone deacetylase
HDR	Homology directed repair
HLA	Human leukocyte antigen
HPFH	Hereditary persistence of fetal haemoglobin
HRI	Heme regulated inhibitor
HSCT	Haematopoietic stem cell transplant

HSPC	Haematopoietic stem and progenitor cell
HSP70	Heat shock protein 70
IDR	Intrinsically disordered regions
IVT	In-vitro transcription
JAK2	Janus kinase 2
KLF	Krüppel-like factor
LCR	Locus control region
Ldb1	LIM domain binding protein 1
LMO2	LIM domain only 2
LRF	Lymphoma/leukaemia related factor
LT-HSC	Long term haematopoietic stem cell
MCC	Micro-Capture-C
MCS	Multispecies conserved sequence
MDS	Myelodysplastic syndrome
MEL	Mouse erythroleukaemia cell line
MEP	Myeloid-erythroid progenitor
MHC	Major histocompatibility complex
MLPA	Multiplex ligation dependent probe amplification
MMEJ	Microhomology mediated repair
MPP	Multipotent progenitor
MNC	Mononuclear cell
MSD	Matched sibling donor
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NHSBT	National Health Service Blood and Transplant service
NTDT	Non-transfusion dependent thalassaemia
NuRD	Nucleosome remodelling and deacetylase
OT	Off-target
PAM	Protospacer associated motif
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
PRC2	Polycomb repressive complex 2
ROS	Reactive oxygen species
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RT-qPCR	Reverse transcriptase quantitative polymerase chain reaction
SCD	Sickle cell disease
SCF	Stem cell factor
SCID	Severe combined immunodeficiency
SCL/Tal1	Stem cell leukaemia/T-cell acute leukaemia protein 1
SD	Standard deviation
SgRNA	Single guide RNA
ShRNA	Short hairpin RNA
SNP	Single nucleotide polymorphism
SsDNA	Single stranded DNA

STFR	Soluble transferrin receptor
TAD	Topological associated domain
TALEN	Transcription activator like effector nucleases
TI	Thalassaemia intermedia
TF	Transcription factor
TFBS	Transcription factor binding site
TPO	Thrombopoietin
TracrRNA	Trans-activating RNA
TSS	Transcriptional start site
TTS	Transcriptional termination site
UCB	Umbilical cord blood
UGI	Uracil DNA glycosylase inhibitor
UTR	Untranslated region
VNTR	Variable number tandem repeat
WAS	Wiskott-Aldrich Syndrome
WT	Wild type
ZBTB7a	Zinc finger and BTB domain containing 7a
ZFN	Zinc finger nucleases
ZNF410	Zinc finger protein 410

Table of Contents

Abstract ii

Acknowledgements iii

Abbreviations iv

1. Introduction 1
   1.1 The haemoglobinopathies - a historical perspective 1
   1.2 Erythropoiesis 2
   1.3 Haemoglobin and its production 5
   1.4 The regulation of α-globin gene expression 7
      1.4.1 Enhancers 7
      1.4.2 The regulatory landscape of the α-globin locus 8
      1.4.3 Dynamic changes through differentiation and transcription 10
   1.5 Disorders affecting the production of Haemoglobin 13
      1.5.1 Epidemiology 13
      1.5.2 α-Thalassaemia 15
      1.5.3 β-Thalassaemia 26
   1.6 Genetic modifiers of severity 37
   1.7 Potential Approaches to restore chain imbalance 43
      1.7.1 HbF induction 43
      1.7.2 Gene addition 44
      1.7.3 Targeted reduction of α-globin strategy 45
      1.7.4 Generating a thalassaemia trait genotype 46
   1.8 Summary and Aims 48

2. Materials and Methods 50
   2.1 Cell culture methods 50
      2.1.1 CD34 extraction 50
      2.1.2 HSPC culture prior to editing 50
      2.1.3 Erythroid differentiation culture from CD34 cells. 51
      2.1.4 HUDEP-2 cell culture in maintenance phase 52
      2.1.5 HUDEP-2 cell erythroid differentiation 52
      2.1.6 HEK293T cell culture 53
      2.1.7 Cytospin slide preparation 53
      2.1.8 Cell count and viability analysis using trypan blue staining 53
      2.1.9 Cell count and viability analysis using Nucleocounter 54
      2.1.10 Mouse xenograft experiments 54
   2.2 Genome engineering methods 55
      2.2.1 Plasmid transfection 55
      2.2.2 RNP transfection 55
      2.2.3 mRNA transfection 57
      2.2.4 Lipofection 57
      2.2.5 Plasmids used 58
      2.2.6 sgRNA sequences used 59
      2.2.7 Oligos used for cell line creation by HDR 60
      2.2.8 sgRNA cloning and production 60
      2.2.9 mRNA production 63
      2.2.10 HUDEP-2 cell line creation 64

2.3 DNA methods 65
2.3.1 DNA extraction 65
2.3.2 Polymerase Chain Reaction 65
2.3.3 Primers used for PCR 66
2.3.4 Sanger Sequencing 67
2.3.5 Next Generation Sequencing library preparation for genotyping 67
2.3.6 Assay for Transposase Accessible Chromatin Sequencing (ATAC-Seq) 68
2.3.7 NF-E2 ChIP-qPCR 69
2.3.8 Next Generation Capture-C 70
2.3.9 Micro-Capture-C 72
2.3.10 Multiplex Ligation Probe Amplification (MLPA) 73
2.3.11 CIRCLE-seq 73
2.4 RNA methods 74
2.4.1 RNA extraction 74
2.4.2 Reverse Transcription 74
2.4.3 Quantitative real-time PCR (qPCR) 74
2.5 Protein methods 75
2.5.1 Isoelectric Focusing (IEF) 75
2.5.2 High Performance Liquid Chromatography (HPLC) 75
2.6 Flow Cytometry methods 76
2.6.1 Immunophenotyping 76
2.6.2 Bulk cell sorting 77
2.6.3 Single cell sorting 77
2.7 Bioinformatics methods 77
2.7.1 ICE 77
2.7.2 Analysis of edits using PERL 78
2.7.3 CRISPResso 78
2.7.4 ATAC and Chromatin Immunoprecipitation sequencing alignment, normalisation and visualisation 78
2.7.5 Chromatin Immunoprecipitation (ChIP) motif analysis 79
2.7.6 ATAC-seq peak analysis 79
2.7.7 NG Capture C analysis 80
2.7.8 Micro-Capture-C analysis 80
2.7.9 DeepHaem enhancer mutagenesis 80
2.7.10 Off-target analyses 80
3. Assessment of the *in-vivo* function of MCS-R2 and an analysis of its sequence 82
3.1 Introduction 82
3.1.1 Upstream deletions cause α-thalassaemia 82
3.1.2 Analysis of the role of MCS-R2 using transgenic approaches 83
3.1.3 Analysis of the role of MCS-R2 using human patient studies 85
3.1.4 Analysis of the molecular signature of MCS-R2 88
3.1.5 Analysis of the sequence of MCS-R2 by in-vitro study 89
3.1.6 Analysis of the sequence of MCS-R2 through evolutionary conservation 94
3.1.7 *In-vivo* studies - Mouse 96
3.2 Outstanding questions 97
3.3 Results 1 - Patient NSE 99
3.3.1 Patient NSE - clinical presentation 99
3.3.2 Patient NSE - genetic diagnosis 101
3.3.3 Patient NSE - assessment of erythroid differentiation and transcription 105
3.3.4 Patient NSE - ζ-globin is not significantly de-repressed 108
3.3.5 Patient NSE - assessment of chromatin accessibility at the globin loci 110

3.3.6 Patient NSE - assessment of chromatin interactions at the α-globin locus 111
3.4 Discussion 1 - Patient NSE 114
3.5 Results 2 - MCS-R2 sequence analysis 117
3.5.1 Derivation of transcription factor binding site motifs within MCS-R2 117
3.5.2 Analysis of MCS-R2 using DNAse-seq footprinting 120
3.5.3 *In-silico* screen of MCS-R2 using Sasquatch to predict effects of mutagenesis on DNAse I footprinting 121
3.5.4 *In-silico* screen using DeepHaem to predict effects of mutagenesis on chromatin accessibility 124
3.6 Discussion 2 - Analysis of the MCS-R2 Sequence 127
4. Development of a therapeutic strategy for ameliorating β-thalassaemia by editing MCS-R2 to reduce α-globin expression 133
4.1 Introduction 133
4.1.1 Viral gene therapy 133
4.1.2 Gene editing 135
4.1.3 Gene editing using CRISPR/Cas systems 136
4.1.4 Gene editing using CRISPR/Cas9: mechanism of action 138
4.1.5 Gene editing - Delivery considerations 139
4.1.6 Gene editing - Adverse off-target editing outcomes 141
4.1.7 Gene editing - Adverse on- and off-target editing outcomes 144
4.1.8 Gene editing for the β-thalassaemias - previous approaches 145
4.1.9 Gene editing for the β-thalassaemias - Reduction of α-globin expression 145
4.1.10 Gene editing strategy to reduce α-globin expression 147
4.2 Results 150
4.2.1 The α/β ratio in non-thalassaemic and α-thalassaemic cells 150
4.2.2 Designing sgRNAs 151
4.2.3 Optimising of GFP positivity 152
4.2.4 Plasmid-based editing 153
4.2.5 Plasmid-based editing - DSB repair outcomes 155
4.2.6 Ribonucleoprotein editing 157
4.2.7 Ribonucleoprotein editing and globin gene expression 160
4.2.8 Ribonucleoprotein editing and assessment of differentiation 161
4.2.9 Ribonucleoprotein editing of HbE/β-thalassaemia patient cells 162
4.2.10 Ribonucleoprotein editing - protein assays 163
4.2.11 Ribonucleoprotein editing and chromatin accessibility 164
4.2.12 Ribonucleoprotein editing and chromatin architecture 167
4.2.13 Ribonucleoprotein editing and Off-target editing 170
4.2.14 Ribonucleoprotein and mRNA editing: combinatorial approaches 171
4.3 Discussion 173
5. Developing a base editing strategy to cure HbE/β-thalassaemia 180
5.1 Introduction 180
5.1.1 Forms of DNA editing 180
5.1.2 Cytosine base editors (CBEs) 182
5.1.3 Adenine base editors (ABEs) 185
5.1.4 Off-target (OT) base editing 185
5.1.5 Editing windows 186
5.1.6 Base editing and the HbE/β-Thalassaemias 187
5.2 Results 191
5.2.1 Plasmid-based editing in HEK293T cells 191
5.2.2 Plasmid-based editing in HUDEP-2 cells 192

5.2.3 Plasmid-based editing in wild type CD34+ cells 195
5.2.4 Plasmid based editing in HUDEP-2 HbE cells 196
5.2.5 Plasmid-based editing in HbE/β-Thalassaemia patient CD34+ cells 198
5.2.6 mRNA-based editing in wild type CD34+ cells 203
5.2.7 mRNA-based editing in HbE/β-Thalassaemia patient derived CD34+ cells 205
5.2.8 mRNA-based editing of long-term haematopoietic stem cells 210
5.2.9 Off-target effects - CIRCLE-seq of base-edited patient cells 213
5.2.10 Off-target effects - assessment of chromatin accessibility in base-edited patient cells 215
5.3 Discussion 219
6. Final discussion and conclusions 227
6.1 MCS-R2 behaves according to a flexible enhancer model 227
6.2 Continuing to develop genome engineering cures for red cell disorders 229
6.3 New ways of assessing off-targets in the clinic 232
6.4 The future of genome editing - logistical and technical considerations 234
Appendix 1 - Laboratory and radiological test results from patient NSE 237
Appendix 2 - Acknowledgement of previously published material 238
References 239

1. Introduction

1.1 The haemoglobinopathies - a historical perspective

The thalassaemias are the most common set of monogenic disorders worldwide, and are caused by a reduction in the production of globin chains that form the essential oxygen-carrying molecule haemoglobin (Piel and Weatherall, 2014). Thalassaemia was formally discovered in 1925 by Thomas Cooley, who published a case series involving five children in the *Transactions of the American Pediatric Society* (Weatherall et al., 2013). The thalassaemias are so called in reference to the ethnic origin of the patients that had been described at the time; in Greek 'thalassa' means 'the sea' referring to the Mediterranean, and anaemia refers to the hallmark clinical finding in individuals with the disease (Whipple and Bradford, 1932).

From these early days, thalassaemia went on to become one of the first genetic diseases for which the molecular basis determined, and its study helped usher in the era of molecular medicine. Numerous scientists including Max Perutz, J.S Haldane, Linus Pauling, James Neel and David Weatherall, to name but a few, made outstanding breakthroughs when working on thalassaemia, making discoveries that had a broad and foundational application to the study of genetics and molecular biology. Transformative insights were made in fields as varied as population genetics, gene ontogeny, genetic and epigenetic regulation of transcription and protein translation, structure and function (Weatherall, 2010b).

Being at the forefront of molecular biology, many new technologies were either developed through the study of the haemoglobinopathies or were taken up early by the field. The thalassaemia research field is now well positioned to offer new curative strategies to patients who, in the past, have been treated by supportive care alone. Over the past few decades broad and detailed research into the genetic regulation of globin genes has delivered on the promise that translational medicine offers, with a wide variety of targeted molecular therapies making their way to the clinic (Thompson et al., 2018, Esrick et al., 2020, Frangoul et al., 2021).

1.2 Erythropoiesis

In adult humans up to 2 million red blood cells are produced every second (Bao et al., 2019). According to the classical model of haematopoietic differentiation, at the apex of the differentiation tree are long-term haematopoietic stem cells (LT-HSCs). These are marked by two key characteristics: 1) self-renewal capacity and 2) multipotency; the ability to reconstitute all the different blood lineages upon transplantation. In humans, it is estimated that there are 3000-10000 LT-HSCs, dividing every 3 months-3 years (Laurenti and Gottgens, 2018). Upon differentiation, LT-HSCs give rise to progenitor cells, which do not have the capacity for long-term self-renewal and multipotency. The earliest committed erythroid progenitor is the burst forming unit erythroid (BFU-E), followed by the colony forming unit erythroid (CFU-E). BFU-Es and CFU-Es are indistinguishable morphologically from other progenitor cells, but their downstream precursors have distinct appearances. Proerythroblasts arise from CFU-Es, and differentiate sequentially into basophilic erythroblasts, polychromatophilic erythroblasts, orthochromatophilic erythroblasts and,

following enucleation, reticulocytes (Figure 1.1). Alongside morphology, cell surface markers can be used to separate different populations of erythroid cells. Cells at the myeloid erythroid progenitor (MEP) stage express CD34 and CD38 (Doulatov et al., 2010). These markers are gradually lost by the CFU-E stage, while expression of CD71 (transferrin receptor) increases (Li et al., 2014). Later in differentiation, surface CD235 expression increases while CD71 expression declines. Greater delineation of subpopulations at the various erythroblast stages can be rendered by comparing α-4 integrin (CD49d) and Band III expression on cells that are both CD71 and CD235 positive (Hu et al., 2013).

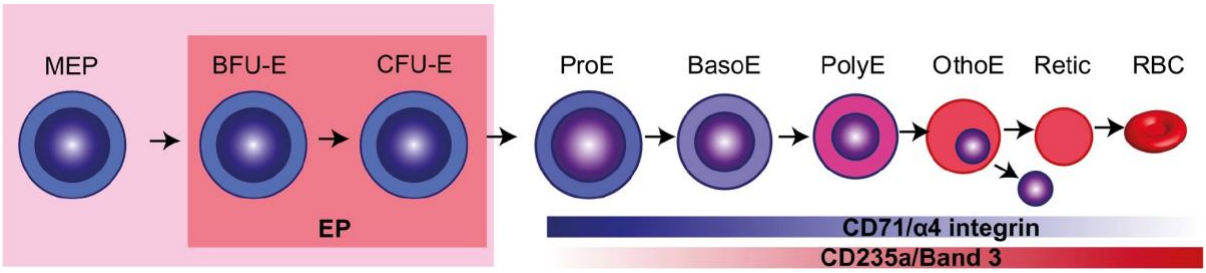


Figure 1.1 - Erythroid Differentiation. Differentiation from the megakaryocyte-erythroid progenitor (MEP) into erythroid progenitors (EP) and more mature erythroid cell types. Erythroid progenitor (EP) cells share a similar immunophenotypic profile to MEPs, but are also CD105 and potentially CD71 positive. Later cell surface marker expression shown below the cells. Adapted from (Nandakumar et al., 2016)

Commitment to the erythroid lineage - and other lineages more broadly - is governed by lineage-specific master transcription factors (TFs) (Pevny et al., 1991). Several key canonical factors form a 'core erythroid network' for transcriptional regulation (Nandakumar et al., 2016). These include DNA binding transcription factors (GATA binding factor 1 [GATA1], stem cell leukaemia/t-cell acute lymphocytic leukemia protein 1 [SCL/Tal1], erythroid Krüppel like factor 1 [KLF1]) or non-DNA binding factors (lim only 2 [LMO2] and lim binding domain 1 [LDB1]). For example, mutations in GATA1 and KLF1 in humans are known to cause

numerous defects in erythroid differentiation and function (Perkins et al., 2016, Crispino and Horwitz, 2017). Conversely, TAL1 overexpression in human CD34+ cells leads to enhanced erythroid and megakaryocytic differentiation, and knock-down in mouse results in impaired erythroid and megakaryocytic lineage specification (Elwood et al., 1998, Hall et al., 2003).

Several of the erythroid transcription factors mentioned above take part in cross-antagonistic interactions with non-erythroid factors early in differentiation, particularly at the unstable multipotent progenitor stage (Laurenti and Gottgens, 2018, Gillespie et al., 2020, Palii et al., 2019). Interestingly, this relationship is shaped not only by the presence or absence of these factors but their relative concentrations within the cell (Kulessa et al., 1995, Gillespie et al., 2020, Palii et al., 2019). For example, the antagonism between GATA1 and PU.1 is detected early in erythroid differentiation, and biases towards erythroid versus myeloid lineage choice respectively (Rekhtman et al., 1999, Nerlov et al., 2000). This is followed by the KLF1:friend leukaemia integration 1 [FLI1] antagonism that governs erythroid versus megakaryocytic cell fate decisions (Frontelo et al., 2007).

Transcription factors within a core network may act individually or in concert. In mouse and human erythroid cells an erythroid 'pentameric' complex comprising of GATA1, SCL/Tal-1, transcription factor E2-alpha (E2A), LMO2 and LDB1 is found at erythroid-specific accessible elements (Wadman et al., 1997, Anguita et al., 2004). The latter two factors are non-DNA binding; LMO2 has a bridging role within the complex, and LDB1 has roles in chromatin looping and interacting with downstream complexes (Deng et al., 2012, Krivega et al., 2014). These local TFs are able to bind other general transcriptional cofactors such as Mediator,

BRD4, Cdk9 and other chromatin modulators which exert influence on various elements of transcription including RNA Pol II recruitment to gene promoters (Kagey et al., 2010, Spitz and Furlong, 2012, Fujisawa and Filippakopoulos, 2017, Bacon and D'Orso, 2019).

1.3 Haemoglobin and its production

Haemoglobin is a 64-kD protein that is the primary oxygen transporting molecule in the body, and is the major component of mature red blood cells. It evolved from an ancient monomer into a heterodimeric tetramer over 400 million years ago, and in human is comprised of two α-like and two β-like globin chains, each of which have a prosthetic heme group (Pillai et al., 2020). Cooperative moderate affinity binding of oxygen by the two paralogous sets of globin chains enables haemoglobin to bind and release oxygen effectively (Storz, 2018).

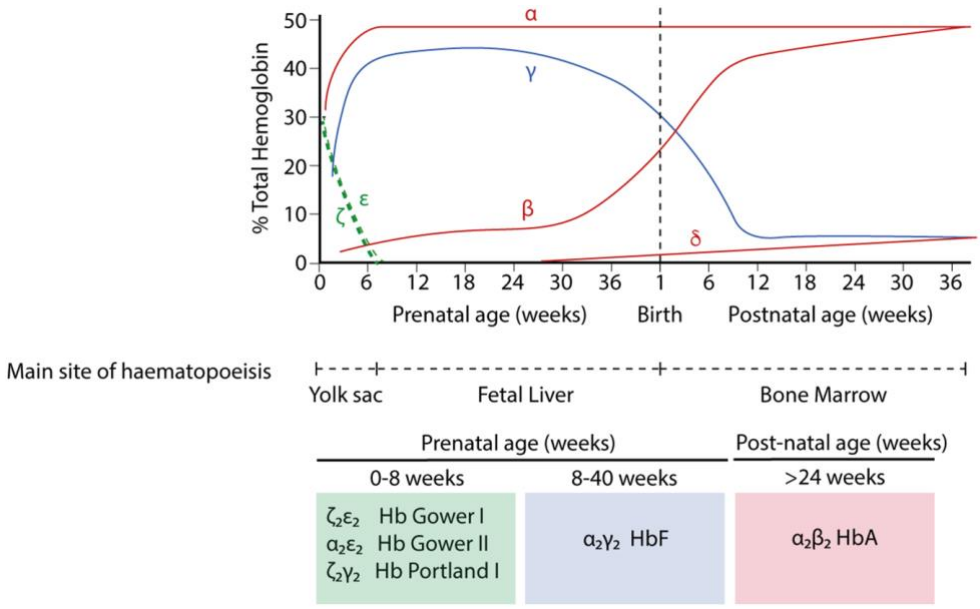


Figure 1.2 - Haemoglobin switching ζ- and ε-globin are the first chains to be expressed, followed soon after by α- and γ-globin. β-globin is expressed at low levels antenatally, but switches with γ-globin at 0-6 months postnatally. Adapted from (King and Higgs, 2018)

Through ontogeny the haemoglobin tetramer is comprised of different globin chain isoforms. In human, at the embryonic stage these are Hb-Gower I ($\zeta_2\varepsilon_2$), Hb-Gower II ($\alpha_2\varepsilon_2$) and Hb-Portland ($\zeta_2\gamma_2$), followed by HbF ($\alpha_2\gamma_2$) in foetal life, and then adult haemoglobin ($\alpha_2\beta_2$) post-partum (Figure **1.2**) (Forget and Hardison, 2009).

The genes from which the two groups of globin chains are synthesised are located at the α-globin and β-globin loci, located at chromosome 16p13.3 and chromosome 11p15.4 respectively. The genes encoding ζ- and α-globin at the α-locus, and ϵ-, γ- and β-globin at the β-locus, along with most of their major regulatory elements have been syntenically conserved throughout mammalian evolution. Both sets of genes are arranged in the order that they are expressed through development. *Cis*-regulation of globin gene expression is conducted by the respective regulatory element clusters. At the α-locus in human there are four enhancers, known as multispecies conserved sequence (MCS) R1-4 (Figure **1.3**A). At the β-locus there are five upstream elements, known as the β-locus control region (LCR)(Figure **1.2**B). Interestingly, although the two loci originally arose through a duplication event, they exist in different chromatin environments and in some regards, are regulated differently (Mettananda et al., 2015).

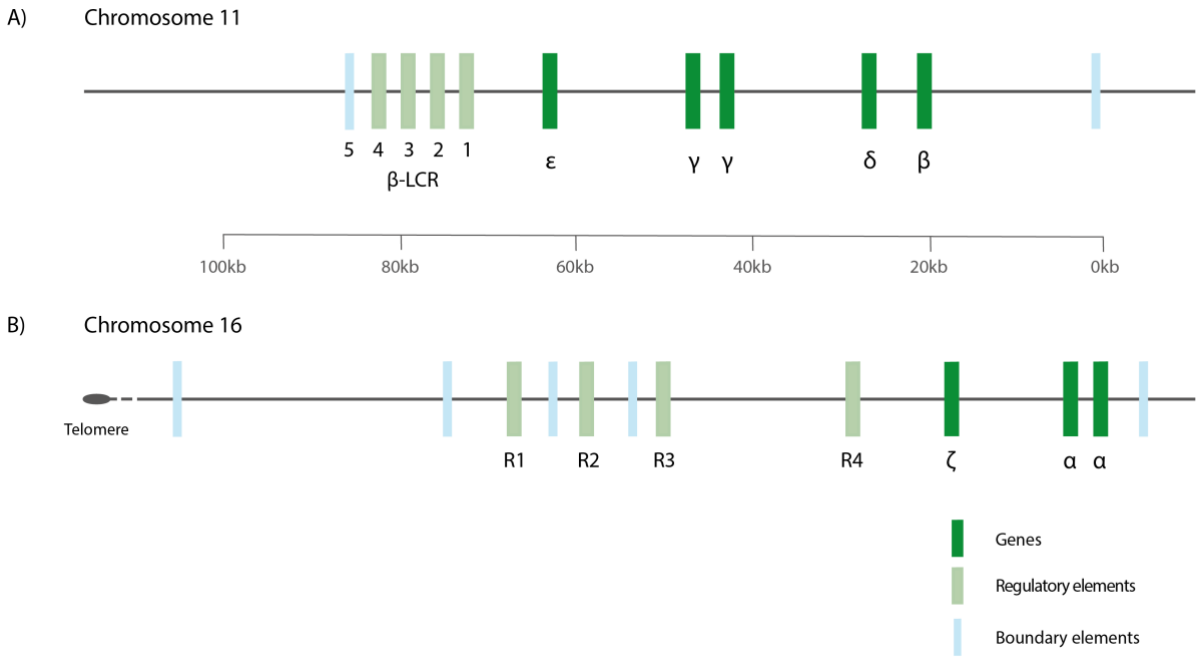


Figure 1.3 - The human globin loci (A) β-globin and (B) α-globin loci. Regulatory elements are defined as in 1.4.1, and boundary elements are loci known to bind CCCTC-binding factor (CTCF) in human

1.4 The regulation of α-globin gene expression

1.4.1 Enhancers

Enhancers are DNA sequences that regulate the transcription of their target genes. They were originally named for their ability to enhance transcription of the rabbit β-globin gene (Banerji et al., 1981). These sequences are accessible and active in a tissue- and developmental-stage specific manner, orientation independent, and can be located within, very near (kb) or very far (Mb) from their associated genes (Long et al., 2016). Individual enhancers bind tissue-specific transcription factors, discussed below, through which they exert their function, using downstream co-activator proteins such as the Mediator complex (Allen and Taatjes, 2015). Multiple enhancers can control the same gene, consistent with

the fact that there are many more enhancers than genes (Shen et al., 2012). Multiple enhancers in close proximity that drive high levels of expression of a target gene have been variously described as stretch enhancers, locus control regions and super-enhancers (Grosveld et al., 1987, Whyte et al., 2013, Parker et al., 2013). It has been postulated that individual enhancers within such clusters may act synergistically to potentiate higher levels of expression than would be expected, although functional work also suggests that individual elements within these clusters may, in some instances, act in an additive fashion (Whyte et al., 2013, Hay et al., 2016). Enhancers can be identified through a variety of methods. When functionally active, they tend to reside in open chromatin, which can be identified through methods which identify nucleosome free regions such as DNase-seq, or ATAC-seq (Consortium, 2012, Buenrostro et al., 2013). This approach can be refined further by intersecting these regions with chromatin modification signatures; enhancers tend to be marked by H3K4me1 and H3K27ac, and less by H3K4me3 which marks promoters (Heintzman et al., 2007, Creyghton et al., 2010). An orthogonal approach is to look for regions of evolutionary conservation in the non-coding genome, as these may have had constraint applied upon them due to their functional significance as regulatory elements (Yue et al., 2014). Finally, once elements have been identified they can be tested functionally using reporter assays and *in-vivo* animal reporter models (Splinter and de Laat, 2011).

1.4.2 The regulatory landscape of the α-globin locus

The human α-globin gene locus lies within an approximately 135kb syntenically conserved region around 150kb away from the chromosome 16p telomere (Figure **1.4**). The enhancers within the cluster are located 10, 33, 40 and 48kb upstream to the ζ-globin promoter and

are known as MCS-R1, R2, R3 and R4 respectively. Three of the elements, MCS R2-4, lie within the introns of the neighbouring gene *NPRL3*, which has remained linked to the α-globin cluster in evolutionary history as far back as the river lamprey (~500 MYA) (Miyata et al., 2020). The upstream regions were initially identified after assessing patients who presented with α-thalassaemia and yet had intact α-globin genes (Hatton et al., 1990, Higgs et al., 1990, Liebhaber et al., 1990, Wilkie et al., 1990b). Characterisation using interspecific hybrid cells, reporter assays and transgenic mice revealed that amongst the DNase I hypersensitive sites within the locus the most important regulatory element is located 40kb upstream of the genes. When absent, there is a reduction in α-globin transcription to ~3-5% of normal (Bernet et al., 1995, Wallace et al., 2007, Vernimmen et al., 2007). Further next generation sequencing (NGS) assays combining functional genomics methods and evolutionary conservation approaches refined the mapping to identify the full complement of enhancers as outlined above (Hughes et al., 2005). Recently reported rare mutations in patients demonstrated that while *in-vitro* the reduction in α-globin transcription is extremely severe in the absence of MCS-R2, its loss can result in a milder phenotype than expected *in-vivo* (Coelho et al., 2010, Mettananda et al., 2017a, Badat et al., 2021).

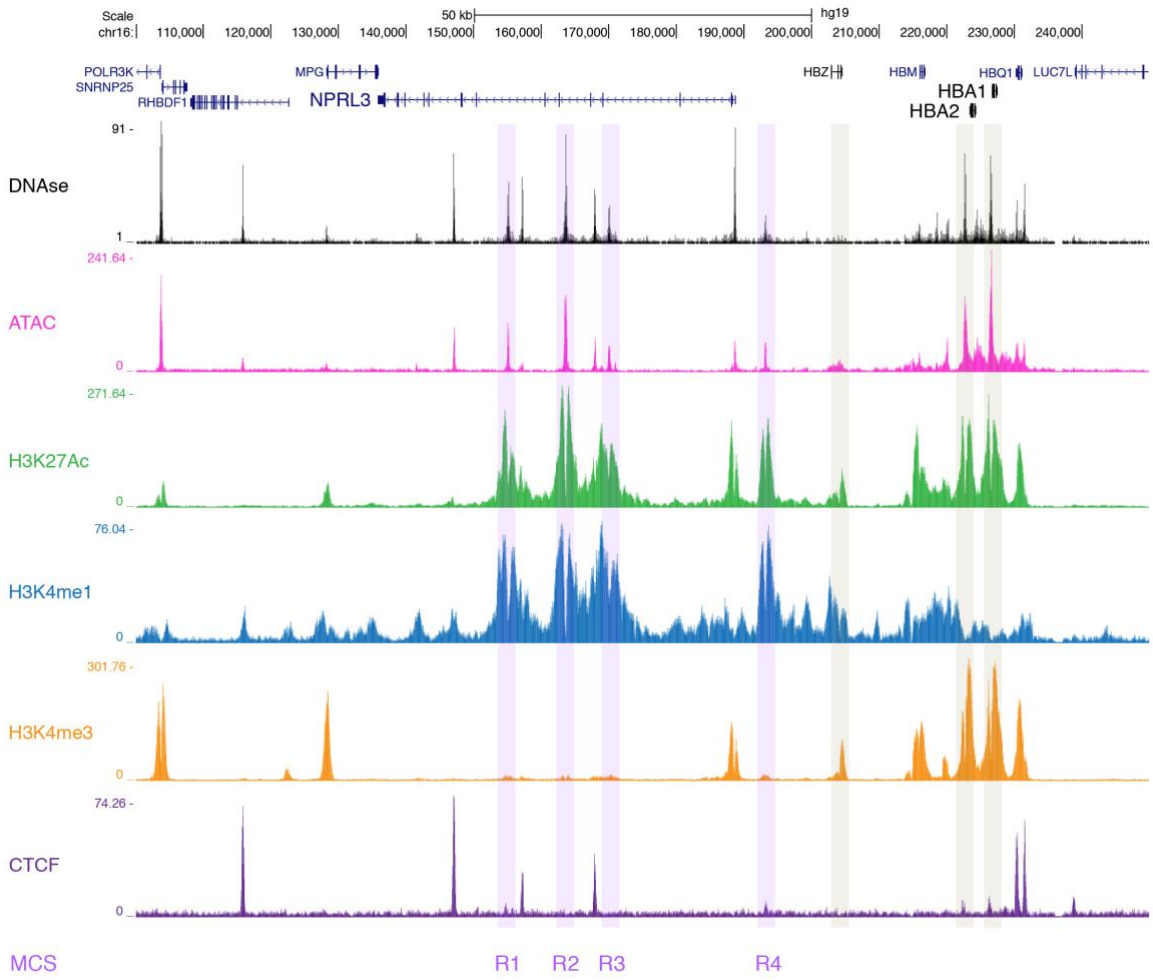


Figure 1.4 - Identifying the regulatory elements of the α-globin locus Characterisation of the locus at day 10 of erythroid differentiation in human CD34+ derived erythroid cells. Positions of enhancers are highlighted by vertical violet bars, and the embryonic and definitive globin genes are marked out by green bars. As described in 1.4.1 DNAse-seq, ATAC-seq and ChIP-seq for H3K27Ac mark out active elements within the locus, ChIP-seq for H3K4me1 and H3K4me3 selectively mark active enhancers and promoters respectively (Data generated by Damien Downes, WIMM, unpublished).

1.4.3 Dynamic changes through differentiation and transcription

The α-globin locus is located in a gene-dense area of chromosome 16 flanked by housekeeping genes. The locus itself is GC rich, and the α-globin promoters are associated with unmethylated cytosine-guanine dinucleotide (CpG) islands (Higgs et al., 2005). The locus is in a region of constitutively open chromatin, and therefore active mechanisms are

required to maintain the strict tissue-specificity in expression that is observed (Craddock et al., 1995, Flint et al., 1997). In non-erythroid cells the locus is marked by H3K27me3, a signature that is associated with the downregulation of nearby genes (Figure 1.4) (Garrick et al., 2008). This mark is deposited by the histone methyltransferase zeste homolog (EZH2), itself part of the polycomb repressive complex (PRC2), a multi-subunit complex that can also recruit histone deacetylases. Indeed, PRC2 subunits including EZH2 are found at the α-globin locus in non-erythroid cells, and so PRC2 and its associated proteins are thought to maintain the locus in a repressive state in HSCs, multipotent progenitors (MPPs) and differentiated non-erythroid tissues (Lynch et al., 2012).

In-vitro and *in-vivo* work over the past two decades has outlined the various processes by which the α-globin locus - and gene loci more generally - are switched on and transcription activated. The following data were primarily generated in mouse except where indicated. At the pluripotent embryonic stem cell stage, the regulatory elements and promoters are not accessible, reflective of the fact that erythroid transcription factors and therefore the α-globin genes are not expressed. Whilst architectural transcription factor CTCF and cohesin do bind to convergent flanking CTCF-motifs and the locus is within a ~165kb topological associated domain (TAD) present in all examined cell types (Figure 1.4) (Oudelaar et al., 2021). At the HSC and MPP stage the locus is more accessible as assayed by ATAC-seq (Downes et al., 2019, Oudelaar et al., 2020). While specific enhancer-promoter interactions aren't present, the enhancers are bound by several erythroid and haematopoietic cell transcription factors such as NFE2, Scl/Tal1 and GATA2 (Anguita et al., 2004). As differentiation proceeds the repressive histone mark H3K27me3 is removed, likely by histone lysine demethylase, as determined in a humanised mouse model (Vernimmen et al.,

2011). This is accompanied by an increase in chromatin modifications signalling transcription activation: deposition of H3K4me3 and H3K4me1 found predominantly at the promoters and enhancers respectively, and acetylation of histone H3 and H4 at both enhancers and promoters (De Gobbi et al., 2007, De Gobbi et al., 2011, Hay et al., 2016). GATA1 replaces GATA2 at the enhancers, the full erythroid pentameric complex is recruited and KLF1 also binds the enhancers and α-globin promoters in human and mouse (De Gobbi et al., 2007, Hay et al., 2016). In committed erythroid cells a smaller ~70kb sub-TAD forms, likely through loop extrusion driven by erythroid-specific cohesin recruitment, possibly mediated by loading at the active regulatory elements (Oudelaar et al., 2020, Hua et al., 2021). This sub-TAD is marked by strong enhancer-promoter contacts in both human and mouse, and as these interactions increase, a concomitant increase in α-globin expression is noted, mediated through loading of RNA PolII and the Pre-Initiation Complex (PIC) at the enhancers and promoters (Vernimmen et al., 2007, Oudelaar et al., 2020).

While very small amounts of α-globin are produced even in stem cells, prior to dynamic chromatin mark deposition, transcription factor binding, chromatin conformation changes and enhancer-promoter interactions, the subsequent coordinated series of events allows prodigious amounts of mRNA and protein to be produced to enable the red cell to be almost completely devoted to producing and filling the red cell with haemoglobin. Unsurprisingly, perturbations in this process, most commonly due to genetic mutations at the locus, result in significant human disease characterised by anaemia.

1.5 Disorders affecting the production of Haemoglobin

The thalassaemias are characterised by a reduction in the production of either α-globin in α-thalassaemia or β-globin in β-thalassaemia. These are most commonly due to genetic mutations that affect the genes or promoters themselves, but can infrequently be caused by mutations involving their regulatory elements or *trans*-acting factors outside their respective gene loci (Higgs, 2013, Thein, 2013). Whilst there are variations in their molecular and clinical pathophysiology (discussed below) there are broad similarities in the demographics of individuals who carry disease-causing mutations.

1.5.1 Epidemiology

The thalassaemias are an emerging and global health burden. Approximately 300,000-400,000 children are born with severe haemoglobinopathies per year, with around 90% of these born in low-middle income settings (Walani and Biermann, 2017, Williams and Weatherall, 2012). Approximately 30-40,000 have severe β-thalassaemia, and around 200,000 have sickle cell disease (SCD) (Weatherall, 2010c, Modell and Darlison, 2008).

There are several explanations as to why these disorders are so common worldwide. Their high frequency was hypothesised to be due to natural selection for heterozygotes carrying globin mutations due to their relative protection against severe malaria (Haldane, 1949). Haldane's hypothesis has been largely borne out, particularly for *P. falciparum* malarial infections, by mechanistic laboratory studies, through comparing the distributions of

malaria and haemoglobinopathy mutations, and clinical studies in affected areas (Weatherall, 2008, Williams and Weatherall, 2012, Premawardhena et al., 2017).

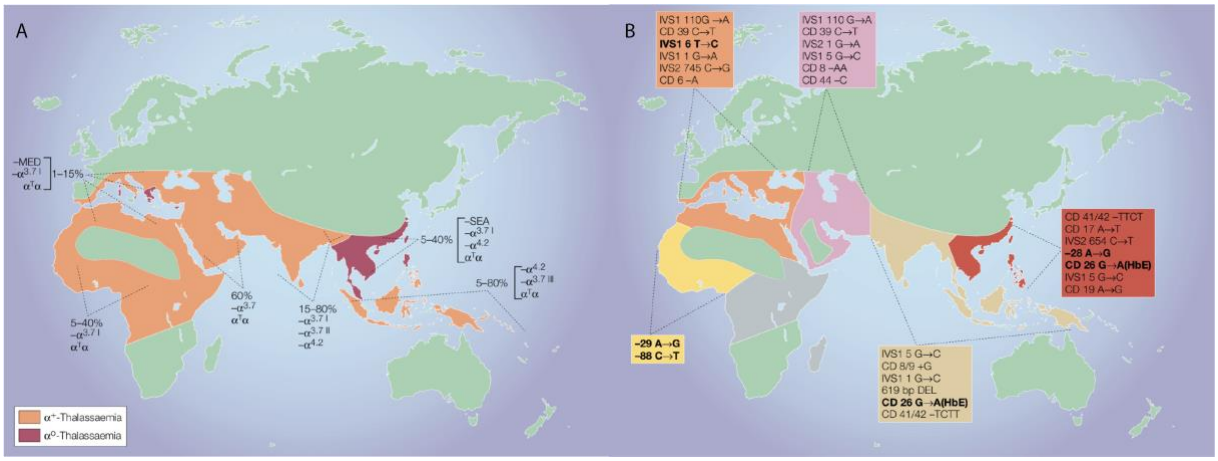


Figure 1.5 - Worldwide distribution of the thalassaemias (A) Distribution of α-thalasssaemia with frequencies and common mutations (B) Distribution of β-thalassaemia with frequencies and common mutations. Adapted from (Weatherall, 2001)

Until relatively recently, individuals with thalassaemia mutations tended to be located in tropical or sub-tropical regions, in keeping with these areas having relatively high rates of malarial infection. Specific mutations in the α- and β-globin genes may have spread over time and established themselves in certain areas due to founder effects (Weatherall, 2010a, O'Shaughnessy et al., 1990). They may also have arisen *de-novo* in different places; for example, the β-globin mutation causing SCD is thought to have emerged in at least three points in Africa and one in India (Stuart and Nagel, 2004). In contrast, the codon 26 β-globin mutation producing HbE tends to be found predominantly in South-east Asia, Bangladesh, Sri Lanka and eastern parts of India (Williams and Weatherall, 2012).

The incidence of thalassaemia in the last few decades has not significantly risen, but many children who would otherwise have died due to the condition now survive to adulthood.

This is and will continue to become an emerging health burden on low and middle income countries due to improvements in sanitation, healthcare systems and local and regional economies (Weatherall, 2010a). Of note, due to the legacies of slavery and indentured servitude, and recent economic migration there are a significant number of patients with the thalassaemias now living and being born in Western countries (Michlitsch et al., 2009, Weil et al., 2020).

1.5.2 α-Thalassaemia

1.5.2.1 Molecular Basis

The normal genotype of the two pairs of duplicated α-globin genes is written as αα/αα. Interestingly within the locus as a whole there is a remarkable amount of normal variation that does not result in a detectable phenotype. This ranges from single nucleotide polymorphisms (SNPs), variations in the number of tandem repeats (VNTRs), copy number variants (CNVs), segmental insertions/deletions and even chromosomal translocations (Higgs, 2013). The majority of mutations that cause α-thalassaemia however tend to be large deletions that remove one or both of the α-globin genes. Infrequently point mutations in or around either α-globin gene cause non-deletional thalassaemia, and very rarely α-thalassaemia can be caused by regulatory element deletions that invariably include MCS-R2 or *trans*-acting factors such as ATRX (Harteveld and Higgs, 2010). Broad categories of mutatations and common genotypes are outlined below; a comprehensive overview of all deletions can be found elsewhere (Higgs, 2009a, Harteveld and Higgs, 2010, Giardine et al., 2014, Kountouris et al., 2014).

1.5.2.1.1 Deletions removing one α-globin gene

The α-globin genes are situated in a highly homologous 4kb duplicated region, which are divided into three homology blocks, X, Y and Z which are themselves separated by non-homology elements numbered I, II and III (Lauer et al., 1980). The most common deletion events occur due to recombination between the X boxes or the Z boxes. Recombination between the Z boxes, which are 3.7kb apart is also known as a rightward deletion and results in a non-thalassaemic triplicated $\alpha\alpha\alpha^{anti3.7}$ allele and a thalassaemic $-\alpha^{3.7}$ allele (Embury et al., 1980, Goossens et al., 1980). Similarly, recombination between the X boxes, known as leftward deletion, results in a $\alpha\alpha\alpha^{anti4.2}$ and $-\alpha^{-4.2}$ allele (Embury et al., 1980, Trent et al., 1981).

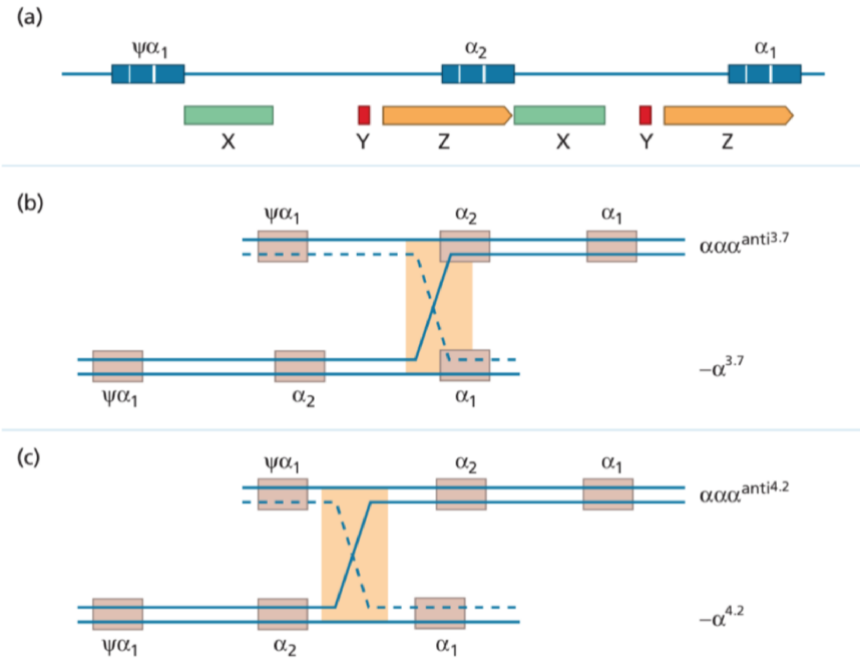


Figure 1.6 - Molecular mechanisms underlying single gene deletions (A) single allele showing the normal structure of the gene locus, with X, Y and Z homology boxes. (B) 3.7kb rightward deletion (C) 4.2kb leftward deletion. Adapted from (Thein and Rees, 2015)

1.5.2.1.2 Deletions removing both α-globin genes

Over 50 types of deletions removing both structural α-globin genes from the same allele have been described. They can be grouped into those that are restricted to the α-globin locus alone and those that are much larger, affecting flanking genes and extending in some cases to nearly a megabase beyond the cluster (Higgs, 2009a). In contrast to deletions removing a single α-gene, these deletions are thought to occur as a result of non-homologous recombination events. Deletion breakpoints may involve short sequences that are partially homologous, such as the Alu-family repeats which make up around 25% of the entire sequence of the α-globin locus (Nicholls et al., 1987).

Common mutations like $--^{MED}$ and $--^{SEA}$ remove both α-globin genes but leave ζ-globin intact, whereas others remove both sets of genes (Harteveld and Higgs, 2010). Homozygotes for the latter mutation would not survive beyond a few weeks gestation due to the inability to make even embryonic haemoglobin. Homozygotes and heterozygotes for ζ-globin sparing mutations should theoretically be able to make the embryonic globin Hb Portland ($\zeta_2\gamma_2$). In practice this tends to be variable, with ζ-globin expressed at very low levels, certainly less than the amount of γ-globin expressed by individuals with similar deletions at the β-globin locus (Higgs, 2009a).

Larger deletions involving other genes have been observed but only in the heterozygous state. Interestingly whilst deletions within the α-globin locus can reach high frequencies in certain populations, these extended deletions are never found at polymorphic frequencies (Higgs, 2013). This is possibly due to these mutations carrying a negative selective advantage

due to gene dosage effects, or unmasking mutations in the retained homolog. In keeping with this, some individuals with monosomy for a large segment of chromosome 16p13.3 have developmental abnormalities and cognitive impairment in addition to α-thalassaemia, a condition known as ATR-16 (Wilkie et al., 1990a, Gibbons, 2012, Babbs et al., 2020).

1.5.2.1.3 Deletions involving regulatory elements

The upstream enhancers are vital for α-globin transcription. They were initially identified through characterising patients presenting with α-thalassaemia. To date over 25 patients have been described who have deletions that involve the upstream elements and spare the α-globin genes (Figure **1.7**). Some of these deletions may be very large; $(\alpha\alpha)^{CMO}$ extends to the tip of chromosome 16p, and involves repair of the end by addition of telomeric repeats to non-telomeric DNA (Flint et al., 1994). In contrast, a recently reported deletion was as small as 742bp (Wu et al., 2017, Huang et al., 2017). Despite these variations, every reported deletion of this kind removes MCS-R2, indicating its crucial role in human α-globin transcription.

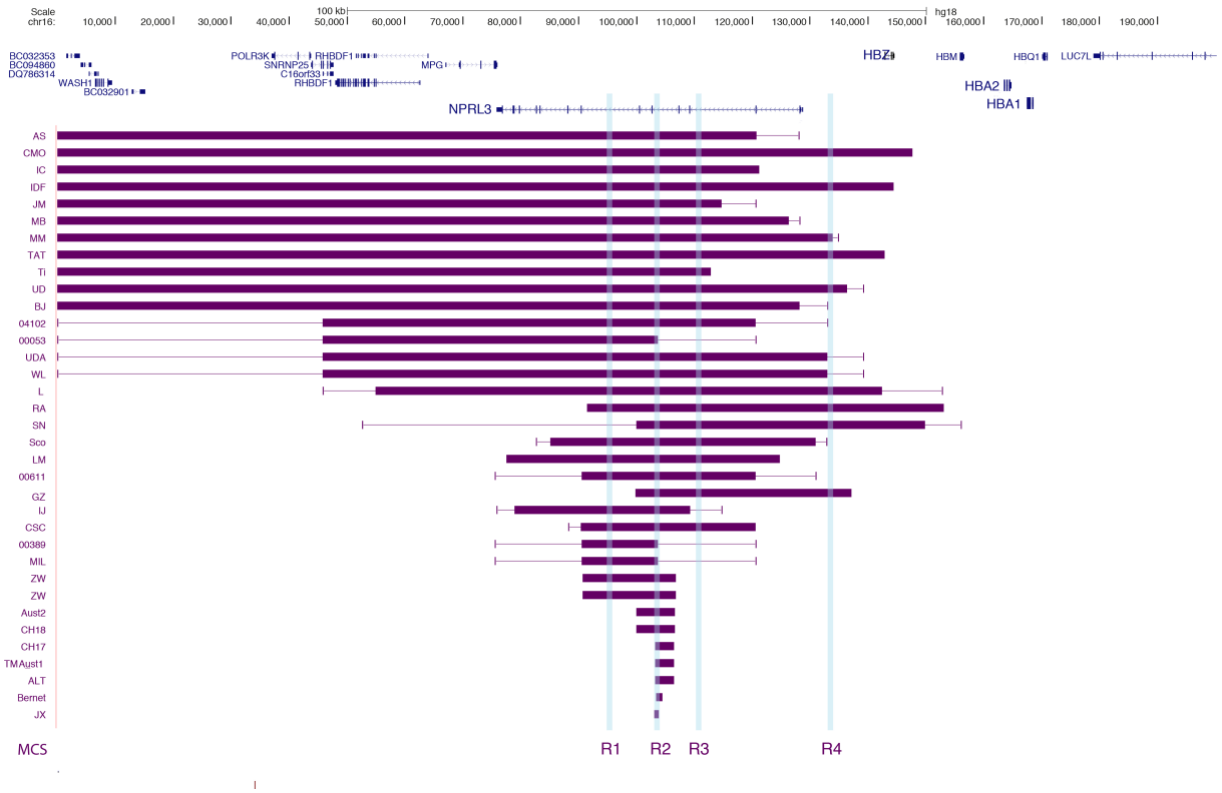


Figure 1.7 - The upstream deletions *The known upstream deletions leaving the α-globin genes intact that cause α-thalassaemia. Every deletion includes MCS-R2. Deletions are annotated with an alphabetic name as reported. Numerical names denote unpublished reports. As most deletions were characterised by multiplex ligation dependent probe amplification (MLPA) and may not have the breakpoint verified by sequencing, the solid bars denote regions which are known to be deleted through loss of probe signal. The maximum potential size of the deletion is denoted by the length of the thin bars, although the breakpoints may be at any point along them.*

1.5.2.1.4 Non-deletional mutations

Numerous small point mutations have been reported that affect gene expression or protein stability. These mutations can involve splice donor and acceptor sites such as $\alpha^{IVSI(-5nt)}\alpha$, polyadenylation site mutations such as α_2^{AATAAG} and termination codon mutations such as the common Hb Constant Spring variant that leads to extended unstable mRNA transcripts (Weiss and Liebhaber, 1994, Higgs, 2009a). Certain mutations within the genes themselves can give rise to haemoglobins such as Hb Quong Zse, Hb Adana and Hb Aghia Sophia that

are highly unstable, rapidly degraded and cause an α-thalassaemia phenotype (Goossens et al., 1982, Chan et al., 1997, Traeger-Synodinos et al., 1999).

Two extremely rare mutations result in an α-thalassaemia phenotype and have shed light on gene regulatory mechanisms. The α(α^{-ZF}) deletion removes α_1 and θ, and brings the downstream gene Luc7L into close proximity to α_2. Whilst α_2, its promoter elements and the upstream enhancers are intact, its expression is silenced due to an anti-sense mRNA from Luc7L transcribing through α_2 and aberrantly methylating intragenic CpG islands, a phenomenon subsequently seen in malignancy predisposition disorders such as Lynch Syndrome (Tufarelli et al., 2003, Barbour et al., 2000, Jeziorska et al., 2017). Secondly, a comprehensive SNP and RNA analysis of samples from Melanesian patients with α-thalassaemia who had intact α-globin genes and enhancers revealed a point mutation between ζ and pseudo-ζ that created a *de-novo* GATA1 binding site, and in so doing formed a promoter unit (De Gobbi et al., 2006). Lying between the upstream enhancers and the α-globin promoters, this new promoter is thought either to compete with the native promoters for the activity associated with the enhancers or to act as an enhancer blocker and thereby reduce α-globin transcription (Bozhilov et al., 2021).

1.5.2.1.5 ATRX and AT-MDS

In addition to the patients described previously with ATR-16, a rare group of male patients has been identified with more severe developmental delay, characteristic dysmorphic facial features and α-thalassaemia with HbH inclusions noted in the peripheral blood. This condition, ATR-X syndrome is caused by mutations in the chromatin remodeller ATRX on the

X-chromosome (Gibbons et al., 1995). ATRX has been implicated in various normal functions, including suppressing the alternative lengthening of telomeres pathway, DNA repair, histone 3.3 deposition and DNA methylation (Gibbons et al., 2000, Goldberg et al., 2010, Clynes et al., 2015). The α-thalassaemia phenotype is potentially explained by its role in resolving G-quadruplexes, which may be formed in G-rich VNTRs found at the α-globin locus (Law et al., 2010). ATRX mutations are also found in a form of acquired α-thalassaemia known as AT-MDS, which presents later in life (median age 68 years) in the context of myelodysplastic syndrome (MDS), a clonal haematopoietic disorder (Gibbons et al., 2003). These cases are usually accompanied by various other mutations commonly seen in MDS.

1.5.2.2 Cellular Pathophysiology

Due to the mutations described above, the primary problem in α-thalassaemia is the underproduction of α-globin chains. In normal erythropoeisis there is usually a mild excess in α-globin mRNA and protein when compared to β-globin. In α-thalassaemia excess unpaired γ-globin chains in the foetus form γ_4 tetramers (Hb Bart's) and in adult life excess β-globin chains form β_4 (HbH). Both Hb Bart's and HbH are very inefficient oxygen delivery molecules due to their high affinity for oxygen and inability to display any Bohr shift (Benesch et al., 1961).

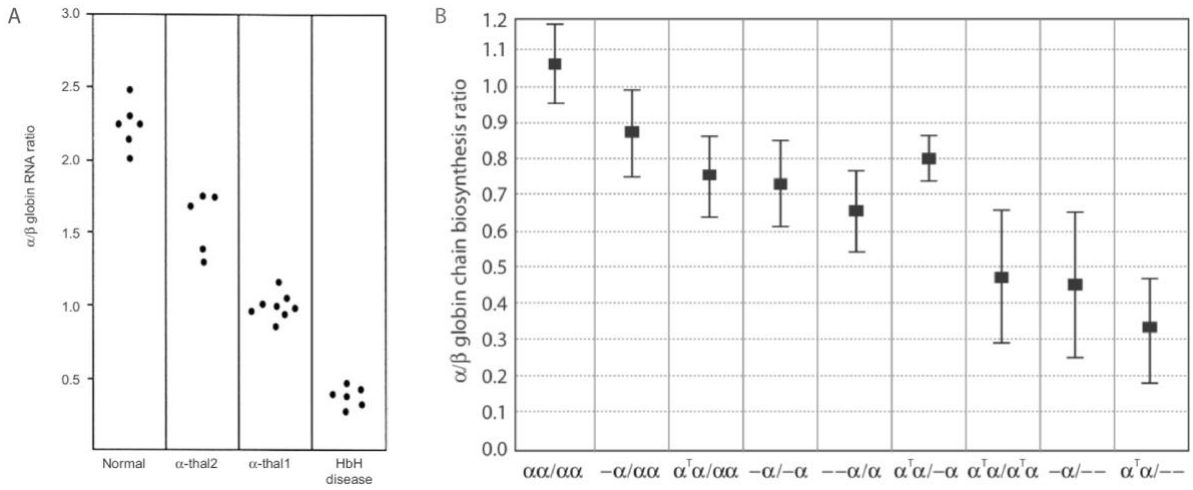


Figure 1.8 - Globin expression by genotype *A) α/β mRNA ratios from peripheral blood of normal individuals and carriers of 3 groups of deletional α-globin mutations. α-thal2 refers to a single α-globin deletion, α-thal1 to two α-globin gene deletions (nomenclature now not in use). B) α/β globin-chain biosynthesis ratios in normal individuals and carriers for deletional and non-deletional mutations. α^T denotes a non-deletional variant and - denotes a deletional variant. Adapted from (Steinberg, 2009)*

The most important consequence of the production of these homotetramers, particularly HbH, is that they precipitate and adhere to the cell membrane of circulating mature red blood cells. This results in oxidative damage which itself causes membrane phospholipid dysfunction. A consequence of these changes is that the membrane becomes more rigid, and an increased exposure of phosphatidylserine and Immunoglobulin G on the cell surface membrane (Schrier et al., 1989). Due to reduced membrane deformability, these abnormal red cells have an increased transit time through the reticuloendothelial systems of the body, particularly the spleen, and are cleared through erythrophagocytosis by resident tissue macrophages (Higgs, 2009b). In contrast to β-thalassaemia, this extravascular haemolysis is the predominant mechanism for the occurrence of anaemia in α-thalassaemia. Ineffective erythropoeisis does occur however, as HbH instability results in oxidation and precipitation in erythroblasts leading to necrosis, apoptosis and expansion of the erythroid compartment in the bone marrow (Papassotiriou et al., 1998).

1.5.2.3 Clinical manifestations

The clinical phenotypes of individuals with α-thalassaemia can be split into three broad categories: α-thalassaemia trait, HbH disease and Hb Bart's Hydrops foetalis. The majority of individuals with an α-thalassaemia mutation have a very mild phenotype and their diagnosis may often be made on incidental blood tests or screening procedures. However as can be expected for a condition whose mutations almost reach fixation in certain parts of the world, the interaction between mutations may produce severe or even lethal phenotypes (Piel and Weatherall, 2014).

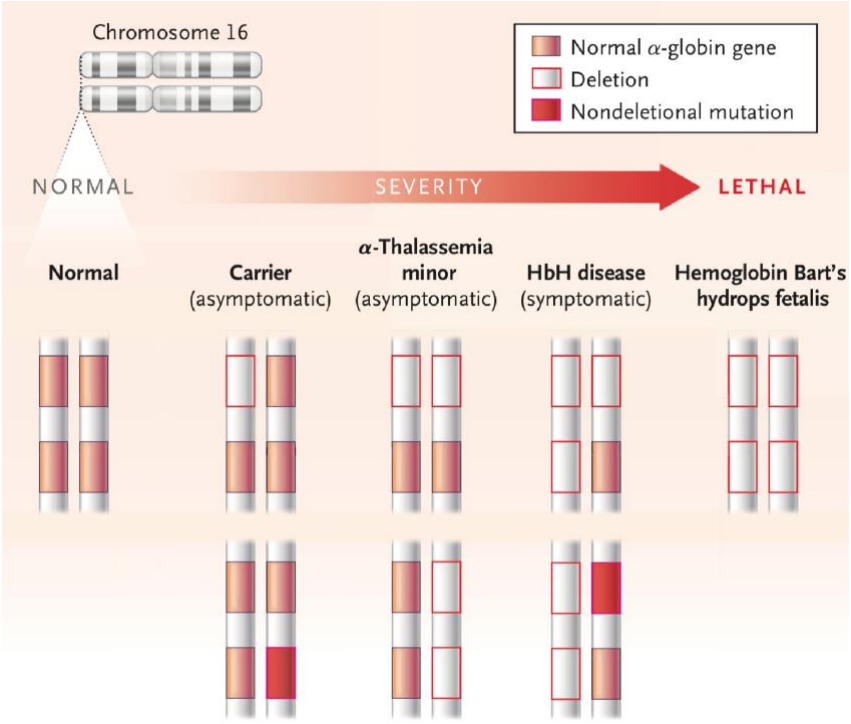


Figure 1.9 - Genotypes of the major α-thalassaemia disorders *The relationship between α-globin deletion and non-deletional mutations is depicted; non-deletional mutations tend to be more severe, and so the coinheritance of two gene deletions and a non-deletional mutation can result in Bart's hydrops fetalis syndrome. Adapted from (Piel and Weatherall, 2014)*

α-thalassaemia trait is most commonly caused by single (-α/αα) and double deletions of the α-globin genes, be it --/αα or -α/-α, and less commonly in those with a single non-deletional variant ($\alpha^T\alpha/\alpha\alpha$). Individuals with this condition are clinically asymptomatic, but their carrier status is potentially of relevance for pregnancy screening purposes and also for their modifying effect on severe β-thalassaemia as discussed in 1.6. Of the 5% of individuals worldwide who are estimated to carry an α-thalassaemia variant, the majority will be in this group (Piel and Weatherall, 2014).

HbH disease is more infrequent, encompassing those with a compound heterozygous deletion (e.g -α/--) or homozygotes for a moderately severe non-deletional variant ($\alpha^T\alpha/\alpha^T\alpha$). The condition is marked by anaemia (26-133 g/L) and HbH detected in the peripheral blood (0.8-40%). As suggested by the ranges of both parameters HbH disease is quite diverse clinically. Symptomatic patients tend to have episodes of haemolytic anaemia that may be severe, requiring transfusion. These episodes may be precipitated by an intercurrent infection or drug administration, particularly sulphonamides (Weatherall and Clegg, 2001). Similarly, episodes of jaundice are common, and around 40% of patients have gallstones (Au et al., 2005). Hepatosplenomegaly is a common feature; the splenomegaly in particular can be severe and cause hypersplenism (Higgs, 2009b).

Disease phenotypes can often be predicted based on the particular deletions involved in the interaction. Of importance, individuals with non-deletional forms tend to have a more severe phenotype as noted in Figure 1.8b (Higgs, 2009b). This may be for a number of reasons. In single-gene deletional α-thalassaemia the remaining gene on the allele displays a compensatory increase in expression, a phenomenon not observed in non-deletional

thalassaemia (Liebhaber et al., 1985). This is compounded by the fact that most non-deletional mutations involve α_2 which is expressed roughly three times as much as α_1. Furthermore, some non-deletional variants result in highly unstable chains that are rapidly degraded post-translation (Wajcman et al., 2008, Higgs, 2009a). Again, most of these unstable variants affect α_2 rather than α_1.

The most severe form of α-thalassaemia is the Hb Bart's Hydrops foetalis syndrome (BHFS). Most commonly it occurs due to the inheritance of no α genes (--/--), though it can arise due to the inheritance of two severe non-deletional alleles ($\alpha^T\alpha/\alpha^T\alpha$) or a compound heterozygote for a severe non-deletional allele and an allele bearing no α genes ($\alpha^T\alpha$/--). The only oxygen carrying molecule that keeps affected foetuses alive *in-utero* beyond a few weeks gestation is Hb Portland ($\zeta_2\gamma_2$), and the condition is marked haematologically by the high amounts of Hb Bart's (γ_4) and HbH (β_4) that are detected. Clinically, the foetuses exhibit prolonged intrauterine oedema and resultant signs of cardiac failure including generalised oedema, pleural and pericardial effusions (Higgs, 2009b). Until recently BHFS was thought to be a universally fatal disorder, but with improvements in screening and antenatal care, an early diagnosis followed by intra-uterine transfusions can keep foetuses alive until term (Beaudry et al., 1986). 26% of the cohort of an international registry of BHFS patients have survived well into their second decade (Songdej et al., 2017). Congenital abnormalities are common, with 48% of survivors having urogenital defects, particularly males, and 16% with limb abnormalities. This may be due to the prolonged period of hypoxia *in-utero*. Remarkably, 45% of the foetuses untransfused *in-utero* were born without hydropic features, potentially due to persistent ζ-globin expression. This is an area of detailed and on-

going study, and may provide therapeutic options for such patients in years to come (King and Higgs, 2018).

1.5.3 β-Thalassaemia

1.5.3.1 Molecular basis

Mutations resulting in quantitative reductions in β-globin cause β-thalassaemia. Around 300 mutations have been identified, which can be broadly described functionally (Thein and Wood, 2009). Individuals with a β^0 mutations are unable to produce any β-globin from the affected allele, and β^+ or β^{++} describe mutations allowing some degree of expression. The vast majority of these classes of variants involve point mutations or small insertions and deletions involving the gene or flanking sequences. Mutations can cause defects in transcription initiation, premature termination/protein truncation, nonsense mediated decay, RNA processing and RNA translation (Thein, 2013, Giardine et al., 2014). In contrast to α-thalassaemia, large deletions are relatively rare and range from ~100bp to 65kb in size. Some large deletions also selectively involve the regulatory elements, and as with MCS-R2, they all involve the enhancer β-LCR2. Mutations in *trans*-acting factors such as XPD, GATA1 and SUPT5H have been reported to cause β-thalassaemia, always as part of a broader disorder (Viprakasit et al., 2001, Yu et al., 2002, Achour et al., 2020). Certain mutations are more common amongst β-thalassaemia patients, and these may be more prevalent in differing geographical locations (Figure **1.5**).

1.5.3.1.1 HbE/β-thalassaemia

Structurally abnormal β-globin variants can cause a reduction in production, giving rise to thalassaemic haemoglobinopathies. Prominent amongst them is the *HBB* codon 26 G>A point mutation resulting from the conversion of the wild-type codon GAG to AAG (Minnich et al., 1954, Chernoff et al., 1954). The resultant structural variant, HbE, is potentially mildly unstable in-vitro (Itano et al., 1954, Frischer and Bowman, 1975). This effect is aggravated in the presence of excess α-globin chains, and countered by the presence of α-haemoglobin stabilising protein (Strader et al., 2016). This instability may not be encountered *in-vivo,* unless in the context of fever, at least in the red cells of non-severe HbE/β-thalassaemia patients where the α-globin excess may not be severe (Rees et al., 1998). More importantly however, this mutation activates a cryptic splice site in exon 1 which competes with the normal site and leads to the transcription of an abnormally spliced non-functional β^E transcript that omits part of exon 1 and generates a novel stop codon (Traeger et al., 1980, Orkin et al., 1982). A reduced amount of correctly spliced β^E mRNA is produced, resulting in a β-thalassaemia phenotype.

The β^E variant is very common in the Indian subcontinent and South-East Asia; In parts of northern Thailand and Cambodia known as the "haemoglobin E triangle" up to 70% of the population are carriers for the mutation (Williams and Weatherall, 2012). Due to the relatively high population prevalence of other β-thalassaemia mutations, the compound heterozygous condition HbE/β-thalassaemia occurs very frequently; in Thailand alone an estimated 100,000 individuals have this disorder with 3000 children born with the condition every year (Flint et al., 1998). Characteristically a clinically heterogenous condition, it

comprises approximately 50% of the world's severe β-thalassaemia cases (Modell and Darlison, 2008).

1.5.3.2 Cellular pathophysiology

The cellular pathophysiological features of β-thalassemia are caused by the relative excess of unpaired α-globin chains in erythroid tissues due to a reduction in β-globin production (Weatherall et al., 1965, Nathan and Gunn, 1966, Fessas et al., 1966). There are two mechanisms that both result in excessive erythroid cell destruction: ineffective erythropoeisis in the bone marrow and extravascular haemolysis of mature red blood cells, primarily in the spleen (Rund and Rachmilewitz, 2005).

Throughout development the relative amounts of α- and β-globin chains are closely linked to avoid any deleterious imbalances. In normal developing red blood cells, free α-globin not bound to β-globin pairs with its cellular chaperone α-haemoglobin stabilising protein (AHSP), which stabilises the chains by promoting folding and preventing aggregate formation (Yu et al., 2007, Kihm et al., 2002, Feng et al., 2005). In normal erythropoeisis AHSP can bind and sequester unpaired α-globin chains reversibly if β-globin chains are not available (Zhou et al., 2006).

In β-thalassaemia, due to reduced β-globin chain production, unpaired α-globin chains exceed AHSP's binding capacity. Free α-globin is highly unstable and forms large aggregates that can be detected by microscopy. α-globin normally undergoes auto-oxidation in the cell forming methaemoglobin and superoxide (Nagababu et al., 2008, Rifkind et al., 2014). This

process occurs on a much larger scale due to the large excess in unpaired chains, leading to the formation of α-globin monomers containing ferric iron (known as α-hemichromes) and the degradation of haem and release of iron. These can then take part in redox reactions which form reactive oxidative species (ROS)(Voon and Vadolas, 2008). The build-up of ROS can lead to erythroid cell destruction in two major ways.

Firstly, a major consequence of ROS formation is the oxidation of membrane lipids and proteins including band 4.1, spectrin and transmembrane band 3, which are also bound by α-hemichromes (Advani et al., 1992, Yuan et al., 1994). This leads to clustering of Band 3, which can cause significant membrane deformation, exposure of neoantigens and eventual extravascular haemolysis of mature red blood cells. (Advani et al., 1992, Yuan et al., 1992).

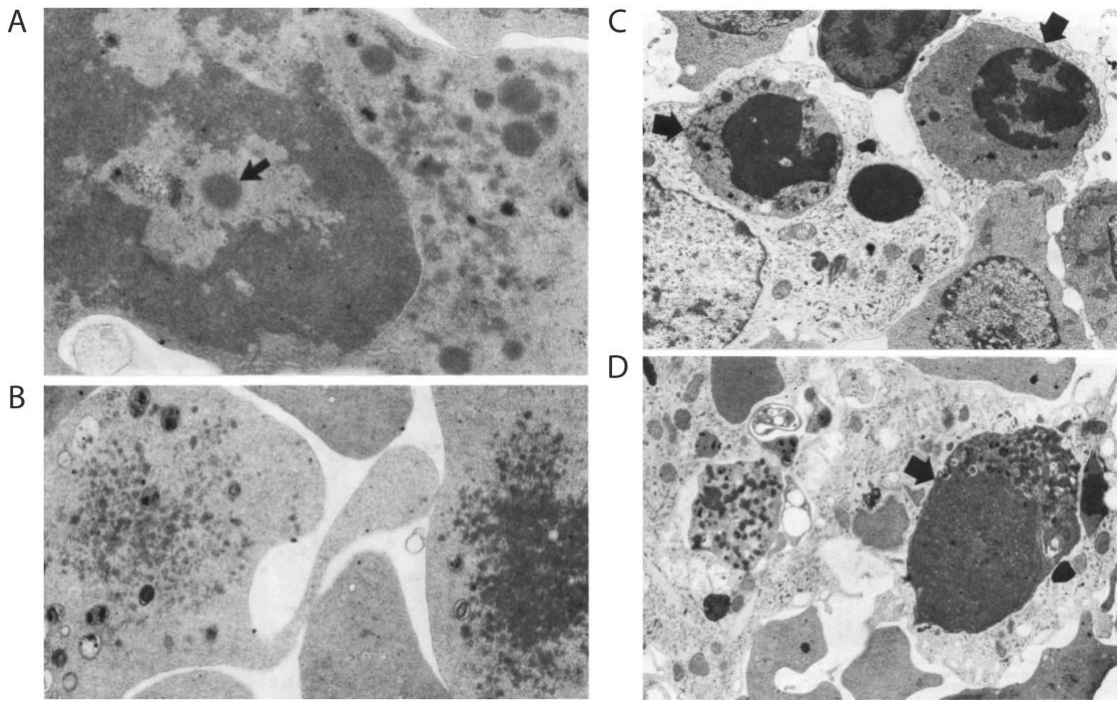


Figure 1.10 - Electron micrographs of erythroid cells from a patient with HbE/β-thalassaemia (A) Polychromatic erythroblast from a patient's bone marrow, showing cytoplasmic and nuclear (arrow) α-globin precipitates. (B) Reticulocytes from peripheral blood showing cytoplasmic precipitates and iron-laden mitochondria (C and D) Macrophages in the bone marrow that have phagocytosed polychromatic erythroblasts, indicated with the arrows. Adapted from (Wickramasinghe et al., 1981)

Secondly, and more prominently in β-thalassaemia, ineffective erythropoeisis is caused by accelerated apoptosis. There is a massive expansion of the erythron, but as many as 80% of these cells undergo premature death (Pootrakul et al., 2000). This seems to peak at the polychromatophilic stage onwards, with bone marrow cells from patients with β-thalassaemia having approximately fifteen times the number of apoptotic cells at the polychromatophilic and orthochromatic erythroblast stages compared to normal controls (Mathias et al., 2000, Centis et al., 2000). The mechanism for this is incompletely understood, but it may involve cell surface interactions of the pro-apoptotic fas cell surface death receptor (FAS) with FAS-ligand, which is upregulated in response to high levels of ROS. Interestingly, ROS upregulate the expression of growth differentiation factor 11 (GDF11), an activin receptor IIA (ActRIIA) ligand (Dussiot et al., 2014). Use of an activin-receptor ligand trap targeting GDF11 decreased ROS and α-globin aggregates at the cell membrane, suggesting it is part of a positive-feedback loop, although recent data suggest that GDF11 may not be the target of these traps (Dussiot et al., 2014, Guerra et al., 2019). Finally, free α-globin chains interact directly with heat shock protein 70 (HSP70), which can no longer chaperone one of its molecular targets, the key erythroid transcription factor GATA1. GATA1 undergoes cleavage by Caspase-3, leading to maturation arrest and apoptosis (Arlet et al., 2014).

The ineffective erythropoeisis and extravascular haemolysis pathways both result in anaemia. Anaemia leads to a compensatory response to increase the erythroid cell mass. This occurs through anaemia causing hypoxia, which is detected by the oxygen sensing component of hypoxia inducible factors (HIFs), a group of transcription factors. HIF-2 in

particular stimulates the production of erythropoietin (EPO) in renal peritubular interstitial fibroblasts (Haase, 2013). EPO causes phosphorylation of JAK2 through its receptor on erythroid cells, leading to an increase in erythroid progenitors and increase in overall erythroid cell mass (Libani et al., 2008). Unfortunately, as explained in the previous section this leads to further ineffective erythropoeisis and haemolysis.

In normal erythropoeisis, an increase in red cell production would be accompanied by a rise in iron absorption, iron being a key component of haemoglobin. The same occurs in β-thalassaemia; the massive expansion in the erythroid compartment results in the suppression of hepcidin, likely through the action of a number of factors such as growth differentiation factor 15 (GDF15), erythroferrone (ERFE) and twisted gastrulation (TWSGI) (Pasricha et al., 2016). Hepcidin is a small peptide hormone secreted by the liver, which binds to the iron transporter ferroportin and causes its removal from the gut enterocyte membrane. As iron is absorbed by enterocytes through ferroportin, hepcidin prevents iron absorption. Its inhibition in β-thalassaemia thus leads to excessive iron absorption and iron overload (Pagani et al., 2019). Furthermore, patients with severe β-thalassaemia receive blood transfusions which worsen the iron overload. Humans have no mechanisms to excrete iron, and lose ~1mg a day. A single unit of blood typically results in 200mg iron being delivered to tissues through red cells (Rachmilewitz and Giardina, 2011). Whether by increased absorption as a result of ineffective erythropoeisis or treatment related side effects, iron overload causes many of the clinical problems faced by patients with severe β-thalassaemia as discussed below.

1.5.3.3 Clinical manifestations

β-thalassaemia can be clinically divided into three categories, β-thalassaemia minor/trait, β-thalassaemia intermedia, and β-thalassaemia major/severe β-thalassaemia, although the accepted clinical classification is now into transfusion dependent and non-transfusion dependent thalassaemia (NTDT) (Taher et al., 2021). To delineate different forms of NTDT I will use the older classification in this work. These clinical phenotypes are generally linked to the genotype, although some conditions such as HbE/β-thalassaemia can be heterogenous and genetic modifiers can alter the expected course of the disease.

β-thalassaemia trait is mainly caused by heterozygous β-thalassaemia mutations. Individuals in this carrier state have minimal haematological abnormalities and may only be detected incidentally on routine blood testing. Interestingly a recent study showed that despite having a normal haemoglobin, individuals with β-thalassaemia trait may experience some symptoms of anaemia, and have an increased risk of developing infections (Premawardhena et al., 2008).

Severe β-thalassaemia on the other hand is due to homozygosity or compound heterozygosity for two severe β-thalassaemia mutations. Patients in this category have severe anaemia that requires lifelong blood transfusions. Anaemia usually develops in the first few months of life, due to the foetal-to-adult haemoglobin switch that occurs as a result of the normal repression of γ-globin transcription. Children may present with poor feeding, pallor, failure to thrive and difficulty in breathing (Weatherall and Olivieri, 2009). Due to the massive erythroid compartment expansion due to ineffective erythropoeisis, children later

develop frontal skull bossing, abnormalities of the long bones and hands, hepatosplenomegaly and occasionally extramedullary haematopoiesis (Nienhuis and Nathan, 2012).

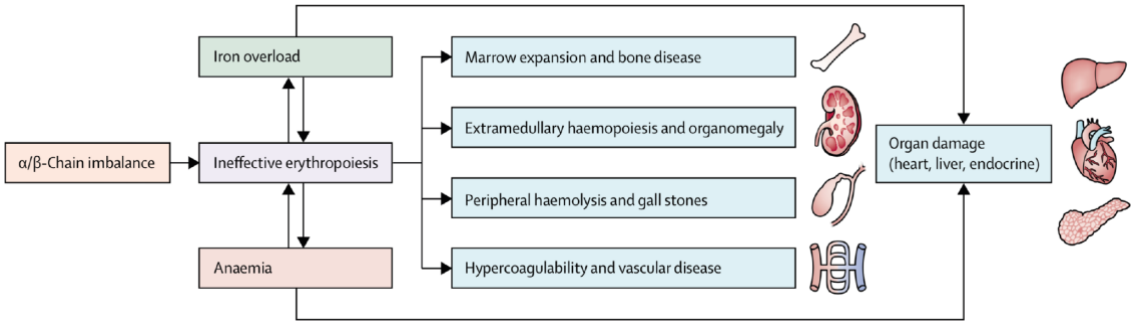


Figure 1.11 - Clinical manifestations of β-thalassaemia driven by underlying pathophysiology
Adapted from (Higgs et al., 2012)

High-output cardiac failure is a common sequela of untreated anaemia, and a major cause of morbidity and mortality. These clinical features are now mainly seen in certain low- and middle-income countries due to its eradication in developed countries as a result of advances in antenatal and post-natal screening and diagnosis. Even in high-income countries with advanced healthcare systems, as patients move into their second and third decades they are at risk of developing significant complications of iron overload which causes most of the morbidity and mortality of the disease (Rund and Rachmilewitz, 2005). Iron is deposited in visceral parenchymal tissues and can result in a wide variety of problems such as cardiomyopathies, chronic hepatitis and cirrhosis, growth delay, hypogonadism, hypothyroidism, hypoparathyroidism, pituitary and adrenal insufficiency and insulin dependent diabetes mellitus (Cao and Galanello, 2010).

In the untreated or undertreated setting, patients often die due to these complications in their first or second decades. With the advent of effective iron chelating drugs, life expectancies have improved dramatically in high and middle income settings, whilst still remaining low elsewhere in the world (Brittenham et al., 1994, Vitrano et al., 2017, Dhanya et al., 2020).

The term 'β-Thalassaemia intermedia' was first used to describe patients with disease severity in between severe β-thalassaemia major and β-thalassaemia minor (Sturgeon et al., 1955). The diagnosis is clinical, and most patients have compound heterozygosity for two β-thalassaemia mutations. Individuals present later in life (2-6 years), and generally only need intermittent transfusions (Higgs et al., 2012). Due to inefficient erythropoeisis and resultant hepcidin suppression patients become significantly iron overloaded, even if they have never been transfused, requiring careful monitoring and treatment. Other significant contributors to morbidity and mortality are hypercoagulability resulting in peripheral and central arterial and venous thromboses, pulmonary hypertension, leg ulcers, gallstones and extramedullary haematopoiesis (Musallam et al., 2012b).

1.5.3.4 Current standard of care

Until very recently treatment for β-thalassaemia and the haemoglobinopathies was in general supportive. Patients with severe β-thalassaemia are given regular red blood transfusions to keep the lowest level of haemoglobin concentration above 90-105g/L (Trompeter and Cohen, 2014). This alleviates hypoxia resulting from the anaemia, which in turn allows better growth and development, reduces the stimulus for bone marrow

erythroid expansion and hepatosplenomegaly. This approach is effective but is ultimately limited due to its complications. As in any patient group, transfusions carry side effects and risks from the common to rare, including acute severe allergic, haemolytic and anaphylactic reactions, a range of bacterial and viral infections, cardiac overload and transfusion associated lung injury (Norfolk et al., 2013, Goel et al., 2019). As mentioned previously, iron overload is a prominent risk both due to increased gut and parenteral absorption. In recent decades iron chelation drugs, namely desferrioxamine, deferiprone and deferasirox, have been introduced and improved upon to mitigate overload. As mentioned in 1.5.3.3 in certain contexts they have made dramatic improvements in morbidity and mortality. Unfortunately, desferrioxamine can only be given parenterally, deferiprone is less efficacious and can cause arthropathy and agranulocytosis, and deferasirox currently has a high purchase cost (Rund and Rachmilewitz, 2005).

Recently activin receptor ligand traps have been licensed in the UK and EU for use in severe β-thalassaemia (Cappellini et al., 2019, Cappellini et al., 2020). Both sotatercept and luspatercept reduce the transfusion burden in patients with β-thalassaemia (Cappellini and Taher, 2021).

Ultimately despite improvements in the quality of care and recent drug breakthroughs, the supportive care on offer to β-thalassaemia patients results in a significant but limited improvement in their quality of life, which still leaves them with numerous complications of the disease and treatment sequelae to manage.

Allogeneic haematopoietic stem cell transplant is the only established curative option for β-thalassaemia. First performed in 1981, over 3000 have now been carried out worldwide (Thomas et al., 1982, Angelucci et al., 2014). If a patient is transplanted with a human leukocyte antigen (HLA) matched sibling donor (MSD), early in childhood at an experienced centre the outcomes are excellent, with over 90% overall survival reported in one Europe wide 10-year retrospective study (Baronciani et al., 2016). However only 30% of patients with severe β-thalassaemia have an MSD, and outcomes outside of this setting are significantly inferior (Angelucci et al., 2014). Even with a MSD, patients with successful transplants can have long term side effects, particularly growth and endocrine insufficiencies and secondary malignancies. The latter tend to be solid organ cancers, and can arise 18-25 years after the transplant, highlighting the conditional nature of the cure on offer (La Nasa et al., 2013, Rahal et al., 2018, Santarone et al., 2018).

As might be expected such an intensive treatment requires careful cell preparation and excellent supportive care for adequate delivery. The costs and facilities required for this - approximately 100,000-150,000 US dollars in the US and Europe - are out of reach for many countries with a sizeable disease burden (Angelucci, 2010, Angelucci et al., 2014). Taking these factors into consideration allogeneic HSCT cannot be considered a feasible therapeutic option for the majority of patients with β-thalassaemia.

1.6 Genetic modifiers of severity

Whilst β-thalassaemia is a monogenic autosomal recessive disorder, there is a large degree of phenotypic variability, even though approximate genotype-phenotype predictions can often be made. This is due to the variability in genetic modifiers that act in various ways.

Even though β-thalassaemia is a monogenic disorder, as mentioned in 1.5.3.1 over 300 mutations at the β-globin locus have been identified to date. These mutations reduce β-globin production variably, and so the primary layer of genetic modification is the disease-causing variant itself and how it interacts with its counterpart on the other allele. Clearly homozygosity for two different β^0 mutations is more likely to result in severe β-thalassaemia (Ho et al., 1998, Winichagoon et al., 2000). However other powerful ameliorating genetic factors may also be present that can attenuate the phenotype of even this severe genotype (Cao and Galanello, 2010).

As discussed in 1.5.3.2 the primary pathophysiological driver of disease in β-thalassaemia is the excess pool of α-globin in developing erythroblasts. It was therefore hypothesised early on that coinheritance for α-globin mutations in those with β-thalassaemia could act as an ameliorating factor (Kan and Nathan, 1970). As can be noted in Figure **1.5** there is a significant overlap between areas of relatively high prevalence for α- and β-thalassaemia mutations in the general population. Many individuals therefore coinherit mutations at both gene clusters. In the Dongguan region of southern China and in Northern Thailand for example, 0.5% and 0.6% of the population co-inherit mutations at both gene loci (Peng et al., 2020, Mankhemthong et al., 2019)

Careful clinical studies over decades demonstrated that the presence of α-thalassaemia acts as a powerful genetic modifier of disease. Whilst in these patients the absolute amount of transcription at the relative loci is reduced, the overall balance between globin-chains is ameliorated or restored. For example, in the case of a patient with a β^{+}/β^{0} and --/-α genotype, the α/β ratio was almost perfectly balanced (0.96)(Kanavakis et al., 2004). This could potentially result in a milder or even asymptomatic phenotype compared to patients with severe β-thalassaemia who have an intact α-globin locus.

The number of α-globin deletions that could potentially improve the β-thalassaemia phenotype range from 1 to 3. Of these, -α/αα (one gene deletion) and --/αα or -α/-α (two gene deletions) are beneficial (Kan and Nathan, 1970, Loukopoulos et al., 1978, Wainscoat et al., 1983a, Wainscoat et al., 1983b, Furbetta et al., 1983, Oggiano et al., 1988, Galanello et al., 1989, Weatherall et al., 1981, Camaschella et al., 1995, Winichagoon et al., 2000). Depending on the β-globin mutation, even three α-globin gene deletions can result in an intermedia phenotype in patients with severe β-thalassaemia genotypes (Wainscoat et al., 1983a). It has generally been held that this balancing effect of α-globin mutations works in β-globin mutations other than those resulting in homozygous β^{0} disease. However even in this instance, depending on the β-thalassaemia mutation reduction in the amount of α-globin can still have a powerful ameliorating effect (Oggiano et al., 1988, Galanello et al., 2009).

The modifying effect of α-thalassaemia is most pronounced in patients with HbE/β-thalassaemia (Olivieri et al., 2010). Coinheritance of HbE, α- and β-thalassaemia was noted by early researchers, and its modifying effect was clear (Wasi et al., 1967, Winichagoon et

al., 1985). Subsequent small and large prospective and retrospective studies have consistently shown that one and two α-globin gene deletions are associated with a much milder disease than having the normal full complement (Winichagoon et al., 2000, Bandyopadhyay et al., 2001, Sripichai et al., 2008b, Sharma et al., 2009, Suresh et al., 2013, Premawardhena et al., 2005a, Charoenkwan et al., 2014, Siriworadechkul et al., 2014, Yamsri et al., 2016, Kalantri et al., 2020). These studies show that coinheritance of α-thalassaemia is associated with a milder or even asymptomatic phenotype, with significant improvements in steady state haemoglobin, transfusion requirements, age at diagnosis, growth and sexual development and liver and spleen size (Sripichai et al., 2008a). In certain parts of the so-called 'haemoglobin-E triangle' and the Indian subcontinent, up to 14% of the local population are carriers for α-thalassaemia determinants. Interestingly, in several HbE/β-thalassaemia patient cohort studies the frequencies of α-thalassaemia deletions tends to be far less than in the matched local population (Winichagoon et al., 1985, Suresh et al., 2013). A Thai analysis of 925 patients with HbE/β-thalassaemia found only 0.22% of the cohort had α-thalassaemia, despite the gene frequency in the local Thai population being 3.6-10% (Sripichai et al., 2008b). A potential explanation for this striking discrepancy is that individuals with two α-globin gene deletions have such a mild phenotype that they are not diagnosed and thus do not attend hospital. Patients in these studies that do have α-globin deletions tend to almost completely be restricted to groups with mild disease; patients with severe disease conversely seldom have concurrent α-thalassaemia (Winichagoon et al., 2000, Bandyopadhyay et al., 2001, Kalantri et al., 2020).

Whilst having α-globin gene deletions can lead to phenotypic improvement in β-thalassaemia, the converse is also true; individuals who inherit more than the normal four

copies of the α-globin gene have a much more severe phenotype (Camaschella et al., 1995, Sripichai et al., 2008b, Sharma et al., 2009, Suresh et al., 2013, Origa et al., 2014, Perera et al., 2019, Theodoridou et al., 2020). This can be the case even if the patient has a heterozygous β-thalassaemia mutation that would otherwise cause asymptomatic β-thalassaemia trait (Premawardhena et al., 2005b, Gurunathan et al., 2020). The potency of this effect is directly due to a harmful increase in the cellular excess of α-globin chains, and thus more marked ineffective erythropoeisis. In normal erythropoeisis these excess chains are stabilised by AHSP, but it is not clear whether this is a genetic modifier of disease severity. A Thai study of patients with HbE/β-thalassaemia found that no particular mutations in AHSP were associated with mild, moderate or severe disease (Viprakasit et al., 2004). However, a more recent study of 54 HbE/β-thalassaemia patients in India showed that specific mutations in exon 3 of *AHSP* were strongly associated with transfusion dependent thalassaemia (Ray et al., 2019).

The other major genetic modifier of disease severity is the co-inheritance of variants that restore chain balance by increasing the amount of γ-globin expression, which can pair with excess α-globin chains to produce HbF (α2γ2). γ-globin expression is normally silenced in the first 6 months of life, and disease onset correlates closely with the timing and degree of this switch (Nienhuis and Nathan, 2012). Due to ineffective erythropoeisis, the small number of F-cells present are afforded a potent selective advantage due to their relative chain balance, and these cells persist in most patients with severe β-thalassaemia. This applies even in patients who don't have any known genetic modifiers due to cell-to-cell differences at the transcriptional level (Khandros et al., 2020).

Hereditary Persistence of HbF (HPFH) can arise due to mutations at the β-globin cluster or at additional loci. At the cluster, numerous SNPs have been found that are associated with a lack of γ-globin repression, the most common of which is the XMN1 polymorphism (Wienert et al., 2018). This SNP at the -158bp site at the ^{G}γ-globin promoter is not thought to be the causal variant but rather in linkage with a mutation that is. Larger mutations, including those causing δβ-thalassaemia, and HPFH can underlie considerable persistence of HbF synthesis and thereby significantly ameliorate the phenotype of β-thalassaemia (Sankaran et al., 2011). What these mutations have in common is likely to be a disruption of functional non-coding sequences that are required for relative γ-globin silencing through transcription factor binding.

Genome wide association studies (GWAS) have revealed SNPs associated with a reduction in expression of certain candidate transcription factors. B-cell lymphoma/leukaemia 11A (BCL11a), which is located on chromosome 2 has been shown in clinical studies, functional *in-vitro* and *in-vivo* work to be a significant element in the γ-globin repressive process (Menzel et al., 2007, Sankaran et al., 2008, Lettre et al., 2008, Uda et al., 2008). BCL11a binds upstream to the γ-globin promoter at the -115 site (Liu et al., 2018a, Martyn et al., 2018). Similarly, Zinc Finger and BTB Domain Containing 7 (ZBTB7a)/Leukemia/Lymphoma-Related Factor (LRF) is also known to bind upstream to the γ-promoter and HPFH mutations at the -200bp site are thought to disrupt its binding (Martyn et al., 2018, Weber et al., 2020). Another variant, at the HBS1L-MYB intergenic region that results in lower MYB expression correlates with higher HbF levels (Thein et al., 2007). Finally, the key erythroid transcription factor KLF may modulate HbF levels through a gain of function mutation (Borg et al., 2010, Wienert et al., 2017).

The aforementioned variants have been described in family studies and validated *in-vitro*, but It is possible that there could be many more naturally occurring mutations in genetic modifiers of disease, as several determinants have been recently discovered through *in-vitro* screens. The erythroid-specific eIF2α kinase heme-regulated inhibitor (HRI) is a repressor of γ-globin expression; upregulating activating transcription factor 4 (ATF4), which itself activates BCL11a transcription. (Grevet et al., 2018, Peslak et al., 2020, Huang et al., 2020). BCL11a is thought to exert its repressive effect through recruitment of the nucleosome remodelling and deacetylase (NuRD) complex (Sankaran et al., 2008, Xu et al., 2013). Engineered mutations in chromodomain helicase DNA binding protein 4 (CHD4), a component of the NuRD complex, and its regulator Zinc Finger Protein 410 (ZNF410) have been shown to increase HbF levels (Sher et al., 2019, Lan et al., 2020). As these variants derepress γ-globin without compromising cell fitness, similar naturally occurring variants in these genes may exist that have not been identified as yet.

The underlying mechanistic feature common to all of these genetic modifiers of β-thalassaemia severity is the reduction of excess α-globin chains in red blood cells. Bearing this in mind, most novel therapeutic strategies for β-thalassaemia seek to harness what we have learned from these variants to restore α- and β-globin chain balance in patients with thalassaemia. These observations form the basis of the approach taken in this thesis to provide an effective functional cure for β-thalassaemia.

1.7 Potential Approaches to restore chain imbalance

1.7.1 HbF induction

Several different modalities have been trialled and are in the process of being tested for HbF induction. Pharmacological agents like hydroxyurea have long been used in sickle cell disease and β-thalassaemia for this purpose (Fibach et al., 1993, Bradai et al., 2003). 5-Azacytidine, decitabine, sodium phenylbutyrate, erythropoietin and recently the JAK2 inhibitor ruxolitinib have also been used but with limited or variable results (Musallam et al., 2013, Taher et al., 2018). Histone deacetylase (HDAC) inhibitors and the immunomodulatory agent pomalidomide may have efficacy in inducing HbF (Shi et al., 2013, Dulmovits et al., 2016, Mettananda et al., 2019, Khamphikham et al., 2020). Unfortunately, these drugs have a wide side-effect profile, so their clinical use is limited.

Lentiviral gene therapy and gene editing approaches to induce HbF have shown more promise. Erythroid-specific knockdown of BCL11a, either through stable integration of a short hairpin RNA or through CRISPR/Cas9 induced disruption of the BCL11a enhancer robustly induced HbF production in preclinical and phase 1/2 studies (Brendel et al., 2016, Canver et al., 2015, Wu et al., 2019, Esrick et al., 2020, Frangoul et al., 2021). Other strategies have been applied to reduce γ-globin repression through direct manipulation of transcription factor binding sites at the γ-globin promoters, which have been successful in *in-vitro* and non-human primate studies (Traxler et al., 2016, Humbert et al., 2019, Weber et al., 2020). Whilst these approaches are promising, longer term safety and efficacy data is required before expanding these approaches for wider use.

1.7.2 Gene addition

The correction of globin chain imbalance by adding wild-type or corrected β-globin variants has been a goal in the field since the 1980s (Mercola and Cline, 1980). Researchers developing this approach have been at the forefront of the developing gene-therapy field, and the work they have done characterising the β-globin LCR, improving viral vectors and overcoming significant safety setbacks have been reviewed in detail elsewhere (Nienhuis and Persons, 2012). Over the last decade patients have been successfully treated by a number of groups using lentiviral vectors, most recently in a multi-centre international phase 3 trial using a product manufactured by Bluebird Bio (Cavazzana-Calvo et al., 2010, Locatelli et al., 2018). In the disease-genotypes treated, the results have been encouraging. Longer-term results are awaited; theoretical safety concerns still persist due to the semi-random nature of lentiviral insertion patterns which could lead to clonal HSC selection. In addition, the extremely high-cost of the treatment - over a million US dollars for the Bluebird bio product - currently puts it out of the reach of most patients, even in high-income countries. A potentially safer and more cost-effective approach is to use non-viral gene-editing methods to deliver a β-globin template into HSCs for homology directed repair (HDR) mediated integration into the native locus. Currently HDR rates are too low to be clinically useful, but technological advances may make this approach feasible in the future (Dever et al., 2016, DeWitt et al., 2016, Vakulskas et al., 2018, Cromer et al., 2021).

1.7.3 Targeted reduction of α-globin strategy

Reduction in α-globin expression to reduce to the pool of free α-globin chains has the potential to cure β-thalassaemia as it directly addresses the cause of the cellular pathophysiology (Mettananda et al., 2015). Given the epigenetic regulation of α-globin transcription as outlined in 1.4.3, a drug screen was performed to search for compounds that might selectively silence α-globin. IOX1, a histone demethylase inhibitor, significantly improved the α/β ratio. However, whilst differentiation and cellular viability was maintained, proliferation of cells in erythroid culture was reduced by 40%, which limits its potential in future clinical studies (Mettananda et al., 2017b).

Genome engineering strategies offer an attractive alternative. Early experimental murine work suggested that this could be a tractable approach. Heterozygote mice for β-thalassaemia mutations displayed severe anaemia that was rescued by introducing a deletion removing two α-globin genes ($\alpha^{++/--};\beta^{+/-}$)(Voon et al., 2007). These double knock-out mice showed near normal red blood cell indices. Due to the high sequence homology between the four α-globin genes, a targeted deletion trying to remove one or two genes may yield unpredictable results, with potentially all four being removed. One approach is to attempt to use a single double strand break in the 3' UTR of *HBA2* to replace this gene with a WT β-globin gene (Cromer et al., 2021). The work leading up to this thesis aimed instead to delete the major α-globin enhancer MCS-R2. Rare mutations that selectively remove just the enhancer in homozygosity - annotated as $(\alpha\alpha)^{ALT}/(\alpha\alpha)^{ALT}$ - show that such individuals have normal growth and development (Sollaino et al., 2010, Phylipsen et al., 2010, Coelho et al., 2010, Wu et al., 2017, Huang et al., 2017). Deletion of the enhancer in these patients

is only associated with a reduction in α-globin production. These small deletions were recapitulated in CD34+ cells using CRISPR/Cas9 in wild type and HbE/β-thalassaemia patient derived cells (Mettananda et al., 2017a). A robust knockdown of α-globin transcription and the α/β ratio was achieved, showing the strategy in principle was technically viable. Whilst this may be the case, homozygous deletion of the enhancer resulted in too severe a reduction in α-globin transcription, which would give the patient severe α- as well as severe β-thalassaemia. A milder and more graduated reduction in α-globin expression would be desirable, discussed further in chapter 4.

1.7.4 Generating a thalassaemia trait genotype

With very few exceptions (co-inherited excess α-globin genes and dominant β-thalassaemia mutations), β-thalassaemia follows a classical autosomal recessive inheritance pattern, requiring a mutation on both alleles for disease to be present (Luo and Chui, 2016). Whilst two severe mutations generally yields a severe phenotype, as mentioned previously having only one results in β-thalassaemia trait (Nienhuis and Nathan, 2012). Numerous cohort studies comprised of various genotypes and ethnic backgrounds demonstrated that carriers for a single mutation have a normal or very mildly lowered haemoglobin concentration, mild microcytosis and hypochromia, and are thus to all intents and purposes phenotypically normal (Rosatelli et al., 1992, Yamsri et al., 2011, Saleh-Gohari et al., 2015, Singha et al., 2019). Therefore, a powerful strategy with functionally curative potential would be to change a single pathogenic mutation on one allele *in-situ* back to its wild-type sequence or a normal variant. Whilst this premise seems straightforward, technological considerations have limited its use. Various non-viral techniques have been utilised to repair mutations,

including with homology directed repair (HDR) of double-strand breaks (DSBs) using CRISPR-Cas9. Whilst there has been success *in-vitro*, the efficiencies achieved are below those that would be required to generate an effective cure, or the techniques used would not be appropriate in the clinic (Li et al., 2001, Lonkar et al., 2009, Ma et al., 2013, Xie et al., 2014, Song et al., 2015, Xu et al., 2015, Wattanapanitch et al., 2018). In certain very limited circumstances CRISPR-Cas9 can be used to highly-efficiently disrupt specific splice-site abberations, but as this involves semi-random error-prone non-homologus end joining repair of DSBs this would not be suitable for the majority of mutations (Xu et al., 2019, Patsali et al., 2019). Excitingly however, the recent development of CRISPR-Cas9-guided cytidine and adenine deaminases potentially allows the seamless correction of specific pathogenic base changes (Gaudelli et al., 2017, Komor et al., 2016, Liang et al., 2017, Zeng et al., 2020). Whilst in its infancy, this technology may offer an elegant and efficient option for repairing a number of β-thalassaemia mutations and thus curing the disease.

1.8 Summary and Aims

Research into the thalassaemias began nearly a hundred years ago, and much of what we know about genes, their regulation and the pathological states that arise when their expression is disrupted comes directly from research into the α- and β-globin loci. Whilst immense strides have been made in our understanding of these genes, and how to manage patients clinically, for the majority a cure is out of reach. However, recent technological advances offer the opportunity to significantly alter the course of thalassaemia treatment by combining precise genome engineering techniques with the wealth of genetic data gathered by numerous researchers in the field. As the central pathophysiological driver of β-thalassaemia, the goal of any curative treatment strategy should be to reduce the cellular excess of α-globin chains observed in erythroid cells.

The specific aim of this thesis is to develop therapeutic strategies, in pre-clinical models and ultimately in patients, to restore globin chain balance without significantly compromising erythropoiesis and haemoglobin synthesis. My initial aim is to address this situation specifically in HbE/β-thalassaemia a genotype which underlies half of the world's patients with severe β-thalassaemia. In contrast to many other approaches, which seek to derepress γ-globin in definitive erythroid cells, I sought to do this in two ways:

1. By reducing α-globin expression in a controlled manner through identifying and manipulating key regulatory sequences within the main α-globin enhancer MCS-R2.
2. By converting β-thalassaemia major to minor through direct *in-situ* repair of the HbE mutation in *HBB.*

Manipulating MCS-R2 for therapeutic applications requires an accurate understanding of its role *in-vivo*. In chapter 3 I characterise a rare mutation causing HbH disease that enables us for the first time to quantitate the relative contributions of the human enhancers to α-globin expression, confirming that in human MCS-R2 is the major regulatory element. This is followed by an analysis of the MCS-R2 sequence using various genomics methods, in order to identify sequences that would be ideal candidates for manipulation to knock down α-globin expression to the appropriate level for clinical benefit. Chapter 4 describes the gene-editing experiments conducted to disrupt the identified sequences, and the functional genomics, transcriptional and protein based assays used to assess the impact of this editing. Chapter 5 addresses the second aim of the thesis by using adenine base editors to convert the pathogenic HbE mutation back to the wild-type sequence or a normal variant. Similar validation experiments were conducted as in chapter 4, and our ability to base-edit long term haematopoietic stem cells is described through serial mouse xenograft experiments.

2. Materials and Methods

2.1 Cell culture methods

2.1.1 CD34 extraction

Human CD34 cells were purified from adult component donation leucocyte cones (NHS Blood and Transplant, Oxford, UK) or from unmobilised peripheral blood venesected from consenting patients. Peripheral blood was diluted 1:1 with PBS (Gibco), leucocyte cones 1:10, and then fractionated by layering 10ml of PBS/blood mixture on 5ml Histopaque-1077 Hybri-Max (Sigma) and centrifuging for 25 minutes at 1800rpm. Mononuclear Cells (MNCs) were washed with PBS and then MACS buffer (2mM EDTA, 0.5% bovine serum albumin [(BSA), PBS). CD34+ cells were purified using anti-CD34+ magnetically-labelled antibodies (Miltenyi Biotech). After a 30-minute incubation with the antibodies at 4°C, the cells were passed through two LS columns (Miltenyi Biotech) and eluted using 5ml MACS buffer. Purified CD34+ cells were either placed directly into culture or stored and -80°C and then liquid nitrogen in freezing solution (90% foetal bovine serum, 10% dimethyl sulfoxide)

2.1.2 HSPC culture prior to editing

HSPCs were placed in HSPC media (Table 2-1), seeded at a density of $0.25x10^6$ cells/ml and cultured for 36-48 hours at 37°C and 5% CO_2.

Table 2-1 - HSPC media

Component	Final concentration
StemSpan SFEM II (Stemcell technologies)	-
Stem cell factor (SCF) (PeproTech)	100ng/ml
Fms-like tyrosine kinase ligand 3 (Flt3L) (PeproTech)	100ng/ml
Thrombopoeitin (TPO) (PeproTech)	100ng/ml
Penicillin/streptomycin (Gibco)	1IU/ml

2.1.3 Erythroid differentiation culture from CD34 cells.

CD34+ cells were differentiated using a published three-stage liquid differentiation protocol (Griffiths et al., 2012, Scott et al., 2020). In all phases cells were incubated at a temperature of 37°C and 5% CO_2.

Table 2-2 - Erythroid differentiation media

Base media - components	Final concentration
Iscove's modified Dulbecco's media (Bioscience UK)	-
human holo-transferrin (HT) (R&D systems)	200µg/ml
Recombinant human insulin (Sigma Aldrich)	10µg/ml
Heparin sodium (Sigma Aldrich)	3IU/ml
Inactivated group AB Plasma (Dept. of Haematology, Oxford University Hospitals Trust)	3%
Erythropoietin (Janssen-Cilag)	3IU/ml
Fetal bovine Serum (Life technologies)	2%
Penicillin/Streptomycin (Life technologies)	1IU/ml

Phase 1 media - components	Final concentration
Base media	NA
SCF (PeproTech)	10ng/ml
Interleukin-3 (Peprotech).	1ng/ml

Phase 2 media - components	Final concentration
Base media	NA
SCF (PeproTech)	10ng/ml

Phase 3 media - components	Final concentration
Base media	NA
Human holo-transferrin (HT) (R&D systems)	500µg/ml

Phase 1 (Day 0 to 7) - HSPCs were seeded at a density of $2x10^5$ cells/ml in Phase 1 media Cells were counted every 48 hours from Day 3.

Phase 2 (Day 7 to 11) - The media was changed by centrifuging cells for 5 minutes at 300g. Cells were seeded and maintained at $2x10^5$ cells/ml in Phase 2 media.

Phase 3 (Day 11 to 19) - Cells were counted and the media changed by centrifuging the cells for 5 minutes at 300g. The cells were seeded and maintained at $5x10^5$ cells/ml in phase 3 media.

2.1.4 HUDEP-2 cell culture in maintenance phase

HUDEP-2 cells were maintained at a concentration of $2.5x10^5$ cells/ml - $1.5x10^6$ cells/ml in StemSpan SFEM (Stemcell technologies) supplemented with 2mM glutamax (ThermoFisher), 1IU/ml penicillin/streptomycin (Gibco), 50ng/ml SCF (PeproTech), 3IU/ml erythropoietin (Janssen-Cilag), 840nM dexamethasone (Hameln) and 2µg/mL doxycycline (Sigma Aldrich). Cells were counted every 48 hours, centrifuged at 300g for 5 minutes and resuspended in fresh media at a concentration of $2.5x10^5$ cells/ml.

2.1.5 HUDEP-2 cell erythroid differentiation

HUDEP-2 cells were differentiated in the same media as described in 2.1.3 with the following modifications: On Day 0 cells were seeded at $0.5x10^6$ cells/ml in CD34+ Phase 1 media with

the addition of 2µg/mL doxycycline. Media changes were performed every 48h, keeping cells at $2.5x10^5$ cells/ml. On Day 5 cells were counted, a total media change performed and then resuspended in CD34+ Phase 3 media (without doxycycline).

2.1.6 HEK293T cell culture

HEK293T cells were seeded at 25% confluency in Dulbecco's modified Eagle's medium (Gibco) containing 10% fetal bovine serum (Gibco), 1IU/ml penicillin/streptomycin. Cells were passaged when they reached 90% confluency. Cells were washed gently with PBS, 0.05% Trypsin (Gibco) was added and the cells were incubated at 37°C for 5 minutes. The trypsin was quenched by adding media and cells were seeded at 25% confluency in fresh media.

2.1.7 Cytospin slide preparation

$2x10^5$ cells were made up to 200µl with PBS and centrifuged at 400 rpm for 5 minutes onto a glass slide. The dried cells were then stained with a modified Wright's stain using a Hematek 2000 slide stainer (Siemens).

2.1.8 Cell count and viability analysis using trypan blue staining

An equal volume of cell suspension was added to 0.4% trypan blue (Sigma) and incubated for 2 minutes. Cells were then manually counted using a haemocytometer.

2.1.9 Cell count and viability analysis using Nucleocounter

47.5µl cell suspension was mixed with 2.5µl solution 13 (Chemometec) and 10µl of the cell-dye mix was added to an A8 cassette lane. Counting and viability assessment was performed by the NucleoCounter NC-3000 automated reader (Chemometec).

2.1.10 Mouse xenograft experiments

$2x10^5$ cord blood CD34+ cells from three independent biological donors were electroporated and kept in HSPC culture medium for 24 hours. Cells were washed and resuspended in PBS + 1% FBS and injected via tail-vein into 10-week-old sub-lethally irradiated NSG (NOD.Cg-PrkdcscidIl2rgtmlWjl/SzJ; Jackson laboratory) mice. Mice were culled for cell harvesting at 16 weeks post-injection. Cells were harvested from vertebral, lower limb and pelvic bones. All experiments were performed under a project license approved by the UK Home Office under the Animal (Scientific Procedures) Act 1986 and in accordance with the principles of 3Rs (replacement, reduction and refinement) in animal research. The murine xenotransplantation work subsequent to cell culture was conducted by Siobhan Rice.

2.2 Genome engineering methods

2.2.1 Plasmid transfection

1ml antibiotic free HSPC media/HUDEP-2 expansion media were added to a 24-well plate well and placed in a 37°C incubator. $1.2x10^6$ cells were collected in a conical 15ml tube and washed with PBS, centrifuging at 300g for 5 minutes. The PBS was removed, the pellet was resuspended and washed in a v-bottom cryovial using 2ml PBS, and centrifuged for 5 minutes at 300g. The PBS was completely removed. 10μg plasmid DNA was added to a 1.5ml Eppendorf tube. The cell pellet was resuspended in 100μl Human CD34+ Cell Nucleofector Solution (Human CD34+ Cell Nucleofector Kit, Lonza), and mixed thoroughly with the plasmid by pipetting. The cell suspension was then added to the Human CD34+ Cell Nucleofector Kit cuvette taking care not to produce any bubbles, and electroporated using program U-08 on the Amaxa Nucleofector IIb machine (Lonza). Electroporated HUDEP-2 cells were added immediately to the pre-warmed media in the 24 well plate. CD34+ cells were kept at room temperature for 10 minutes before adding to HSPC media. GFP positivity was assessed after 24-48 hours using an inverted fluorescence microscope or flow cytometry.

2.2.2 RNP transfection

250μl antibiotic-free HSPC media was added to a flat bottomed 96-well plate and placed in an incubator. SgRNAs (Synthego) were thawed, mixed and left on ice. Cas9 Nuclease V3 (IDT)

was diluted from the stock concentration (10µg/µl) to an appropriate working concentration in sterile nuclease free H_2O. Editing was performed using the P3 Primary cell Nucleofector kit (Lonza).

Ribonucleoproteins were formed by mixing the required amounts of sgRNA and Cas9 enzyme together in a PCR tube and incubating at 23°C for 10 minutes in a thermocycler and then kept at 4°C until the transfection. SgRNAs and Cas9 were kept at a molar ratio of 1:2.5, but the absolute amounts varied between optimisation and final experiments. The sgRNA used was a 30µM working solution that had been diluted from the 100µM stock solution by adding sterile nuclease free H_2O. An example set of conditions are summarised in Table 2-3.

Table 2-3 - example setup for RNP assembly

sgRNA mass (µg)	sgRNA volume (µl)	Cas9 mass (µg)	Cas9 working concentration	Cas9 volume (µl)
1.6	1.6	3	3µg/µl	1
3.2	3.2	6	6µg/µl	1

Human CD34+ cells were washed in the same way as described for plasmid transfection. 0.5-2x10^5 cells were selected per transfection. After both washes the cells were resuspended in P3 Primary Cell Nucleofector Solution (suspension volume = 20µl x number of replicates), taking care not to create any bubbles. Using the same pipette tip 20µl of the cell suspension was aspirated, mixed with the RNP complex and then added to a single well in the 16-well nucleocuvette strip. Once all the reactions had been added, the cells were electroporated using the DZ-100 setting on the 4D Nucleofector (Lonza). Post-electroporation the cells were added to the pre-warmed HSPC media.

2.2.3 mRNA transfection

250µl antibiotic free HSPC media was added to a flat bottomed 96-well plate and placed in an incubator. SgRNAs (Synthego) and base editor mRNA were thawed, mixed in sterile PCR tubes and left on ice. An example set of conditions are summarised in Table 2-4.

Table 2-4 - example setup for mRNA editing

Reaction	Cell no.	Amount mRNA (g)	Conc. mRNA (g/L)	Vol mRNA (L)	Amount sgRNA (mol)	Vol sgRNA
1	$2.00x10^{5}$	$2.5x10^{-6}$	1.70	$1.47x10^{-6}$	$5.x10^{-11}$	$1.67x10^{-6}$
2	$2.00x10^{5}$	$2.5x10^{-6}$	1.70	$1.47x10^{-6}$	$1x10^{-10}$	$3.33x10^{-6}$
3	$2.00x10^{5}$	$2.5x10^{-6}$	1.70	$2.94x10^{-6}$	$5x10^{-11}$	$1.67x10^{-6}$
4	$2.00x10^{5}$	$2.5x10^{-6}$	1.70	$2.94x10^{-6}$	$1x10^{-10}$	$3.33x10^{-6}$

Human CD34+ cells were prepared for transfection and electroporated as in 2.2.2.

2.2.4 Lipofection

HEK293T cells were seeded in 12 well plates at $1.5x10^{5}$ cells/well in 1ml HEK293T DMEM media. Lipofection was carried out using the FuGene 6 transfection reagent (ProMega). Each reaction had a 100µl final volume. 1µg plasmid DNA was prepared by mixing molar equivalent amounts of sgRNA and base editor plasmid. Opti-MEM media (ThermoFisher) was added to a U-bottom 96 well plate. FuGene 6 reagent was added to the media at a ratio of 6µl to 1µg of plasmid DNA. The media-reagent mix was incubated at room temperature for 5 minutes and 1µg of the plasmid mix was added. After a further incubation for 15 minutes at room temperature 100µl of the Opti-MEM/DNA/reagent solution was layered

onto the seeded HEK293T cells. GFP positivity was assessed after 24-48 hours using an inverted fluorescence microscope.

2.2.5 Plasmids used

Table 2-5 - Plasmids used in thesis

Plasmid name	Addgene ID	Notes
pSpCas9(BB)-2A-GFP (px458)	48138	
pLKO5.sgRNA.EFS.GFP	57822	Lentiviral plasmid used for sgRNA expression only in initial base editing experiments.
pBS_hU6-sgRNA_EF1a_eGFP	-	Custom in house minimal plasmid containing sgRNA expression cassette, GFP and AMP resistance gene only
pCMV-BE3	73021	
BE4	100802	
ABE(7.10)	102919	
xCas9(3.7)-ABE(7.10)	108382	
pCMV_AncBE4max	112094	
pCMV_ABEmax	112095	
pCMV_AncBE4max_P2A_GFP	112100	
pCMV_ABEmax_P2A_GFP	112101	
xCas9(3.7)-ABEmax	120398	
evoAPOBEC1-BE4max	122611	
evoAPOBEC1-BE4max-NG	125616	
NG-ABEmax	124163	
ABE8.13_m	136296	
ABE8.17_m	136298	
ABE8e	138489	
NG-ABE8e	138491	
ABE8e(TadA-8e V106W)	138495	
pCMV-ABEmax-NRCH	136923	
AAV ABE N-terminal	137177	
AAV ABE C-terminal	137178	

2.2.6 sgRNA sequences used

The following sgRNA sequences were either used for cloning into plasmids and then transfected into cells or transfected directly as synthetically produced sgRNAs.

Table 2-6 - sgRNA sequences used in thesis

sgRNA name	sgRNA site	Sequence
SM2	R2 GATA 5'1	CGACCCTCTGGAACCTATCA
MB2	R2 GATA 5'1	GGCTGACTGTGGTCCCTGAT
MB3	R2 CACC 5'	TCTGCCCAAGCCAAGGGTGG
MB4	R2 CACC 5'	CAGCTGCATGCCTCCACCCT
SM11	R2 GATA 5' 2	CACTTGAGGGAGCAGATAAC
SM3	R2 GATA 5' 2	ACTTGAGGGAGCAGATAACT
SM5	5' NFE2	TCCAGAAGCACTGAGTCA
SM6	3' NFE2	AGGACTGCTGAGTCATCCTG
MB13	R2 CACC 3'	GGGGGTGGAGGTGGGACA
SM12	R2 GATA 3' 2	CCCCTCCTGTTTATCTGAGA
MB19	R2 GATA 3' 2	GCCTTCCCTCTCAGATAAAC
HbSS Creation	HBB Codon 6	CATGGTGCATCTGACTCCTG
HbE creation	HBB Codon 26	GTGGATGAAGTTGGTGGTG
SCAC	HBB Codon 6	CATCTGACTCCTGAGGAGAA
HBE01	HBB Codon 26	TGGTGAGGCCCTGGGCAGGT
HBE02	HBB Codon 26	GGTGAGGCCCTGGGCAGGT
HBE01a	HBB Codon 26	TGGTAAGGCCCTGGGCAGGT
HBE03a	HBB Codon 26	GGTAAGGCCCTGGGCAGGT

Lower case bases indicate cloning overhangs, included in table to provide clarity on differences between highly similar sgRNA sequences.

2.2.7 Oligos used for cell line creation by HDR

Table 2-7 - Oligos used for cell line creation by HDR

Cell line	Oligo sequence
HbSS	A*A*CCTTGATACCAACCTGCCCAGGGCCTCACCACCAACTTCATCCACGTTCACCTTGC CCCACAGGGCAGTAACGGCAGACTTCTCCACAGGCGTCAGATGCACCATGGTGTCTGT TTGAGGTTG*C*T
HbE	C*A*AGAGTCTTCTCTGTCTCCACATGCCCAGTTTCTATTGGTCTCCTTAAACCTGTCTT GTAACCTTGATACCAACCTGCCCAGGGCCTTACCACCAACTTCATCCACGTTCACCTTGC CCCACAG*G*G

* denote phosphorothioated bonds. Underlined bases mark deviations from the reference sequence to be introduced by HDR.

2.2.8 sgRNA cloning and production

Short oligos containing the sgRNA sequence were cloned into pre-linearised stock plasmid backbones and transformed into competent bacteria, from which plasmid DNA was isolated.

Plasmid linearization - 10µg of the target plasmid backbone was linearised with either BbsI or BsmBI and gel purified using a QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer instructions and eluted into 100µl ddH_2O.

Oligo preparation - oligos were dissolved to 100µM in H2O. The oligos were combined in the follow reaction to anneal them together. The samples were heated to 99°C for 5 minutes and then cooled at 0.1°C/second until they reached 23°C.

Table 2-8 - sgRNA oligo hybridisation reaction

Reaction component	Volume (µl)
Forward oligo	10
Reverse oligo	10
NEB 2.1	10
Water	70

Double stranded oligonucleotides (dsODNs) were precipitated by addition of 30µl 3M Sodium Acetate (NaOAC, pH 5.6) and 300µl 100% ethanol (ETOH). Samples were centrifuged at 21,000g for 30 minutes at 4°C. The supernatant was removed, and 1ml 70% ETOH was added to wash the pellets, followed by another centrifugation at 21,000g for 10 minutes at room temperature. The supernatant was completely removed and the pellet was allowed to air dry. The oligos were resuspended in 100µl H_2O and quantitated by Nanodrop. The dsODNs were phosphorylated in the following reaction by incubating them at 37°C.

Table 2-9 - sgRNA dsODN phosphorylation reaction

Reaction component	Volume (µl)
dsODN	Volume for 1µg
T4 ligase buffer (NEB)	2
Water	to 19
T4 PNK (NEB)	1

The dsODN-Ps were ligated into the linearised plasmid backbone by incubating in the following reaction for 1 hour at room temperature:

Table 2-10 - sgRNA dsODN-P plasmid ligation reaction

Reaction component	Volume (µl)
dsODN-P (50ng/µl)	0.5
T4 ligase buffer (NEB)	1
Linearised plasmid backbone (25ng/µl)	1
Water	6.5
T4 DNA Ligase (NEB)	1

Bacterial transformation - 3µl of ligated plasmid was transformed into DH5-α competent E. coli (NEB) according to the manufacturer's instructions. 20µl and 50µl of the cells were plated on Ampicillin-containing agarose selection plates, which were incubated for 16 hours in a 37°C incubator.

Colony selection - Colonies were selected using a pipette tip, which was swirled in 10µl H_20 and added to a tube containing 3ml Luria-Bertani (LB) broth. The LB samples were agitated at 220rpm, 37°C. The following PCR was run to confirm insertion of the dsODN-P.

Table 2-11 - sgRNA plasmid insertion PCR

Reaction component	Volume (µl)
Human U6 promoter Forward Primer	1
Bottom sgRNA oligo Reverse Primer	1
Water containing bacteria	2
Platinum PCR Supermix HiFi (Invitrogen)	13

Step	Temperature (°C)	Time (s)
1	95	60
2	95	30
3	62	30
4	68	15
5	Repeat steps 2-4 29 times	
6	4	∞

100µl of the 3ml LB samples matched to the positive PCR were added to 100ml LB containing 100µg/ml Ampicillin and agitated at 220rpm, 37°C overnight. Plasmid DNA was isolated using a plasmid midi kit (Qiagen) according to the manufacturer's instructions.

2.2.9 mRNA production

mRNA was produced through *in-vitro* transcription (IVT) of the desired base editor plasmid.

Linearisation - Plasmids were linearised through AgeI digestion, ensuring that a T7 promoter was at the start of the transcript and the restriction site was immediately at the end of the transcript.

Table 2-12 - plasmid linearisation reaction for mRNA transcription

Component	Volume (µl)
Cutsmart 10x buffer (NEB)	5
Plasmid (30µg)	x
H20	x to 49µL
AgeI-HF (NEB)	1

Cut products were run on a 1% TAE agarose gel, excised and purified using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions.

Transcription - IVT was conducted using the mMESSAGE mMACHINE T7 ULTRA Transcription Kit (ThermoFisher). 1µg of linearised plasmid was used for the transcription reaction, conducted according to the manufacturer's instructions. As the transcripts were longer than 5kb, 1µl GTP was added. Clean-up was carried out with LiCl provided and the mRNA was resuspended in water. Samples were quantified using a Nanodrop spectrophotometer and run on an RNA tapestation (Agilent) to confirm a product of the correct size (~6kb) had been produced. mRNA was stored at -20°C.

2.2.10 HUDEP-2 cell line creation

HUDEP-2 lines carrying the HbE and HbS mutations were created as described in (Nussbaum et al., 2018). Briefly, HUDEP-2 cells were edited as outlined in 2.2.1 using the sgRNAs and ssODN templates in Table 2-6 and Table 2-7. GFP+ Cells were single-cell sorted as in 2.6.3. Colonies were transferred from u-bottom to flat-bottom 96 well plates on day 7 and split when confluent. When colonies were 100,000 cells, half the cells were stored in freezing media at -80°C, and half used for genotyping as below

Multiple 96-well plates of cells were lysed as in 2.3.1.2. 2 rounds of PCR were perfomed. The first involved locus specific primers with the M13 sequence added to the 5' end and the second PCR involved 8 forward and 12 reverse primers bearing the M13 sequence and an 8bp barcode sequence. Thereafter the NEBnext Ultra (NEB) library preparation protocol was followed to ligate adaptors and add indices. One index was used per plate, so several 96 well plates could be screened at the same time. Sequencing was performed on the Miseq platform with a 500-cycle nano kit (Illumina). Screening for edited cells was performed using an in-house pipeline, and positive clones were matched with the corresponding frozen cells, thawed and resequenced by Sanger sequencing.

2.3 DNA methods

2.3.1 DNA extraction

2.3.1.1 Genomic DNA extraction from moderate to large cell numbers (>100,000 cells)

Cell pellets from cells in culture were resuspended in 200µl PBS. Genomic DNA was extracted using the Purelink genomic DNA mini kit (ThermoFisher) according to the manufacturer's protocol.

2.3.1.2 Genomic DNA extraction from small cell numbers (<100,000 cells)

Cells were washed with PBS by centrifuging for 5 minutes at 300g. The pellets were resuspended in lysis buffer (50mM Tris pH 8.0, 1mM EDTA pH 8.0, 0.5% Tween 20, 0.002 U/µl Proteinase K, nuclease-free water), and placed in a 37°C incubator overnight. To denature residual proteinase K and reduce viscosity the lysate was heated for 10 minutes at 95°C and used directly in PCR reactions.

2.3.2 Polymerase Chain Reaction

Routine PCRs were carried out using the Herculase II Fusion Polymerase kit (Agilent). A typical reaction was set up on ice as outlined in Table 2-13.

Table 2-13 - Herculase PCR reaction setup

Reaction component	Volume (µl)
5x Herculase II reaction buffer	5
Nuclease free water	17.25
Forward primer 10µM	0.5
Reverse primer 10µM	0.5
dNTP	0.25
DNA template	1
Herculase II fusion DNA polymerase	0.5
Total	25

Step	Temperature (°C)	Time (s)
1	95	120
2	95	10
3	60	20
4	72	30
5	Repeat steps 2-4 29 times	
6	72	30
7	4	∞

5µl of PCR product was mixed with 1µl 6x loading dye (NEB) and run on a 2% agarose gel in TAE buffer at 110V for 45 minutes. Gels were visualised using the Gel-doc 2000 (BioRad laboratories). PCR products were cleaned up using the QIAquick PCR purification kit (Qiagen) according to the manufacturer's instructions.

2.3.3 Primers used for PCR

Table 2-14 - Primers

Name	Primer	Sequence	Notes
MCS-R2	Forward	TGGTCCTGAAGGATGAGAAG	Routine R2 amplification
MCS-R2	Reverse	AGCAACAGTCCTTTCTCTGG	
MCS-R2 375 NGS	Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCTTCTGCAACCATGATGAC	Used for 500 cycle NGS library prep
MCS-R2 375 NGS	Reverse	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTTGCCCCTGAAGCCTG	

MCS-R2 300 NGS	Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCA GATAACTGGGCCAACC	Used for 300 cycle NGS library prep
MCS-R2 300 NGS	Reverse	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGAT GACTCAGCAGTCCTGT	
HBB - PromoterF	Forward	GAGGGAGGGCTGAGGGTTTGA	Routine HBB exon 1 amplification
HBB - 3'β2	Reverse	CAAAGGACTCAAAGAACCTC	
HBENGS	Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCAA CCTCAAACAGACACCA	Use for 300 cycle NGS library prep
HBENGS	Reverse	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAAA GGACTCAAAGAACCTC	
NGS HbSS screen	Forward	GTAAACGACGGCCAGTttagacctcaccctgtggagc	For screen to make HbSS line
NGS HbSS screen	Reverse	AGCGGATAACAATTTCACACAGGAggtgggaaaatagac caataggca	
M13	Forward	GTAAACGACGGCCAGT	For cell line screening
M13	Reverse	AGCGGATAACAATTTCACACAGGA	
Adaptor	Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	For NGS library preparation
Adaptor	Reverse	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
Human U6	Forward	GAGGGCCTATTTCCCATGATTCC	Plasmid sequencing

2.3.4 Sanger Sequencing

Cleaned up PCR products were sequenced by the MRC WIMM Sequencing facility, using the ABI-3730 DNA analyser (Applied Biosystems). Sequences were provided as ab1 files which were analysed using Macvector software (version 15.5.4, Macvector, inc.).

2.3.5 Next Generation Sequencing library preparation for genotyping

A modified 2-PCR version of the NEBnext Ultra (NEB) library preparation protocol was followed. In the first PCR, set up as in Table 2-13, primer pairs with NEB adaptor sequences 5' to the locus-specific primer sequence were used for amplification, ensuring the products had the adaptor sequences added by the end of the PCR. 5µl of this was used directly

without clean-up for the indexing PCR, set up as in Table 2-15. This added either single or dual end indices taken from a NEBNext Multiplex Oligos for Illumina kit (NEB).

Table 2-15 - NGS PCR 2 for index addition

Reaction component	Volume (µl)
5x Herculase II reaction buffer	10
Nuclease free water	23.5
Universal primer/i5 index primer (NEB)	5
i7 index primer (NEB)	5
dNTP	0.5
DNA template from PCR 1	5
Herculase II fusion DNA polymerase	1
Total	50

Step	Temperature (°C)	Time (s)
1	98	30
2	98	10
3	65	30
4	72	30
5	Repeat steps 2-4 6 times	
6	72	5minutes
7	4	∞

A single clean-up was then performed using AMPure XP beads (1.8x). Prior to sequencing individual samples were diluted to 4nM and pooled.

2.3.6 Assay for Transposase Accessible Chromatin Sequencing (ATAC-Seq)

Libraries were prepared as in (Buenrostro et al., 2013). Briefly, $2x10^5$ cells were lysed, separated into 3 aliquots and the nuclei of each were tagmented using Tn5 transposase (Illumina) by incubating at 37°C for 30min. DNA was purified using a MinElute reaction clean-up kit (Qiagen) according to the manufacturer's protocol and eluted in 23µl elution buffer. Recovered DNA was amplified using the NEBNext 2x Mastermix (NEB) for 6 cycles

using custom barcoded primers. The profiles of libraries were examined using a D1000 tapestation analyser kit (Agilent), and quantified using a KAPA library quantification kit (Roche). Samples were sequenced on the NextSeq platform (Illumina) using a high-output 75 cycle kit (Illumina) with paired end reads.

2.3.7 NF-E2 ChIP-qPCR

This work was carried out by Damien Downes as outlined in (Hua et al., 2021). Chromatin was crosslinked with 1% formaldehyde (Sigma) by the addition of 1 mL 10x crosslinking buffer (50mM HEPES, 1mM EDTA, 0.5mM EGTA, 100mM NaCl, 10% formaldehyde) to 10^7 cells in 9 mL of media and incubation at RT for 10 minutes. Crosslinking was quenched with 130mM glycine, and cells were washed with cold PBS before snap freezing pelleted cells. Immunoprecipitation was performed using an Agarose ChIP Assay Kit (Merck Millipore). 10^7 cells were lysed by incubation on ice with 130µL lysis buffer for 15 minutes. Lysed cells were transferred to Covaris microtubes and sonicated on the Covaris M220 to generate 200-400bp fragments. Insoluble material was removed by centrifugation (15,000g, 15 min, 4ºC) and soluble material was diluted to 4 mL with dilution buffer. 2 mL diluted chromatin (equivalent to $5x10^6$ input cells) was incubated with antibodies for NF-E2 (0.2 µg sc-22827 X, lot: B1304; Santa Cruz) overnight. Chromatin binding to Protein A/agarose slurry, washes and elution were performed according to the manufacturer's instructions. DNA was purified by phenol-chloroform extraction with PhaseLock tubes (5Prime) and ethanol precipitation with NaOAc, and 2 µL GlycoBlue (Invitrogen). ChIP enrichment was determined by RT-qPCR with KAPA SYBR Fast Universal reagents (Merck) and custom primers (Table 2-16)

Table 2-16 - NF-E2 ChIP-qPCR primers

Target	Forward	Reverse
MCS-R1	CCTGCATCCTGACAATGACTAT	GACTAAGCTTGTGCACATTTCTG
MCS-R2	CTCCTGTTTATCTGAGAGGGAAG	AAGAACGTATTTACTGTCTGGGT
Intergenic	GGGTTTCCCTGTTACAGAAATG	CTCTGGTCATCAGGTCAGTATG

2.3.8 Next Generation Capture-C

Capture-C was performed on primary human erythroid cells as previously published (Davies et al., 2016).

Fixation - cells were fixed in 2% formaldehyde, quenched in 1M glycine, followed by lysis in 10mM NaCl (Invitrogen), 10mM Tris pH 8 (Invitrogen), 0.2% Igepal CA-630 (Sigma), complete protease inhibitor cocktail (Sigma). Cells were snap frozen on dry ice and ethanol and stored at -80°C.

DpnII digestion - cells were digested with 1000U of DpnII in DpnII buffer (NEB) overnight.

Ligation and decrosslinking - digests were ligated with T4 DNA ligase (Life technologies) overnight, after which Proteinase K (ThermoFisher) was added to de-crosslink chromatin.

DNA extraction - DNA was extracted through a phenol-chlorophorm isolation, and precipitated through addition of NaAc, ETOH and a 45-minute centrifugation at 4°C. After washing the pellet with 70% ETOH the DNA was dissolved in water and stored at -20°C. Digestion efficiency was assessed by running samples on a 1% agarose gel.

Sonication - 5µg of 3C library was sonicated to 200bp on a Covaris S220 focused-ultrasonicator in 120µl Covaris microTUBE AFA Fiber pre-split snap-cap 6x16mm microtubes

(Covaris) using the following settings: duty cycle - 10%; intensity - 5; cycles per burst - 200; time - 360 seconds; set mode - frequency sweeping.

Library preparation and indexing - Library preparation was carried out using a NEBNext DNA library Prep Reagent kit (NEB) according to the manufacturer's instructions. NEBNext Multiplex Oligos for Illumina Primers and a Herculase II Fusion polymerase kit were used for indexing.

Capture Hybridisation- Indexed libraries were pooled and captured using the Nimblegen SeqCap EZ Hybridisation and wash kit (Roche) according to the manufacturer's protocol. Biotinylated DNA probes used were designed to the promoters of the targeted gene (Sigma) as outlined in Table 2-17. Recovery of hybridised material was performed using M-270 streptavidin Dynabeads (Life technologies). The captured multiplex DNA sample was amplified using POST-LM_PCR Oligos 1 and 2 according to the Nimblegen SeqCap Library SR protocol. Following an AMPure XP bead clean-up the capture-hybridisation was performed a second time. Resulting multiplexed libraries were quantified using a Qubit dsDNA HS kit (ThermoFisher). Sequencing was performed on a MiSeq (Illumina) using a v2 Reagent Kit (Illumina) with 150bp paired-end reads.

Table 2-17 - probes for Capture-C

Probe target	Sequence
HBA1	GATCCCGCTGGAGTCGATGCGCGTCCAGCGCGTGCCAGGCCGGGGCGGGGGTGCGGGCTGACTTTCTCCC
HBA2	GATCCCGCTGGAGTCGATGCGCGTCCAGCGCGTGCCAGGCCGGGGCGGGGGTGCGGGCTGACTTTCTCCC
HBB	GATCCCCAAAGGACTCAAAGAACCTCTGGGTCCAAGGGTAGACCACCAGCAGCCTAAGGGTGGGAAAATA
SLC25a37	GATCAGGCCTCTACAGACCTTATGGGCCTATCAAACCCAAAGAAAATTTCCTGGAGGTGCCTACTCAGAG

2.3.9 Micro-Capture-C

MCC was performed as previously described in (Hua et al., 2021) Briefly:

Cell fixation - HUDEP-2 cells were grown, differentiated, and harvested ($20x10^6$ cells/replicate) on Day 5 of differentiation. Cells were aliquoted into 30ml of media in 50ml falcon tubes and formaldehyde (37%, Sigma) was added to a final concentration of 2% to crosslink DNA. Cells were incubated on a roller for 10 minutes at room temperature, and 1M cold glycine was added to quench the reaction.

Cell permeabilisation - The cells were washed with PBS and resuspended in 1ml PBS. 0.5% Digitonin (Sigma) was added to permeabilise the cells to a final concentration of 0.005% (wt/vol). The cells were snap frozen on dry ice mixed with 100% ETOH and stored at -80°C.

MNase digestion - Samples were thawed, washed in PBS and resuspended in Micrococcal nuclease buffer (Tris-HCl pH 7.5 10 mM; $CaCl_2$ 1 mM). Micrococcal nuclease (NEB) was used to digest the chromatin by incubation at 37 °C for 1 hour. The digestion was quenched with ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetra acetic acid (EGTA, Sigma) and the cels were washed twice with PBS.

DNA repair, phosphorylation and ligation - Cells were resuspended in DNA ligase buffer (ThermoFisher) with added dNTPs (ThermoFisher) and EGTA. For DNA end repair and phosphorylation of the DNA T4 Polynucleotide Kindase (PNK, NEB) and DNA Polymerase I Large (Klenow) Fragment (NEB) were added and the reaction was incubated for 1 hour at 37°C. The DNA fragments were then ligated using T4 DNA ligase (ThermoFIsher) and incubating overnight at 16°C in a heating block set at 800rpm.

Decrosslinking and DNA extraction - The following day the DNA was decrosslinked by the addition of Proteinase K and the DNA was extracted using a Qiagen DNeasy blood and tissue

kit according to the manufacturer's instructions. The library profiles were assessed using a D100 tapestation (Agilent) looking for characteristic nucleosomal patterning and an increase in fragment size in the ligated samples.

The remainder of the Capture-C protocol was the same as for conventional Capture-C as outlined in 2.3.8.

2.3.10 Multiplex Ligation Probe Amplification (MLPA)

MLPA was carried out with a specific probe mix for the human alpha globin locus (MRC-Holland), and was conducted according to the manufacturer's instructions. Briefly, 100ng input DNA was heat denatured and then hybridised with the probe-mix overnight. The next day adjacent hybridised probes were ligated and amplified by PCR. Products were Sanger sequenced and dosage quotients calculated and visualised using Coffalyser.Net software (MRC-Holland).

2.3.11 CIRCLE-seq

CIRCLE-seq was carried out using the previously described protocol (Lazzarotto et al., 2018). OT editing using the NF-E2 5' sgRNA was assessed using a single biological replicate, and OT editing using the HbE targeting sgRNA was assessed in triplicate, using the HUDEP-2 HbE line and cells from two patients with HbE/β-thalassaemia. CIRCLE-seq sequencing data was analysed using the circleseq Python package (https://github.com/tsailabSJ/circleseq)

2.4 RNA methods

2.4.1 RNA extraction

RNA was extracted using the RNeasy mini kit (Qiagen) according to the manufacturer's instructions. RNA was generally eluted in 25µl water and stored at -80°C.

2.4.2 Reverse Transcription

cDNA was produced using the SuperScript III First-Strand synthesis Supermix for qRT-PCR (ThermoFisher) according to the manufacturer's instructions and stored at -20°C.

2.4.3 Quantitative real-time PCR (qPCR)

Pre-designed and validated Taqman probes (Applied Biosystems) and Taqman Universal PCR mastermix (ThermoFisher) were used in all qPCR assays. All Taqman probes spanned exon junctions and are listed in Table 2-18. Reactions were setup in 20µl in technical triplicate and run on the Quantstudio 3 Real Time PCR System (Applied Biosystems). Gene expression was calculated using the delta delta CT method. RPL13a was used as the reference for cells in culture due to its previously reported constant nature throughout erythroid differentiation and for consistency with previous work (Merryweather-Clarke et al., 2011, Mettananda et al., 2017a).

Table 2-18 - Taqman probes

Gene ID	Gene Name	Taqman assay ID
HBA	α-globin	Hs00361191_g1
HBB	β-globin	Hs00747223_g1
HBG	γ-globin	Hs00361131_g1
HBZ	ζ-globin	Hs00923579_m1
RPL13a	Ribosomal protein L13a	Hs03043885_g1

2.5 Protein methods

2.5.1 Isoelectric Focusing (IEF)

IEF was performed as per the manufacturer's instructions (Resolve; PerkinElmer), running for 45 minutes at 300V; 15°C. $0.5x10^6$ cells were used per sample.

2.5.2 High Performance Liquid Chromatography (HPLC)

RP-HPLC and CE-HPLC were performed as per the manufacturer's instructions.

For samples analysed by RP-HPLC $2x10^6$ cells were used, for CE-HPLC $7x10^6$ cells were used.

For CE-HPLC testing, the Bio-Rad variant haemoglobin testing system was used with the Variant II β-thalassaemia short program pack. For RP-HPLC the Agilent 1260 infinity II quaternary stem was used with a Phenomex C18 column.

2.6 Flow Cytometry methods

2.6.1 Immunophenotyping

$1x10^5$ cells were washed with PBS and labelled for 30 mins at 4°C with the antibodies in Table 2-19 or

Table 2-20. Flow cytometry was performed on the Attune NxT Flow cytometer (Invitrogen) and analysed on FlowJo software (V. 10.7.1). Flow cytometry for human cell engraftment in murine xenotransplants was conducted by Peng Hua.

Table 2-19 - antibodies used to assess erythroid differentiation

Antibody	Conjugate	Supplier
CD235a	PE	BD Bioscience
CD71	PerCP Cy5.5	Biolegend
CD49D	APC	BD Bioscience
CD34	PE/Cy7	Biolegend
CD233	FITC	IBGRL
CD36	APC/Cy7	Biolegend
Hoechst 33258	-	Invitrogen

Table 2-20 - antibodies used to assess human cell engraftment and differentiation

Antibody	Conjugate	Supplier
Fixable Viability	Ef506(alpha)	Thermo Fisher
mCD45.1	APC-Cy7	eBio
hCD45	AF700	BD Bioscience
CD3	BV711	Biolegend
CD235a	PerCP-Cy5.5	Biolegend
CD56	BV605	BD Bioscience
CD34	BV421	Invitrogen
CD19	APC	Biolegend
CD33	PE	Biolegend

2.6.2 Bulk cell sorting

Cells were washed with PBS and then resuspended in 400µl sorting buffer (2% BSA (Gibco) in PBS). Cells were stained with Hoechst 33258 to exclude non-viable cells, and were run on the Sony MA900 cell sorter (Sony) using the accompanying software. Cells were sorted for GFP positivity or for viability into either flat or round bottom 96 well plates containing 50µl media.

2.6.3 Single cell sorting

Samples were prepared as for bulk cell sorting, but were sorted exclusively into round bottomed 96-well plates and placed into boxes containing 50ml milli-Q water (Merck) to avoid evaporative losses. Cells were kept in a humidified box and examined after 7 days.

2.7 Bioinformatics methods

2.7.1 ICE

Cas-9 indel generation efficiencies were assessed initially using the ICE online web-tool (Synthego) found at https://ice.synthego.com/. Sanger sequencing ab1 files for control and edited samples were used to calculate an indel percentage and spectrum of editing outcomes.

2.7.2 Analysis of edits using PERL

Paired-end Fastq files generated from edited and non-edited cells were analysed using a custom PERL script. Adaptor sequences were trimmed using Trim Galore (0.3.1) and reads were combined using flash (1.2.8). Flashed reads were converted into FASTA format and compared to a custom reference genome comprising of the MCS-R2 enhancer using Blat (35). The resulting PSL file was analysed by the script, which parsed lines in the file corresponding to each read, building up a profile of where deletions and insertions were created. These piled up 'base counts' were used to create the indel plots.

2.7.3 CRISPResso

The CRISPresso command line tool (2.0.31) with the following command syntax: CRISPResso -r1 [read1.fastq] -r2 [read2.fastq] -a [amplicon sequence] -g [sgRNA sequence]. For base editing experiments the following flags were added: -wc -16 -w 8 --base_editor_output.

2.7.4 ATAC and Chromatin Immunoprecipitation sequencing alignment, normalisation and visualisation

Data were analysed using an in-house pipeline (Telenius and Hughes, 2018). Aligned reads were normalised to reads per kilobase per million (RPKM) and mapped using deepTools. Data tracks were visualised using the University of California Santa Cruz (UCSC) publically available genome browser.

2.7.5 Chromatin Immunoprecipitation (ChIP) motif analysis

Peaks were called using the 'Model-based Analysis of ChIP-Seq' tool (MACS2,2.0.10) on the filtered bam files for immunoprecipitated (IP) and input samples. The resulting bed files were curating using the bedtools (2.25.0) intersect function to remove shared peaks from the IP file. These peaks were visualised using the Multi-image Genome (MIG) viewer (McGowan et al., 2013). A cut-off for peak height was selected, and this was applied to the list of called peaks in the IP file. The Multiple Em for Motif Elicitation (MEME) tool was used to derive Position Weight Matrices (PWMs) from the positive peaks. (Bailey and Elkan, 1994). The 'Find Individual Motif Occurrences' (FIMO) tool (Grant et al., 2011) was used by applying the generated PWMs to the R2 enhancer.

2.7.6 ATAC-seq peak analysis

ATAC-seq peaks in edited and non-edited samples were called using MACS2. Peaks found in more than one sample were compiled into a single file along with respective read-depths for each peak using the bedtools multicov function. Peaks were compared using DESeq2 with variance stabilisation applied and a false discovery rate of 0.05.

2.7.7 NG Capture C analysis

Data were analysed using an in-house custom pipeline (J. Davies, J. Telenius) and as previously described (Davies et al., 2016). Tracks were visualised using the UCSC genome browser.

2.7.8 Micro-Capture-C analysis

MCC analysis was conducted as previously described using a custom-designed pipeline (Hua et al., 2021). These scripts are publically available at https://github.com/jojdavies/Micro-Capture-C. Data tracks were visualised using the UCSC genome browser.

2.7.9 DeepHaem enhancer mutagenesis

DeepHaem (code available at https://github.com/rschwess/deepHaem) was used to establish the chromatin accessibility profiles for a 1kb region across MCS-R2, mutating every base individually to all 4 possible bases. A damage score was generated and visualised as shown in Chapter 3.

2.7.10 Off-target analyses

OT sites were generated using CasOFFinder and CRISPOR (Bae et al., 2014, Concordet and Haeussler, 2018). These sites were intersected with sites identified by CIRCLE-seq assays.

For DeepHaem analysis, all OT sites with an adenine in the editing window were used. For each site with the WT sequence and edited sequence a probability score between 0 to 1 was assigned by deepHaem to predict whether it was likely to be an open chromatin site. A damage score was then calculated by subtracting the reference probability from the OT-edited probability.

3. Assessment of the *in-vivo* function of MCS-R2 and an analysis of its sequence

3.1 Introduction

3.1.1 Upstream deletions cause α-thalassaemia

The vast majority of cases of α-thalassaemia arise due to mutations at the α-globin genes including deletions of one (-α/αα) or both (--/αα) genes, or less commonly due to non-deletional variants. In rare cases however, patients have presented with α^0 thalassaemia trait or HbH disease due to deletions from upstream of the α-globin cluster that leave the α-globin genes intact. Three decades of careful work has shown that these deletions all involve the major α-globin regulatory element; an enhancer known variously in human as hypersensitivity site -40 (HS-40) or multispecies conserved sequence regulatory element 2 (MCS-R2). This work has involved forward and reverse genetic approaches involving interspecific hybrid cells, stable transfectants, gene edited/transiently transfected primary cells, transgenic mouse models and clinical observational studies. Many insights about the α-globin locus and gene regulation more generally have been gained through the careful study of these rare and valuable cases of alpha thalassaemia. Through characterising one such patient, presented in this chapter, I aim to further analyse the role of MCS-R2 and its role in controlling α-globin transcription, as well as examining its underlying sequence more closely.

3.1.2 Analysis of the role of MCS-R2 using transgenic approaches

The presence of *cis*-regulatory elements proximal to the α-globin cluster was initially suspected after the characterisation of two sets of patients bearing deletions -28 to -67kb (patient IJ) and +10 to -52kb (patient RA) in relation to the ζ-globin mRNA cap site (Liebhaber et al., 1990, Hatton et al., 1990). Both patients presented with a low MCV, MCH and HbH inclusions noted in peripheral blood. Using a mouse erythroleukaemia (MEL) cell line, interspecific hybrids were created, bearing either a copy of human chromosome 16 with the 62kb deletion from RA or RA's wild-type chromosome 16. No human mRNA was detected in the MEL-hybrid line containing the upstream deletion, in contrast to hybrids containing the WT chromosome. The α-globin genes *in-cis* were found to be normal on sequencing, and so the presence of upstream regulatory elements analogous to those found at the β-globin cluster was suspected (Grosveld et al., 1987).

Due to the large nature of the deletions, further work was carried out to dissect the 62kb region involved in the $(\alpha\alpha)^{RA}$ deletion by creating expression vectors containing variably sized fragments of the region and *HBA2*, which were stably transfected into MEL cells (Higgs et al., 1990). Analysis of human α-globin expression as a percentage of endogenous α-globin found that the only fragments that had any detectable expression included the genomic region -37 to -47kb upstream of the α-globin cluster. When these constructs containing the α-globin gene with the positive regulatory elements were introduced into transgenic mice, high levels of erythroid-specific expression were noted. To further identify the location of the regulatory element, DNase I hypersensitive assays were performed in human primary erythroid and non-erythroid cells. It has been shown that DNase I hypersensitive sites (DHS)

map to tissue-specific regulatory elements that bind *trans*-activating transcription factors (Galas and Schmitz, 1978). These assays found that the two most prominent hypersensitive sites were at -40kb and -33kb (HS-40 and HS-33). HS-40 was localised further using an expression plasmid system to a 350bp region as shown in Figure 3.1 (Jarman et al., 1991).

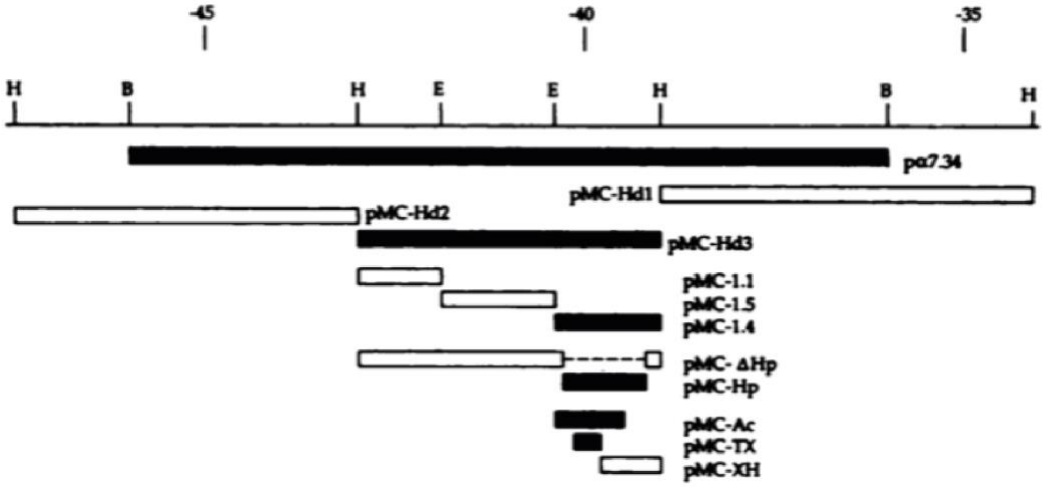


Figure 3.1 - Localisation of HS-40 activity to 350bp fragment *The regulatory sequence of HS-40 was localised to a 350bp fragment (pMC-TX). Adapted from (Jarman et al., 1991)*

Whilst the aforementioned studies identified HS-40 as the major regulatory element at the α-globin cluster, further work sought to clarify the function and possible redundancy of the other known hypersensitivity sites HS-33 and HS-10 (later known as MCS-R3 and MCS-R4 respectively) that had been identified previously (Higgs et al., 1990). These were performed in MEL cells as well as transgenic mice (Sharpe et al., 1992). Constructs containing a HS-40-α-globin backbone with combinations of HS-33, HS-10 and ζ-globin showed that globin expression was tissue and developmental-stage specific, but the addition of hypersensitivity sites to HS-40 did not increase expression or confer any qualitative changes to regulatory activity (Sharpe et al., 1993a, Sharpe et al., 1993b, Gourdon et al., 1993). This regulatory behaviour stood in contrast to work carried out at the β-globin locus (Forrester et al., 1989, Talbot et al., 1990).

Hitherto study on HS-40 had used chromosomes bearing large deletions derived from patient tissue, or expression constructs. *Bernet et al* were the first to engineer a deletion of HS-40 in an intact human chromosome 16, itself stably transfected into a human-MEL interspecific hybrid, to rule out the presence of undiscovered regulatory sites (Bernet et al., 1995). When globin mRNA was measured using an RNA protection assay with mouse α-globin as an internal control, human α-globin expression was found to be 3% compared to endogenous murine α-globin, whereas in wild type clones the ratios were equal. The other DHSs were unchanged despite the absence of HS-40, and so this and a further study provided evidence that, in human, there were no other strong regulatory elements and HS-40 did not have a significant chromatin remodelling function (Craddock et al., 1995).

These findings were re-examined in greater precision later using a humanised mouse model engineered to replace the entire syntenic murine α-globin locus with its human counterpart (Wallace et al., 2007). Human α-globin expression in mice in which the human locus had a 1.1kb deletion removing HS-40 was <2% of the amount of murine α-globin produced by the murine locus. Furthermore, embryos homozygous for the human α-locus with the HS-40 deletion died by E11.5, confirming that HS-40 was sufficient and critical for α-globin expression *in-vivo*.

3.1.3 Analysis of the role of MCS-R2 using human patient studies

In parallel to the above experimental approaches, clinical studies centering on naturally occurring mutations in MCS-R2 and the surrounding locus helped shed light on the relative importance of the various regulatory elements. Initially the observed deletions tended to

be large and involved multiple elements and surrounding genes. Aided by technological advances, over time, highly specific deletions have been characterised involving MCS-R2 alone as summarised in Figure 3.2, allowing us a clearer insight to its importance in its native physiological state.

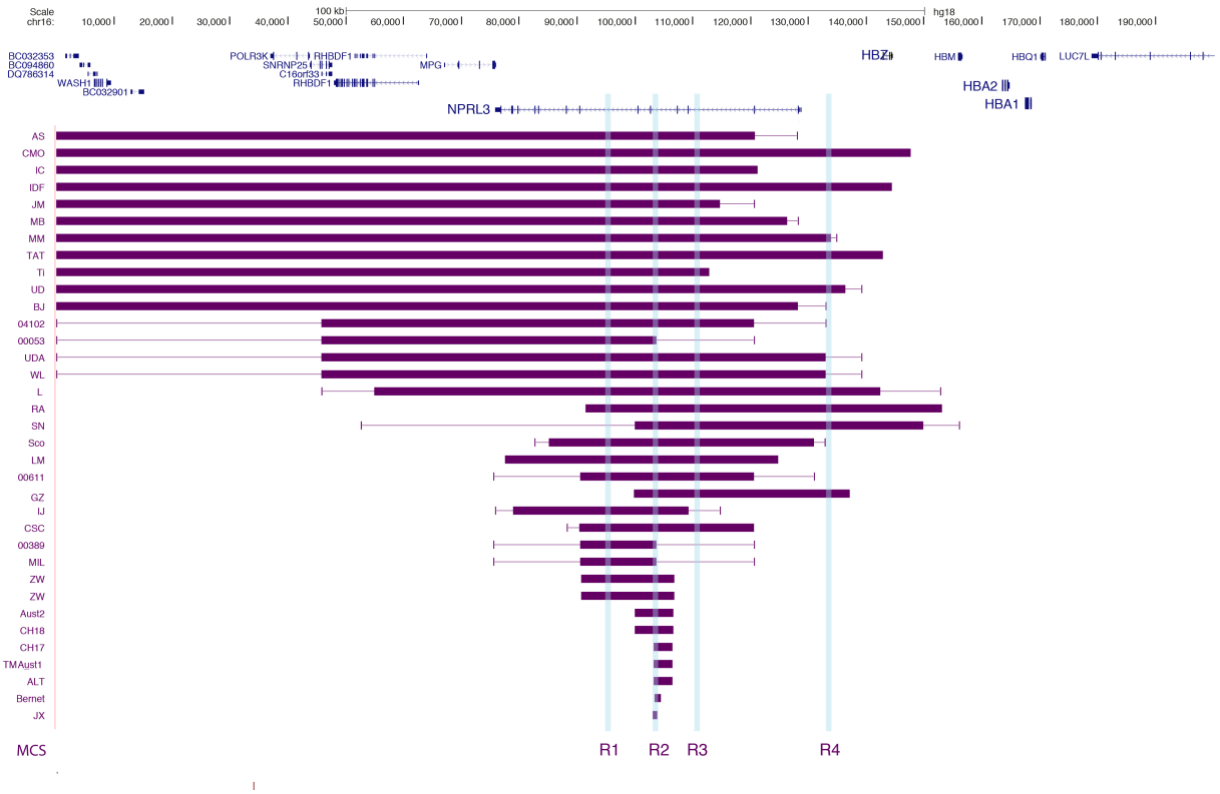


Figure 3.2 - Upstream deletions involving MCS-R2 *Published and unpublished upstream deletions involving MCS-R2. Each line represents a deletion identified in a patient, as annotated on the left bar. The vertical blue bars identify the four upstream enhancers.*

Two patients, IJ and RA, that formed the foundation of early studies have been mentioned above. Several patterns of deletions were observed; smaller deletions within or extending out of the α-globin locus and larger deletions that arose from truncations of the short arm of chromosome 16, with stabilisation by addition of a telomeric repeat (Liebhaber et al., 1990, Romao et al., 1992, Romao et al., 1991, Wilkie et al., 1990b, Hatton et al., 1990, Flint et al., 1994, Flint et al., 1996, Viprakasit et al., 2003, Viprakasit et al., 2006). Both sets of

deletions presented in heterozygosity or as compound heterozygotes with other more common α-thalassaemia deletions such as $-\alpha^{(3.7)}$. Patients with either type of inheritance pattern displayed a reduction in MCH and reduced α/β synthesis ratio; heterozygotes having the phenotype of α^0 thalassaemia and compound heterozygotes having HbH disease. These findings suggested that a deletion including HS-40 was equivalent to removing two α-globin genes.

Recently, a series of studies have reported homozygous or heterozygous deletions removing MCS-R2 alone. The 3.3kb $(\alpha\alpha)^{ALT}$ deletion was initially described in three patients in the same region of Portugal, one in homozygosity (Coelho et al., 2010). The haematological findings were similar to previous studies, with cases exhibiting a low MCV and MCH. Patient J who presented in homozygosity had a surprisingly mild reduction in haemoglobin (108g/L) and intriguingly at the site of the deletion there was a 39bp orphan sequence that has no single homologous sequence in the human genome. There is a possibility that it is the result of a complex rearrangement involving two segments of DNA from chromosome 3 (chr3:144280515-144280533 and chr3:157269186-157269203)(Phylipsen et al., 2010, Ferrao et al., 2017). Apart from a potential Sp1 binding site there are no consensus binding sites for TFs such as NF-E2, GATA1, SCL/Tal1 or KLF1 in the insert, and ChIP studies have not revealed SCL/Tal1 binding or the presence of an active H4Ac mark (Coelho et al., 2010, Mettananda et al., 2016). Nevertheless, a hypothesis for the mild phenotype is that this 39bp has functional activity to modulate α-globin expression. This is supported by two further sets of patients with homozygous MCS-R2 deletions. The first study described an Italian boy who had a compound heterozygote mutation; on the paternal allele had a deletion of MCS-1-3 (MCS-R1 refers to a recently discovered upstream DHS, also known as

HS-48) and the maternal allele had a 3.3kb deletion with a 39-nucleotide insertion, presumably the same as the ALT deletion (Sollaino et al., 2010). In contrast to the Portuguese patients, this case had severe HbH disease. The second set of studies from China describe a different, smaller deletion removing 742bp including MCS-R2, known as $(\alpha\alpha)^{JX}$ (Wu et al., 2017). In heterozygosity, *in-utero* and neonatal Hb Bart's levels corresponded with that of patients with α^0 thalassaemia. However, in homozygosity the case was found to have severe anaemia, requiring blood transfusions (Huang et al., 2017). Notably the $(\alpha\alpha)^{JX}$ deletion was not associated with an orphan insert, potentially leading to the noted severe phenotype.

3.1.4 Analysis of the molecular signature of MCS-R2

With the development of sensitive genomics assays, detailed analyses of the chromatin at the upstream enhancers could be conducted. Using a ChIP-on-chip approach De Gobbi *et al* analysed H4acetylation as well as binding of various TFs in humanised mouse cells, K562 cells and primary human erythroblasts (De Gobbi et al., 2007). H4acetylation, a marker of active chromatin, was noted at MCS-R2 the other upstream elements and α-globin promoters in primary human erythroid cells, reaching a peak at the basophilic erythroblast stage. MCS-R2 was found to bind the TFs GATA1, SC/Tal1, E2a, ldb, LMO2 (the components of the erythroid pentameric complex), as well as NF-E2. This complex is specific for erythroid cells and known to bind to erythroid specific elements explained in section 1.2. Interestingly, MCS-R3 and -R4 could bind the pentameric complex but not NF-E2, in contrast to MCS-R1 and -R2 which could. Given that previous clinical observational studies and experimental

approaches had shown MCS-R2 to be the most powerful enhancer, this finding pointed to the importance of NF-E2 to its activity.

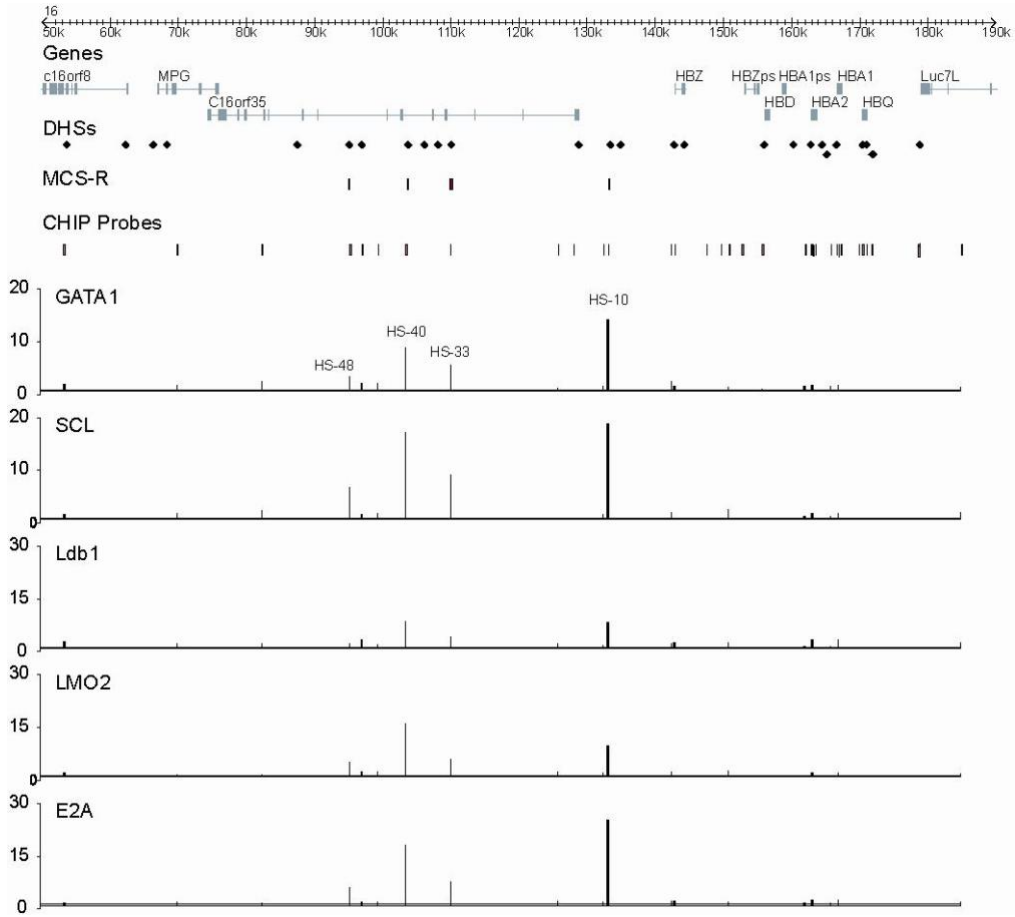


Figure 3.3 - Erythroid pentameric complex at MCS-R2 ChIP-qPCR results showing binding of the erythroid pentameric complex to MCS-R2 in K562 erythroid cells (De Gobbi et al., 2007).

3.1.5 Analysis of the sequence of MCS-R2 by in-vitro study

While the large-scale role and function of MCS-R2 in the α-globin cluster was initially being explored, the fine details of its sequence were also examined (Jarman et al., 1991). The ability of DNase I to preferentially digest open chromatin was used to identify regions of the active enhancer that are bound by proteins (Weintraub and Groudine, 1976). *In-vitro* DHS footprinting assays revealed five distinct areas of protection or footprints across this region

(FP I-V). The erythroid specific areas, FP-I, -IIa, -IV and –V coincide with GATA1 consensus binding sites [(T/A)GATA(A/G)]. Excess oligonucleotide corresponding to confirmed GATA1 motifs competed effectively with these footprint sites. A similar pattern was seen in FP-IIb and IIIa with AP-1 sites [TGA(C/G)TCA]. These findings were largely confirmed with electrophoretic mobility shift assays (EMSA), with super-shifts suggesting that the protein involved at the AP-1 sites was NF-E2. Other sites observed in the DNase and EMSA assays involved CACC boxes in FP-IIIb, though a specific binding protein was not identified.

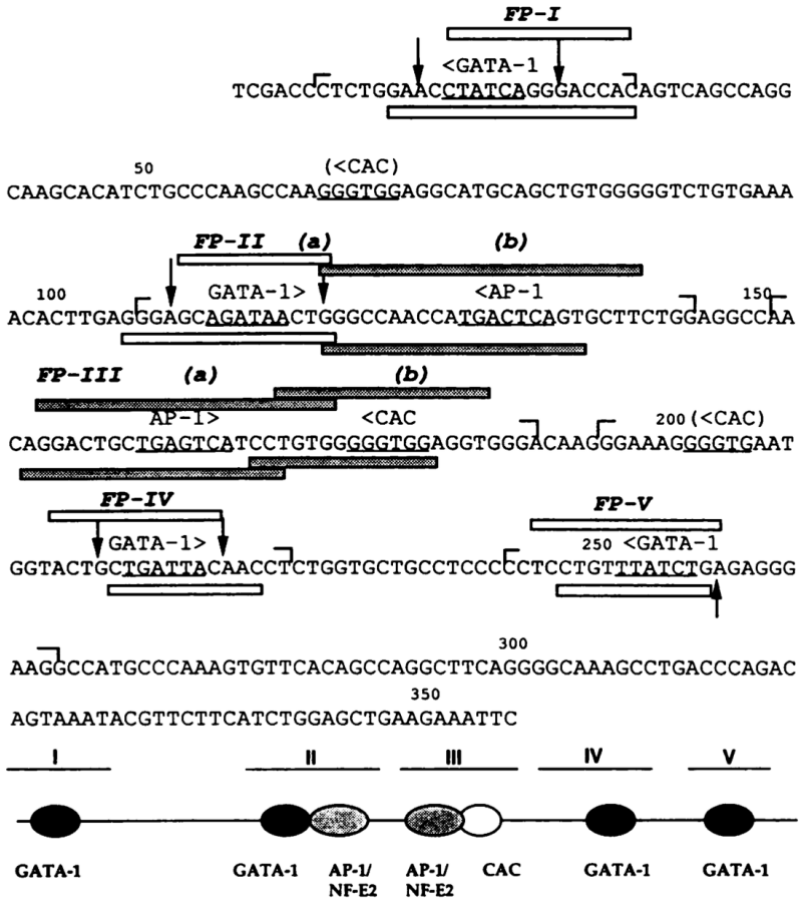


Figure 3.4 - Sequence of the 350bp fragment in Jarman et al 1991 Annotation of the 350bp fragment named pMC-TX. Five footprints (FP) are annotated, with arrows indicating very strong DNAseI hypersensitivity. Open boxes are erythroid specific footprints, stippled boxes indicate footprints also found in MEL, HEL and HeLa nuclei. A summary schematic including presumed cognate proteins is included at the bottom.

In-vivo footprinting using GA-LMPCR confirmed these findings, providing further evidence for the involvement at MCS-R2 of GATA1 and NF-E2, the latter involved at the two anteverted AP-1 sites identified in the earlier study (Strauss et al., 1992). It was noted that the upstream motif differed from the downstream and proposed consensus motif by one base: ACTGAGTCA versus GCTGAGTCA. Once again, the CACC elements were found to bind unidentified proteins, but hypersensitivity was noted between the distal GATA1 sites. Analysis of this site concluded that the hypersensitivity may have reflected local chromatin structures or interaction between the GATA1 molecules and other proteins that may not directly contact DNA.

As the regulatory elements, HS-40 chief among them, were found to be located upstream of the entire α-globin gene cluster, its role in influencing ζ-globin activity was examined (Zhang et al., 1993). An artificial reporter construct was created involving the HS-40 core element, the ζ-globin promoter and a human growth hormone (GH) reporter. Activity of HS-40 was assessed in K562, primary human erythroblasts or non-erythroid cells through quantifying secreted GH levels or primer extension assays for GH mRNA. HS-40 was found to act as an erythroid-specific enhancer of the ζ-globin promoter, acting in an orientation and position-independent manner. To explore which sites within the HS-40 core region were functionally significant, various fragments were created with truncating mutations extending from the 5' or 3' extremity. This approach found that the extreme 3' GATA1 site and the two proximal 5' sites were not major determinants for enhancer activity. When the truncating deletions from the 5' side reached the 5' NF-E2 site or from the 3' side reached the GATA1 (c) site, there was a significant drop in reporter expression as shown in Figure

3.5. This indicates either that these two sites or the sites and the sequence between them are vital for enhancer function

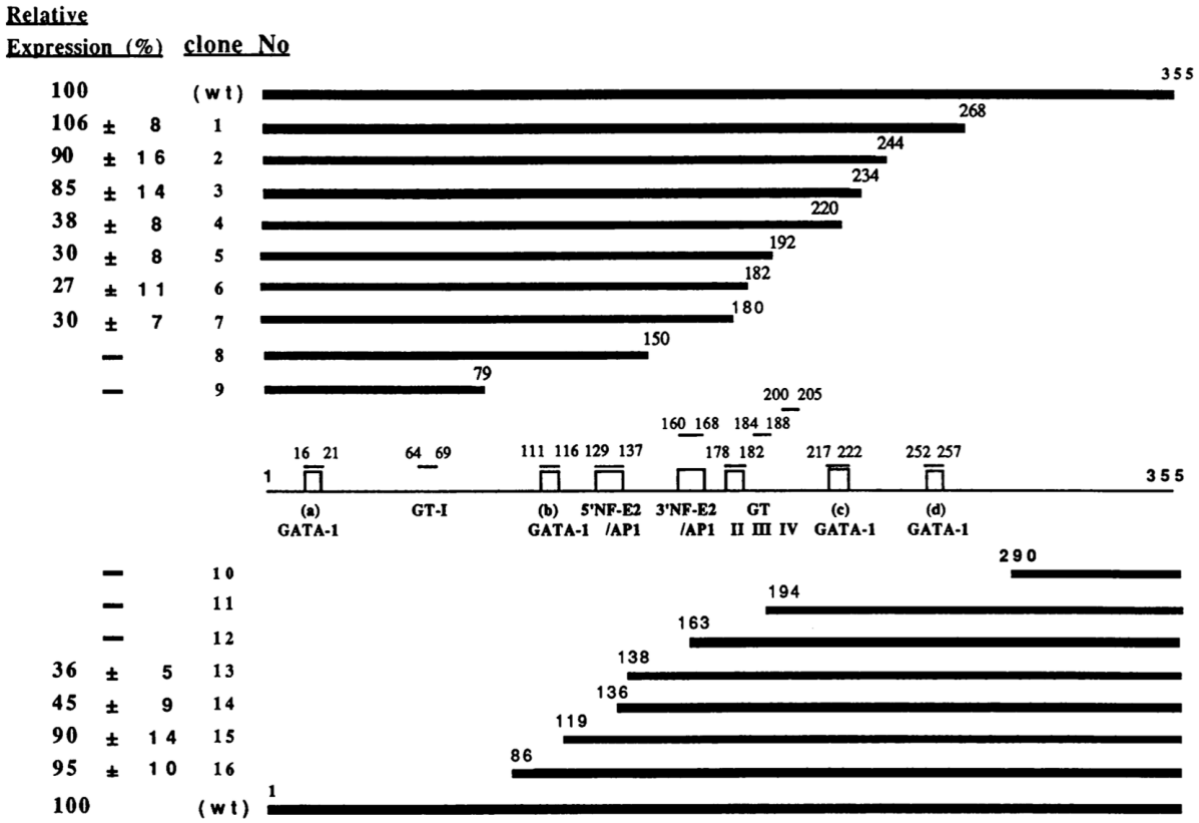


Figure 3.5 - Effects of truncation mutations of HS-40 Constructs involving the ζ-globin promoter, HS-40 fragments and the human growth hormone gene were transfected into K562 cells, and expression quantified by primer extension assay. Adapted from (Zhang et al., 1993)

Following on from this study, each TF binding motif in MCS-R2 was mutated in K562 cells, using a mutant HS-40/ζ-globin promoter/GH reporter construct in K562 cells and a mutant HS-40/α-globin promoter/GH report construct in MEL cells (Zhang et al., 1995). In MEL cells, mutations in any of the transcription factor binding sites resulted in a fall in activity, whereas the results in K562 cells, which are closer to recapitulating the haematopoietic landscape in fetal erythroid development, were more intriguing. The only GATA1 mutant site that resulted in a reduction in activity was GATA1 (c). When 3bp mutations were introduced into

either NF-E2 site or a 1bp mutation into the 5' NF-E2 site, reporter expression went down. However, when a 1bp substitution was inserted into the 3' NF-E2 site, expression increased. The proposed hypothesis was that this 1bp change abolishes NF-E2 but not AP-1 binding; in embryos AP-1 Is more abundant in the cell than NF-E2. The corollary of this hypothesis is that in fetal development as NF-E2 concentration increases it competes with AP-1 at the 5' and 3' binding sites and may somehow mediate a switch from ζ-globin to α-globin expression.

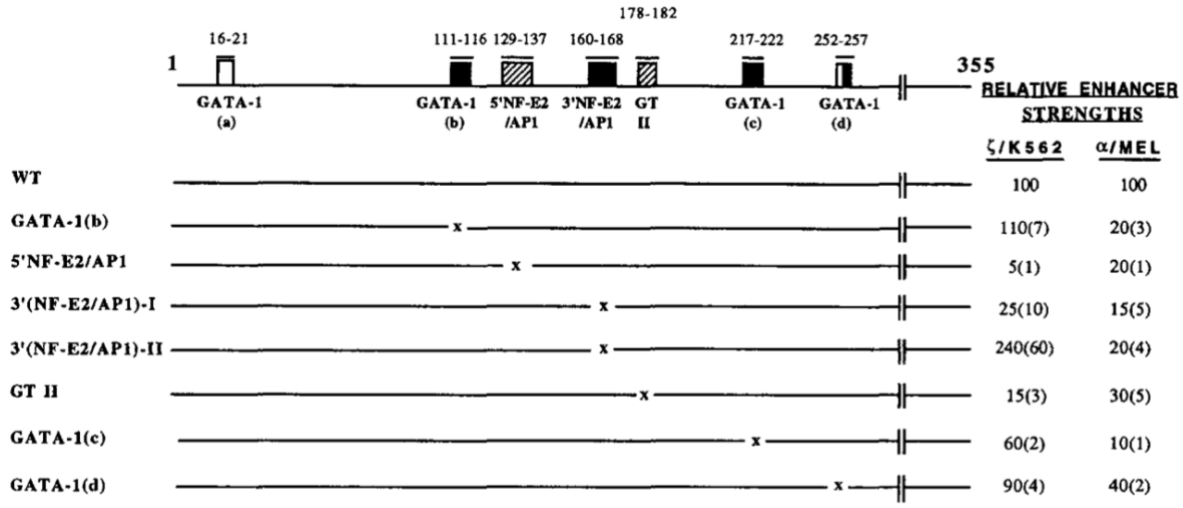


Figure 3.6 - Roles of individual TFBS in HS-40 Gene expression upon mutagenesis of individual transcription factor binding sites (TFBSs) using the HS-40- ζ-globin promote rα-globin promoter-GH reporter construct and primer extension assay in K562 and MEL cells Adapted from (Zhang et al., 1995)

Different models had been proposed throughout the 1990s to describe how enhancers function, most notably the 'enhanceosome' describing the human IFN-β locus (Thanos and Maniatis, 1995). Given that HS-40 has multiple transcription factor binding motifs, Wen *et al* attempted to characterise its functional properties (Wen et al., 2000). Through using the familiar HS-40 mutant/α-globin promoter/GH reporter construct in K562 cells they found

that when the 5' NF-E2 site, GATA1 (c) site and 3'-NF-E2 site were mutated, these specific sites lost transcription factor (TF)-binding as shown by DNase I footprinting. When the 'GT site' - the previously identified area of protection between the distal GATA1 sites - was mutagenized, there was loss of all TF-binding at the sites mentioned. Finally, when either of these sites were mutated, gene expression was reduced by 50-90%, leading the authors to characterise MCS-R2 as an enhanceosome.

3.1.6 Analysis of the sequence of MCS-R2 through evolutionary conservation

One notable difference in the works described above were the variations in TF-binding sites studied. This may have been due to the unknown significance of certain motifs at the time, but also the lack of certainty as to which motifs when identified were significant, especially as MCS-R2 has multiple potential binding sites for the same protein. For instance, whilst there are around 7 million GATA1 motifs in the human genome, studies using ChIP-seq have found occupancy to be as low as 0.07% (Katsumura et al., 2013). One powerful approach to help resolve this issue is to look for regulatory sequences that have been conserved throughout evolutionary history. Selection pressures may have kept these sequences intact as they could have a functional role. Hughes *et al* bioinformatically examined a 47-235kb syntenic region at the α-globin cluster, including 22 species spanning 500 million years of evolution (Hughes et al., 2005). The synteny of the α-like globin genes was preserved from human to fish. Due to the high conservation of the regulatory hypersensitive elements between species, they were reannotated as being multispecies conserved regulatory elements. In addition to the three known regulatory elements HS-40 (MCS-R2), HS-33 (MCS-R3), HS-10 (MCS-R4), another element was found in mammals, HS-48 (MCS-R1). MCS-R2

was the most highly conserved, with more than 69% conservation over 256bp between human and mouse. In keeping with previous work, this study showed that the core MCS-R2 element consisted of a series of TFB sites in the following order: GATA1, CACC, GATA1, NF-E2, NF-E2, GATA1, GATA1. Of these the most highly conserved was the GATA1 site immediately 5' of the NF-E2 sites as shown in Figure **3.7**.

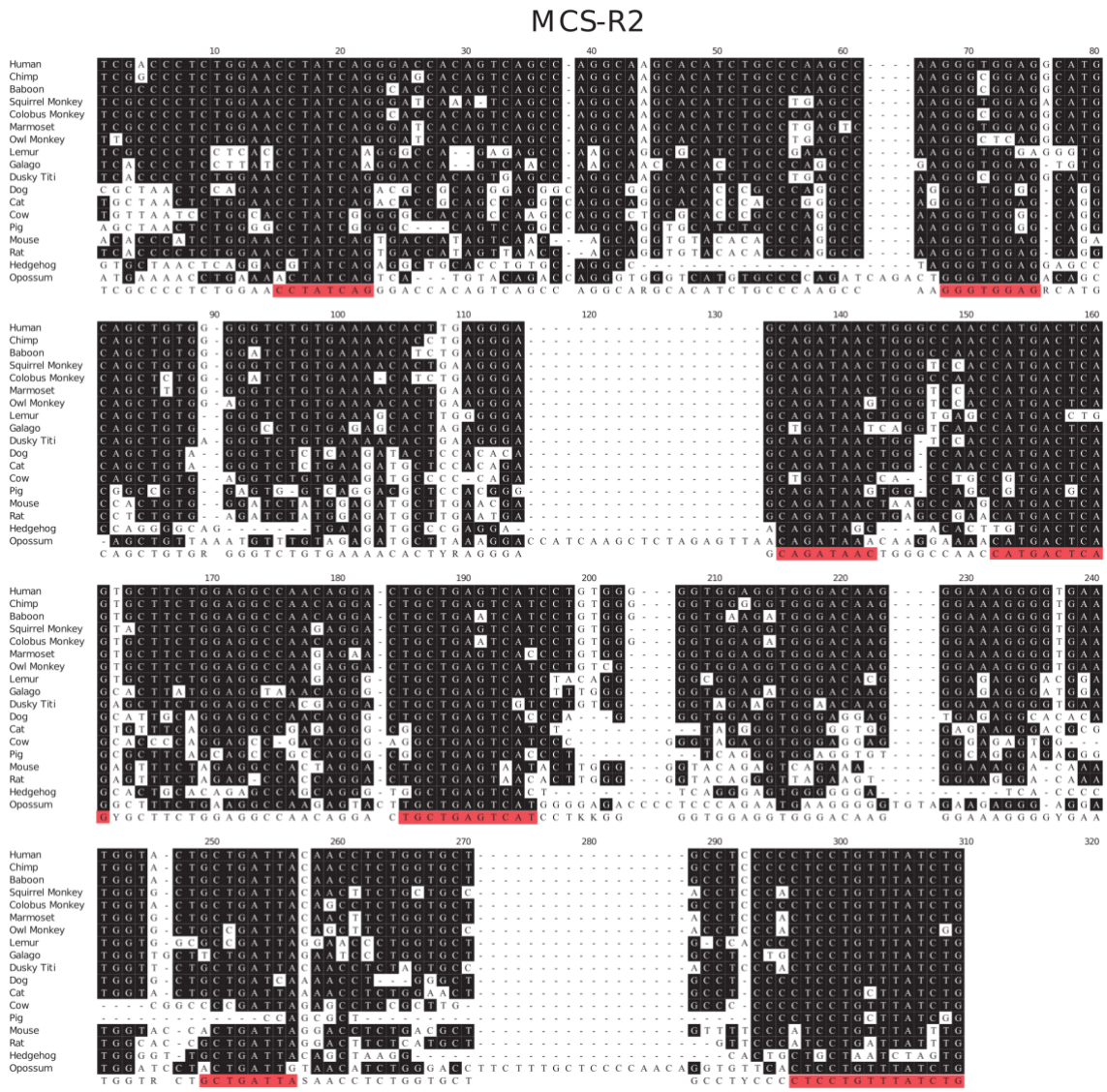


Figure 3.7 - Evolutionary conservation at MCS-R2 Sequence conservation in 18 mammalian species spanning approximately 180 million years of evolution. Black boxes denote bases identical to human, white boxes indicate bases that are different. The red coloured bases denote likely TFBSs, notable by their relative sequence conservation compared to other regions. Adapted from (Hughes et al., 2005)

3.1.7 *In-vivo* studies - Mouse

The majority of the studies discussed above had the human chromosome 16 or human cells as their prime focus. The α-globin cluster is conserved in a syntenic manner in mouse with five upstream elements. Four of them, MCS 1-4, have counterparts in human, with a murine specific enhancer MCS-Rm located between MCS-R3 and R4. Given the similarity in locus structure, insights gained in mouse may be applicable to human. The erythroid pentameric complex binds MCS-R2 in erythroid cells when analysed by ChIP (Anguita et al., 2004). As in human, NF-E2 was found to bind to MCS R1-2 but not the other enhancers. These findings were confirmed using ChIP-seq in a more recent study, which also showed that KLF1 binds to the upstream enhancers and promoters (Hay et al., 2016).

The murine model is limited in terms of the conclusions we can draw due to important differences. MCS-R1 is far more important in mouse than in human. Mice that had a homozygous deletion of MCS-R2 had an unexpectedly mild phenotype with a normal haemoglobin (Anguita et al., 2002). Individual homozygous deletions of the enhancers showed that MCS R1 and MCS-R2 contribute approximately 30% and 50% to α-globin transcription, in contrast to human where MCS-R2 is clearly the major enhancer (Hay et al., 2016).

3.2 Outstanding questions

The data from experimental systems and naturally occurring mutations all strongly indicate that MCS-R2 is the major enhancer of α-globin, critically required for its transcription. Nevertheless, several points of contention and areas for inquiry remain:

1. Patients homozygous for the $(\alpha\alpha)^{ALT}$ and $(\alpha\alpha)^{JX}$ mutations removing MCS-R2 have an unexpectedly mild phenotype; patient MC who has the former genotype had a haemoglobin concentration of 103g/L at diagnosis and has never received a transfusion. Individuals with the $(\alpha\alpha)^{JX}$ genotype seem to have a more severe phenotype but this is still milder than expected: given *in-vitro* α-globin transcription is reduced to <5%, a clinical picture similar to the Bart's hydrops fetalis syndrome would be expected. What might explain these discrepancies, both between the genotypes and between *in-vivo* and *in-vitro* human studies? It is possible that the 39-nucleotide orphan sequence involved in the $(\alpha\alpha)^{ALT}$ mutation may have some hitherto undetectable regulatory activity? This might explain why the $(\alpha\alpha)^{JX}$ mutation is more severe; the 39-nucleotide insertion is not present at the breakpoint. Another possibility is that there is compensatory activity from the other enhancers, or other unidentified elements, allowing α-globin transcription to be maintained. Finally, it must be noted that many of the assays used to study the human enhancers have been in highly artificial contexts, such as interspecific hybrid cells. α-globin transcription and translation may occur differently in whole multi-organ systems.

2. The relative contributions of the upstream enhancers have not been fully established in human. The role of each enhancer has been precisely quantified in genetically-engineered mice, but in human the same has not been perfomed *in-vitro* or of course *in-vivo.*
3. Within MCS-R2 the functional landscape of TFBSs has not been definitively mapped. Early DNase I hypersensitivity and mutagenesis assays offered detailed and occasionally surprising results, but these were conducted in highly artificial systems which partly contradicted the evolutionary conservation work performed later. Further work using more appropriate cellular systems as well as modern genomics approaches may be able to refine our knowledge of the active TFBSs that contribute to MCS-R2's activity.

In this chapter I address the first two points by characterising in detail the erythroid cells from an individual with severe HbH disease who is a rare compound heterozygote at the α-globin locus. I then used various bioinformatic and genomic assay-based techniques as well as the previously described research to arrive at a schematic for proposed functional TFBSs in MCS-R2.

3.3 Results 1 - Patient NSE

3.3.1 Patient NSE - clinical presentation

NSE is a Dutch female born in Surinam from parents of Indonesian descent. She has a haematologically very severe form of α-thalassemia (Figure 3.8). She initially presented, aged 8 years, complaining of shortness of breath during sports activities and occasional tiredness. She was otherwise well; her growth and development were normal. She has received two blood transfusions, prompted by concurrent infections during childhood. Assessment for hypomenorrhea during puberty did not reveal any abnormalities. At the time of referral, she was 26 years old and in full-time employment as an office worker. On examination, she did not have any evidence of skeletal dysplasia or facial abnormalities. An echocardiogram did not find any evidence of pulmonary hypertension and an abdominal ultrasound found no hepatomegaly though the spleen was enlarged (16.7cm x 14.3cm x 7.2cm). She had a normal bone mineral density, and T2* MRI did not find any evidence of cardiac or hepatic iron overload.

Test	NSE	Mother of NSE	Father of NSE	Reference HbH Values
Hb (g/L)	69	133	157	84 ± 16
MCV (fL)	40	72	73	73.8 ± 6.7
MCH (pg)	12.7	22.8	24	20.2 ± 1.8
HbH on BCB	30.7%	-	-	15.9% ± 8.0

***Figure 3.8 - Haematological indices**. Reference ranges for HbH values are drawn from (Chen et al., 2000)*

Hematological analysis revealed a severe hypochromic, microcytic anaemia, with anisopoikilocytosis noted on blood film examination. Brilliant cressyl blue staining of peripheral blood showed HbH inclusions in 30.7% of cells (Figure 3.9). The predominant haemoglobin molecule on cation-exchange high performance liquid chromatography (CE-HPLC) testing was HbA, a fast-moving band consistent with HbH and a small amount of HbA_2/HbE (Figure 3.10). Glycolytic enzyme analysis of NSE's blood indicated the RBC population was composed of young cells with hexokinase 5.8 [normal range 0.8-1.5 units/ml], pyruvate kinase 27.3 [6.1-12.3 units/g Hb], reticulocyte count $127x10^9$ cells/L, RBC $5.85x10^{12}$ cells/L).

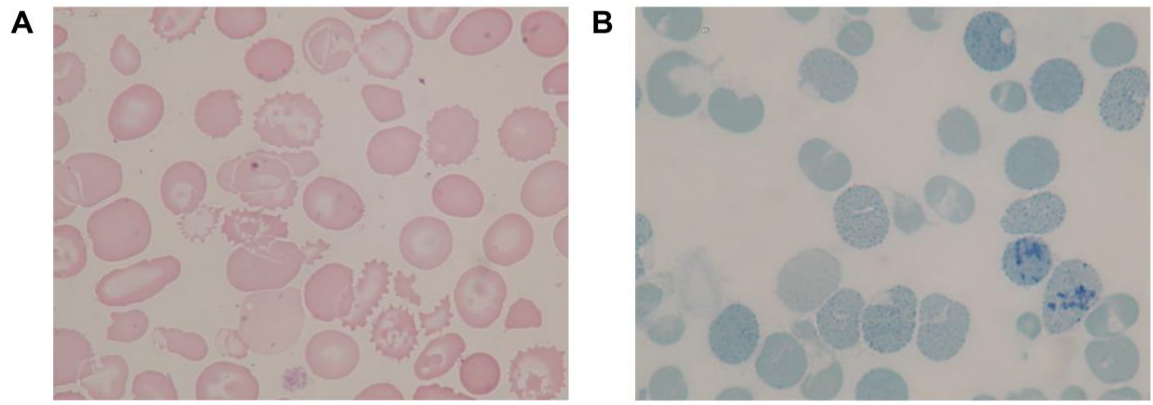


Figure 3.9 - NSE Peripheral blood film examination A) Peripheral blood film, Giemsa stained, showing marked hypochromia and anisopoikilocytosis consistent with a diagnosis of severe HbH disease B) Hemoglobin H film, Brilliant Cresyl Blue stained, exhibiting typical cells containing HbH inclusions.

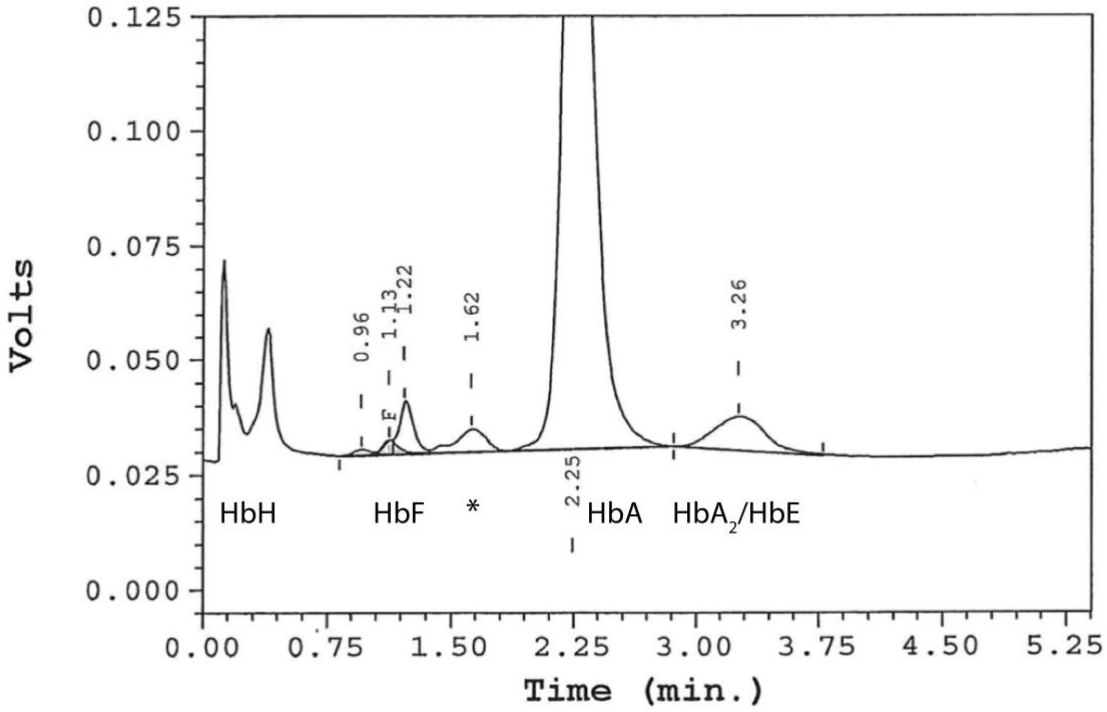


Figure 3.10 - CE-HPLC results from peripheral blood *HPLC results showing haemoglobin profile in peripheral blood from NSE. The different haemoglobin molecules represented by the peaks are annotated below the trace. The majority of the haemoglobin is comprised of by HbA eluted at 2.25 minutes, and HbE/A_2 at 3.26 minutes. * denotes a small peak comprised of degraded HbA.*

3.3.2 Patient NSE - genetic diagnosis

MLPA analysis (Figure 3.11) and agarose gel (Figure 3.12) analysis showed that both α-globin genes are deleted from one allele ($--^{SEA}$) inherited from her mother. Her paternal allele is affected by the previously reported mutation [$(\alpha\alpha)^{ALT}$] in which MCS-R2 is deleted (Coelho et al., 2010).

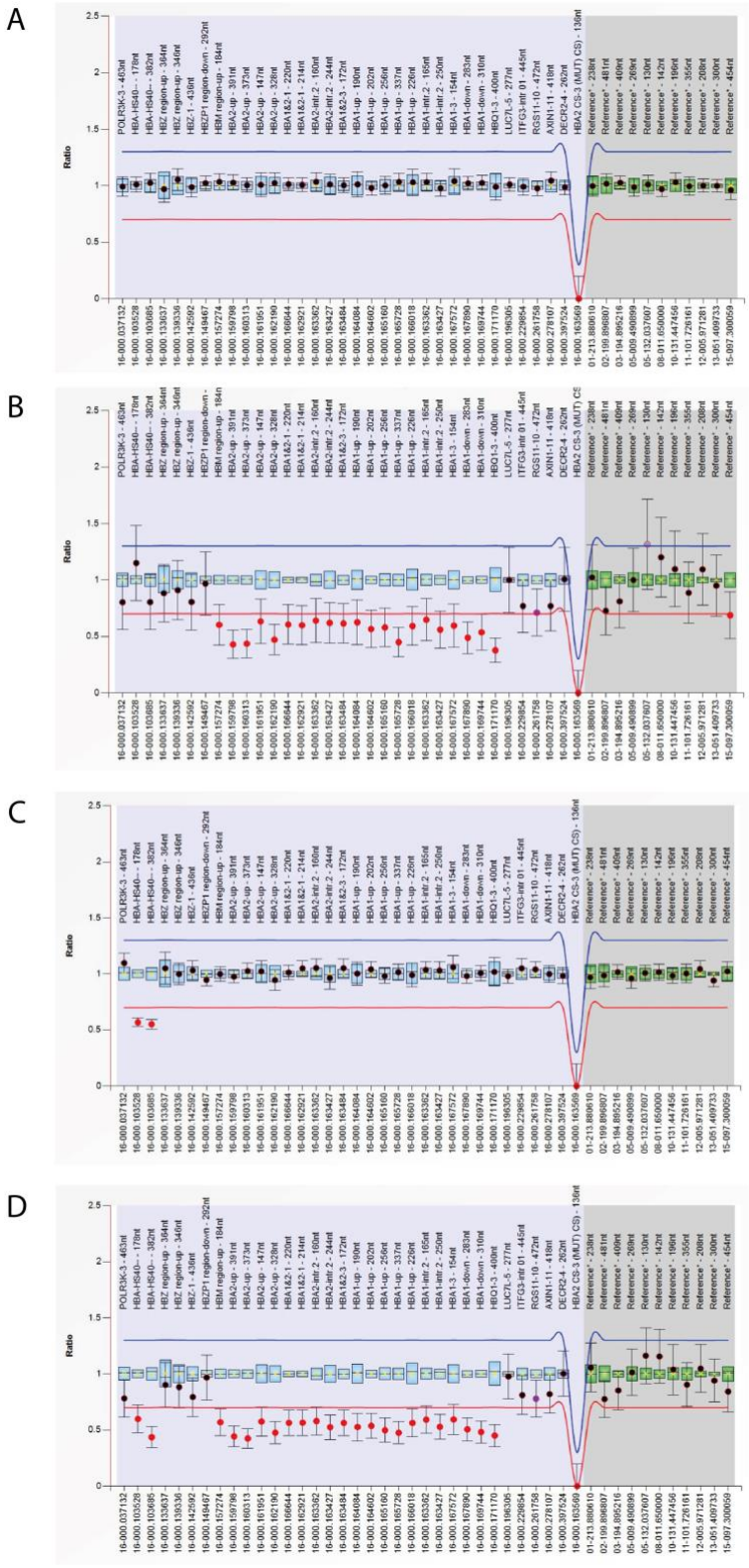


Figure 3.11 - MLPA analysis of the α-globin locus *MLPA probes in the blue shaded region are ranged across the α-globin cluster. A ratio of 1 indicates normal biallelic binding of a probe, a ratio of 0.5 indicates heterozygous loss and a ratio of 0 indicates homozygous loss. A) Normal control B) Mother of NSE. Ratio of 0.5 over the α-globin genes sparing ζ-globin representing --^{SEA} C) Father of NSE. Ratio of 0.5 over MCS-R2 indicating heterozygous loss. D) NSE. Ratios indicating heterozygous loss of MCS-R2 and --^{SEA}.*

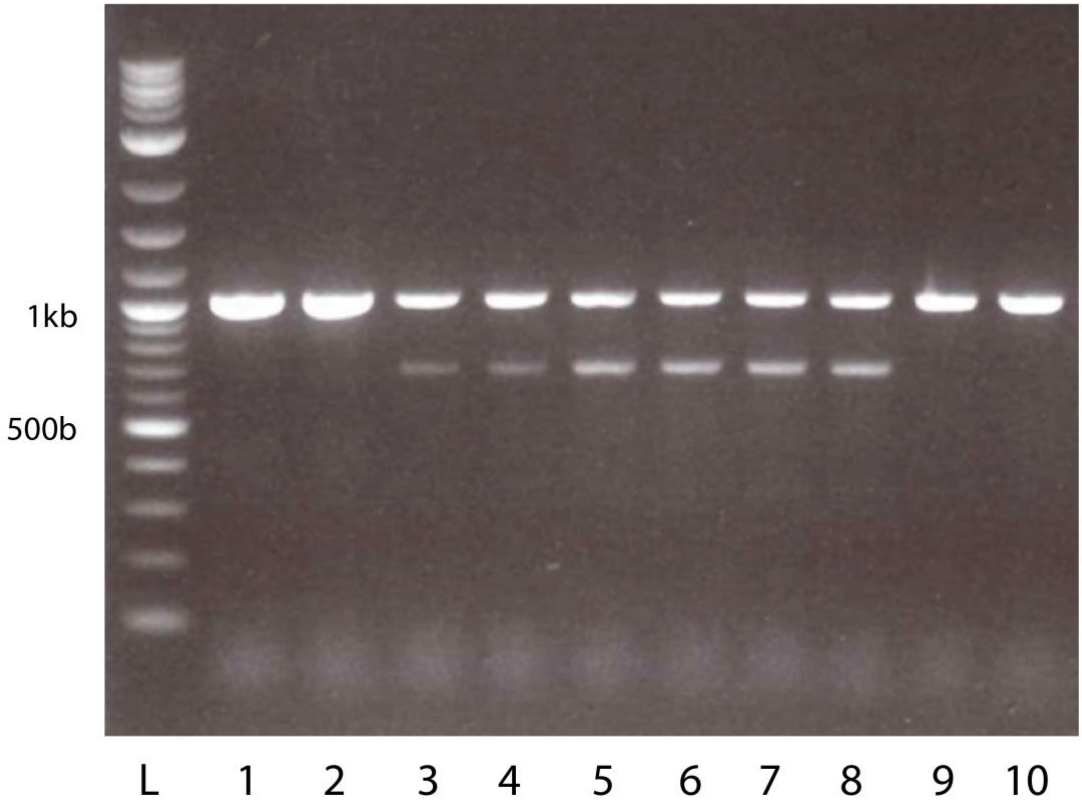


Figure 3.12 - SEA deletion agarose gel analysis *Multiplex PCR run on NSE and her parents to diagnose the $--^{SEA}$ deletion. A single band at ~1kb identifies a WT allele, and a band ~700bp identifies an allele with the mutation. Lanes annotated as follows: L - ladder; 1, 2 - negative control; 3, 4 - positive control; 5, 6 - mother of NSE; 7, 8 - NSE; 9, 10 - father of NSE.*

The deletion breakpoints were mapped to hg19:163,192-166,554, showing a deletion of 3.3kb with a 39-nucleotide orphan segment insertion in keeping with the $(\alpha\alpha)^{ALT}$ genotype (CAGTGGAGAGAGACTGTGCCCTCCACTGCTCTCTCCAGT). NSE is also heterozygous for the common codon 26 β-globin mutation causing HbE, and therefore her genotype is broadly consistent with that seen in patients with AE-Bart's disease (Chaibunruang et al., 2014). No other mutations were detected at the globin loci.

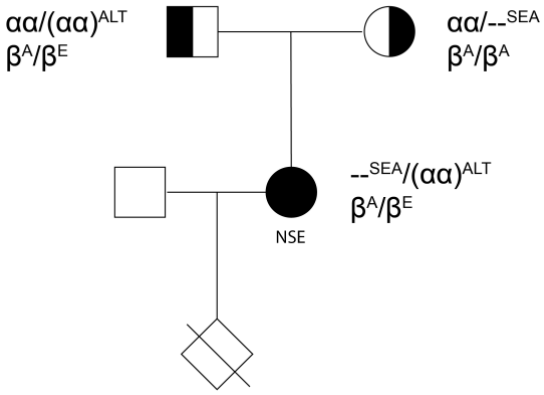


Figure 3.13 - Family pedigree of NSE with annotated genotypes

The consequence of this genotype is that all α-globin transcription in this patient comes from her paternal allele affected by $(\alpha\alpha)^{ALT}$ on which MCS-R2 is deleted. From this paternal allele, presumably MCS-R1 alone drives expression of the two α-globin genes *in-cis*. NSE thus produces only 50% of the α-globin produced by the previously described $(\alpha\alpha)^{ALT}/(\alpha\alpha)^{ALT}$ homozygote with severe HbH disease: the total α-globin output compared to normal (αα/αα) would be predicted to be 0% from the maternal ($--^{SEA}$) allele, and less than 5% from the $(\alpha\alpha)^{ALT}$ allele. Given such a severe predicted reduction in α-globin expression, it was surprising that the patient survives without transfusions.

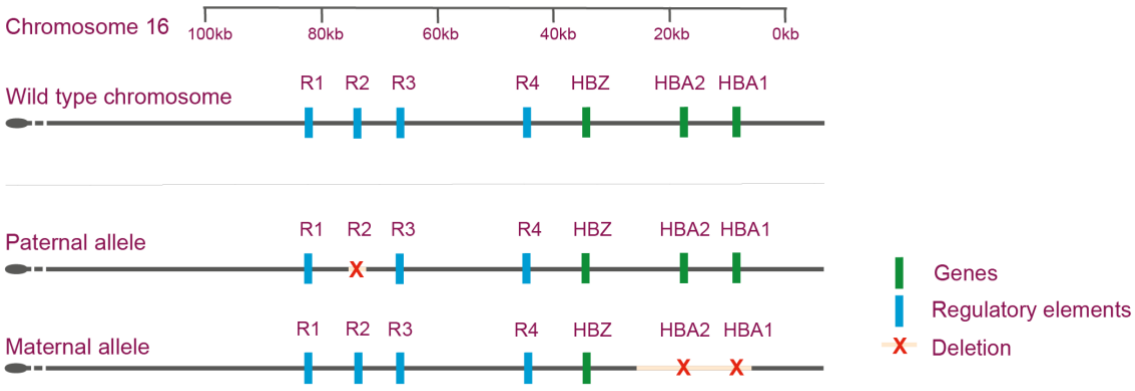


Figure 3.14 - Schematic of NSE's genotype at the α-globin cluster

3.3.3 Patient NSE - assessment of erythroid differentiation and transcription

To investigate further, an initial assessment of transcription at the globin loci was performed. CD34+ cells were selected from peripheral blood and differentiated along the erythroid lineage. Cell counts were similar to control cells throughout differentiation (Figure 3.17). Immunophenotyping and morphology assessed by cytospin showed normal erythroid differentiation (Figure 3.15 and Figure 3.16). At peak globin production, at day 10 and day 13 of differentiation, the α/β-globin mRNA ratio was markedly reduced when compared to WT cells and those of a patient carrying the $--^{SEA}$ deletion ($\alpha\alpha/--^{SEA}$) as assessed by qPCR. Even so, the reduction in α-globin expression was not as great as predicted: the α/β ratio was 0.17 in NSE's cells compared to WT (Figure 3.18A). Interestingly the α/β-globin mRNA ratio in NSE was more balanced in peripheral blood than in CD34+ cell derived erythroid precursors (Figure 3.18B).

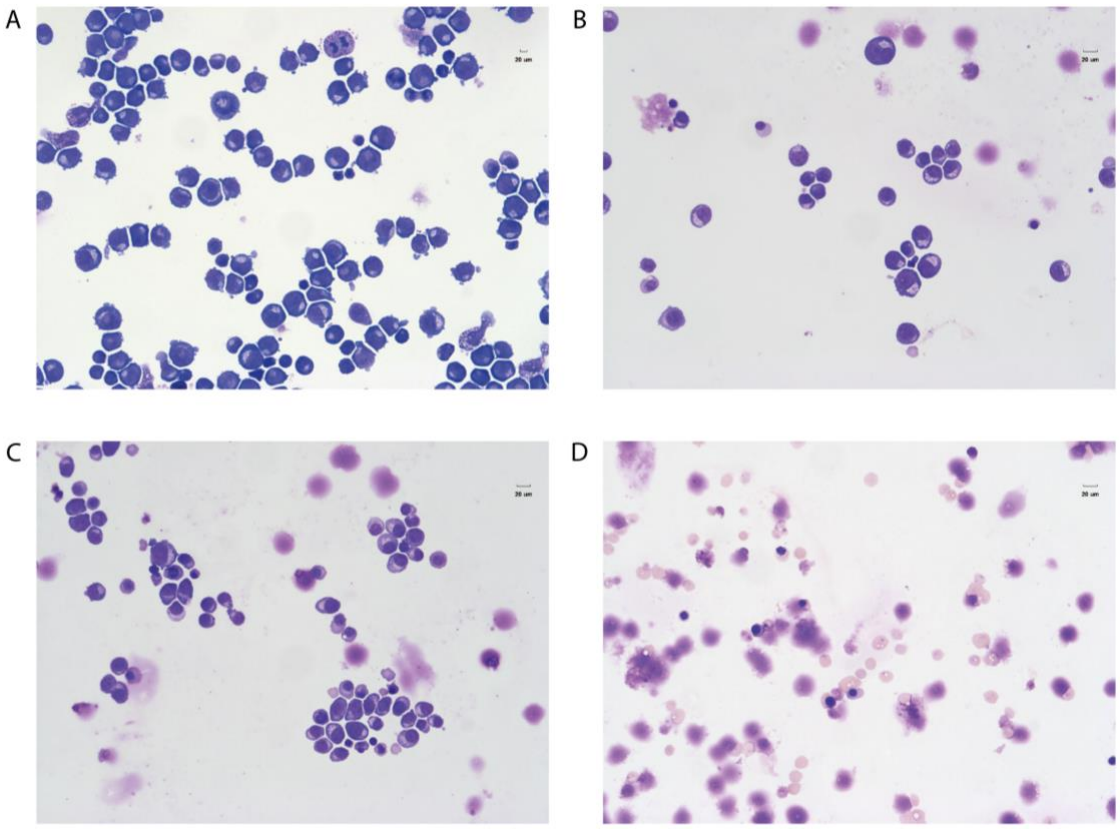


Figure 3.15 - Cytospins through differentiation *Cytospins of erythroid cells derived from CD34+ cells isolated from NSE. A) Day 7 B) Day 11 C) Day 13 D) Day 17. There was no delay in differentiation as compared to normal controls as shown in (Scott et al., 2020).*

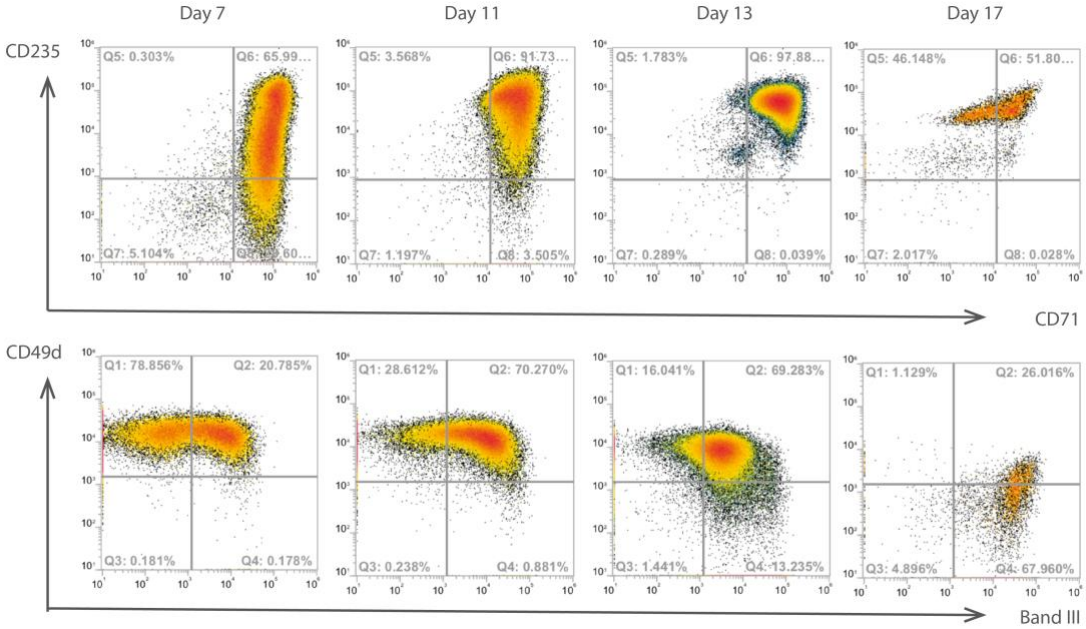


Figure 3.16 - Immunophenotyping of patient cells *Flow cytometry of differentiated patient erythroid cells from day 7 through to 17. The top panel shows CD71 and CD235, both gated on live cells. The bottom panel shows CD49d and Band III both gated on the CD235 positive population. There was no delay in differentiation as compared to normal controls as shown in (Scott et al., 2020).*

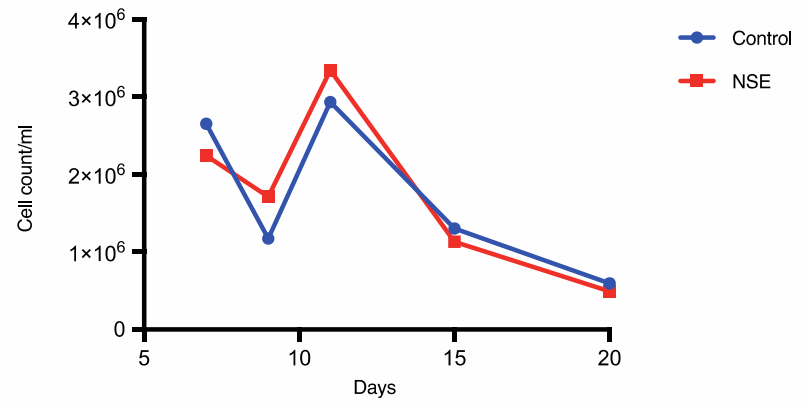


Figure 3.17 - Cell counts in culture *Cell concentrations given in cell count/ml. Cells not counted early in differentiation due to low cell number.*

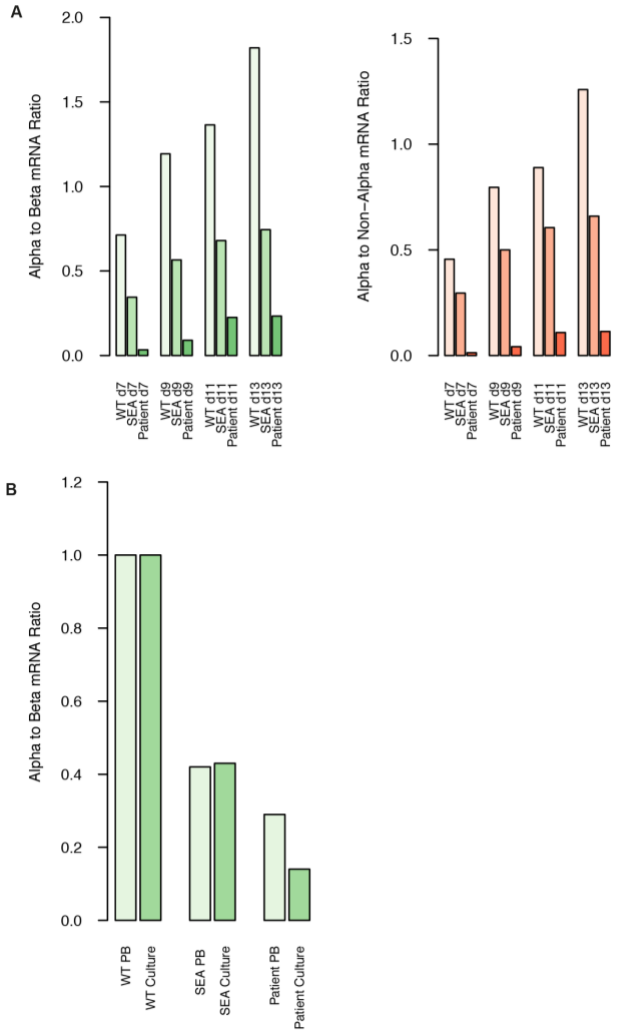


Figure 3.18 - Globin gene expression from NSE's cells in culture and in peripheral blood *RT-qPCR assessment of globin expression normalised to RPL13a. A) α/β and α/non-α globin ratios throughout erythroid differentiation in culture. Gene expression from NSE is compared to an individual carrying the $--^{SEA}$ mutation in heterozygosity and WT. B) α/β ratios comparing globin expression in culture and in peripheral blood from individuals that are WT, carrying the $--^{SEA}$ mutation in heterozygosity and NSE.*

3.3.4 Patient NSE - ζ-globin is not significantly de-repressed

One potential explanation for the mild phenotype is that ζ-globin, the embryonic form of α-globin, could have been de-repressed in the same way that γ-globin is in patients with HPFH. ζ-globin transcription was assessed by qPCR; only a small amount of ζ-globin was expressed, as reflected by the low α-like/β-like ratio (Figure 3.19A). Next, protein levels were assessed. When paired with β-globin in haemoglobin molecules, ζ-globin forms haemoglobin Portland II ($\zeta_2\gamma_2$). Isoelectric focusing (IEF) was performed on cell lysate from NSE. Compared to a reference control only a small amount of Hb Portland II was detected (Figure 3.19B). RP-HPLC was conducted to look for individual ζ-globin chains. A small amount of ζ-globin was detected, comprising 0.57% of all globins produced in the sample from NSE, which was more than 10-fold less than that observed in a ζ-globin expressing control (Figure 3.19C). These assays together showed that ζ-globin was not significantly activated to compensate for the loss of α-globin.

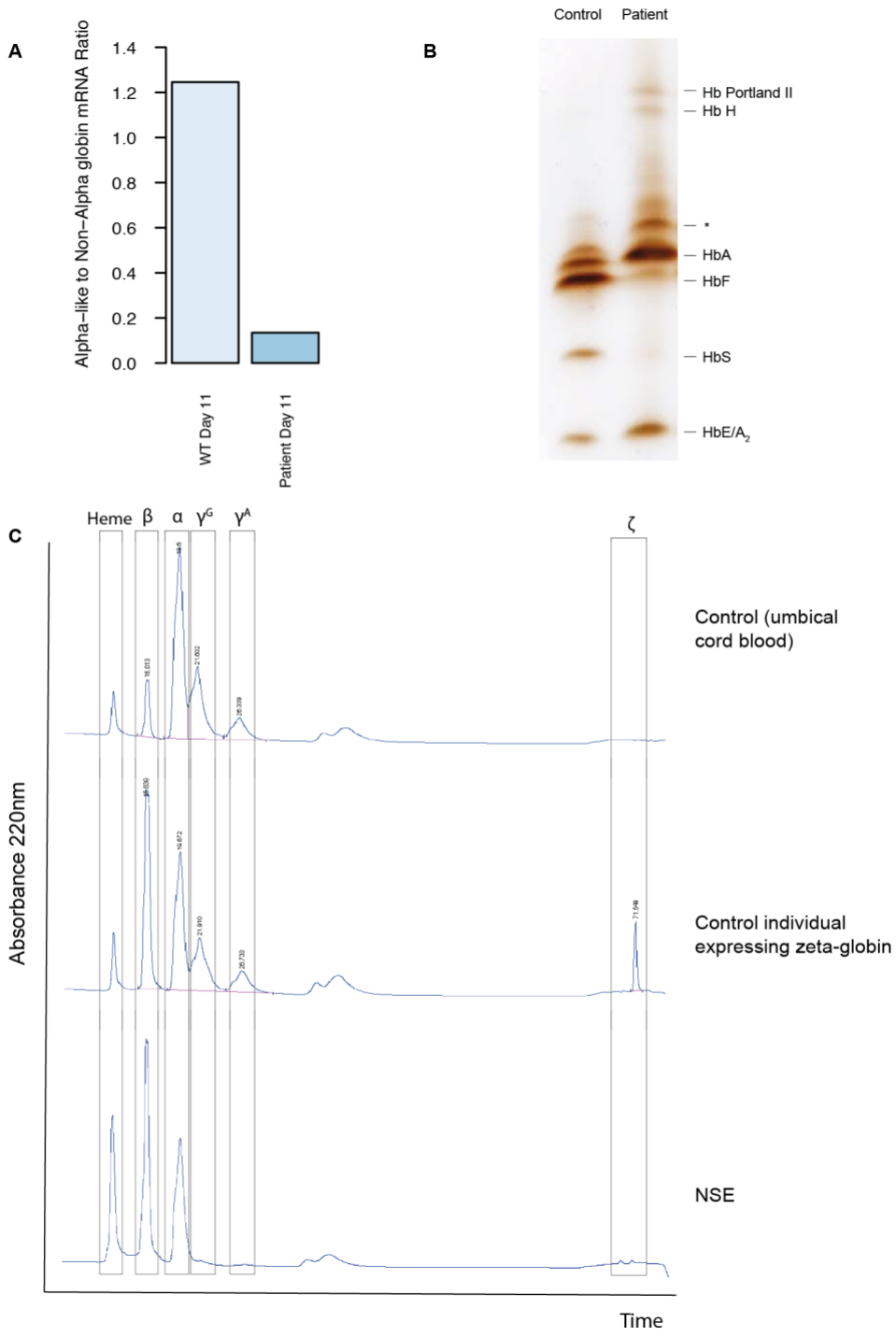


Figure 3.19 - Characterisation of mild ζ-globin derepression *A) (α+ζ)/(β+γ)-globin mRNA ratios at day 11 compared to WT, normalised to RPL13A, measured by RT-qPCR. B) IEF analysis of cell lysate from NSE comparing normal and variant globin production to a reference control sample. The majority of the globin produced by the NSE is HbA and HbE ($\alpha_2\beta^E_2$) in keeping with the genotype. A small amount of Hb Portland II ($\zeta_2\beta_2$) is detected. * denotes an artefact observed when samples age during storage and processing. C) RP-HPLC quantifying individual globin chains produced by a WT control, a control individual expressing ζ-globin and NSE.*

3.3.5 Patient NSE - assessment of chromatin accessibility at the globin loci

Whilst α-globin expression was severely reduced, it was still higher than expected. One potential explanation for this was that there may be compensatory changes in the chromatin landscape at the α-globin locus. This might be in the form of increases in accessibility of the remaining enhancers (R1, R3 and R4), or the emergence of previously unknown enhancers on the $(\alpha\alpha)^{ALT}$ allele. Enhancers can be identified by their accessibility to transposases using ATAC-seq, reflecting their accessibility to TFs *in-vivo*. Compared to WT and a patient with the $--^{SEA}$ deletion there was a reduction in the ATAC peak corresponding to MCS-R2, in keeping with its deletion from one allele. However, we did not observe any new peaks forming in a 150kb region including and surrounding the α-globin genes, or any compensatory increase in the peak heights of the remaining enhancers, although functional alterations in TF binding without changes in accessibility cannot be ruled out. Consistent with the reduction in α-globin transcription, there was a reduction in ATAC signal over the α-globin genes compared to control. Importantly, in all replicates there were no reads in the pre-mapped data that contained the 39-nucleotide insert, supporting previous work indicating that this orphan sequence did not have any functional capacity (Mettananda et al., 2017a). No significant changes were observed at the β-globin locus.

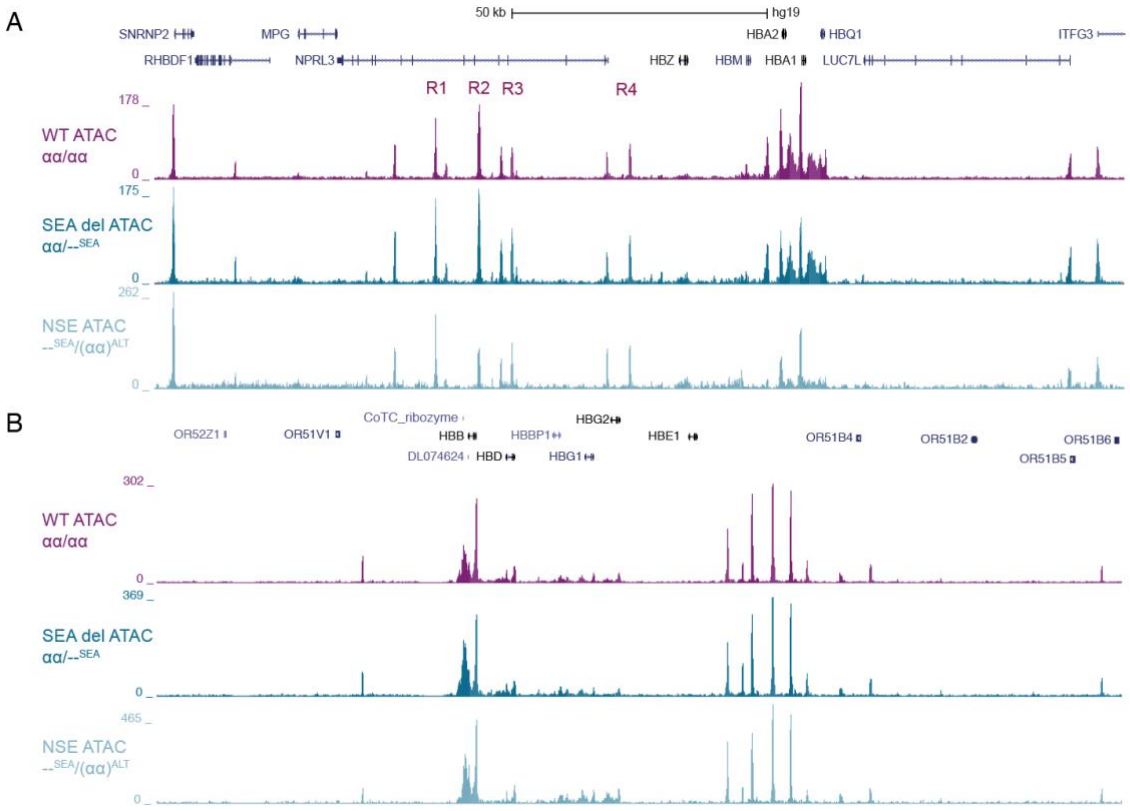


Figure 3.20 - Chromatin accessibility at the globin loci *ATAC-seq in primary erythroid cells on Day 13 of erythroid differentiation. A) The top track is generated from cells derived from three unrelated wild-type controls (αα/αα), the middle is from a carrier of the SEA mutation (--^{SEA}/αα) and the bottom is from NSE (--^{SEA}/(αα)^{ALT}). B) ATAC-seq at the β-globin locus.*

3.3.6 Patient NSE - assessment of chromatin interactions at the α-globin locus

Even through there were no major perturbations in the accessibility landscape, chromatin interactions between the intact enhancers and the promoters may have increased. Whilst these might not have been detected by ATAC-seq, these potential alterations could have increased the amount of α-globin transcription. To investigate this possibility Capture-C, a sensitive assay to detect physical interactions between regions of chromatin, was performed (Davies et al., 2016). When capturing from the intact α-globin promoters,

present only on the $(\alpha\alpha)^{ALT}$ allele *in-cis* with the MCS-R2 deletion, no interactions with MCS-R2 were detected. This clearly shows that the single copy of MCS-R2 on the $--^{SEA}$ allele was not interacting with the α-globin genes on the other allele and therefore not driving α-globin expression in-*trans*. Surprisingly, when compared to three WT controls, interactions between the α-globin promoters and the intact MCS-R1, -R3 and -R4 enhancers on the $(\alpha\alpha)^{ALT}$ allele were also reduced rather than increased as might be expected if they were compensating for the loss of MCS-R2.

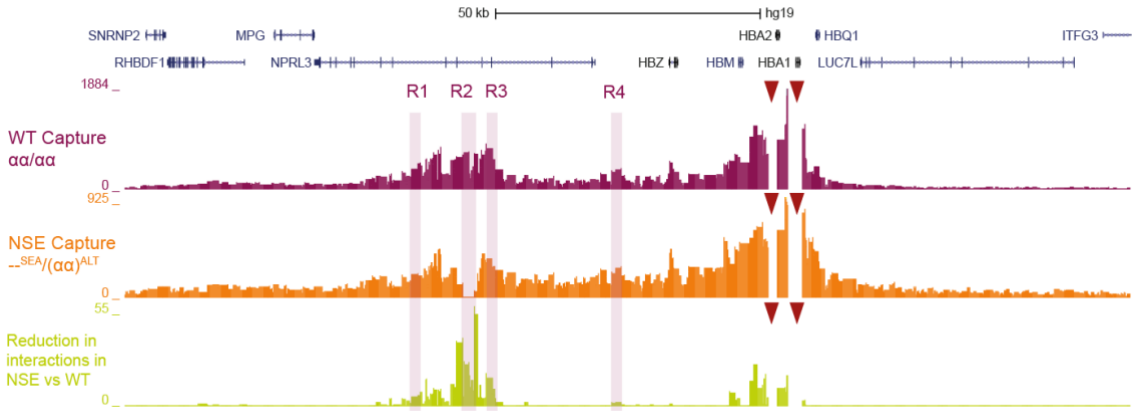


Figure 3.21 - Chromatin interaction profiles in NSE and WT erythroid cells *Chromatin interaction profiles between the α-globin promoters (arrow heads) and surrounding chromatin in primary erythroid cells (Capture-C). Peaks represent interactions with the α-globin promoters. The top track depicts the mean interaction profile observed in cells from three wild-type controls (αα/αα) The middle track is the interaction profile observed in NSE's cells, showing an absence of interactions between the α-globin promoters and MCS-R2, in keeping with its deletion. The bottom track depicts the reduction in interactions when comparing NSE and the WT controls, represented as log adjusted p-values. There is a highly significant reduction in interactions between the α-globin promoters and MCS-R2 due to its deletion, and a modest reduction in interactions between the promoters and MCS-R1, -R3, and -R4.*

In the light of this unexpected reduction in interactions with the remaining enhancer elements on the $(\alpha\alpha)^{ALT}$ allele, the ATAC-seq data was re-examined to look for a reduction in allele-specific chromatin accessibility as measured by ATAC-Seq using variants between alleles. Just one SNP was found - within MCS-R3 – and, in keeping with the Capture-C data,

there appeared to be a reduction in accessibility *in-cis* to the MCS-R2-deletion, with only 38% of the reads coming from this allele. Therefore, in NSE there was no increase in chromatin interactions on the $(\alpha\alpha)^{ALT}$ allele to compensate for the deletion of MCS-R2.

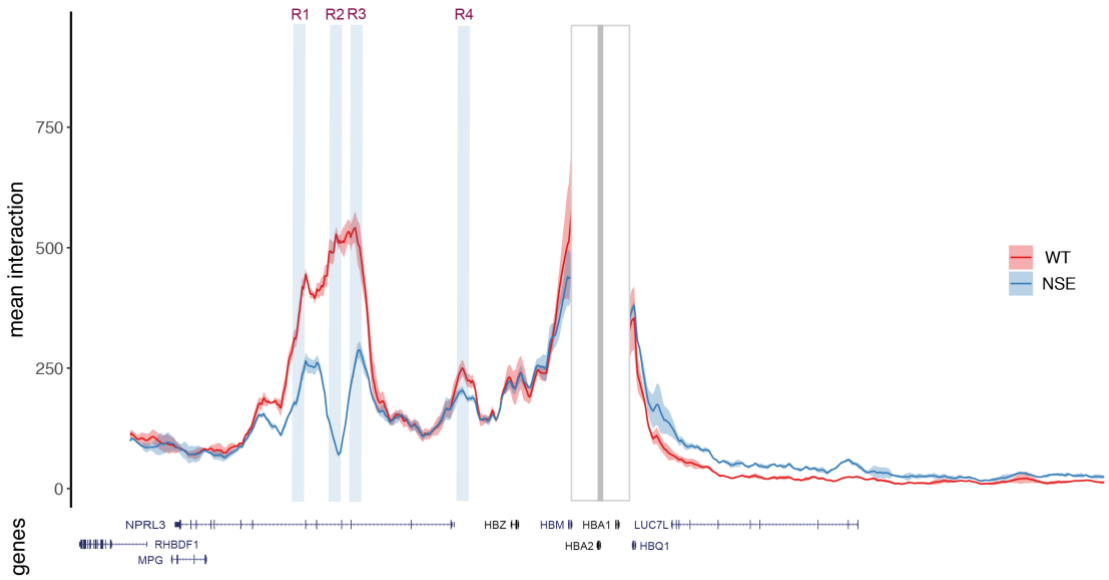


Figure 3.22 - Alternative mean interaction profile at the α-globin locus *Data are presented as mean interactions per restriction fragment plus and minus one standard deviation of sliding 5kb windows. The grey vertical bar indicates the position of the viewpoint, with exclusion zone either side to allow for proximity effects. A reduction in interactions in NSE's cells compared to WT is observed at MCS-R1, -R2 and -R4, as well as a modest increase in interactions with chromatin 3' to the α-globin genes.*

3.4 Discussion 1 - Patient NSE

The characterisation of individual patients who carry rare mutations has helped establish and further our understanding of gene regulation, particularly at the α- and β-globin loci. This has deepened our knowledge of how disease phenotypes manifest, in predictable and unexpected ways. The study of NSE confirms some of what was previously known about gene regulation at the α-globin cluster. MCS-R2 is clearly functionally the most important enhancer, and its loss results in severe HbH disease. This is in contrast to mouse, where it is known that MCS-R1 is highly functional, the relative contributions towards α-globin expression being 40% (MCS-R1), 50% (-R2), <2-3% (-R3) and 10% (-R4) (Hay et al., 2016). In human, these data show that the in human, the enhancers at α-globin locus are likely to contribute approximately 10% (MCS-R1), 90% (-R2), <2-3% (-R3) and <2-3% (-R4). This differs from previous *in-vitro* work outlined in 3.1.2, which would indicate that MCS-R2 contributes >97% to α-globin expression (Bernet et al., 1995, Wallace et al., 2007). This discrepancy may be explained by the model systems used. In these studies, expression was assessed in interspecific MEL cells and a humanised mouse respectively. Whilst the humanised mouse is less artificial than the former, *trans*-acting factors and the native chromosome are still murine, and so may affect gene regulation at the humanised α-globin locus. In this study, primary human cells were examined with no genetic manipulation involved, which is likely to recapitulate gene expression at the human α-globin locus more faithfully. Therefore, MCS-R1 may in fact may be more functionally important in human than previously thought, though still clearly not as active as in mouse. One potential objection to this finding is that the 39-nucleotide insertion may have some functional activity which could be falsely attributed to MCS-R1. As discussed in 3.1.3 and in this study, no evidence has been

found to support this theory. Unfortunately, the resolution of Capture-C data is not fine enough to check whether there was a significant increase in contacts between the α-globin promoters and this orphan sequence. However, the fact that no reads in the ATAC-seq data set involved it strongly argues against the sequence having a functional role.

Why was the phenotype not as severe as expected? NSE appears to have a well-compensated haemolytic anaemia with no evidence of significant ineffective erythropoiesis or haemolysis. The α/β-globin mRNA ratio was lower than reported from most patients with HbH disease (Harteveld and Higgs, 2010). ζ-globin was not significantly activated to compensate for the loss of α-globin. One ameliorating factor may have been the co-inheritance of the β^E mutation which acts as a mild form of β-thalassemia, thereby reducing globin imbalance. Furthermore, NSE's genotype is similar to that seen in individuals with AE-Bart's Disease, in which a severe reduction in α-globin synthesis occurs in combination with the β^E mutation. Individuals with α-thalassemia who co-inherit β^E-thalassemia may have a milder phenotype than otherwise expected (Chaibunruang et al., 2014).

In the absence of evidence for compensation via the remaining enhancers, an only mildly raised RBC ($5.85x10^{12}$ cells/L) and a normal erythropoietin level (14IU/L) at diagnosis, it seems most likely that the relatively mild clinical phenotype occurs due to selection of red cells with the least globin chain imbalance. Interestingly the α/β-globin mRNA ratio in the patient was more balanced in peripheral blood than in CD34+ culture-derived erythroid precursors. Since there will be a normal distribution of the α/β-globin ratios around the mean in any population of erythroid precursors this selective process at the cellular level may play an important role in this remarkable clinical phenotype.

These findings suggest that although at some loci compensation in gene expression may be mediated by enhancer redundancy, other pathophysiological mechanisms may compensate in whole organ/cell systems. Clearly, when evaluating the effects of enhancers on gene expression it is important to consider other influences on the ultimate cellular and organismal phenotype.

3.5 Results 2 - MCS-R2 sequence analysis

3.5.1 Derivation of transcription factor binding site motifs within MCS-R2

Since MCS-R2 is the most highly conserved and powerful enhancer, it provides an excellent model for understanding in more detail how the syntax of such a sequence relates to its function. A key feature of enhancers is their ability to bind transcription factors in a tissue- and developmental-stage specific manner (Heinz et al., 2015). Multiple identical TFBSs may be found within the same region; approximately 2% of the human genome is thought to be composed of these homotypic clusters of TFBSs (Gotea et al., 2010). One strategy to uncover important functional sequences in an enhancer is to identify motifs within its sequence that are potential binding sites for the TFs known to bind to it. Various computational models are available for predicting and defining binding site sequences, curated in databases such as JASPAR in the form of position weighted/frequency matrices (PWMs/PFMs) (Sandelin et al., 2004, Wasserman and Sandelin, 2004).

Analysis using ChIP-seq has shown that MCS-R2 in human is bound by the erythroid pentameric complex, NF-E2 and KLF1 (De Gobbi et al., 2007, Su et al., 2013). Despite this, publicly available curated PWMs fail to identify binding sites in MCS-R2 for certain TFs. An explanation for this may lie in the type of datasets used to generate these motifs; if they were in the wrong cell type or developmental stage the binding patterns of TFs could potentially be different (Starick et al., 2015). This may particularly be the case when comparing the binding patterns of TFs when they bind alone and when they bind as part of a complex (Jolma et al., 2015, Cohen et al., 2018, Lin et al., 2020).

Furthermore, even though KLF1 is known to bind to MCS-R2, and activates ZBTB7A/LRF expression in erythroid cells, suppressing γ-globin expression, there are no curated binding motifs in human available on JASPAR (Norton et al., 2017). KLF1 is known to bind to CACCC sequences within the murine and human β-globin promoters (Miller and Bieker, 1993). This motif was later refined to CCM CRC CCN, albeit in mouse (Tallack et al., 2010). Despite the high sequence homology between mouse and human KLF1, this motif does not exist in MCS-R2 (Bieker, 1996).

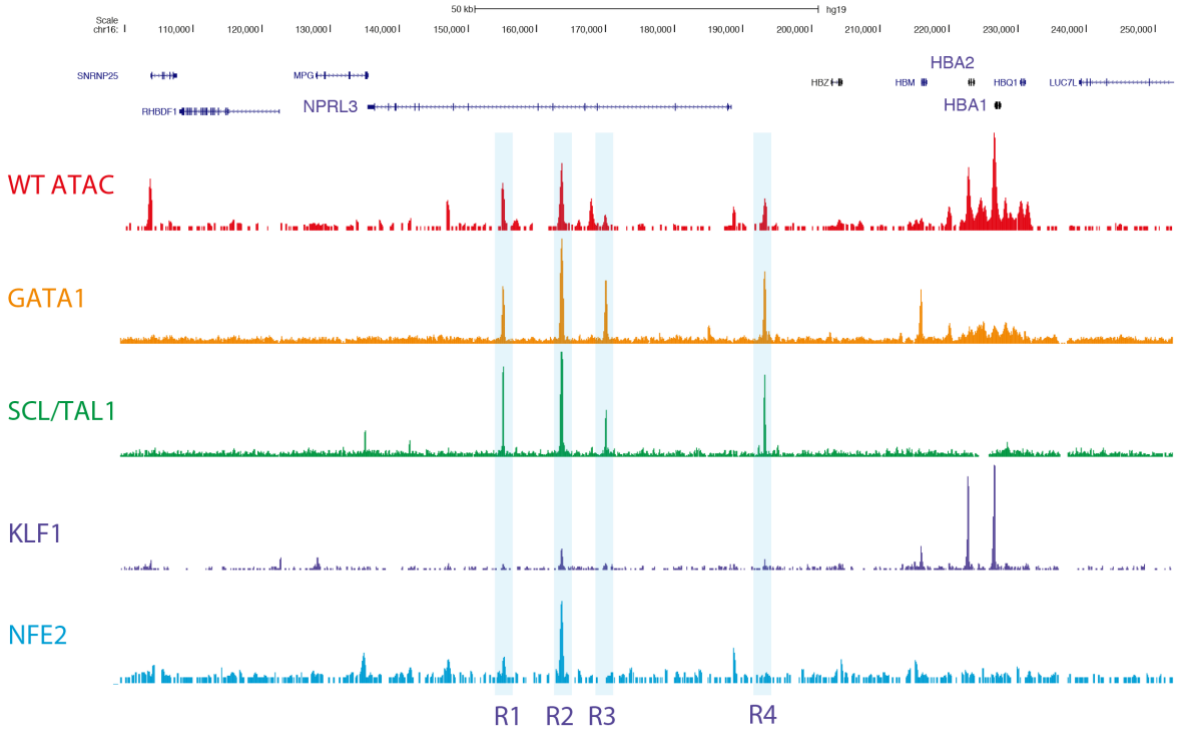


Figure 3.23 - Transcription factor binding at MCS-R2 *TFs at MCS-R2 in primary erythroid cells (with exception of KLF1 track which is derived from the HUDEP-2 cell line). These data are the same as used to derive the PWMs for motif analysis.*

Given this apparent discrepancy, publicly and privately available ChIP-seq datasets from primary human erythroid cells were gathered and analysed. The only exception was a ChIP-seq dataset for KLF1; the only publicly available data in primary cells showed a potentially non-specific pattern of binding at the α-globin locus, so data from the erythroid progenitor

cell line HUDEP-2 was used (Norton et al., 2017). Positive binding peaks in these cell types were analysed to generate PWMs for each DNA-binding transcription factor as outlined in chapter 2. These PWMs were then used to identify highly significant binding motifs within the MCS-R2 sequence.

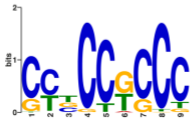
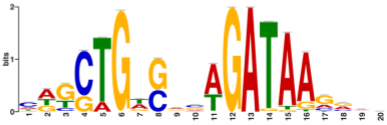
TF	PWM	Sequence in R2	P-value/Glam2 score
GATA1		AGATAAAC*	$1.38x10^{-3}$
		AGATAACT	$6.01x10^{-3}$
		TGATAGGT*	$8.93x10^{-3}$
KLF1		CCTCCACCC*	$1.05x10^{-4}$
		CTGCCTCCC	$2.75x10^{-5}$
		CCTCCACCC	$1.05x10^{-4}$
NFE2		GCTGAGTCA*	$0.3.31x10^{-6}$
SCL/TAL1		CTTGAGGGAGCAGATAACTGG*	8.31

Figure 3.24 - PWMs and identified putative TFBS in MCS-R2 *TF motifs derived from ChIP-seq datasets in erythroid cells using MEME. Binding sites for these motifs were identified using FIMO/GLAM tools, with p-value/Glam2 score included in the table. TFBSs that are regions of high evolutionary conservation are marked with an asterix.*

Each TF has several motifs within the enhancer, the overall composition of which is noted in Figure 3.24. Unsurprisingly, there are several GATA1 sites, in keeping with several canonical GATA1 core motifs being present in the sequence as well as one atypical motif (GATTA). Interestingly, only the 3' site of the two anteverted NF-E2 motifs present in the centre of the enhancer matches the derived motif. This is likely due to the 5' site having a single base pair difference within it. The KLF1 motif yielded three potential binding sites; the 5' site lies over an area previously shown to be highly conserved between mammalian

species (Hughes 2005). TFBSs that are regions of high evolutionary conservation are marked in **Figure 3.24** potentially indicating that these may be more functionally relevant than others.

3.5.2 Analysis of MCS-R2 using DNAse-seq footprinting

Whilst it may be known that particular TFs bind MCS-R2, positive binding motifs alone do not reveal whether they are actually bound by their cognate proteins. This is particularly the case when multiple motifs are present, as in MCS-R2. One method to look into this further is DNAse I hypersensitivity footprinting.

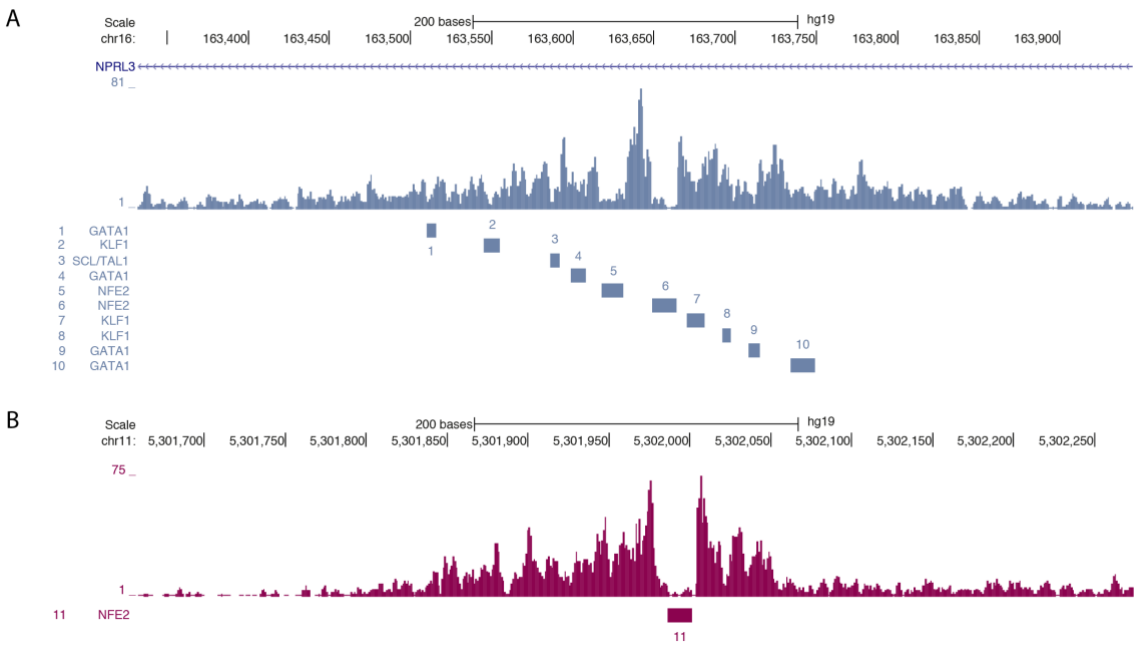


Figure 3.25 – DNAse I hypersensitivity footprints at the major globin enhancers *DNAseI hypersensitivity data from human primary erythroid cells. Footprints mark out regions potentially bound by TFs and not cut by DNAseI. A) The DNAse-seq peak at MCS-R2. Bars under the track indicate sites identified as TFBSs, either through previous work or motif analysis in 3.5.1. B) The DNAse-seq peak at LCR-2, the main β-globin enhancer, included here as a comparator. Likely NF-E2 binding site shown, other potential TFBSs not included. Data generated by Damien Downes, Maria Suciu (WIMM, unpublished).*

As DNase I cuts efficiently at sites of open chromatin, regions that aren't cut indicate sequences that are bound by proteins – likely TFs – that sterically inhibit DNAse I's endonuclease activity. High-throughput sequencing has allowed the development of various techniques to establish footprinting of greater resolution and accuracy in an *in-vivo* context (Levy and Noll, 1981, Crawford et al., 2006, Boyle et al., 2008). A high-resolution DNase-seq dataset derived from primary human CD34+ cells that had been differentiated down the erythroid lineage was examined (Figure 3.25). The deepest footprints are in the centre of the enhancer, corresponding to the greatest area of H3K27me1 and H3K27ac depletion, and highest areas of ATAC-seq signal. These striking footprints correspond exactly with the two central convergent NF-E2 motifs. A similar pattern is seen over the main *HBB* enhancer LCR-2, although this only has one NF-E2 site. Moving outwards from the central NF-E2 motifs there are several other footprints that correspond well to the TF motifs identified previously or those uncovered by motif analysis in 3.5.1. Whilst DNAse I footprints alone cannot definitively prove that a particular TFBS is vital to enhancer function, they do strongly indicate that certain sequences are bound by a protein *in-vivo* in that particular cell type, and so provide surrogate evidence that these presumed TFBSs contribute to enhancer activity.

3.5.3 *In-silico* screen of MCS-R2 using Sasquatch to predict effects of mutagenesis on DNAse I footprinting

Bioinformatic tools using NGS data have been developed to predict the effects of sequence variants in non-coding regions of the genome. Being able to reliably assess the impact of SNPs in regulatory elements is vital as they can help to decipher the impact of disease-

associated variants uncovered in large scale GWAS studies. Sasquatch is a computational tool that uses tissue- and developmental-stage specific DNAse I footprinting data to assess and visualise the impact of variants within a sequence of a given length, usually 5-7bp (known as a *k-mer* sequence) (Schwessinger et al., 2017). Sasquatch mutates every base within a *k-mer* to each of the three alternative bases and calculates a 'damage-score' based on their footprint profiles in other hypersensitive sites across the genome within which they are found. When these *k-mers* are tiled across an enhancer, the resulting damage scores predict which sites are highly sensitive to variation and therefore may have functional and regulatory activity. An intermediate damage score is defined as being between 0.5 and 1, and a severe damage score is a result above 1 (validation of the tool against SNPs known to have negative functional activity found these variants had scores between 1 and 2). The central 300bp of the enhancer sequence, which includes the majority of the DNase I hypersensitivity and ATAC-seq peaks, was analysed using Sasquatch and is shown in Figure 3.26.

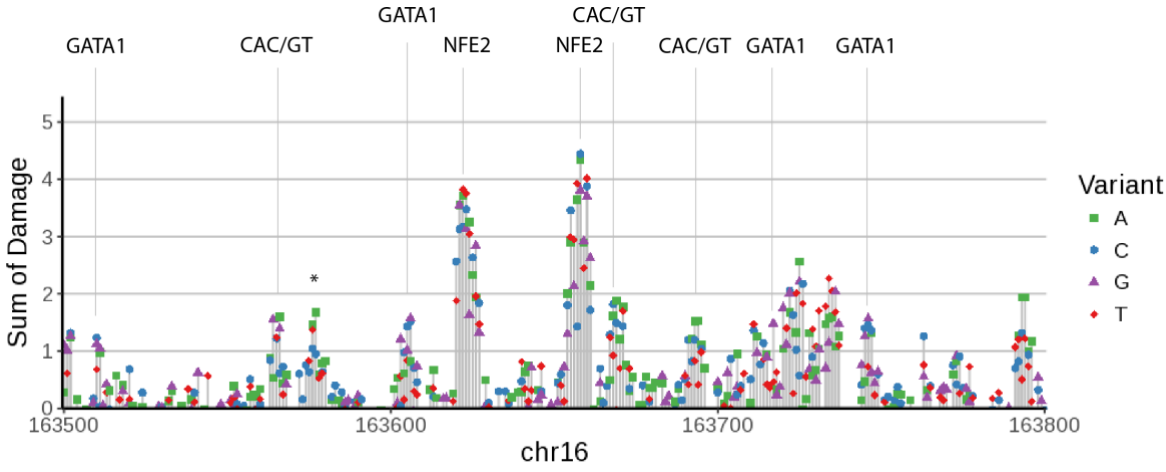


Figure 3.26 - Sasquatch analysis of MCS-R2 *Damage scores across the central 300bp of MCS-R2. Peaks with raised damage scores are annotated with identified TFBSs above the plot. The peak marked with an asterix is not associated with any known TFBSs and is not a region of evolutionary sequence conservation.*

In the centre of the enhancer sequence there were two striking peaks with high damage scores of over 3.5. Numerous other peaks were found with scores of over 1.5. Many of these coincided with motifs identified in earlier work on the MCS-R2 sequence, annotated above the peaks in Figure 3.26 (Jarman et al., 1991). These included the two major central damage score regions, which involve the two convergent NF-E2 binding sites, as well as four GATA1 motifs and three 'CAC/GT' sequences. These sites were also identified in the DNAse I hypersensitivity footprints mentioned earlier. There is a large cluster of damage scores between 1.5 and 2.5 in between the 3' GATA1 sites that coincide with a region noted in previous work to be hypersensitive, either due to local chromatin structure or non-DNA bound protein (Strauss et al., 1992, Zhang et al., 1993).

Overall, these results generally confirm the findings from previous *in-vitro* hypersensitivity work and DNase-seq as described in 3.5.2. The only peak not previously found in these assays is marked by the asterix in Figure 3.26. It does not contain any previously described motifs for erythroid TFs and is not a region of high evolutionary conservation (Hughes et al., 2005). Interestingly, some damage scores are higher than others, particularly the central NF-E2 sites. Whilst the damage score is a function of genome-wide DNase I footprinting, which is itself a surrogate marker for regulatory sequences, these high peaks may indicate that there is a hierarchy of TFBSs within MCS-R2, and NF-E2 sites are highly important elements in it.

3.5.4 *In-silico* screen using DeepHaem to predict effects of mutagenesis on chromatin accessibility

Sasquatch relies upon a single dataset in a given cell type and so may be affected by the quality and nature of the input data. Deep learning-based algorithmic frameworks that utilise convoluted neural networks take a different approach by integrating multiple groups of chromatin-profiling data and predicting the effect of sequence variants on chromatin characteristics *de-novo* (Zhou and Troyanskaya, 2015, Kelley et al., 2016). This approach was used to develop deepHaem, a deep convoluted neural network using 936 chromatin-accessibility and epigenetic datasets including ATAC-seq, DNAseI-seq and ChIP-seq to predict chromatin accessibility in haematological cells using the underlying DNA sequence alone (Downes et al., 2019, Schwessinger et al., 2020). An *in-silico* mutagenesis screen was performed on MCS-R2 using deepHaem to look for sequences that when mutated would be likely to abrogate chromatin accessibility over the enhancer as a whole, as opposed to modulation of individual footprints as with Sasquatch.

The deepHaem output in Figure 3.27 is shown alongside that of Sasquatch for comparison. The results are similar, but with a few key differences. As in the Sasquatch output, the central NF-E2 sites still have the highest damage score, and many of the sites previously found to have raised scores also do so here, as annotated in Figure 3.27. However, the GATA1 site at the extreme 5' end of the enhancer and the cluster of hypersensitivity between the 3' GATA1 sites both have a lower relative damage score using deepHaem. In

contrast, the second GATA1 site immediately 5' to the NF-E2 sites seems to be far more important in this analysis, with a damage score almost as high as the two NF-E2 sites. This is in keeping with its high evolutionary conservation as shown in Figure 3.7. Similarly, the CAC/GT site 3' to the NF-E2 sites and the GATA1 site at the extreme 3' end of the enhancer also have relatively higher damage scores. Finally, the peak marked with an asterix in Figure 3.26 is not detected in the deepHaem output.

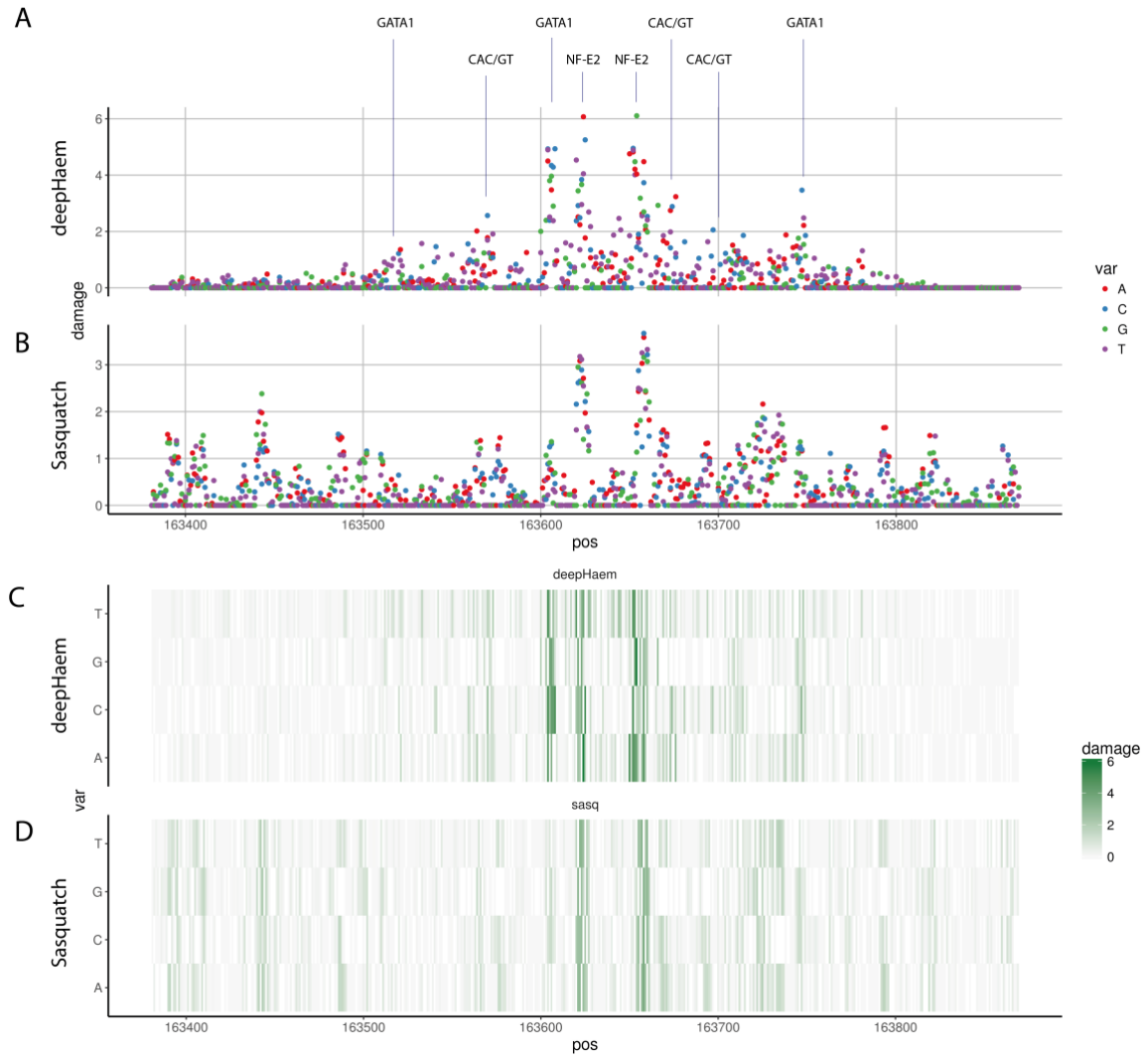


Figure 3.27 - deepHaem analysis of MCS-R2 *deepHaem results, shown alongside Sasquatch output. A and B visualise the damage scores on the y-axis and C and D show the results as a heat-map. A and C are the deepHaem outputs and B and D are the sasquatch results. TFBS as annotated in* ***Figure 3.26*** *are shown in A.*

The data shown above generated by both Sasquatch and deepHaem summarise variants resulting in a positive damage score, corresponding to an *in-silico* reduction in footprint depth and chromatin accessibility respectively. The corollary of this approach is that some variants might create new functional sites, which would be represented by a negative damage score. These variants using both computational approaches are summarised in Figure 3.28. While the vast majority of variants either have a neutral or positive damage score in both approaches, there are relatively far more negative scores in the Sasquatch analysis compared to deepHaem. Therefore, while base-pair variants potentially result in novel footprints, they are not predicted to result in larger enhancer-scale chromatin accessibility changes.

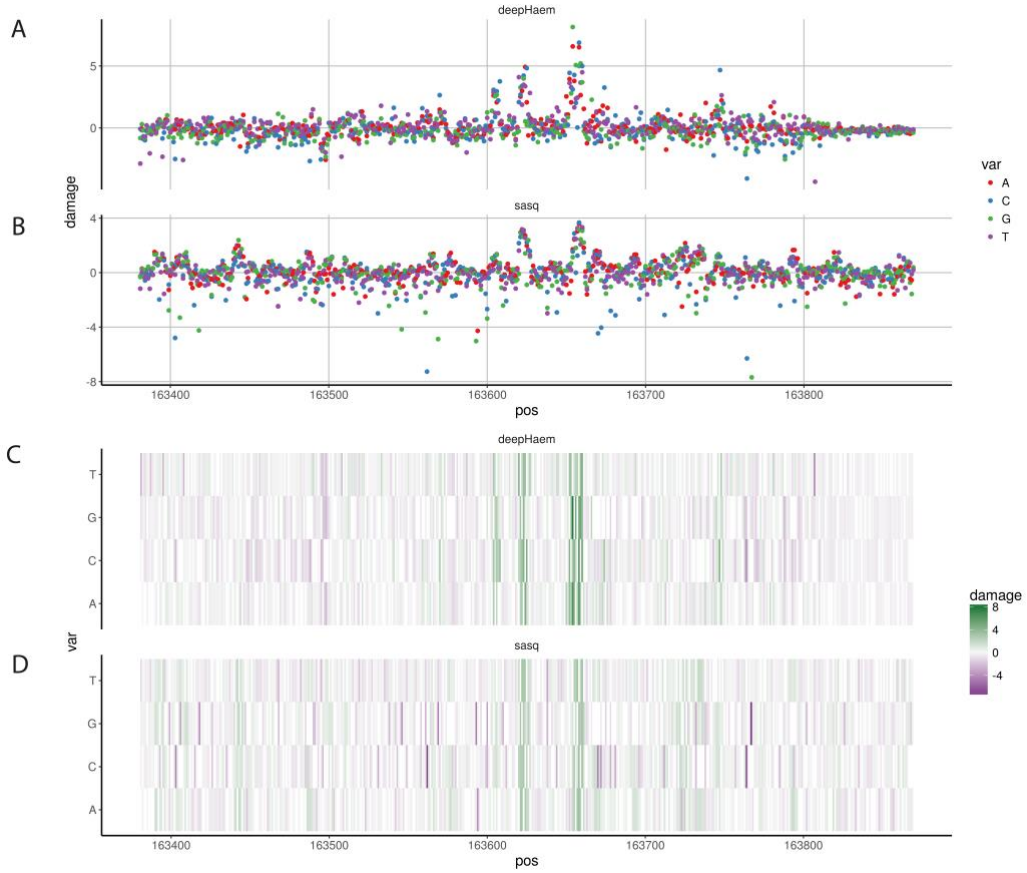


Figure 3.28 - Positive and negative in-silico damage scores *deepHaem and sasquatch results laid out as in Figure 3.27 (A and C are the deepHaem results and B and D are the sasquatch results). Positive*

damage scores indicating disruption of a footprint and negative damage scores indicating creation of a footprint are shown across the y-axis (A and B) and as a heat-map (C and D).

3.6 Discussion 2 - Analysis of the MCS-R2 Sequence

The identification and characterisation of enhancers has long been an area of scientific inquiry. Multiple assays can be used to identify enhancers, and different approaches have been developed to interrogate their underlying sequences. Previous work looking at MCS-R2 used the following approaches to identify important sequences: 1) DNAse I protection assays 2) Mobility shift assays 3) targeted mutagenesis with artificial reporter constructs 4) assessment of evolutionary conservation at base-pair level. These approaches are informative, and assessment of evolutionary conservation between mammalian species remains a powerful tool to examine regulatory elements. However, the approaches used have several important limitations. Due to the technology available at the time, several studies were performed in artificial contexts; cell lines used were either immortalised erythroleukaemia (HEL, K562) or interspecific murine-human lines (J3-8B) (Jarman et al., 1991, Strauss et al., 1992, Zhang et al., 1993, Zhang et al., 1995, Wen et al., 2000). Although the underlying sequence of the enhancer was the same in all these studies, the *trans*-acting environment was very different, either owing to species or cell-type specific variations. The targeted mutagenesis experiments also involved reporter constructs consisting of the HS-40 core element, a ζ-globin or α-globin promoter and a GH reporter gene, again inserted in MEL or K562 cell lines (Zhang et al., 1993, Zhang et al., 1995, Wen et al., 2000). Even if these reporter constructs were used in primary erythroid cells they would not have been able to take into account the powerful role local chromatin context plays in gene expression, as discussed in chapter 1.

This study sought to address these shortfalls by taking advantage of advances in NGS assays and computational tools to use these as orthogonal approaches to interrogate MCS-R2's sequence at high resolution. There are several sites within MCS-R2 that are likely to bind to a transcription factor. Previous work has shown that multiple erythroid TFs bind this region and the locus as a whole, but the ChIP-seq assays used are unable to tell us how many TFs actually bind MCS-R2 along its length (De Gobbi et al., 2007). The MCS-R2 screen here shows that there are likely to be four GATA1, one Scl/Tal1, two NF-E2 and two KLF1 binding sites within MCS-R2. These results are entirely expected given the importance of α-globin transcription to erythroid cell function, and fits with the concept that enhancer activation and function requires the binding of multiple specific transcription factors to their underlying sequence (Long et al., 2016). In one study, approximately half of the experimentally validated human enhancers that were examined were classified as homotypic clusters of TFBSs, within which there were multiple binding sites for the same TF as in MCS-R2 (Gotea et al., 2010). The presence of multiple binding sites for the same TF is thought to promote high-affinity cooperative binding, aiding in local diffusion of a TF within a regulatory region and allowing for functional redundancy for a particular TF in case of an inactivating mutation at one of the sites (Weingarten-Gabbay and Segal, 2014). The binding of multiple, different TFs is also a characteristic feature of enhancers; TFs are known to co-operate, directly or indirectly, to facilitate nucleosome eviction and downstream activator recruitment; it has been proposed as a general rule that increased numbers of TF binding sites are positively correlated with increased target gene expression (Long et al., 2016).

Perhaps unsurprisingly the techniques used previously to characterise the MCS-R2 enhancer, and in this study, significantly overlap, which provides some degree of validation for these approaches. Each approach used asks a slightly different question: motif analysis aims to identify a consensus motif for TFs known to bind to MCS-R2 in primary erythroid tissue and identify the likely binding site within the enhancer. DNAse I digital footprinting on the other hand can look for regions of chromatin that are potentially bound by TFs and co-factors *in-vivo*, corroborating motif analysis results. As can be noted in Figure 3.25, these results intersect well. *In-silico* mutagenesis tools can follow on from these descriptive methods by predicting molecular outcomes we might encounter when the underlying sequence, the primary determinant of TF binding, is perturbed. Sasquatch generates base-pair level outcomes by predicting what will happen to individual 7-bp footprints when single bases are mutated. The two NF-E2 sites in the centre of the enhancer are predicted to have the highest damage score compared with other sites, though several GATA1 sites and a KLF1 site are also represented in positive scores. Taken independently, Sasquatch predicts the consequence of changing a TFBS for a single TF in a particular cell type. Thus, it tells us that the NF-E2 motif in human erythroid cells is fairly stringent, which is notable as the two sites do vary by 1bp. Changing core bases in the GATA1 motif on the other hand does not have the same pronounced effect on its footprint. This is interesting as it has a highly stereotypical motif when looking at the PWM. It may be that where GATA1 is located in open chromatin in erythroid tissues it is present as part of homotypic clusters of TFs as discussed above. When a single base is mutated GATA1 may still be present at the enhancer and generate a DNAse I footprint due to its local diffusion, facilitated by a relatively higher concentration on account of it being bound by the other GATA1 sites. DeepHaem also performs an *in-silico* mutagenesis but utilises convoluted neural networks to synthesise large amounts of

chromatin accessibility data. In this way DeepHaem reports on the effects of single base changes on the accessibility of the entire (300bp) enhancer, as opposed to each individual TF footprint. This is a fundamentally different approach, as it moves beyond descriptive data to predicting functional changes. DeepHaem predicts that a small perturbation in the NF-E2 sites leads to significant loss of accessibility, in keeping with the findings using other assays. However, the GATA1 site immediately 5' to them also has a similar damage score, which is quite distinct compared to the Sasquatch results. This is consistent with the evolutionary conservation data, which shows that not only the motif sequence but importantly the distance between them is highly conserved, which is not the case for the other TFBS. Thus, while one study has shown that GATA1 and NF-E2 do not have a fixed distance in K562 cells, it may be that, in this case, their proximity has some functional relevance (Guo et al., 2012). This could be due to these TFs forming a complex with one another through direct binding. Alternatively they may indirectly cooperate and attract secondary transcriptional co-factors. These results indicate that there may be a functional hierarchy to MCS-R2. Whilst there is evidence to point to functional sites existing elsewhere in MCS-R2, the central NF-E2 and GATA1 sites may be significantly more important for enhancer accessibility, which is of course a key characteristic of active, tissue-specific regulatory elements.

The profiling of MCS-R2 conducted in this study may give us some insight into which enhancer paradigm best describes it. Different models have been proposed to describe how internal enhancer grammar operates. On one end of the spectrum, the enhanceosome model posits that motif grammar including sequence, spacing and orientation is highly stringent, and any deviation from the internal structure of the enhancer results in a large reduction in enhancer activity. The hallmark example of this is the human interferon-β

enhancer (Thanos and Maniatis, 1995). On the opposite end of the spectrum, the flexible billboard model is far more permissive, holding that while enhancers do consist of heterogenous clustering of TFBS, robust enhancer activity is resistant to alterations in TFBS orientation, order and spacing (Levo and Segal, 2014). Similar to the billboard model is the 'TF collective' model. TFs bind their TFBSs as in the billboard model, but are also recruited through protein-protein interactions (Levo and Segal, 2014, Panigrahi and O'Malley, 2021). Whilst in the billboard model enhancer output is additive in terms of how many TFs are bound to underlying DNA sequences, in the collective model certain TFs may need to bind at specific sites to allow higher-order protein-protein interactions to occur. These interactions then allow the enhancer to function fully. These and previous results suggest that MCS-R2 falls somewhere between the two flexible enhancer models rather than behaving as an enhanceosome. MCS-R2 has been linked to the α-globin in species as far back as opossum, indicating that its presence is important for α-globin expression (Hughes et al., 2005). Whilst the TFBSs are highly conserved, in general the spacing between them is not, in contrast to the enhanceosome model. Looking at the results from Sasquatch and DeepHaem we can see that there is a gradation of damage scores when bases across the enhancer are altered, which contrasts with the more binary behaviour of enhanceosomes. However, clearly there is a predicted hierarchy of TFBS in the enhancer; evolutionary conservation shows that the central region of MCS-R2 is conserved both in sequence but also in spacing between the TFBSs. DNAse I footprints are deeper for the central TFBSs compared the peripherals sites, and deepHaem results suggest that the TFBSs in this region when modulated are far more likely to results in a loss of accessibility than other sites, which is at variance with the billboard model but consistent with a TF collective model. Therefore MCS-R2's enhancer behaviour likely locates it somewhere between the TF collective and the

billboard: some regions may be constrained in sequence and spacing to allow the formation of complexes that have steric requirements, like the erythroid pentameric complex, but also flexible enough to allow for other non-complex bound proteins to bind, like NF-E2.

Characterising the determinants of internal enhancer design and their effect on gene expression is a broad and open question at this time. Whilst there are some broad features that are shared by many enhancer architectures, predicting effects on gene expression based on design principles alone is likely to give us limited answers (Levo and Segal, 2014). Enhancers often behave in a highly tissue and developmental-stage specific manner, and so *in-vivo* manipulation is required to fully explore MCS-R2's regulatory landscape. This is the main approach used in the next chapter; the TFBSs identified here are manipulated in primary human erythroid cells, with variations in gene expression used to inform a potential therapeutic strategy to ameliorate β-thalassaemia.

4. Development of a therapeutic strategy for ameliorating β-thalassaemia by editing MCS-R2 to reduce α-globin expression

4.1 Introduction

4.1.1 Viral gene therapy

The development of targeted gene therapies for patients with genetic disorders has been a goal for medical researchers for nearly five decades (Friedmann and Roblin, 1972). Haematological diseases, particularly the haemoglobinopathies, have been at the forefront of this endeavour. The major issue has been how to insert genetic material into stem cells to correct the mutation directly (gene correction) or to add healthy genes alongside mutated ones (gene addition). As early as 1961 this was shown to be possible, albeit inefficient, by adding free DNA to cultured cells from bone marrow taken from a patient with sickle cell disease (SCD) (Kraus, 1961). Gene therapy took a huge step forward in the early 1980s when it was shown that genetic material could be more efficiently delivered into HSPCs by replication-deficient retroviruses (Mann et al., 1983, Williams et al., 1984). These γ-retroviruses were used in the first successful gene-therapy cure, in a 4-year old girl with severe combined immunodeficiency (SCID) (Blaese et al., 1993, Blaese et al., 1995). The first individual with a haemoglobinopathy to be successfully cured with a viral gene therapy had HbE/β-thalassaemia. The patient's cells were transduced with a lentivirus, and the patient became transfusion independent at 33 months (Cavazzana-Calvo et al., 2010). Since then, ten phase I/II trials have been conducted using lentiviral vectors to treat β-

thalassaemia and SCD using a gene addition or shRNA knockdown of BCL11a approach (Ferrari et al., 2021). Several of these approaches have shown promising results (Ribeil et al., 2017, Thompson et al., 2018, Marktel et al., 2019).

Despite these successes, viral gene therapy has been beset by safety concerns. 5 of 20 patients with X-SCID and 6 of 10 with Wiskott-Aldrich Syndrome (WAS) developed T-cell acute lymphoblastic leukaemia, with a further patient in the WAS trial developing acute myeloid leukaemia (AML) following gene therapy with γ-retroviruses (Hacein-Bey-Abina et al., 2003, Hacein-Bey-Abina et al., 2008, Braun et al., 2014). Furthermore, 3 patients with chronic granulomatous disease developed myelodysplastic syndrome (MDS) following gene therapy, one of whom went on to die of AML (Stein et al., 2010). These transformations to leukaemia were due to insertional mutagenesis at transcriptionally active oncogenes. Retroviruses insert at thousands of sites in host cells. Importantly, this occurs in a non-random manner; some oncogenic mutations result at recurrent sites of integration. Worryingly, these effects were not observed in pre-clinical murine and large animal studies with long-term follow up (Kiem et al., 2004, Nienhuis et al., 2006). Lentiviral vectors were developed to circumvent these problems. However, dominant clonal haematopoeisis linked to activation of a gene - *HMGA2* - has been reported in an individual with HbE/β-thalassaemia treated with a lentivirus (Cavazzana-Calvo et al., 2010). Recently a trial for lentiviral treatment of SCD was halted after two patients developed AML and MDS respectively. Whilst Bluebird bio - the company conducting the trial - felt these outcomes were very unlikely to be caused by the vector, they did detect cargo insertion at the *VAMP4* gene in leukaemic blast cells (2021). Given these findings, concerns about insertional mutagenesis associated with viral vectors remain.

4.1.2 Gene editing

Gene editing is an alternative technique that relies on guided endonucleases to effect a double-stranded break (DSB) at a target locus. In the main, the cell is able to repair this break in one of three ways. The most efficient pathway is non-homologous end joining (NHEJ). Following a DSB, the cell either removes or inserts several bases leaving the repair with a 'scar' (Adli, 2018). The editing outcomes can be partially predicted depending on cell type and chromatin context, but in general whilst NHEJ is an attractive option for disrupting sequences, it is not ideal for repairing mutations precisely (Shen et al., 2018, Schep et al., 2021). In contrast homology directed repair (HDR) can resolve the DSB and change the original sequence using a donor template. Whilst this can potentially repair genetic mutations, it is not as efficient as NHEJ (Komor et al., 2017). Furthermore, HDR preferentially occurs in G2-S phase in actively cycling cells, a state that LT-HSCs are unlikely to be in (Heyer et al., 2010). The third method of repair is microhomology mediated end joining (MMEJ). If short microhomologies exist up and downstream of a DSB site, the two regions may anneal to one another, with removal of the intervening sequence (McVey and Lee, 2008).

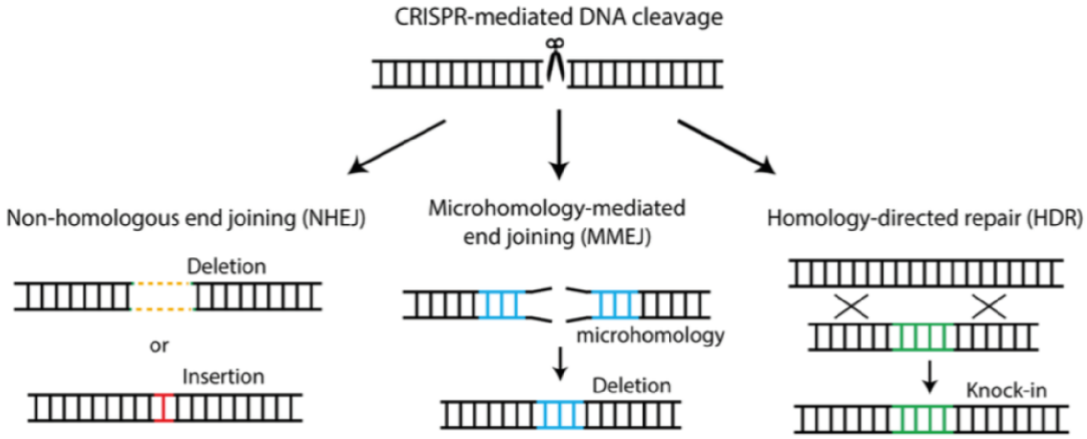


Figure 4.1 - Cellular repair processes after a DSB *Adapted from (Jang et al., 2020)*

Various endonucleases have been developed. Meganucleases are large enzymes with a long DNA recognition site that were used early on to create DSBs (Rouet et al., 1994). Due to their extremely stringent sequence specificities they did not have broad applicability. To address this zinc finger nucleases (ZFNs) and transcription activator like effector nucleases (TALENs) were developed. ZFNs are homodimers, with both components of the dimer pairing an array of DNA binding C2H2 zinc fingers to a FokI endonuclease (Kim and Kim, 2014). Upon dimerisation, a specific DNA site can be cleaved. Despite the relative difficulty of designing and constructing a ZFN, low on-target editing efficiency and a paucity of target sites in the genome (approximately 1 per 100bp), there has been one successful clinical trial using ZFNs to treat individuals infected with HIV (Tebas et al., 2014). TALENs are similar to ZFNs in that they are also homodimers utilising FokI, but instead are composed of 33-35 amino acid arrays which are complementary to a single DNA base. TALENs can therefore theoretically target almost any DNA sequence in the genome. Whilst they have high on-target efficiency they are also difficult to clone and synthesise due to their repetitive sequence, which again has prevented widespread scientific and therapeutic uptake (Komor et al., 2017).

4.1.3 Gene editing using CRISPR/Cas systems

Gene editing took a significant step forward with the adaptation of the CRISPR/Cas9 system for genome engineering. The clustered regularly interspersed short palindromic repeat (CRISPR)/CRISPR associated proteins (Cas) system is part of the anti-phage adaptive immune response in bacteria and archaea. First described in 1993, it was found to utilise *cas* endonucleases to cut and neutralise previously encountered phage DNA (Mojica et al.,

1993, Barrangou et al., 2007). Short RNAs complementary to the phage sequence are encoded in a series of repeats (the CRISPR component) that are found within the bacterial genome. CRISPR RNA (crRNA) forms the RNA complementary to the phage sequence which can be as short as 20bp, and is known as the protospacer. These sequences are taken from the phage genome by the adaptive module of the system, which encodes the enzymes Cas1 and Cas2, and integrates them into the bacterial genome for future use. Another component of the bacterial system, trans-activating RNA (tracrRNA), is transcribed and hybridises to pre-crRNA to form an RNA duplex which then allows mature crRNA to perform its targeting function (Deltcheva et al., 2011). In 2012, two groups used recombinant Cas9 and *in-vitro* transcribed tracrRNA and crRNA to show that the CRISPR/Cas9 system could be reprogrammed to cut DNA, and that the two RNA molecules could be combined artificially to create a single guide RNA (sgRNA) (Jinek et al., 2012, Gasiunas et al., 2012). This work earned Jennifer Doudna and Emmanuelle Charpentier the Nobel prize for chemistry in 2020. The potential bioengineering and therapeutic applications were noted at the time, and the following year the system was adapted for use in mammalian cells (Cong et al., 2013). The advantages of CRISPR/Cas9 over ZFNs and TALENs are its high on target efficiency, relatively short targeting sequence, simple design protocol, low synthesis cost and large number of targeting sites within the genome, approximately 1 every 8bp.

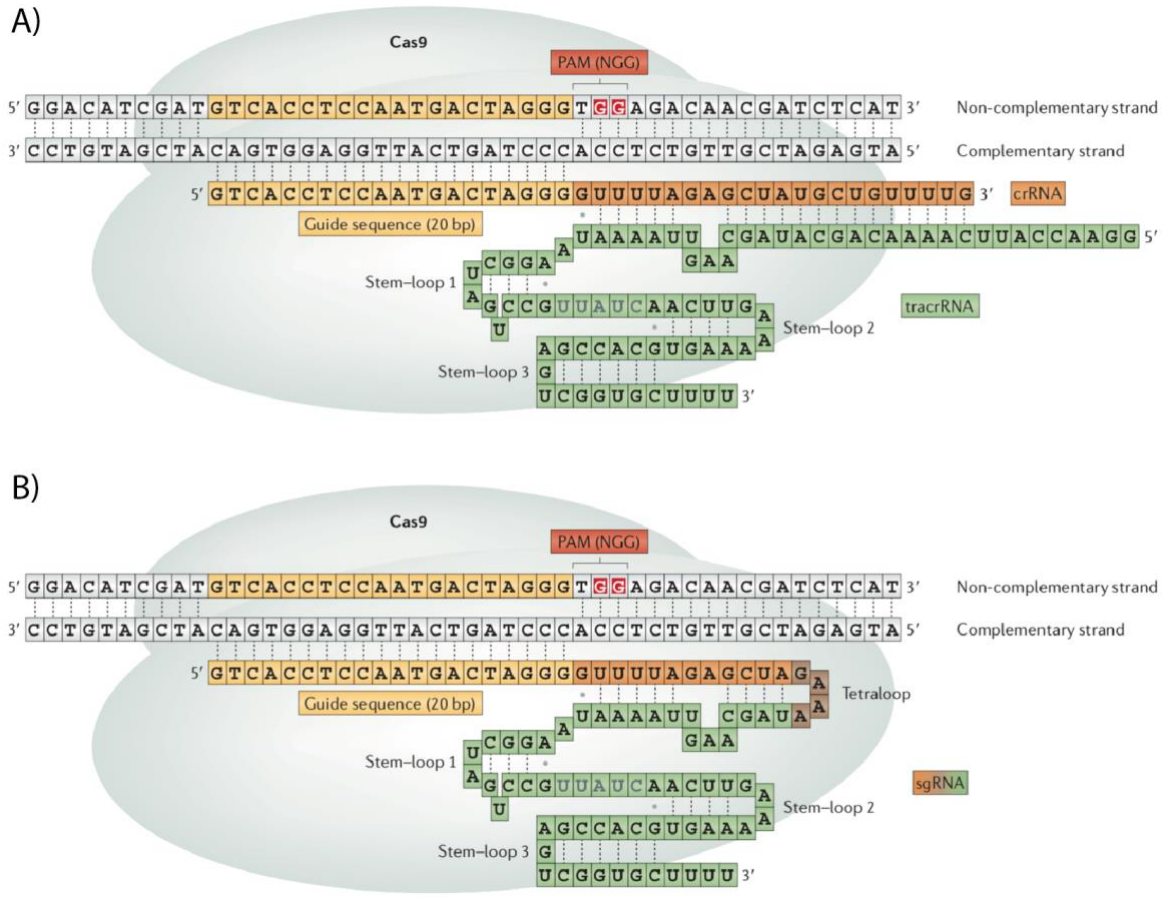


Figure 4.2 - The CRISPR/Cas9 system A schematic of Cas9 loaded with a DNA target site and a fully formed R loop A) In this conformation the crRNA and tracrRNA are separate molecules and are hybridised to one another. B) This depiction shows the artificial sgRNA. Adapted from (Kim and Kim, 2014)

4.1.4 Gene editing using CRISPR/Cas9: mechanism of action

There are two broad classes of CRISPR/Cas systems; type I and II. Type I systems have multi-subunit effector nucleases, while type II have one. Streptococcus Pyogenes Cas9 (SpCas9) is the most well-known, and is a type II system. Cas9 is a bi-lobed protein that is able to bind the sgRNA and DNA (Kim and Kim, 2014). There are two grooves that interact with unwound single stranded DNA (ssDNA), RuvC (present in 2 of 3 type II effectors) and HNH (present in all type II effectors). Before Cas9 can create a DSB at the target site, two stages of 'licensing'

have to occur (Bolukbasi et al., 2016). The first is through the recognition of a 3bp protospacer associated motif (PAM) immediately 3' of the 20bp protospacer. This sequence is NGG for SpCas9. If an appropriate PAM is present Cas9 unwinds the DNA double helix and an R loop forms between the sgRNA and the complementary ssDNA strand. The second step of licensing involves extension of the R-loop if the sgRNA is fully or nearly fully complementary to the ssDNA. Following this Cas9 creates a blunt ended DSB 3bp 5' to the PAM. The HNH groove cleaves the DNA strand complementary to the sgRNA and RuvC cleaves the non-sgRNA-binding strand.

4.1.5 Gene editing - Delivery considerations

Editing machinery can be delivered into the cell in one of three broad ways: electroporation, viral and non-viral based carrier techniques. Electroporation is a membrane-disruption based approach first used in 1982 (Neumann et al., 1982). Cells and the material to be electroporated are held in suspension between two electrodes. Electrical pulses are passed to disrupt the membrane, with variations in voltage, duration, waveform and frequency creating different patterns of pores through which material can be passed. Whilst there is an element of cytotoxicity, partly depending on the biomolecules being transferred, this has been a popular technique as it is rapid, applicable to a diverse set of primary cells, has a defined period of transfection, is straightforward to perform and is not immunogenic (Stewart et al., 2016).

Viral and non-viral carrier-based techniques involve packaging a cargo and delivering it into the cell, potentially into a specific intra-cellular compartment. Viral-vector transduction is

most commonly performed with lentiviruses or adeno-associated viruses (AAV). The viral-vector is usually pseudotyped with glycoproteins that recognise host cell-surface antigens, which specify the tropism of the vector. Upon binding to the surface ligand, viral transmembrane proteins change in conformation, allowing fusion to the cellular surface membrane and viral entry. Genes for editing machinery in the vector cargo are then transcribed (AAV), or integrated into the host genome (lentivirus). Viral approaches can be highly efficient as they use the cell's own processes to deliver cargo, and there is a wealth of clinical experience available from over a decade of clinical trials in different cell types. However, viral-vectors have a limited cargo, and so multiple vectors may be required (Naldini, 2011). Certain viral-vectors might elicit an immune response, which may be experienced at the time of delivery or on subsequent exposures. Aside from the concerns associated with integrating viruses, discussed in 4.1.1, viruses in general require safety measures that result in their manufacture being time-consuming and expensive.

Non-viral carrier-based techniques involve the use of lipoparticles, polymers or other non-organic nanoparticles to enclose editing machinery and deliver it to the cell through fusion with the cell membrane (Yin et al., 2014). These potentially circumvent the genotoxicity of viruses, and could lend themselves to *in-vivo* therapy through either ligand-based targeting or the formulation of the nanoparticles themselves (Cheng et al., 2020). This is a nascent field of research, with cell-dependent toxicities and kinetics of delivery yet to be fully explored. However, recent successful non-human primate studies have been performed targeting hepatic cells, indicating this could be a promising avenue for future therapeutic applications (Musunuru et al., 2021).

CRISPR/Cas9 components be delivered into the cell in the form of DNA, RNA or ribonucleoproteins (RNPs). Nucleic acids can be delivered into the cell using viruses or as free molecules in the form of plasmids or mRNA. Plasmids have been used for decades in a wide variety of cell types *in-vitro*. Construction is relatively straightforward and cheap, but in order to deliver Cas9 and sgRNA plasmids need to be moderately large and so cell based toxicities are frequently encountered, partly due to the host cell immune response to bacterial DNA (Hornung and Latz, 2010). Plasmids are also known to persist in cells for several days post-delivery, meaning that there is the potential for continuous production of Cas9 and off-target (OT) events (Liang et al., 2015). mRNA on the other hand is more rapidly degraded within the cell thus reducing activity time of nucleases and OT effects. This can be a double-edged sword however, as the production and delivery of stable long mRNAs is difficult, reducing overall on-target efficiency (Kowalski et al., 2019). RNPs are delivered as mature Cas9-sgRNA complexes with immediate nuclease activity at high efficiency, but are also degraded by 24-36h, leading to a favourable OT profile (Kim et al., 2014, Liang et al., 2015). Due to the rapid degradation, they are less likely to elicit MHC class I responses against the Cas9 protein itself in whole-body systems (Mout et al., 2017). Clinical trials using RNPs for haemoglobinopathy patients are currently in progress (Frangoul et al., 2021).

4.1.6 Gene editing - Adverse off-target editing outcomes

Whilst genome editing offers an alternative to viral integration-based methods, there are still safety concerns surrounding adverse editing outcomes, which we are learning more about as experience in the field grows. These can be split into adverse on-target and off-target (OT) outcomes.

OT events are noted with all gene-editing nucleases to varying degrees, including CRISPR/Cas9 (Kim and Kim, 2014). OT editing generally occurs at sites that bear some sequence homology to the 20bp protospacer specific to the target site. It is difficult to predict where OT events will occur across the genome, especially as occupancy of Cas9 on DNA as assessed by ChIP-seq does not correlate with indel formation at predicted OT-sites (Kuscu et al., 2014). In two early studies, editing events were noted at loci which had as many as 5 mismatches in the sgRNA sequence (Hsu et al., 2013, Fu et al., 2013). Mismatches in some sgRNA bases are tolerated more than others; the PAM sequence is very stringent, and among profiled OT sites the proximal 10bp 5' to the PAM were conserved (the so-called 'seed sequence'), indicating mismatches distal to the PAM were tolerated (Kuscu et al., 2014). Evidence suggests that Cas9 editing occurs in a more stochastic manner across the genome compared to lentiviral integration, although there is likely to be a preference for binding at euchromatin (Naldini, 2011, Chechik et al., 2020).

Initial bioinformatic efforts to predict OT sites relied on aligning on-target sgRNA sequences across the genome with widely available alignment tools such as Bowtie (Zhang et al., 2020). Unfortunately, these methods have a significant false negative and positive rate, one study finding that the number of validated OTs was ten-fold less than predicted (Tsai et al., 2015, Cameron et al., 2017). Tools have been developed to take into account the observation that mismatches at different protospacer positions are tolerated differently, but these methods are still limited by relying on *in-silico* alignment (Zhang et al., 2020). Recently several machine-learning based algorithms have been developed to predict OTs, relying on large datasets to train the prediction-model (Wang et al., 2020). These models take into account

the rules mentioned previously as well as chromatin accessibility, DNA geometry and energetics, with good results.

Whilst these *in-silico* methods are easy to use, the gold standard method to assess OTs for a particular sgRNA and cell-type remains the use of an *in-vitro* assay. Several have been developed, the most prominent being CIRCLE-seq, SITE-seq, Digenome-Seq and Change-Seq. These all depend on Cas9 cutting purified genomic DNA. The identification of cut sites is carried out in different ways. Digenome-Seq uses whole genome sequencing to identify identical 5' DNA fragments (Kim et al., 2015). This method requires high sequencing depth, which can impact its sensitivity. Other methods enrich reads that involve OT edits. SITE-seq involves insertion of biotinylated adapters at cleavage fragments, followed by sequencing (Cameron et al., 2017). Whilst this improved sensitivity, the majority of reads still do not contain any detectable OT events. To remove these non-specific reads CIRCLE-seq and more recently CHANGE-Seq involve the circularisation of DNA after fragmentation, followed by *in-vitro* digestion and linearisation of DNA if OT editing has occurred (Tsai et al., 2017, Lazzarotto et al., 2020).

These assays are highly sensitive, but represent the OT editing landscape in chromatin-free systems. Cell-based assays have been developed to take into account editing in cell-type dependent chromatin contexts. GUIDE-seq involves the cell repairing DSBs with insertion of a double stranded oligodeoxynucleotide (dsODN), which can be detected using NGS (Tsai et al., 2015). DISCOVER-seq on the other hand performs ChIP-seq for the MRE11 subunit of the MRN complex, which is tightly distributed at Cas9 cut-sites, and subsequent filtering of the data to look for cleavage events (Wienert et al., 2019). A recent study comparing cell

based (GUIDE-seq) and non-cell based (SITE-seq, CIRCLE-seq) assays found that they all performed similarly in detecting true validated OT-events (Chaudhari et al., 2020).

4.1.7 Gene editing - Adverse on- and off-target editing outcomes

OT events are concerning as they may result in dysregulation of nearby genes, but increasing evidence has shown that the formation of DSBs per se, whether at the on- or off-target site, can result in major cytotoxicity. When a DSB is formed the majority of editing outcomes will be mediated by NHEJ. Most of these outcomes are deletions under 50bp (van Overbeek et al., 2016). However, large deletions can be generated that extend to several megabases, and even involve chromosomal losses in human embryos (Kosicki et al., 2018, Cullot et al., 2019, Chakrabarti et al., 2019, Zuccaro et al., 2020, Turchiano et al., 2021, Boutin et al., 2021). Beyond simple deletions, more complex local and large-scale rearrangements have been observed, ranging from small inversions to chromothripsis (Mettananda et al., 2017a, Kosicki et al., 2018, Leibowitz et al., 2021). Chromosomal translocations between the on-target and off-target sites are also possible, particularly in applications involving multiplex CRISPR/Cas9 targeting, and models recapitulating mixed lineage leukaemia have been generated using this technique (Reimer et al., 2017, Jeong et al., 2019).

DSB formation following Cas9 editing has been shown to activate the cellular p53 DNA damage response, with potential cell cycle and growth arrest (Haapaniemi et al., 2018, Ihry et al., 2018). In a large therapeutic graft that is efficiently edited, cells with an intact p53 response could be at a selective disadvantage compared to any cells with mutations in the p53 system which could predominate over time. Even in cells with WT p53, DSB formation

and the p53 response can lead to clonogenic and repopulation defects in HSCs, although this is dependent on the burden of DSBs formed and whether AAVs are used for transduction (Schiroli et al., 2019).

4.1.8 Gene editing for the β-thalassaemias - previous approaches

The β-haemoglobinopathies are attractive candidates for the use of CRISPR/Cas9 gene editing. Decades of research has revealed how the globin loci are structured and regulated, and thus illustrated the ways in which genome engineering can manipulate globin gene expression. Two major strategies have been developed, both to upregulate HbF levels:

1. BCL11a repression - BCL11a is a developmental-stage specific repressor of γ-globin, and can itself be repressed by editing its erythroid-specific enhancer (Sankaran et al., 2008, Canver et al., 2015, Wu et al., 2019). This has been used with success in phase I/II studies in patients with β-thalassaemia and SCD (Frangoul et al., 2021).
2. Direct γ-globin promoter mutation - BCL11a and other repressors are known to bind to the γ-globin promoters (Liu et al., 2018a, Martyn et al., 2018). Mutations of these binding sites are known to cause hereditary persistence of fetal haemoglobin. These sites have been targeted *in-vitro* and in non-human primates to induce HbF (Traxler et al., 2016, Humbert et al., 2019, Weber et al., 2020).

4.1.9 Gene editing for the β-thalassaemias - Reduction of α-globin expression

Whilst these pre-clinical and early clinical studies have shown promising results, long term efficacy and safety data is lacking, and there is a potential need for alternative strategies to

ameliorate β-thalassaemia. One approach that seeks to functionally recapitulate a naturally occurring genotype is through the reduction of α-globin expression. β-thalassaemia is caused by an α-globin chain excess within erythroblasts. Naturally occurring, co-inherited α-globin mutations are a powerful genetic modifier because the balance in intracellular globin chains is restored due to the reduction in α-globin expression (Taher et al., 2021). This ameliorating effect is especially pronounced in individuals with HbE/β-thalassaemia, who comprise half of the world's severe cases of β-thalassaemia (Olivieri et al., 2010). Numerous studies summarised earlier have shown that these individuals have a mild haematological phenotype, and importantly are otherwise phenotypically normal making this an attractive treatment strategy.

How many α-globin gene deletions result in the best improvement in disease phenotype? One to three α-globin genes could be deleted (four results in minimal α-globin transcription and the Hb Bart's hydrops fetalis syndrome). We can look to patient cohort studies to see how many deletions were associated with the best outcomes. Whilst there is some variability due to the β-thalassaemia mutation and other co-inherited modifiers of disease, one or two gene deletions seem to have the most positive effect (Sripichai et al., 2008b, Sharma et al., 2009, Kalantri et al., 2020). This is the case both in individuals with HbE/β-thalassaemia, but also in those with some β^o genotypes (Oggiano et al., 1988, Galanello et al., 1989, Camaschella et al., 1995, Zhong et al., 2018). In several cohort studies in areas with a high prevalence of α-globin deletions, the number of cases with two α-globin gene deletions was well below expected, and sometimes not seen, suggesting that these individuals were so well they did not present to medical services (Bandyopadhyay et al., 2001, Sripichai et al., 2008b, Suresh et al., 2013). Coinherited β-thalassaemia and three α-

globin gene deletions is uncommon, but studies suggest the amount of α-globin reduction may be too severe (Melis et al., 1983, Kanavakis et al., 2004, Zhong et al., 2018). Therefore, a therapeutic strategy aiming to cure β-thalassaemia should aim for a reduction in α-globin expression commensurate with a deletion of one to two α-globin genes.

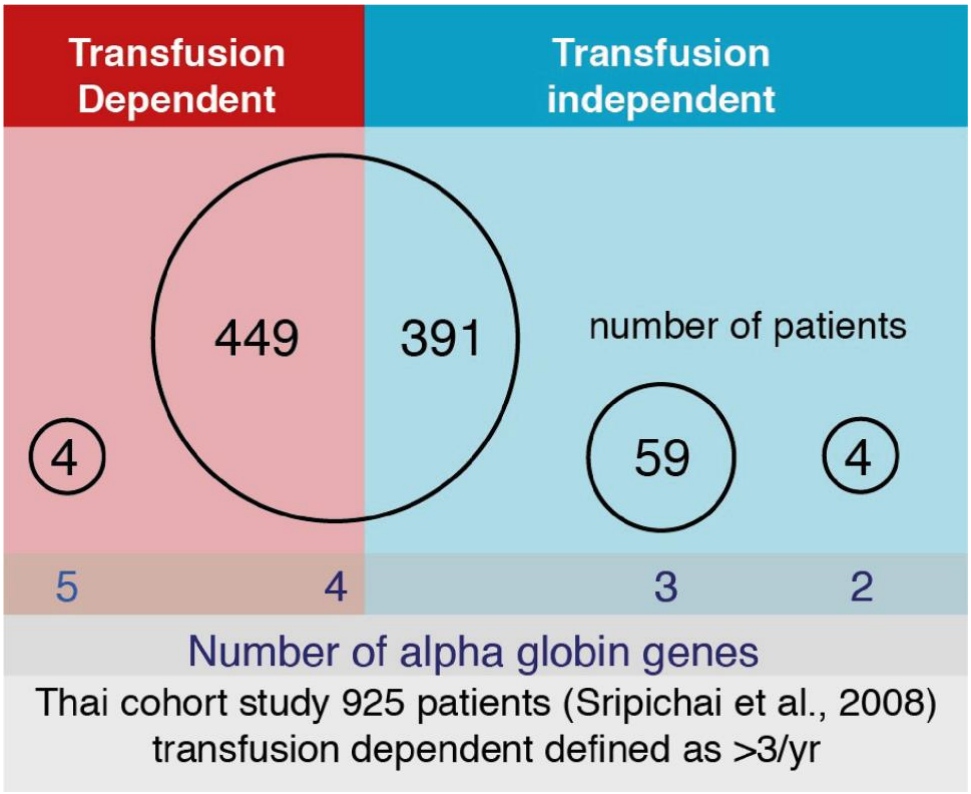


Figure 4.3 - Improved phenotype with α-globin gene deletions in HbE/β-thalassaemia

4.1.10 Gene editing strategy to reduce α-globin expression

The α-globin genes and their flanking sequences are highly homologous, meaning that trying to excise one or two genes precisely is difficult. Furthermore, two DSBs would need to be generated, which increases the likelihood of cell cycle arrest and on-target structural rearrangements. This is a particular risk given the sequence homology and repetitive nature

of the α-globin locus. Previous work in the lab instead tried to recapitulate a naturally occurring mutation that removes the main α-globin enhancer, MCS-R2 (Coelho et al., 2010). CD34+ cells were transfected with two plasmids, both containing Cas9 and a sgRNA (Mettananda et al., 2017a). Each of these was directed either side of the enhancer resulting in its deletion. This efficiently deleted the enhancer in selected cells, but perhaps unsurprisingly homozygous deletion of the enhancer resulted in an excessive reduction of α-globin expression. Thus, while this exact strategy would not be appropriate in the clinic, it provided a powerful proof of principle for using manipulation of the MCS-R2 to modulate α-globin expression.

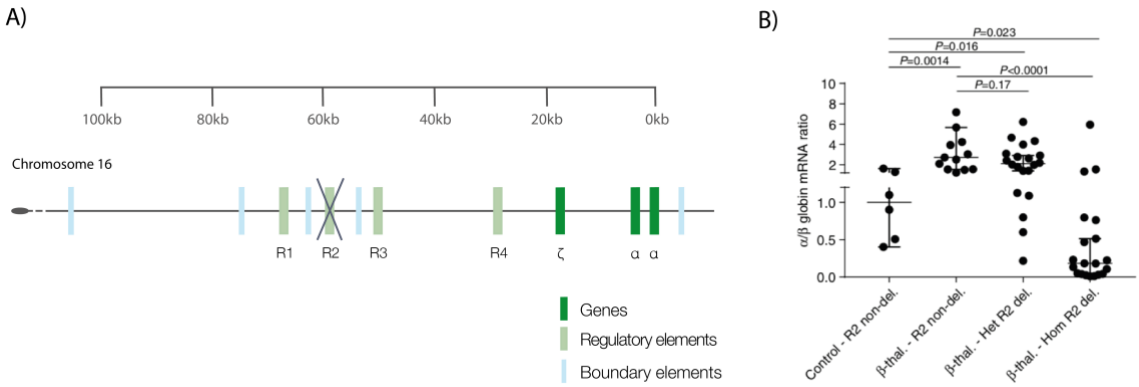


*Figure 4.4 - Previous MCS-R2 editing strategy A) The enhancer was deleted from the locus through CRISPR/Cas9 editing. B) Excessive knockdown of α-globin expression in cells homozygous for MCS-R2 deletions. Figure **4.4**B adapted from (Mettananda et al., 2017a)*

As discussed in chapter 3, enhancers can be considered as landing pads for transcription factors, and are in part composed of transcription factor binding sites (TFBSs). Earlier in this work the underlying DNA sequence of MCS-R2 was reviewed and interrogated using various assays. Through this a number of likely TFBSs were identified, and partly stratified in terms of their importance in maintaining chromatin accessibility at the enhancer. Given that some TFBSs may be more important than others - the central NF-E2 and GATA1 site seem likely

candidates - a plausible hypothesis is that single DSBs generated at these sites may be able to disrupt enhancer function of MCS-R2 and reduce α-globin expression in a graduated manner. If the complement of identified TFBSs in MCS-R2 could be profiled in this way, one might be identified that when edited reproducibly reduces α-globin expression to a level of 50-75%, commensurate with one to two α-globin deletions found in patients with co-inherited α- and β-thalassaemia mutations.

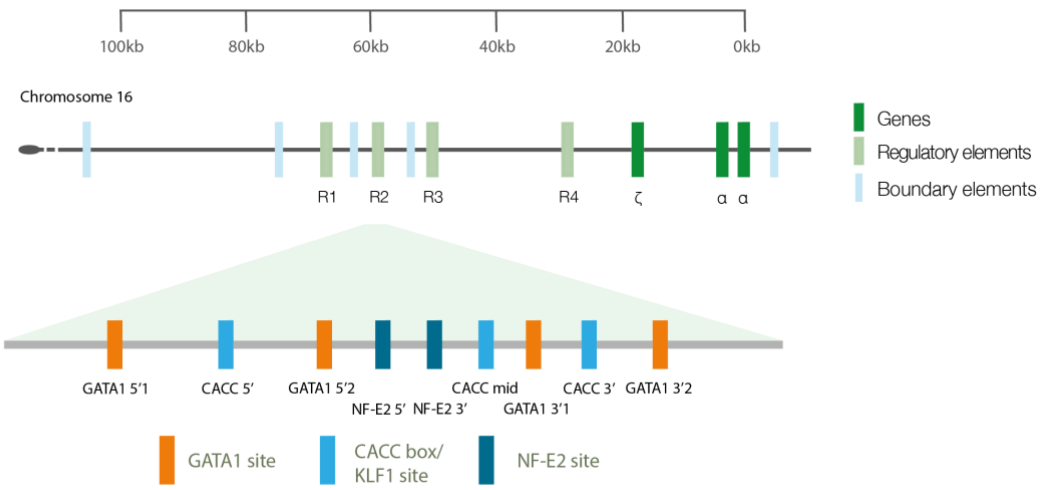


Figure 4.5 - Current MCS-R2 targeting strategy *The lower half of the figure shows a zoomed in schematic of MCS-R2 with individual TFBSs that can potentially be targeted.*

This chapter describes the development of a genome-editing strategy to ameliorate β-thalassaemia by reducing α-globin expression to ~50% of normal by mutating TFBSs within MCS-R2, thus restoring globin chain balance. This was initially done by designing sgRNAs across the MCS-R2 sequence, and editing the enhancer using plasmid vectors. Due to their preferred therapeutic potential RNPs were subsequently used. Primary human WT and patient cells were analysed for globin expression as well as for markers of differentiation. I subsequently analysed protein from edited erythroid cells, followed by an assessment of the OT-profile of one of the sgRNAs used in the editing screen.

4.2 Results

4.2.1 The α/β ratio in non-thalassaemic and α-thalassaemic cells

In order to assess whether edits had reduced α-globin expression to the desired level, the α/β ratio in control and patient cells was measured. The α/β ratio in peripheral blood from patients depends on the type of α-thalassaemia genotype (Figure **4.6**) (Hunt et al., 1980). When one α-globin is deleted, the mRNA ratio is approximately 0.71 when normalised to WT, whereas when two are deleted the ratio is approximately 0.44 (Higgs, 2009b). Later work using RT-qPCR found that the ratio in patients with the αα/--^{SEA} and α$^{-3.7}$/α$^{-3.7}$ genotypes was 0.4 in peripheral blood, although there was a higher degree of variation in this data (Chaisue et al., 2007).

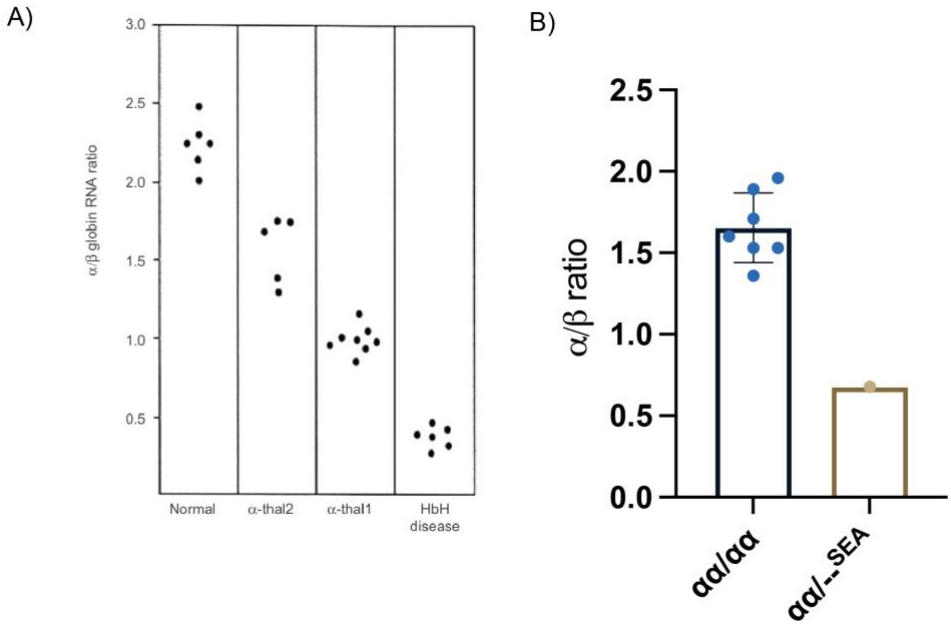


Figure 4.6 - Defining α/β ratio ranges A) α/β mRNA ratios in a variety of α-thalassaemia genotypes. α-thal2 refers to a single α-globin deletion, α-thal1 to two α-globin gene deletions (nomenclature now not in use). Adapted from (Hunt et al., 1980) B) α/β mRNA ratio in control cells vs an individual with the αα/--^{SEA} genotype. Individual data points shown with mean and SD.

As I planned to edit cells differentiated *in-vitro* from CD34+ cells isolated from peripheral blood, WT cells from seven individuals were differentiated into erythroid cells (day 10) using the same differentiation system as planned for the editing screen. The α/β-globin ratio was found to be 1.65 (standard deviation [SD] ± 0.13), and the α/non-α-globin ratio was 1.59 (SD ± 0.2). Differentiated erythroid cells from an individual with the αα/--^{SEA} mutation were also assessed; the α/β-globin ratio was 0.44 and the α/non-α-globin ratio was 0.59. These data were consistent with those generated earlier, and so an approximate target of 0.45 was set for two α-globin gene deletions and 0.7 for one α-globin gene deletion.

4.2.2 Designing sgRNAs

SgRNAs were designed to target the TFBS identified earlier (Figure **4.7**). Due to the size of TFBSs, most sites only had one sgRNA that might possibly target them. SgRNAs were cloned into the commonly used pX458 plasmid which contains the sgRNA expression cassette under the control of a human U6 promoter and a Cas9 attached to an eGFP for selection. Unfortunately, the NF-E2 3' site, predicted to be important for the function of MCS-R2, did not have a PAM close enough to allow a DSB in its consensus sequence (predicted cleavage site 5bp away), though this may now be possible with new Cas9 variants. The closest sgRNA had a commonly occurring variant within its PAM that was frequently found in the samples received from NHSBT. The GATA1 3'2 site did not have a sgRNA that created a DSB within the canonical GATA1 motif, though in this case the DSB site was 1bp away. The only site predicted to have significant off-target effects was the CACC mid sgRNA as assessed by the Sanger Institute off-target assessment tool, probably due to its repetitive sequence (Hodgkins et al., 2015).

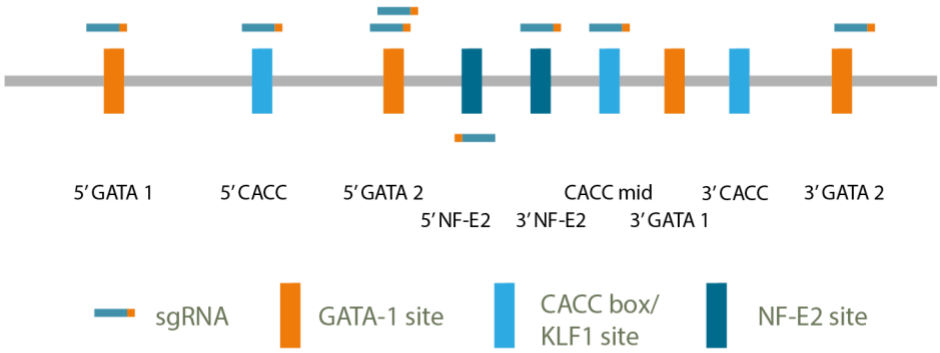


Figure 4.7 - sgRNA schematic *Available sgRNAs targeting the top and bottom strands are shown. Several sites do not have a sgRNA due to the lack of an NGG PAM (GATA1 3'1, CACC 3') and the GATA1 3'2 sgRNA cut-site is not within the GATA1 canonical motif.*

4.2.3 Optimising of GFP positivity

The draft experimental workflow involved selection of cells by GFP positivity. As the cells may have been sensitive post-electroporation, sorting by flow cytometry was not advised within the first 24 hours, but leaving them until full recovery risked missing cells that had lost their fluorescence. CD34+ cells were electroporated with px458, and GFP fluorescence was measured daily using flow cytometry, as shown in Figure 4.8. Using this curve, I decided to sort these transfected cells at 24 hours.

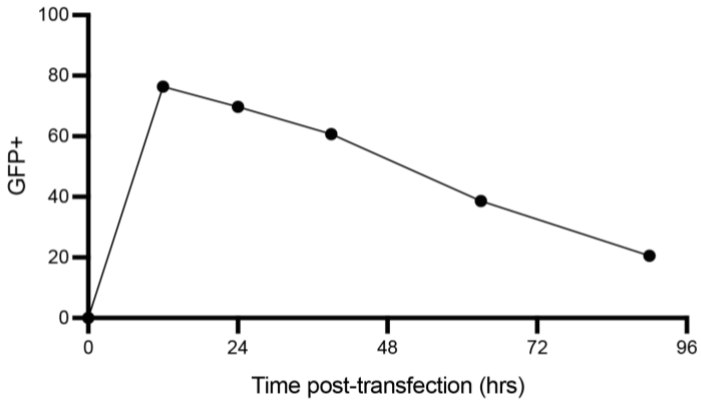


Figure 4.8 - GFP positivity over time *GFP positivity of CD34+ cells as assessed by flow cytometry*

4.2.4 Plasmid-based editing

CD34+ cells were electroporated using an established protocol that had been previously optimised (Mettananda et al., 2017a). CD34+ cells were isolated from leucocyte cones purchased from the local NHS Blood and Transplant Centre. These cells were isolated by apheresis from healthy individuals: demographic data was not available. Cells positive for GFP were sorted using flow cytometry at 24 hours into erythroid differentiation media. Cells were differentiated and harvested for downstream analysis on day 10, as that stage, the majority of the cells are erythroblasts that express specific erythroid cell surface markers, and the globin genes are expressed at high levels (Scott et al., 2020).

Plasmids targeting the GATA 5'1, GATA 5'2, NF-E2 5', NF-E2 3' and GATA 3'2 sites were transfected into cells using the Amaxa 2b platform with an average of 36% (SD ± 13.3%) viability and 28.7% GFP positivity (SD ± 12.7%). Transfection with the newer Amaxa 4d machine gave slightly higher viabilities (41.6% SD ± 19.6%) with much lower GFP positivity (5.1% SD ± 5.86). An initial screen using a T7 endonuclease assay to assess editing efficiencies showed unexpectedly low levels of indel formation. NGS was used to precisely quantitate the indel frequencies using two guides over three replicates; using the NF-E2 5' sgRNA and GATA 5'2 sgRNA. For the GATA 5'2 site the editing efficiency was 7.18% and for the NF-E2 5' site 19.09%. As this analysis was in selected cells, the editing efficiency was too low to be therapeutically relevant and also too low to assess the sites' importance to α-globin transcription. RT-qPCR was conducted using validated Taqman probes for the globin genes using RPL13a as a reference. Unsurprisingly there was no significant difference in the α/β ratios between transfected cells and non-transfection controls (NTC).

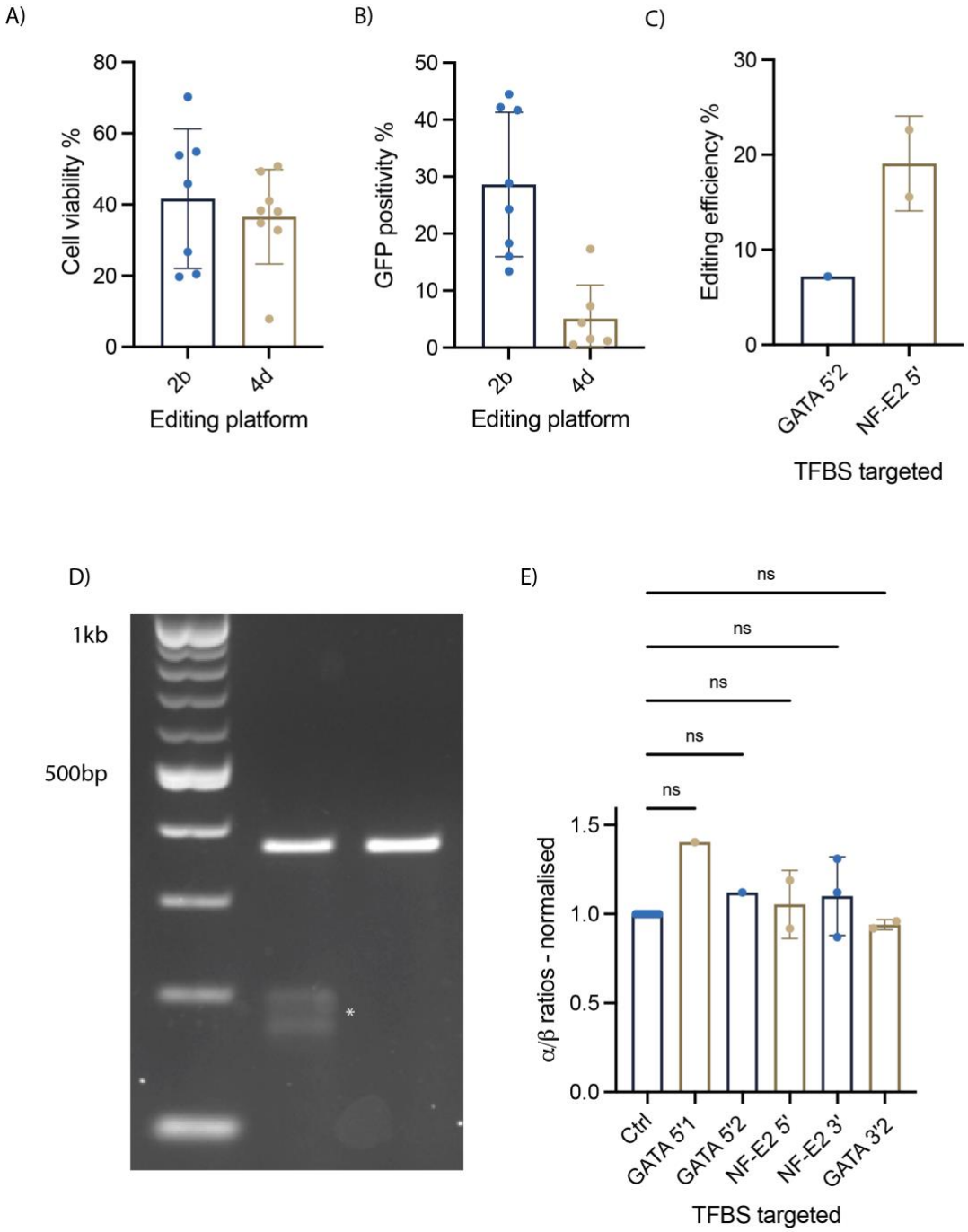


Figure 4.9 - Plasmid based editing results *A and B) Viability and GFP positivity when using the Amaxa 2b and 4d systems C) Editing efficiencies at the GATA1 5'2 and NF-E2 5' sites as assessed by NGS. D) Representative T7 endonuclease assay for targeting the NF-E2 5' site. Lane 1 - 100bp ladder, Lane 2 - edited sample, Lane 3 - unedited control. Most amplification products are WT, with a faint double band (*) indicating T7 restriction of edited amplicons. E) RT-qPCR showing no significant change in expression.*

4.2.5 Plasmid-based editing - DSB repair outcomes

The DSB repair outcomes for the cells that were edited was examined. For the NF-E2 5' sample, the overwhelming outcome was a deletion at the cut site with a small number of insertions. The GATA1 5'2 site in contrast had a much larger number of insertions at the cut site. These findings indicate that repair outcomes post-cleavage may be sequence context specific. The deletion sizes were quantified using this dataset; he majority of deletions were less than 50bp in size, with 75.1% under 20bp, and 83% under 50bp. This confirms that NHEJ is a good method to disrupt sequences in non-coding areas.

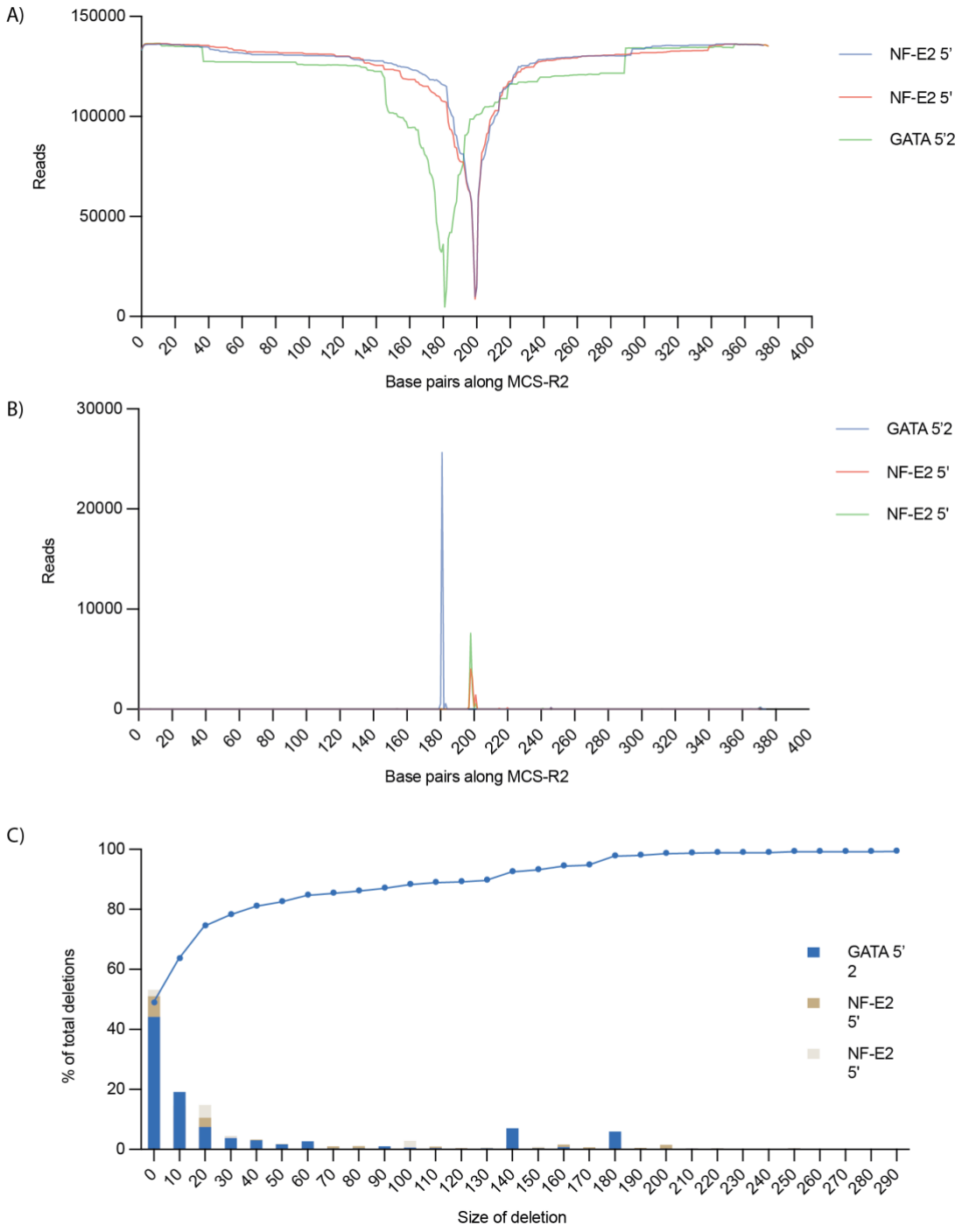


Figure 4.10 - The editing landscape *A) The pattern of deletions at two targeted sites in MCS-R2. A reduced number of reads at a given base indicates a deletion. B) The pattern of insertions at both sites, C) Summary of the size of deletions. The bars represent the percentage of deletions in bins of 10bp for each experiment. The line shows the cumulative count of the deletion percentage.*

4.2.6 Ribonucleoprotein editing

Due to the limited efficacy of editing when using plasmids, alternative CRISPR/Cas9 vectors were used. RNP use involved the incubation of Cas9 protein and synthesised sgRNAs followed by transfection into primary CD34+ cells. Two electroporators were initially used: the Invitrogen Neon and Lonza Amaxa 4D platforms. In contrast to plasmid-based workflows, cells were not selected by GFP fluorescence. Cas9 protein (IDT) and synthetic sgRNAs (Synthego) for several TFBSs were purchased. Editing efficiencies with either of the platforms were far superior to plasmid. Due to its ease of use and higher indel efficiencies (as assessed by Sanger sequencing trace deconvolution tools) all further experiments used the Amaxa 4D. Optimisation experiments were conducted to define how much Cas9 to use per transfection; 3.2µg was found to deliver superior editing efficiencies to 1.6µg. The editing results for the screen of the enhancer used 100,000 cells per transfection. Scrambled sgRNAs that did not target any sequence in the human genome were used for control transfections.

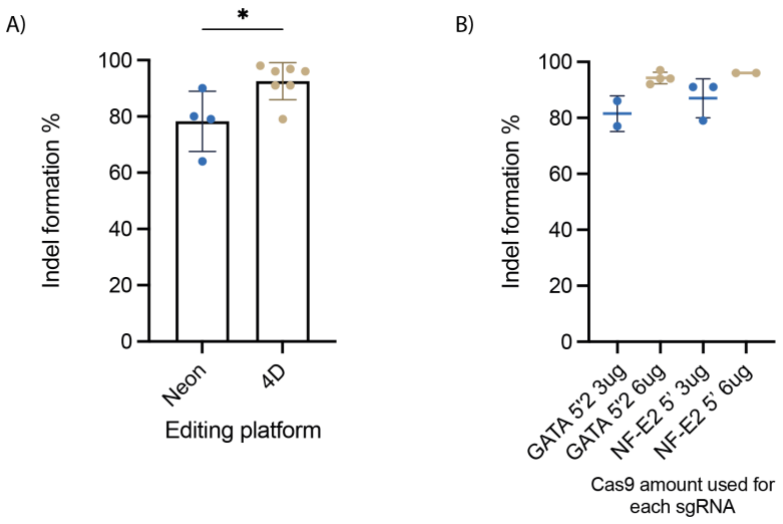


Figure 4.11 - Optimisation of RNP editing *A) Editing using the Neon and Lonza Amaxa 4d. Data shown for the NF-E2 5' sgRNA, mean ± SD shown, significance calculated using two-tailed Mann-Whitney test. * denotes p-value <0.05*

Indel formation for the various sgRNAs used were mostly in excess of 90% as measured by NGS, meaning that most cells had a homozygous or compound heterozygous mutation at the target. Certain sgRNAs did not result in highly efficient editing, suggesting that this was due to limitations in the protospacer sequence. RNP editing was tolerated well, with viabilities at 24h on average 85.6% as measured by trypan blue exclusion.

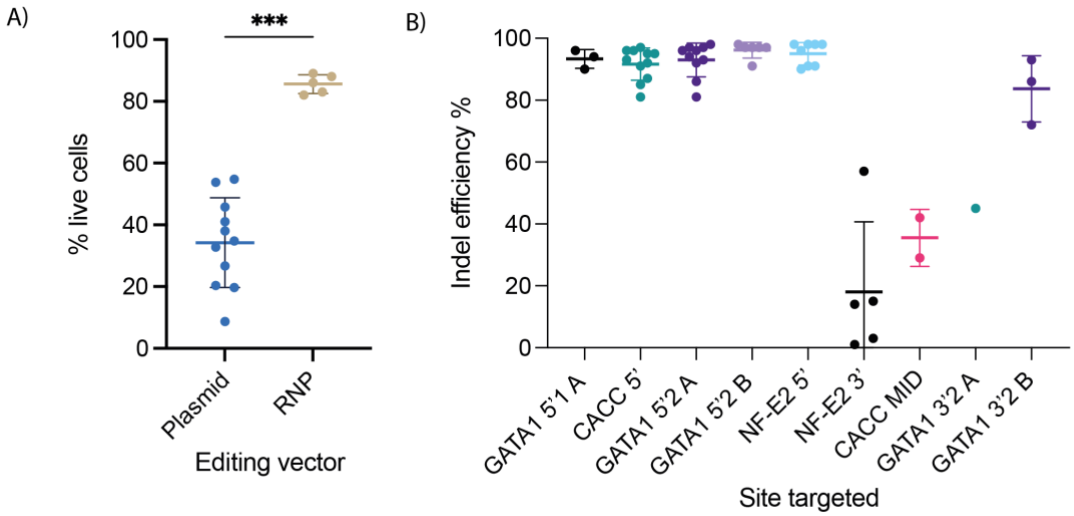


Figure 4.12 - RNP editing results *A) The viability of edited cells at 24h was higher with RNP than plasmid B) Summary of editing efficiencies with all the sgRNAs used to target MCS-R2. The sites are arranged along x-axis in the order that they are found in the enhancer. sgRNAs suffixed with A or B indicate TFBSs with two available sgRNAs.*

An assessment of the deletion profiles using NGS data was made. Compared to plasmid editing there was a clear reduction in size of the deletions generated, with 79.8% of deletions under 10bp and 95.0% under 20bp for the GATA 5'2A sgRNA (Figure 4.13A). These profiles were highly reproducible, and specific for each site, with each site having a slightly different deletion size profile (Figure 4.13A&B) and pattern (Figure 4.13C&D).

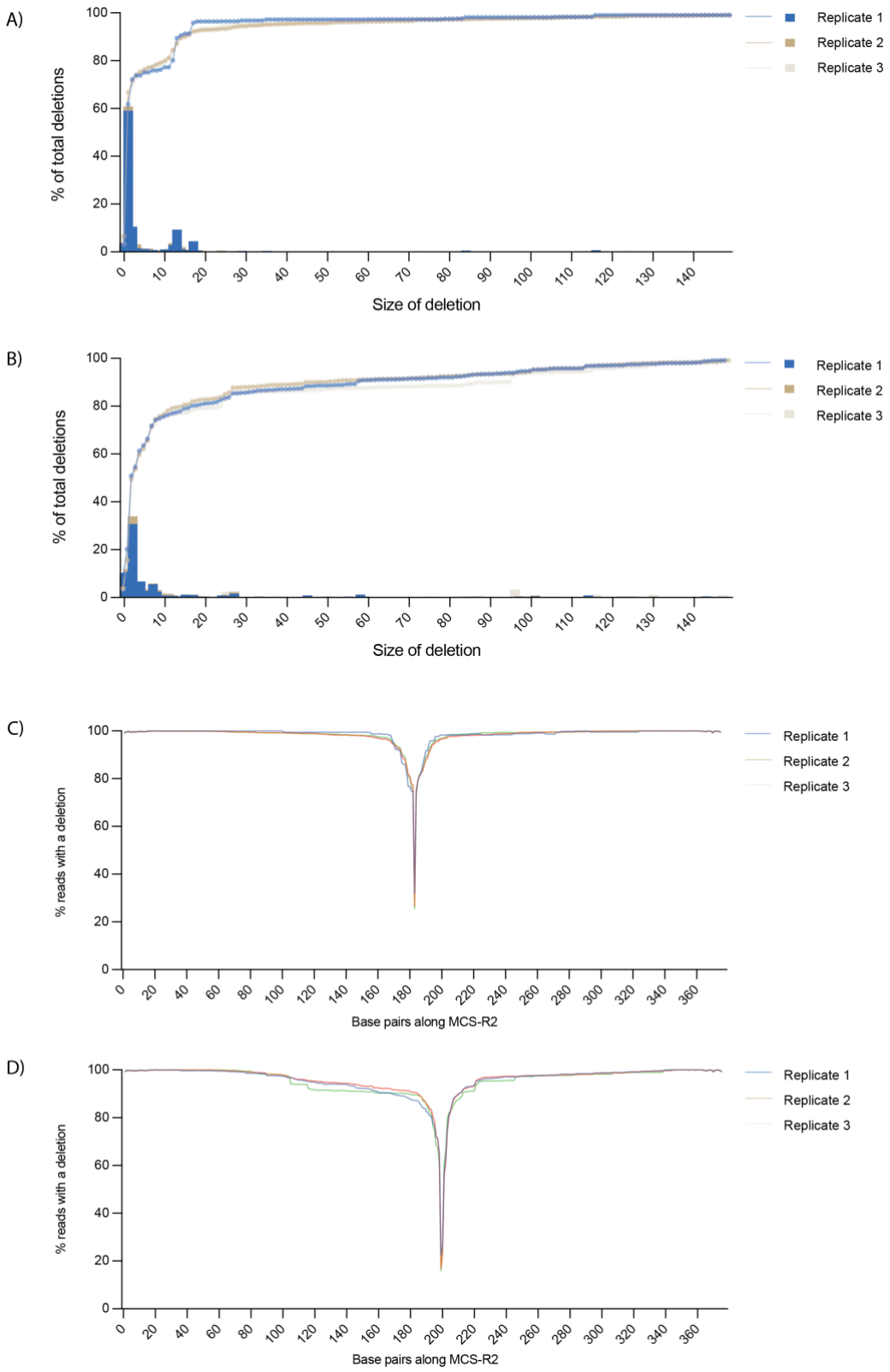


Figure 4.13 - RNP editing profiles A) and B) Summaries of the size of deletions. A) shows data targeting the GATA 5'2 site and B) targeting the NF-E2 5' site. Histograms show individual deletion sizes; the line shows the cumulative count of deletions. C) and D) Deletion profiles across MCS-R2 for the GATA 5'2A sgRNA (C) and the NF-E2 5' site sgRNA (D).

4.2.7 Ribonucleoprotein editing and globin gene expression

For each mutated site in the enhancer, globin gene expression was measured by RT-qPCR at day 10 of erythroid differentiation. No individual site abrogated α-globin transcription completely. There was however a reproducible variation when editing different sites. The greatest reduction was at the NF-E2 5' and GATA 5'2 sites, as predicted, with a reduction in α/β ratio to 0.66 and 0.72 respectively. This editing efficiency was less marked in the peripheral sites. The reduction in α-globin expression was not a function of the transfection alone, as the raw α/β ratios in non-transfection controls and scrambled guide controls were compared and found to be similar. The α/β ratios and α/non-α ratios were found to produce consistent results (NF-E2 5' 0.66, GATA1 5'2 0.68).

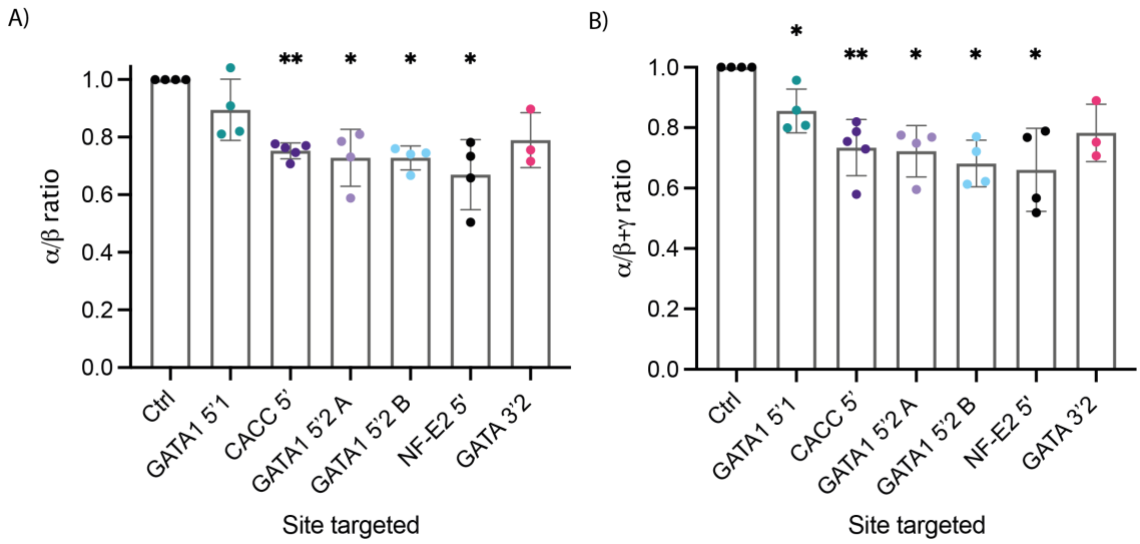


*Figure 4.14 - Globin gene expression with single site edits A) α/β mRNA ratios B) α/β+γ mRNA ratios. Both sets of data measured by RT-QPCR with SD shown. Ctrl - scrambled sgRNA control. Statistical significance calculated using Mann-Whitney U test. * denotes p < 0.05, ** denotes p <0.01.*

The reduction in α/β ratio was commensurate with a 20-30% reduction in a-globin expression and so would potentially be of therapeutic benefit. Whilst not targetable in this

work, it would be expected for the NF-E2 3' site to result in a similar knockdown or greater, based on the work outlined in chapter 3.

4.2.8 Ribonucleoprotein editing and assessment of differentiation

Control and edited cells were compared using microscopy and Immunophenotyping to determine whether editing had any effect on differentiation.

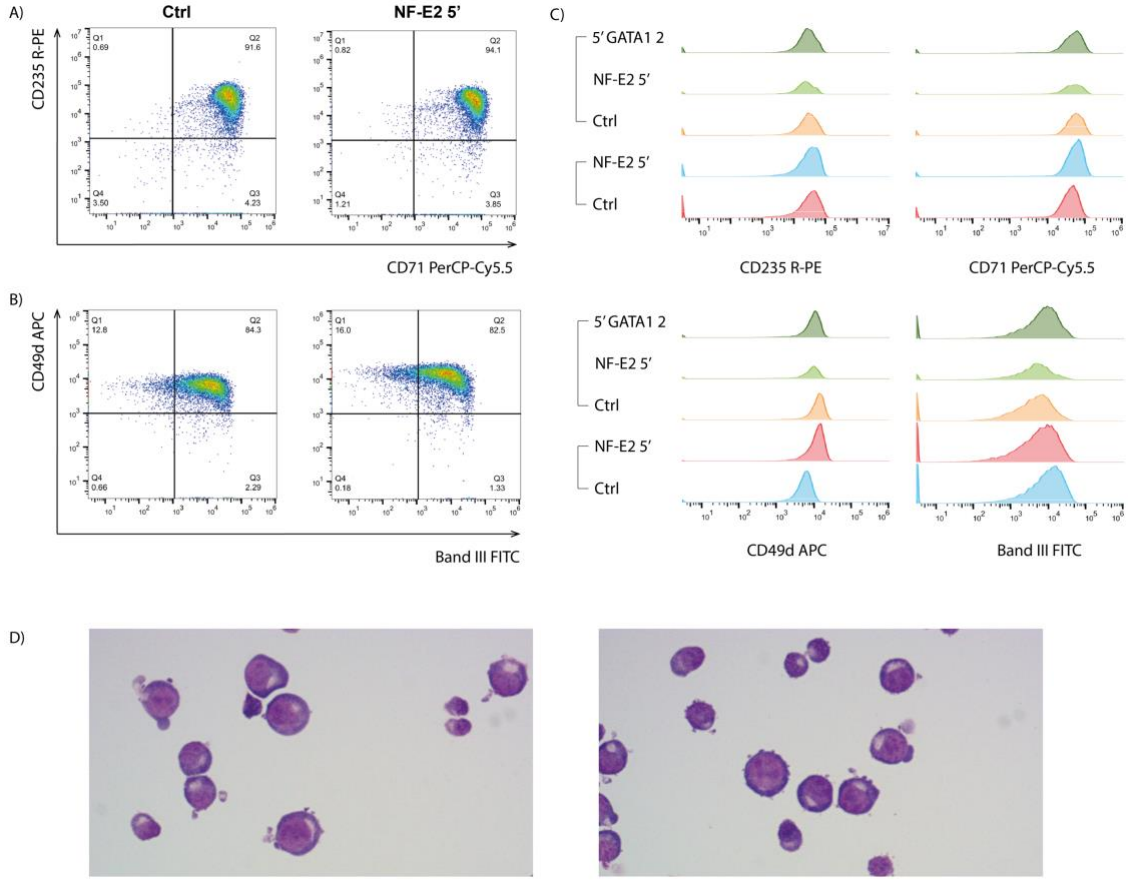


Figure 4.15 - Erythroid Differentiation *Representative data showing that there was no perturbation in erythroid differentiation when editing MCS-R2. Data for the NF-E2 5' site are shown, but the same immunophenotypic and morphological patterns were seen at the other sites. A and B) Control and edited cells (sgRNA targeting NF-E2 5' site) were at similar stages of erythroid differentiation as measured by CD71-CD235 and Band III-CD49D. C) Histogram flow cytometry data showing that cells edited with different sgRNAs and different biological donors were at the same stage of differentiation. D) Representative cytospins showing erythroid precursors from control (left) and edited (right, NF-E2 site targeted) samples.*

There were no morphologic differences seen at day 10, or differences in the expression of key erythroid cell surface markers including CD235, CD71, Band III and CD49D as assessed by flow cytometry (Figure **4.15**A-C).

4.2.9 Ribonucleoprotein editing of HbE/β-thalassaemia patient cells

To investigate whether the MCS-R2 targeting strategy would be therapeutically relevant, the top three performing sgRNAs targeting the CACC 5', GATA1 5'2 and NF-E2 5' sites were transfected separately into CD34+ cells isolated from the peripheral blood of a patient with severe transfusion-dependent HbE/β-thalassaemia. Due to the scarcity of local donors and the viability of cells transported from abroad this was only performed with one biological replicate. The patient had the common IVS1-5 (G>C) mutation on the thalassaemic allele. Editing efficiencies for two of the sites were in excess of 95%, but the CACC 5' site was not edited well, with 65.7% of reads showing an indel as measured by NGS. RT-qPCR showed a reduction in the normalised α/β-ratio to 0.74 for the GATA1 5'2 site and 0.75 for the NF-E2 5' site. As expected, due to the inefficient editing, the CACC 5' targeted cells showed only a slight reduction in the normalised α/β ratio to 0.95.

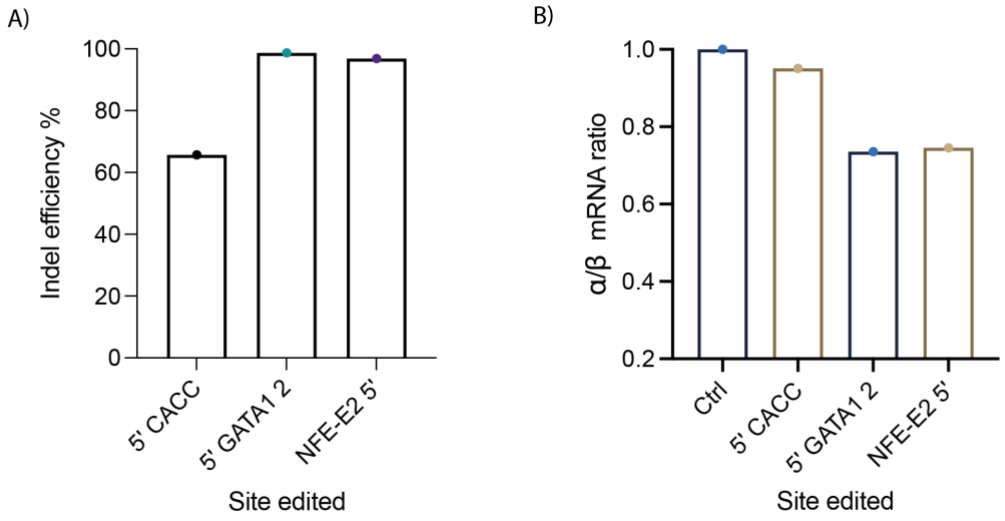


Figure 4.16 - Editing in HbE/β-thalassaemia patient cells A) Editing efficiencies as quantitated by NGS B) α/β-ratio in patient cells, showing a reduction upon editing at the NF-E2 5' and GATA 5'2 sites but not the CACC 5' site.

4.2.10 Ribonucleoprotein editing - protein assays

As the pathophysiology in patients with β-thalassaemia is caused by an excess in α-globin chains, an assessment of protein chain production was performed. Isoelectric focusing (IEF) was carried out to see whether any unexpected haemoglobin molecules were formed (Figure **4.17**A). This did not show any abnormal or unidentified bands. Reverse-phase high performance liquid chromatography (RP-HPLC) to quantify the amount of α-globin chains was attempted, but over two replicates no difference in the α/β chain ratio was detected between control and any edited samples (Figure **4.17**B). The samples included had edits at the GATA1 5'2, NF-E2 5' and CACC 5' sites, all of which had mRNA ratios of between 0.75 and 0.8. A positive control of cells with a partially efficient removal of MCS-R2 and an α/β-mRNA ratio of 0.39 was included, which also showed a balanced α/β globin chain ratio. It may be that this was due to excess β-globin chains being destroyed via proteolysis and/or

precipitation. Further development of the RP-HPLC assay and optimisation was not possible for technical reasons related to the analyser.

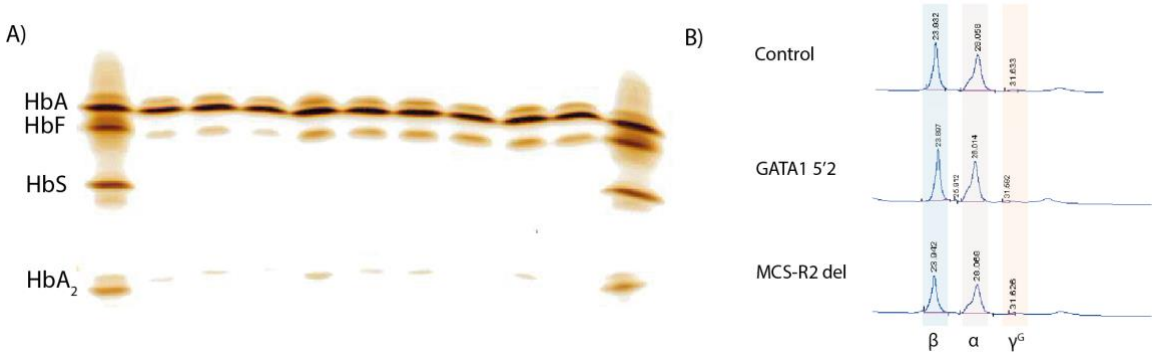


Figure 4.17 - Protein assays *A) Representative IEF gel of samples from one replicate showing production of normal haemoglobins. Lanes at either end are reference samples B) Representative RP-HPLC traces of a control sample, sample edited at the GATA1 5'2 site and MCS-R2 enhancer deletion.*

4.2.11 Ribonucleoprotein editing and chromatin accessibility

To examine the effect of editing on chromatin accessibility at the α-globin locus, ATAC-seq was performed on control samples and samples edited at the CACC 5', GATA1 5'2 and NF-E2 5' sites. In edited samples chromatin accessibility at MCS-R2 was maintained, and no broader perturbations across the α-globin locus was detected (Figure **4.18**A). This was in contrast to the data shown in chapter 3, which resulted in broader reductions in chromatin accessibility at the remaining enhancers when MCS-R2 was deleted. Looking closer at the enhancer, at cut sites there was a relatively large depletion in read depth of about 100-150bp (Figure **4.19**B). This was in spite of the majority of indels using RNP being below 10-20bp depending on the sgRNA used.

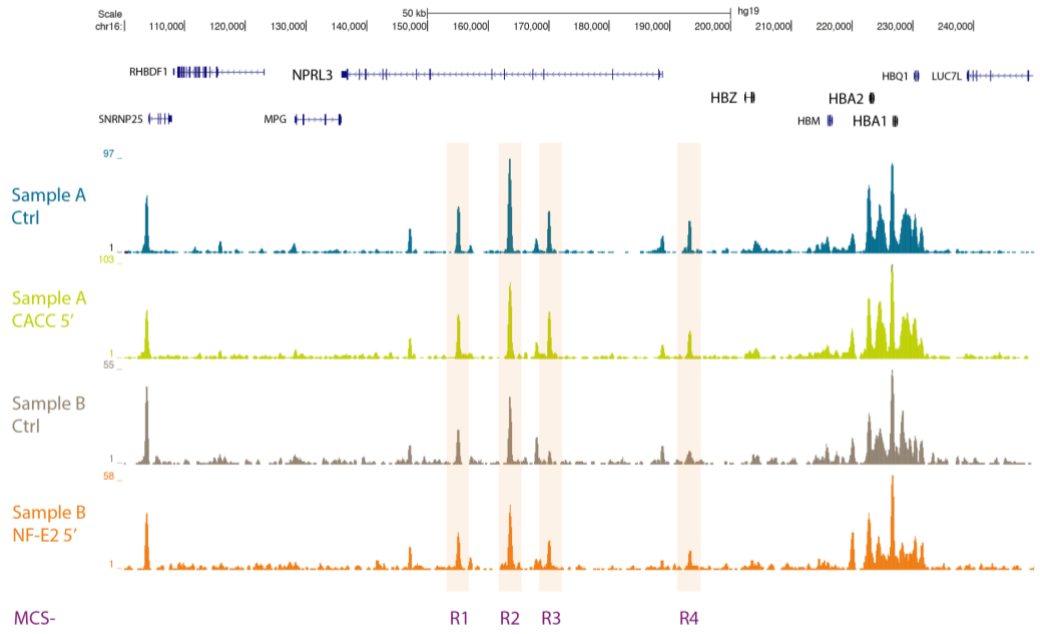


Figure 4.18 - ATAC-seq at the α-globin locus in edited cells *The ATAC peak at the MCS-R2 enhancer was slightly reduced in edited cells, but not observably over the other active elements at the α-globin locus.*

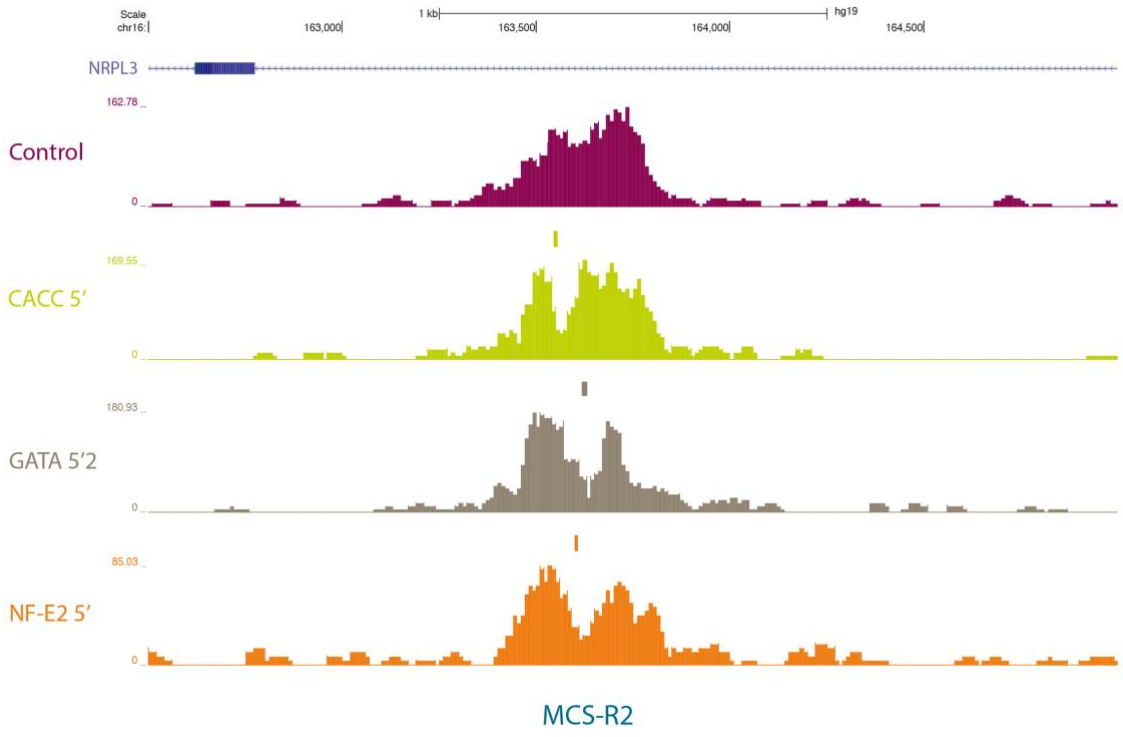


Figure 4.19 - ATAC-Seq at MCS-R2 in edited cells *The small vertical bar above the edited peaks indicates the TFBS motif that was edited.*

Genome-wide accessibility of chromatin was assessed to see whether editing caused any global or specific changes. ATAC-seq was performed on differentiated erythroid cells from one healthy donor, comparing control to cells targeted with the NF-E2 5' sgRNA. Principal component analysis (PCA) did not show that the replicates clustered strongly according to whether they were edited or not (Figure 4.20A), and there were no peaks that were significantly increased or decreased when assessed by DEseq2, meaning that there were no significant changes in chromatin accessibility in edited cells compared to control cells (Figure 4.20B) (Love et al., 2014). PCA is a relative indicator of variance within an experiment, and so the variance shown in Figure 4.20A may not have been particularly large in absolute terms. This sample was compared to another control-edited pair, from a different biological donor and targeting a different site. In contrast to previously, there was a large degree of variance in the first principal component (PC, 91%). This relates to the difference between biological replicates, even if they are control samples, and is represented by the x-axis dimension in Figure 4.20C. There was a much smaller degree of variance in the second PC as represented by the y-axis (6%), which relates to whether the sample was the control or edited sample, suggesting that the differences in the PCA seen in the NF-E2 5' targeting experiment were small in absolute terms, indicating further that editing does not lead to major changes in chromatin accessibility.

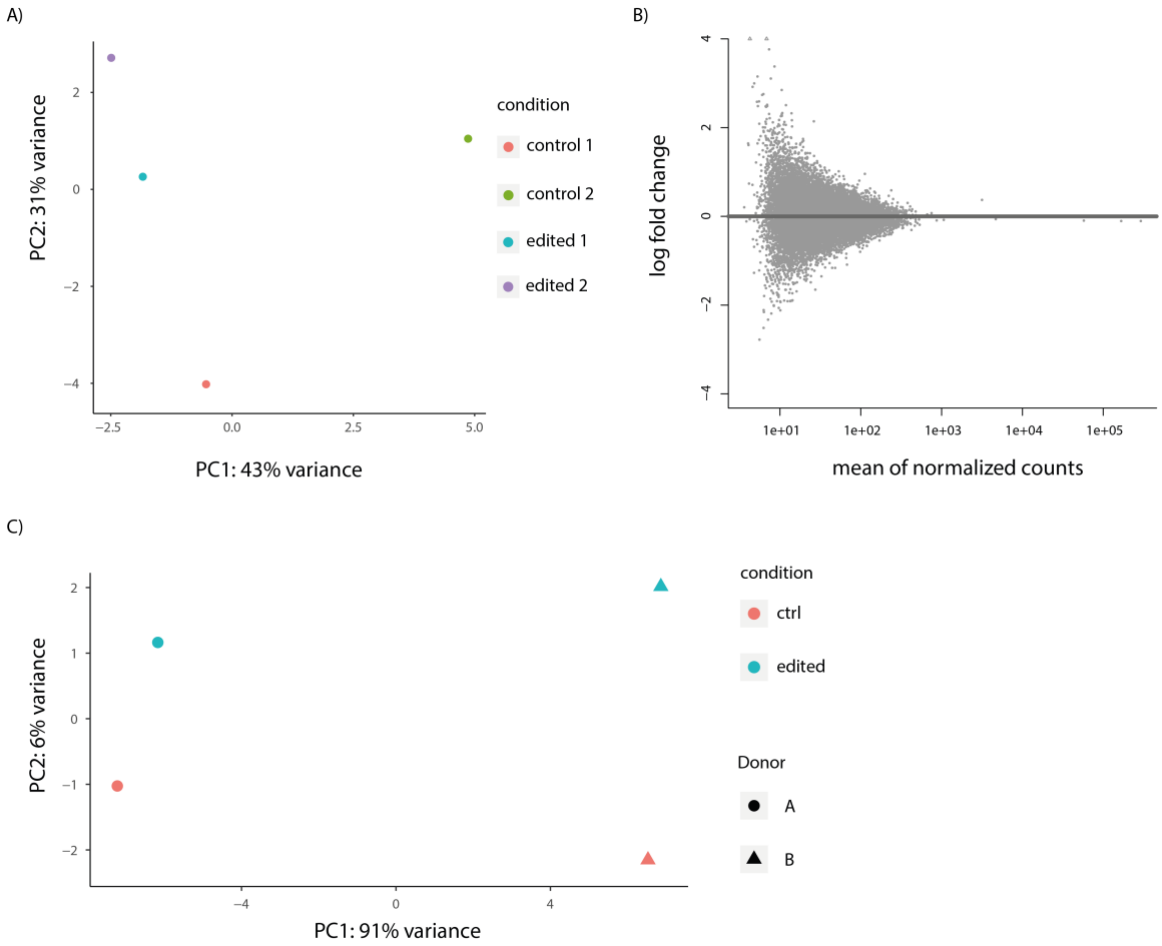


Figure 4.20 - Global ATAC-seq analysis A) PCA of the NF-E2 targeting experiment B) DESeq2 output of the NF-E2 targeting experiment, showing no significant changes in peak heights (significant changes would be shown in red). C) PCA of the ATAC-seq profile between two separate biological replicates (donor A shown with a circle, donor B shown with a triangle).

4.2.12 Ribonucleoprotein editing and chromatin architecture

The reduction in α-globin expression caused by editing TFBSs is presumed to be due to a reduction in cognate TF binding. To investigate further, the NF-E2 5' site was edited in HUDEP-2 cells (Figure **4.21**A). These NF-E2 5' mutant cells were differentiated along the erythroid lineage and ChIP-qPCR was performed for NF-E2 at MCS-R2 (Figure **4.21**B). There was a clear reduction in binding of NF-E2 at MCS-R2 compared to MCS-R1. Binding was not

completely abrogated due to the presence of the NF-E2 3' site which lies 26bp away. Micro-Capture-C, a highly precise chromatin conformation assay capable of detecting changes in chromatin interactions at base-pair resolution was performed. MCC allows assessment of contacts between the α-globin promoters and the individual TFBSs within the enhancers, including MCS-R2. This showed that the strength of interaction is significantly reduced in the NF-E2 5' site mutant line (Figure **4.21**C&D). These interactions are specifically reduced at the mutated NF-E2 5' site, and the footprint of interactions alters, suggesting that NF-E2 is unable to bind there (Figure **4.21**E). There were no other reductions in interactions detected with other active elements at the α-globin locus (Figure **4.21**F).

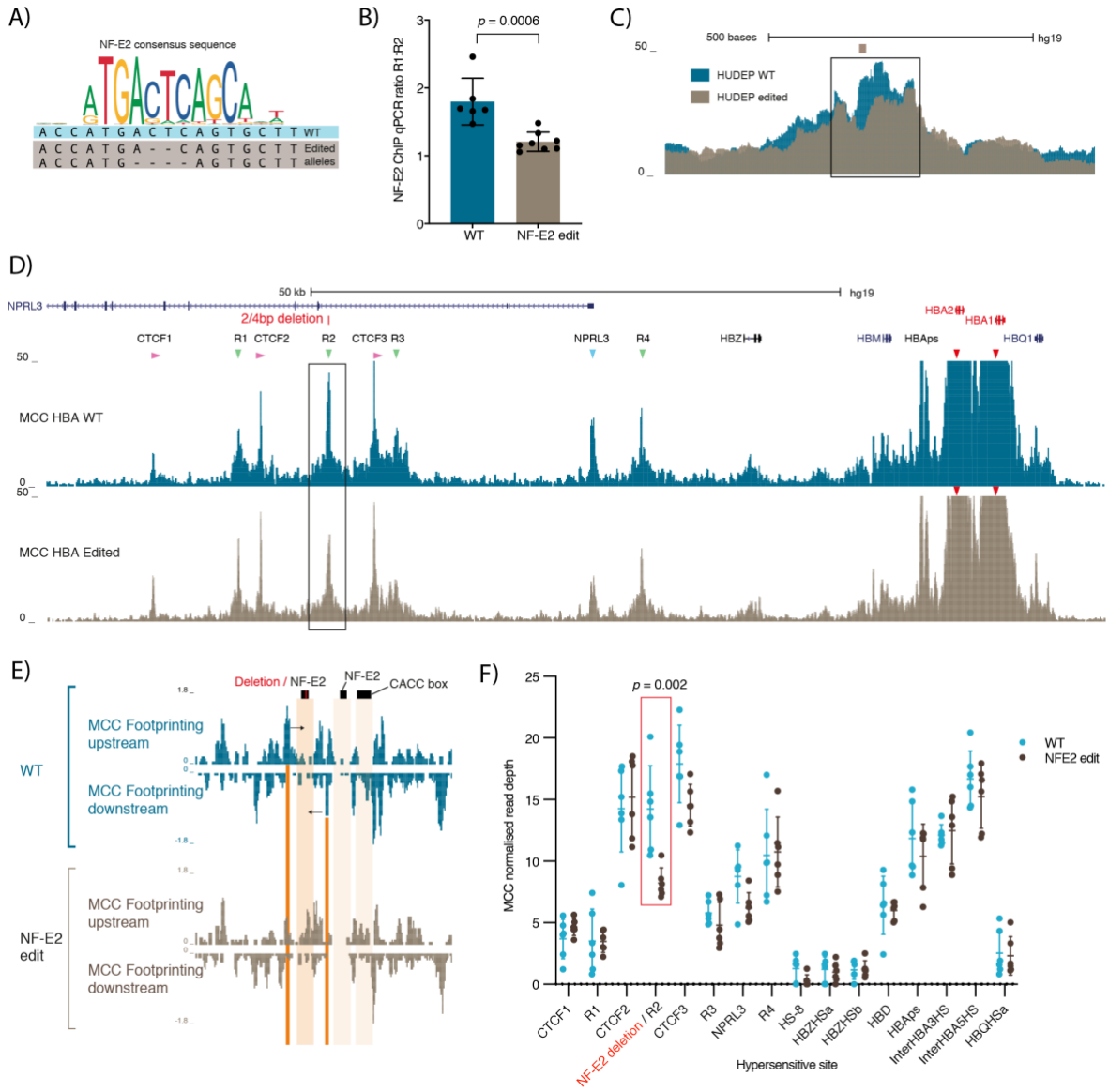


Figure 4.21 - Micro-Capture-C in NF-E2 5' edited cells A) HUDEP-2 cells were engineered with a compound heterozygous mutation in the NF-E2 5' site. B) Binding of NF-E2 is significantly reduced at MCS-R2 in NF-E2 5' edited cells C) and D) MCC profiles in edited and non-edited cells, showing a reduction in peak height over the editing site. The site of the edit is marked with a green bar in C). E) The MCC footprint is specifically altered at the NF-E2 5' binding site in mutant cells. F) Quantification of read depth over the α-globin locus. The only statistically significant change is over MCS-R2. Statistical significance calculated using a two-tailed Mann-Whitney U-test. Adapted from (Hua et al., 2021)

4.2.13 Ribonucleoprotein editing and Off-target editing

As the sgRNA targeting the of the NF-E2 5' site was the best performing and most likely candidate for therapeutic use, an assessment of the OT editing landscape was performed. CIRCLE-seq was carried out using the NF-E2 5' sgRNA. Several OT sites were identified (Figure **4.22**). As expected, the majority of mismatches are found at the 5' end of the protospacer, furthest away from the PAM. When examining ATAC-seq data, no changes in chromatin accessibility were noted between control and edited erythroid cells at any of these OT sites.

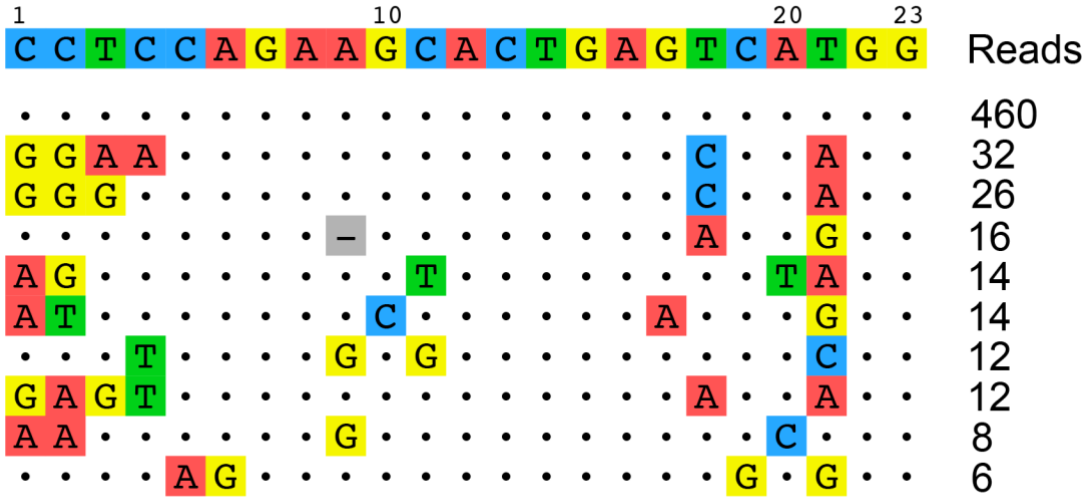


Figure 4.22 - CIRCLE-seq results *List of OT sites identified using CIRCLE-seq. Dots represent bases identical to the protospacer. Coloured identified bases show mismatches. The three bases on the 3' end of the sequence represent the PAM. As expected, the majority of reads are represented by the on-target editing site, shown at the top with 460 reads.*

4.2.14 Ribonucleoprotein and mRNA editing: combinatorial approaches

As each central site seemed to contribute to MCS-R2's enhancer activity, some preliminary combinatorial experiments were conducted to see if further reductions in α-globin expression could be achieved. CD34+ cells were edited as described above, allowed to rest for 24h and then edited again using a different sgRNA to prevent deletion of intervening sequences and complex local rearrangements. There was a reduction in α-globin expression seen with all combinations with the reduction approximately additive to that when the individual sites are edited alone (Figure **4.23**A&B). Sequencing showed the individual sites were edited specifically, without perturbing non-targeted sites.

Base editing, described in detail in chapter 5, allows simultaneous multiplex editing without having to create a DSB. CD34+ cells were transfected using mRNA instead of RNP, and adenines within the GATA1 5'2 and NF-E2 5' sites were targeted for mutation to guanine. As shown in Figure **4.23**C, these bases (marked with an asterisk) within the canonical motifs were efficiently deaminated, resulting in a robust and reproducible reduction in the α/β globin ratio (Figure **4.23**D&E).

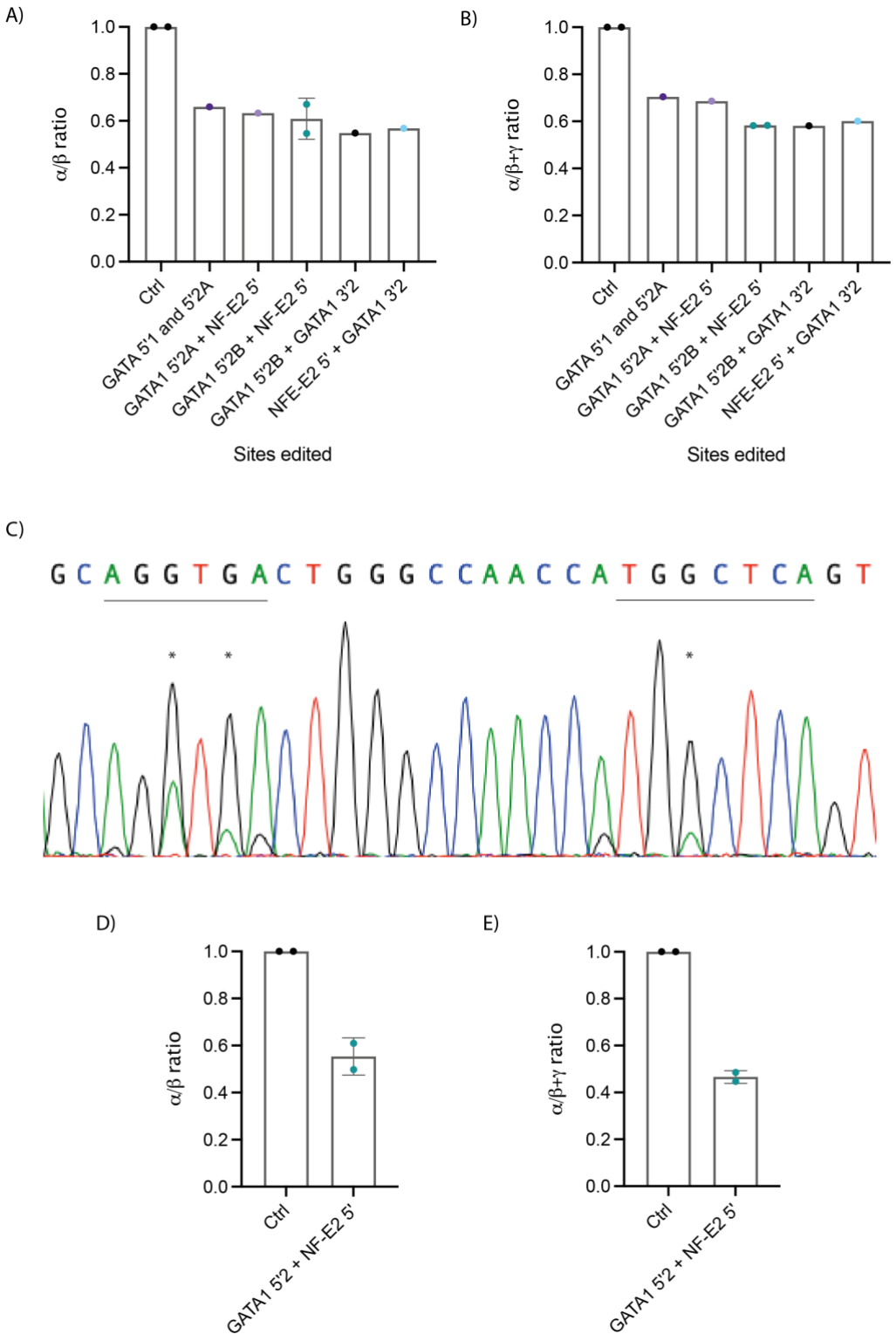


Figure 4.23 - Combinatorial editing *A and B) Globin mRNA ratios for cells serially edited with RNP C) Sanger sequencing for the combinatorial base-edited TFBSs D and E) Globin mRNA ratios for base-edited cells.*

4.3 Discussion

Since it was first characterised the CRISPR/Cas9 system has proved revolutionary in the fields of basic scientific research, biotechnology and clinical therapeutics. Whilst targeted endonucleases had been developed before, CRISPR/Cas9 offered a step-up in ease of use, target-site resolution, and efficiency. This technique has been used in pre-clinical and early clinical work to treat β-thalassaemia by editing sequences in the BCL11a erythroid enhancer and γ-globin promoters and thereby increase expression of fetal haemoglobin. The current work takes an orthogonal route in trying to directly reduce the amount of excess α-globin by reducing α-globin expression itself through editing its main enhancer (MCS-R2). Following the bioinformatic screens discussed in chapter 3, nine transcription factor binding sites (TFBSs) were identified for further targeting. Due to the stringency enforced by the PAM sequence, several of these could not be targeted. Several edited sites showed an appreciable and significant reduction in α-globin transcription as measured by the α/β and α/non-α mRNA ratio. Editing the central TFBSs, NF-E2 5' and GATA1 5'2 produced significant reduction in α-globin expression, with the CACC 5' site also showing a reduction. These findings were reproducible between biological donors, as well as in cells derived from a patient with HbE/β-thalassaemia.

These findings are important in several ways. Editing was performed with high efficiency in primary human cells. When this study began, editing with the highest efficiencies were either performed using plasmid transfection with GFP selection or viral transduction. This work confirms and establishes RNP transfection as an easy to use, reproducible and well tolerated vector for editing primary human HSCs. This seems to have been the case with

most sgRNAs tested here; they had excellent editing efficiencies and certainly high enough to be clinically relevant. In murine models of β-thalassaemia, transplantation and chimerism of normal HSCs as low as 10-20% resulted in phenotypic improvements, likely due a selective advantage of WT cells (Persons et al., 2001). This is consistent with human data showing that donor engraftment of 20-25% results in significant phenotypic benefit in patients with β-thalassaemia undergoing allogeneic stem cell transplant (Andreani et al., 2008). With editing efficiencies as high as those obtained, even modest chimerism in human transplant recipients should have an appreciable therapeutic effect. Inducing a single DSB is also more efficient, straightforward to use and safer than other approaches used to reduce α-globin expression, both of which use AAV transduction and one of which relies on two sgRNAs to excise genome sequences (Pavani et al., 2021, Cromer et al., 2021). Interestingly, RNP-editing deletion profiles were smaller than those using plasmid. This may possibly be due to the lack of re-cutting at the target site upon repair of the DSB to the WT sequence, as plasmids allow for continued expression of Cas9 whilst RNPs are degraded rapidly.

Using the strategy of targeted knockdown of α-globin expression, an ideal reduction would need to be equivalent to that associated with deletion of one or two α-globin genes. Work leading up to this project produced a reduction in excess of this, and would have been similar to a reduction seen in cases of HbH disease (Mettananda et al., 2017a). The normalised α/β globin mRNA ratio seen in cases with two α-globin gene deletions is approximately 0.45 and 0.7 with one deletion. Editing the GATA1 5'2 and NFE2 5' sites results in reduced α/β globin RNA ratios of 0.72 and 0.66 respectively, approximately that seen in a single gene deletion. When editing the HbE/β-thalassaemia patient cells, the normalised α/β-thalassaemia ratio for both of these sites reduced to approximately 0.74,

and a non-normalised ratio from 6.14 (non-edited) to 4.5. The mean non-normalised α/β ratio in WT cells was 1.65, and so this reduction in ratio to 4.5 may have therapeutic effect, as the ratio in β-thalassaemia trait cells would be also expected to be raised compared to WT cells. This comparison would have to be tested formally using cells isolated from individuals with β-thalassaemia trait, using the same culture and qPCR methods.

Whilst this approach would have therapeutic potential when assessing mRNA, this significant reduction in α-globin expression was not recapitulated when measuring globin chains by RP-HPLC for technical reasons. This is a limitation of the current work that needs addressing before testing it in further pre-clinical or clinical work. However, there is reason to believe that globin chain analysis would correspond to the mRNA results. There is no known feedback-mechanism specific to α-globin that alters the rate of transcription, mRNA degradation or translation. Indeed, mRNA ratios have been shown in multiple independent studies to correspond very well to globin chain ratios and haematological parameters in patients (Chaisue et al., 2007, Higgs, 2009b). Measuring the rate of synthesis rather than accumulated protein levels (as in HPLC) would be more instructive, as the latter may be modified by proteolysis and precipitation.

In order to confirm that a gene-editing strategy is therapeutically applicable, it must be demonstrated that the editing of LT-HSCs at the target locus is possible. The gold-standard assay for pre-clinical work is using serial murine xenotransplantation assays, although one group has recently used non-human primate models (Humbert et al., 2019). The current work did not perform these assays for two related reasons. It has been established beyond doubt that it is possible to robustly edit LT-HSCs using the CRISPR/Cas9 system using a

number of vectors including plasmid, adenovirus, mRNA and RNPs (Dever et al., 2016, Mettananda et al., 2017a, Xu et al., 2017, Wu et al., 2019, Pavel-Dinu et al., 2019, Newby et al., 2021). RNP editing was used in the pre-clinical work that led to the first CRISPR/Cas9 clinical trials for β-thalassaemia and SCD, with adequate engraftment noted in all patients at the time of writing (Frangoul et al., 2021). Furthermore, the work leading up to this project that involved a 3kb deletion of MCS-R2 showed persistent editing in LT-HSCs. It would be extremely unlikely that making a single DSB at a single TFBS would result in poor engraftment of edited cells. Therefore, I did not feel that serial xenotransplantation experiments were warranted in this instance.

An important safety concern with the use of programmable nucleases is OT effects. These were assessed in primary human CD34+ cells using CIRCLE-seq. This OT profile is likely to underrepresent the full scope of potential OT events; the original CIRCLE-seq paper identified between 21 and 124 OT sites when assessing 6 sgRNAs. In our profiling, 9 sites were identified. It may be that the NF-E2 5' sgRNA is particularly specific, but it is more likely that the sequencing needs to be performed at greater depth to capture more uncommon OT events. Once this *in-vitro* screen is performed targeted capture at these identified sites and further deep amplicon sequencing will need to be conducted to assess whether any editing had in fact occurred in primary cells. it is highly likely that the actual *in-vivo* OT editing will be far smaller in magnitude. This is because *in-vitro* assays involve targeting of free DNA which is more accessible to nucleases, whereas cellular DNA is more likely to be inaccessible, particularly when in regions of heterochromatin. Previous work targeting MCS-R2 with two sgRNAs found no detectable OT events using this targeted capture based NGS method for profiling candidate OT sites (Mettananda et al., 2017a).

Whilst this work is pending, orthogonal efforts can try and assess the potential consequences of OT editing. OT editing could theoretically be deleterious to the cell through perturbations in normal gene expression through mutations in gene bodies, promoters or distal regulatory elements. These could cause major changes to cell function, or subtle alterations that may confer a selective advantage to the HSC over time, and thereby result in clonal outgrowth. By their nature, OT assays described here and in 4.1.6 are purely descriptive records, and make no significant predictions about functional effects unless they lie within gene bodies or highly annotated regulatory elements. A key area for research and safety in the clinic will be to arrive at a consensus on what descriptive but also what predictive assays should be performed to keep the risk of deleterious OT events at a minimum. To investigate chromatin changes at OT sites, ATAC-seq was conducted, in erythroid cells, at the sites identified by CIRCLE-seq to see whether editing at OT sites actually resulted in changes which might be associated with perturbed function. No gross alterations in chromatin accessibility at these sites was detected and there were no significant changes in global chromatin accessibility in erythroid cells between control cells and cells edited at the NF-E2 5' site. Interestingly there was a greater variability in chromatin accessibility between individual donors than between control and edited cells. This may make biological sense; whilst the cells were sourced from healthy individuals via NHSBT, donors were anonymous and could have been of either gender, and of ages several decades apart. This may play a more significant role in erythroid function than a theoretically precise guide that targets several base pairs of an enhancer that only acts on the α-globin genes. Whilst these results are reassuring and offer an alternative functional approach to assessing OT effects, additional biological replicates targeting the same site are required to confirm the reproducibility of these findings. In addition, ATAC-seq would also have to be performed

in different haematological cell types, including HSCs, as negative effects of OT events may not be detected in erythroid cells. This latter point is addressed in chapter 5.

An important feature of this study is that it demonstrates that enhancers can be manipulated in a tunable manner to alter the expression of their target gene for therapeutic purposes, as well as deepening our understanding of gene regulation. A targeted knock-down of gene expression via enhancer manipulation has not been attempted in gene-editing strategies to cure haematological or non-haematological genetic disease as far as I am aware; the most prominent example of editing an enhancer for haemoglobinopathies have involved trying to completely abolish BCL11a expression in erythroid cells. This offers an important proof of principle that regulatory elements may be used to manipulate gene expression with granularity.

Editing MCS-R2 addresses how enhancers may work *in-vivo*. There are two prevailing models of enhancer grammar; the enhanceosome, which requires stringent motif identity and arrangement, and the billboard/poster-board model which allows greater flexibility (Long et al., 2016). The evidence presented here suggests that MCS-R2 follows the billboard model, whereby there are a number of different TFBS which all contribute to enhancer function, but aren't individually vital. Unlike with the BCL11a enhancer which has a central GATA1 site crucial for function, none of the TFBSs had the same function here. Although the experiments were preliminary in nature, combinatorial editing either through introducing a DSB or by using base editing seemed to result in an additive effect. When the NF-E2 5' and GATA1 5'2' sites were edited individually a 25% knockdown in α-globin transcription was observed. When both were mutated, expression was reduced to approximately 50-60%

which appears to be additive; consistent with the billboard model of enhancer behaviour. MCS-R2 may have evolved this redundancy in TFBSs as it allows for robust function; if one or even two sites carry homozygous deletions, α-globin transcription can still continue. This is vital as α-globin is a key erythroid gene, and a significant reduction in expression causes disease. In contrast, BCL11a's erythroid enhancer does not require this level of redundancy; if it is mutated and its γ-globin repressing function is disrupted, the outcome is not particularly deleterious. It is interesting to note that to date no SNPs within the core MCS-R2 sequence have been found that result in an alteration in a haematological phenotype, whereas BCL11a was identified as a determinant of HbF through a genome wide association study examining quantitative trait loci in individuals with higher the normal numbers of F cells (erythrocytes containing a measurable about of fetal haemoglobin) (Menzel et al., 2007).

In summary, this work shows that gene expression can be reduced in a graduated manner by rational targeting of key sequences within regulatory elements, and that this is a potential therapeutic strategy for diseases where an 'all-or-nothing' approach in targeting is not appropriate. This approach is a potentially attractive alternative to other treatments, as an adjunct or a stand-alone strategy. Despite these encouraging findings, the safety concerns regarding forming DSBs remain, both for approaches currently in clinical trials and for the targeting of MCS-R2 described here. The next chapter describes the use of base editing; a modular system built upon CRISPR/Cas9 that avoids the creation of DSBs, which may be the technology of choice used in future therapeutic gene editing applications.

5. Developing a base editing strategy to cure HbE/β-thalassaemia

5.1 Introduction

5.1.1 Forms of DNA editing

The creation of double strand breaks (DSBs) with Cas9 to disrupt specific DNA sequences through non-homologous end-joining (NHEJ) has proved to be highly efficient (Vakulskas et al., 2018, Bak et al., 2018, Wu et al., 2019). Editing outcomes are to an extent predictable (van Overbeek et al., 2016, Shen et al., 2018). NHEJ-mediated DSB repair may therefore be mechanistically suitable for certain therapeutic applications if the repair profile is permissive, for example restoring a normal splicing pattern in forms of β-thalassaemia caused by aberrant splice sites (Xu et al., 2019, Patsali et al., 2019). However, this approach is unsuitable for repairing mutations where the repair outcome doesn't resolve the mutation at high efficiency. It is possible to repair mutations using single or double-stranded donor oligo repair templates through homology-directed repair (HDR)(Lin et al., 2014, Paquet et al., 2016). However, this process is far less efficient, particularly in non-cycling haematopoietic stem cells, as HDR preferentially occurs in S and G2 phases of the cell cycle, and competes with the more efficient NHEJ process to repair DSBs (Cox et al., 2015). The act of generating double-strand breaks can lead to large and complex rearrangements, translocations and megabase-level chromosomal truncations and chromothripsis (Zhang et al., 2015, Shin et al., 2017, Kosicki et al., 2018, Adikusuma et al., 2018, Owens et al., 2019, Cullot et al., 2019, Zuccaro et al., 2020, Leibowitz et al., 2021). Beyond these immediate

repair-related effects, CRISPR/Cas9 induced DSBs activate the p53-DNA damage response, potentially selecting for p53 mutated cells, and have also been shown to have general cellular effects such as slowing cell cycle progression (van den Berg et al., 2018, Haapaniemi et al., 2018, Ihry et al., 2018, Schiroli et al., 2019, Enache et al., 2020).

Human genetic disease is caused by an array of different mutations, and an analysis of the ClinVar database has shown that by far the most common variants are due to point mutations (Figure **5.1**A) (Landrum et al., 2016, Rees and Liu, 2018). Unfortunately, the aforementioned DSB-based methods would likely be ineffective at reversing such variants as the majority of repair events would involve indels rather than conversions of one base to another.

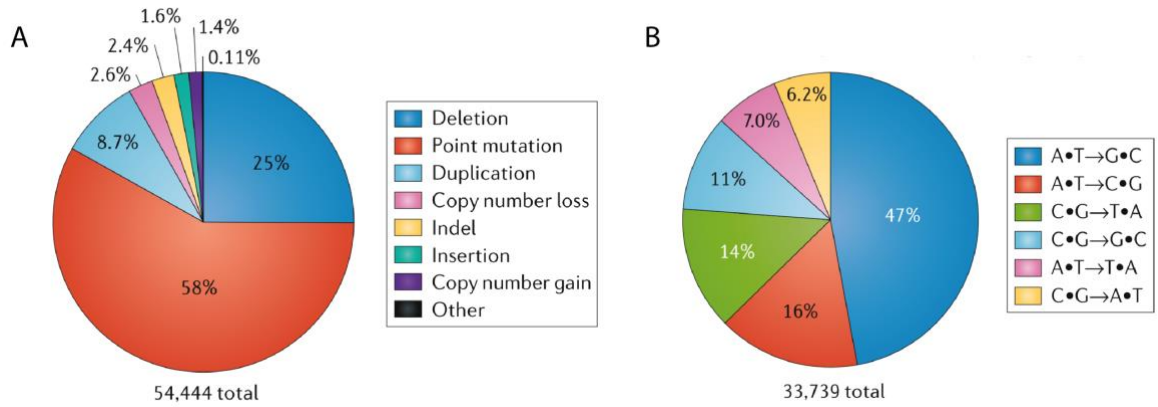


Figure 5.1 – Human pathogenic genetic variants *A) Genetic disease variants from the ClinVar database represented by class of mutation B) Distribution of base pair changes required to repair the pathogenic disease variants back to wild type. In the pie chart transitions are shown in white, transversions are shown in black. Adapted from (Rees and Liu, 2018).*

These concerns were directly addressed by the development of a new class of genome-engineering tools; base editors. Base editors (BEs) are evolved versions of naturally occurring deaminases fused to a targetable cas9-nickase that are able to irreversibly change

one DNA base to another *in-situ* without the need for DSBs. They have been described and characterised in various mammalian species for over two decades and catalyse the cytosine (C) to thymine (T), and adenine (A) to inosine (I) transitions (Harris et al., 2002, Bass, 2002). BEs can edit both RNA and DNA, and have been implicated in processes such as antigen-dependent antibody diversification (AID), innate immunity against invading retroviruses (APOBEC3) and post-transcriptional mRNA processing (APOBEC1, ADARs) (Conticello, 2008, Nishikura, 2010). Given their ability to mutate DNA, the AID and APOBEC homologs have been found to play a role in cancer development and tumour evolution (Okazaki et al., 2003, Burns et al., 2013, Petljak et al., 2019, Roper et al., 2019).

5.1.2 Cytosine base editors (CBEs)

For genome engineering, CBEs were the first to be developed. Variants of these naturally occurring enzymes were added to the cas9 toolkit through the fusion of the base editing protein via a linker sequence to a dead-Cas9 (dCas9), which is a cas9 that has been mutated within each of its DNA-binding grooves to remove its nuclease activity. The BE is directed to a genomic region of interest through binding of a sgRNA (Komor et al., 2016, Nishida et al., 2016). For clarity, the description and work in this chapter only involves base editors developed by Liu and co-workers. CBEs are based on rat APOBEC1 which only recognises single-stranded DNA. Deamination can only occur when an R-loop is formed following local denaturation of DNA upon sgRNA:dCas9 binding (Jore et al., 2011, Jiang et al., 2016).

Cytidine deaminase converts cytosine into uracil (U) through the action of the aforementioned APOBEC1-dCas9 fusion protein, known as BE1. Normally the cell recognises

that this uracil lies opposite a guanine on the non-edited strand, and repairs it back to cytosine through the action of uracil DNA glycosylase, preventing efficient base-editing. Addition of a uracil DNA glycosylase inhibitor (UGI) to the C-terminus of the dCas9 (known as BE2) prevents this process from occurring. The U-G mismatch would normally be repaired by the base-excision pathway, with the non-excised base being used as the template for repair. To improve efficiency by driving repair using uracil as the template, the catalytic His840 residue in the dCas9 was restored to WT so that the non-edited DNA strand held by the HNH domain, which contains the unwanted G is nicked (BE3)(Figure 5.2B). This nick favours the excision to occur on that strand resulting in repair from the edited uracil, resulting in a U-A pairing from an original C-G. Further downstream DNA repair or DNA replication replaces the U with a T, completing the process of converting C-G to T-A.

This first iteration of the CBEs converts C to T within a target window of bases 4-8, where the PAM lies at positions 21-23 (Komor et al., 2016).

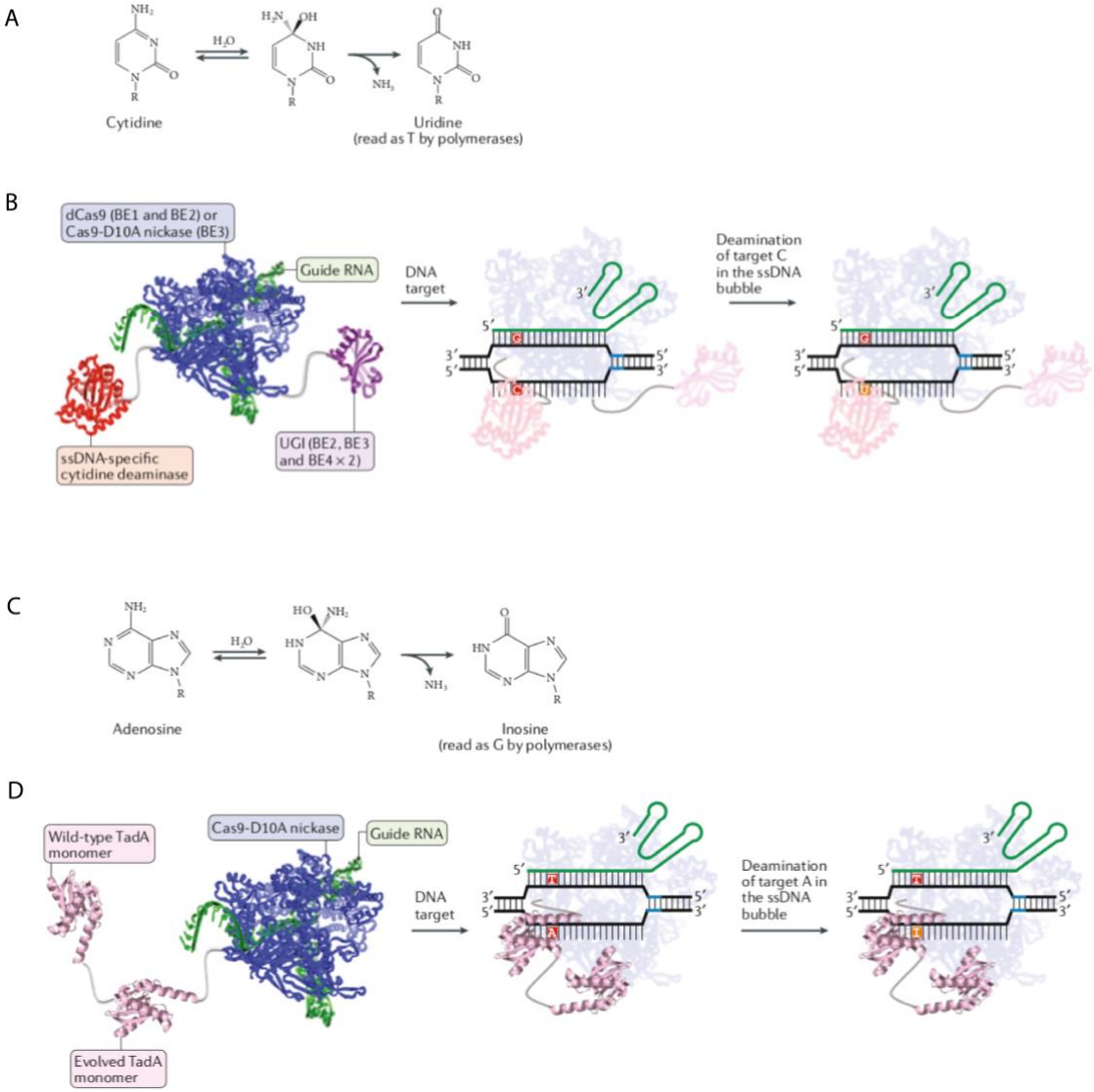


Figure 5.2 - Base Editors A) CBEs deaminate the exocyclic amine on the cytosine base to Uracil B) Composition of BE3 4. C) ABEs deaminate an exocyclic amine on the adenine base to convert it to inosine. D) Composition of ABE7.10. Adapted from (Rees and Liu, 2018)

5.1.3 Adenine base editors (ABEs)

Cytosine deamination occurs at a high rate spontaneously, with around 100-500 cytosines deaminated to thymines per cell every day, resulting in a C-G pairs being frequently converted to a T-A (Lindahl, 1993). Approximately 47% of pathogenic SNPs result from conversions of C-G to A-T pairs (Figure 5.1B). As adenine is also deaminated by naturally occurring enzymes, albeit in tRNA, Liu and co-workers developed an adenine base editor (Gaudelli et al., 2017). An *E. coli* tRNA adenosine deaminase, TadA, was chosen due to its high sequence homology with the APOBEC1 previously used. TadA was fused to dCas9 and evolved to recognise ssDNA, creating TadA*-dCas9. In *E. Coli* tadA functions as a homodimer, so a WT TadA was attached to the evolved TadA*, resulting in a TadA-TadA*-Cas9 nickase fusion construct. The final editor (ABE7.10) efficiently catalysed A to I transitions in mammalian cells, with an editing window of protospacer positions 4-7.

5.1.4 Off-target (OT) base editing

As with any nuclease-based DNA editing platform, the creation of off-target changes (OTs) are a potential confounding factor when using BEs. These may arise due to two mechanisms. The first is dependent on sgRNA-based targeting of the BE to OT loci in the genome. The initial BE reports used sgRNAs that had been validated in early CRISPR/Cas9 experiments, with well characterised OT loci. Whilst at a lower rate than WT Cas9, BEs were still able to cause detectable OT effects when assessed by biased targeted sequencing-based assays (Komor et al., 2016, Gaudelli et al., 2017, Kim et al., 2017a, Rees et al., 2017). The increased fidelity compared to nuclease-based editors that use the same sgRNA is in part because not

every sgRNA has a C or A within the targeting window. Further work has been conducted to perform a more unbiased analysis of OT effects using whole genome sequencing (WGS) based methods such as Digenome-seq and EndoV-seq (Kim et al., 2019, Liang et al., 2019). These have revealed that OT deaminations do occur, but in keeping with previous results, ABEs seem to have a much lower genome-wide OT profile than CBEs (Kim et al., 2017a, Ryu et al., 2018, Liu et al., 2018b, Jin et al., 2019, Zuo et al., 2019, Kim et al., 2019, Liang et al., 2019).

The second mechanism of generating OT effects is via non-sgRNA-dependent processes. As BEs were developed from RNA editing molecules, they can potentially edit ssDNA/RNA in a non-sgRNA-dependent fashion. CBEs, but not ABEs were found to have OT deaminations unrelated to their sgRNAs when whole-genome sequencing was performed (Jin et al., 2019, Zuo et al., 2019). CBEs and to a lesser extent ABEs have also been found to induce sgRNA-independent transcriptome-wide RNA editing (Grunewald et al., 2019a, Zhou et al., 2019, Rees et al., 2019). Through the introduction of structure-guided mutations in deaminase regions CBEs and ABEs have been developed to mitigate the amount of RNA OT events, with significant reductions observed (Rees et al., 2017, Lee et al., 2018).

5.1.5 Editing windows

BEs have an editing window (approximately between bases 4-8) within which bases may be deaminated. This raises several issues. Firstly, multiple editable bases may be present in the window, leading to bystander conversions. This is problematic when editing coding regions, as non-synonymous mutations can lead to deleterious amino acid changes. Whilst some

editors have been developed with narrower editing windows (as narrow as 1-2 bases) they will not be applicable to all editing contexts (Kim et al., 2017b). Secondly, even if the bases around the target are not susceptible to editing, their sequence context still exerts an influence on editing outcomes. This is particularly the case for CBEs, where the base immediately 5' to the target influences how well it will be edited (Komor et al., 2016, Kang et al., 2018). Interestingly this has been used to mitigate bystander edits; BEs have been developed that only efficiently edit cytosines when preceded by a thymine (Gehrke et al., 2018). The sequence context also extends to epigenetic effects; the original APOBEC CBEs edited poorly at methylated CpG dinucleotides, which led to the adoption of a different APOBEC for ncreasing the genome-targeting scope and precision of base editing with engineered GC-rich regions (Wang et al., 2018).

5.1.6 Base editing and the HbE/β-Thalassaemias

Haemoglobin E/β-thalassaemia causes around 50% of all severe thalassaemia worldwide; equating to around 30,000 births per year (Modell and Darlison, 2008). HbE is caused by a point mutation in codon 26 of the human *HBB* gene (GAG; glutamate → GAA; lysine). HbE is not only a structural mutant but it also reduces production of β-globin chains, acting as a β-thalassaemia allele. When inherited in combination with other β-thalassaemia mutations on the other allele, severe HbE\β-thalassaemia can develop (Fucharoen and Weatherall, 2012). Current gene therapy developed for haemoglobinopathies using lentiviruses to introduce normal β-globin cassettes into the genome have significant safety concerns because the constructs are integrated randomly into the genome at thousands of different sites (Ribeil et al., 2017). This leads to the potential to cause malignancy as a result of

insertional mutagenesis. Several methods have been described for genome editing for the treatment of β-thalassaemia. The most promising approach involves reactivation of fetal haemoglobin expression (HbF) either through mutagenesis of the promoters of the HBG genes or the erythroid enhancer of BCL11A which leads to de-repression of the γ-globin genes and increased levels of HbF ($\alpha_2\gamma_2$)(Traxler et al., 2016, Weber et al., 2020, Metais et al., 2019, Humbert et al., 2019, Frangoul et al., 2021). These methods all rely upon gene editing with DSB formation, which come with all the aforementioned risks. Base editing has also been used to edit the BCL11a enhancer and a specific β-thalassaemia promoter mutation (Gehrke et al., 2018, Zeng et al., 2020). However, editing efficiencies using these approaches has been far more modest than the equivalent strategies using NHEJ. Zeng and colleagues attempted to circumvent this by electroporating CD34+ cells twice in 24 hours which did improve editing efficiencies but is unlikely to be clinically appropriate. Even with this intervention it has been challenging to raise HbF levels above 30%, well below levels produced by the NHEJ based approach, and the level needed to result in an asymptomatic phenotype in a patient with β-thalassaemia (Musallam et al., 2012a, Wu et al., 2019, Zeng et al., 2020). Chapter 4 describes an approach that involves editing the main enhancer at the α-globin locus to reduce its expression. This ameliorates HbE/β-thalassaemia by restoring globin chain imbalance and reduces the toxicity caused by excess α-globin chains. Although this is a key component of the pathophysiology in β-thalassaemia it may not always result in a normal phenotype.

An alternative approach is to use base editing to reverse the HbE G>A mutation that is the variant that causes HbE/β-thalassaemia. As β-thalassaemia is a classic autosomal recessive disorder, repairing a mutation on one allele should have a potent therapeutic effect. Such

individuals who have 'β-thalassaemia trait' have a normal or very mildly lowered haemoglobin concentration, mild microcytosis and hypochromia, and are phenotypically normal (Weatherall and Olivieri, 2009).

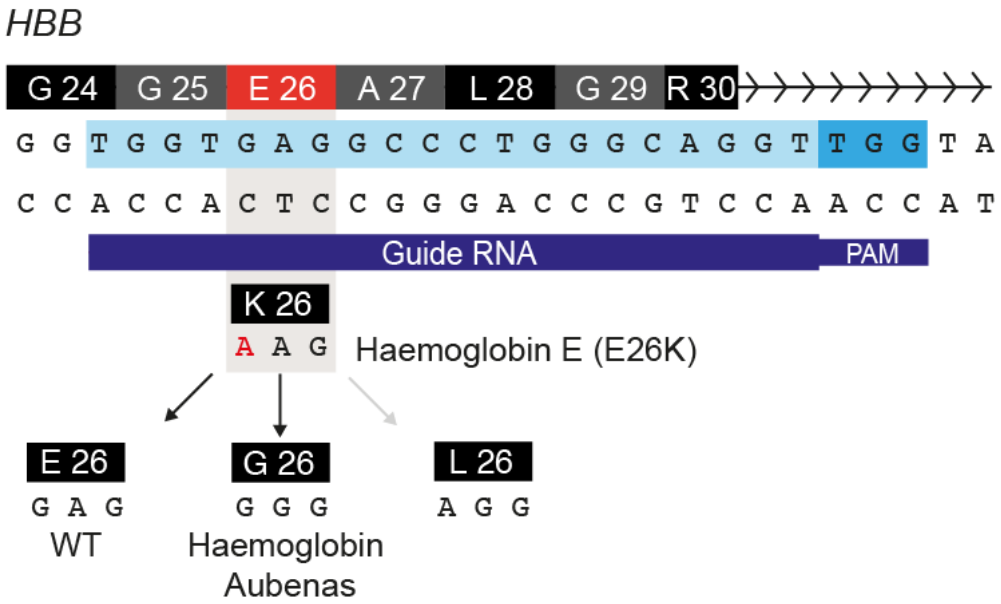


Figure 5.3 - Base editing strategy *The WT sequence is given in the figure, with codon 26 highlighted in red. The targeting sgRNA to repair the HbE (AAG) mutation is shown below the bottom strand of the HBB gene sequence. There are four potential editing outcomes shown: editing to GAG (WT), failed editing with the codon remaining as AAG, repair to GGG (encoding $\beta^{Aubenas}$) or to the unknown GGG codon.*

Codon 26 can be targeted by a sgRNA using a canonical TGG PAM (Figure 5.3). The target A is at position 5 (hereafter referred to as A_5), which is potentially editable by adenine base editors as it lies within the ABE targeting window. A potential problem is bystander editing of the adjacent A at position 6 (A_6), which would change the codon to GGG. This would change the HbE codon that codes for lysine to glycine, instead of the wild-type glutamic acid. Codon 26 in the β-globin chain lies close to the α:β interface, and so bystander editing of A_6 could potentially lead to deleterious effects. This β-globin variant causes an extremely rare haemoglobin variant in the tetramer form called Hb Aubenas (Lacan et al., 1996). A

single individual has been described with this mutation in the heterozygous state who was phenotypically normal, having haematological parameters completely in the normal range and no clinical features of haematological disease. Even though the mutation is in the same codon as HbE, the novel splice site that causes the reduced synthesis of β^E is absent and thus the β-globin biosynthesis rate was found to be normal, 0.96 vs 1.06 in the control. Whilst isopropanol stability testing showed that Hb Aubenas is mildly unstable compared to HbA, this doesn't lead to a haemolytic syndrome. Another potential outcome of editing this region is that A_5 left unedited and only A_6 is deaminated, resulting in an AGG codon which would result in a lysine to arginine conversion. This codon has not been described in published literature or on the HbVar database (last accessed 18th April 2021)(Giardine et al., 2014).

The aim of this chapter of the thesis is to explore the possibility of using base editing to develop a therapeutic strategy to cure HbE/β-thalassaemia, repairing the mutation to wild-type or a normal variant and replicating the asymptomatic β-thalassaemia trait genotype and phenotype. Base editors were first tested in human cell lines and then human CD34+ cells. Different vector modalities were tested to develop a clinically appropriate delivery protocol, which was then used to edit LT-HSCs as assessed by serial xenograft transplantation assays. Finally, the off-target profile of base editing experiments was assessed using biased and unbiased assays; a key safety step needed before proceeding with human clinical studies.

5.2 Results

5.2.1 Plasmid-based editing in HEK293T cells

In the original report, base editing was performed through the lipofection of plasmids containing the BE and sgRNA into HEK293T cells. This protocol was therefore replicated as closely as possible. The initial step involved designing a sgRNA to target the mutated adenine (A_5) in *HBB* in codon 26. The WT codon is GAG and the mutated codon is AAG. In the initial experiments, WT HEK293T cells were used, so A_6 alone was targeted. The editing window for the initial iteration of ABEs is 4-8 in the protospacer (where the PAM is at positions 21-23), so this base was suitable for editing. The only contemporary ABE available used a S. pyogenes dCas9-nickase and so there was only one potential PAM available. Two sgRNAs were created which used this PAM but had different lengths to cater for described differences in transcription of sgRNAs from plasmids depending on the length and first base of the protospacer (Ran et al., 2013). RNA Pol III prefers to commence transcription with a G, so this was added to one of the sequences (protospacer sequence in capitals, added G in red and target A in bold, sgRNA1 - caccgTGGTG**A**GGCCCTGGGCAGGT). Truncated sgRNAs have also been described as being effective, so the other sgRNA was shortened (sgRNA2 - caccgGGTG**A**GGCCCTGGGCAGGT). Cells were lipofected with two plasmids, one containing the base editor construct and another which contained independent sgRNA expression and GFP cassettes. GFP positivity was in excess of 80% and so cells were not sorted (Figure **5.4**A). Once the bulk transfected cells had grown for 72 hours they were harvested, genomic DNA extracted using locus specific primers for HBB and Sanger sequenced (Figure **5.4**B).

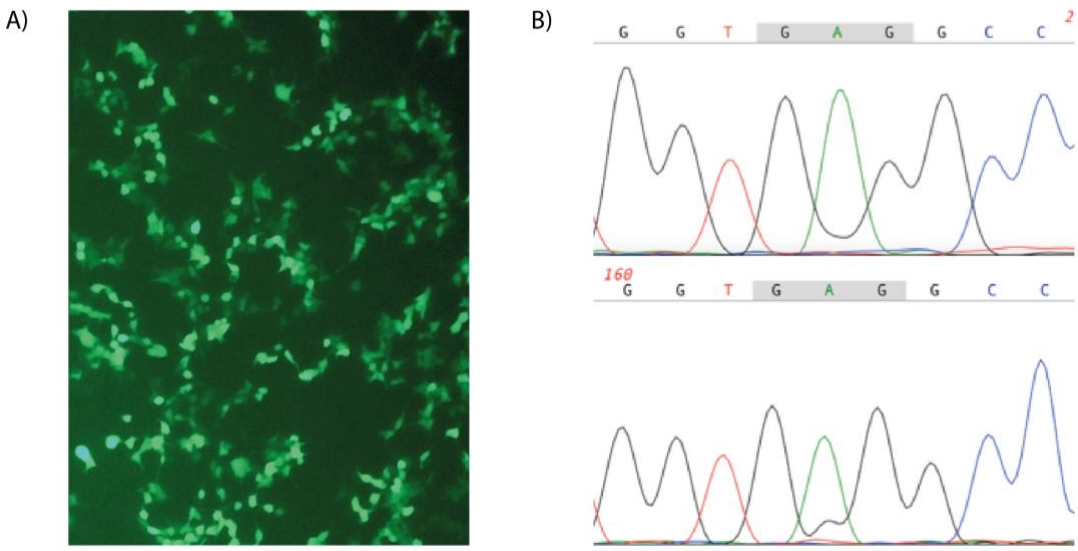


Figure 5.4 - Base editing in HEK293T cells A) Representative inverted microscope images showing lipofected HEK293T cells at 24h post-lipofection. B) Representative Sanger sequencing results from edited HEK293T cells.

Editing efficiency in these initial experiments as assessed by amplicon-based NGS analysis was encouraging, with an average base conversion efficiency of 12.4% (SD±4.28%). SgRNA1, which was not truncated, edited more efficiently than sgRNA2.

5.2.2 Plasmid-based editing in HUDEP-2 cells

Whilst the results in HEK293T cells were encouraging, editing in a more appropriate cell type was required and so these sgRNAs were tested in the erythroid cell line HUDEP-2 (Kurita et al., 2013). HUDEP-2 cells, like CD34+ cells are resistant to lipofection and have been used extensively in work aimed at ultimately editing CD34+ cells (Canver et al., 2015, Traxler et al., 2016, Grevet et al., 2018). Both sgRNAs described above were tested. HUDEP-2 cells were transfected using electroporation of the same plasmids used in HEK293T cells. Once again, sgRNA1 performed better than sgRNA2, and transfecting 10µg of total plasmid

yielded better results than 5µg (Figure **5.5**A, C). Editing results were confirmed by targeted NGS (Figure **5.5**B).

The initial HEK293T experiments used the initial iteration of the ABEs, namely dCas9-ABE7.10. Novel editing constructs were developed which involved pairing the same ABE7.10 to xCas9, a form of cas9 that had been evolved to have more permissive PAM requirements (Hu et al., 2018). In HUDEP-2 cells dxCas9-ABE7.10 was able to edit A_6 more efficiently than dCas9-ABE7.10 had in HEK293Ts (Figure **5.5**D). This is unlikely to have been due to enhanced activity due to the construct, as the ABE was the same. It is more likely a cell type effect, possibly due to chromatin in the β-globin locus being more accessible in HUDEP-2 cells.

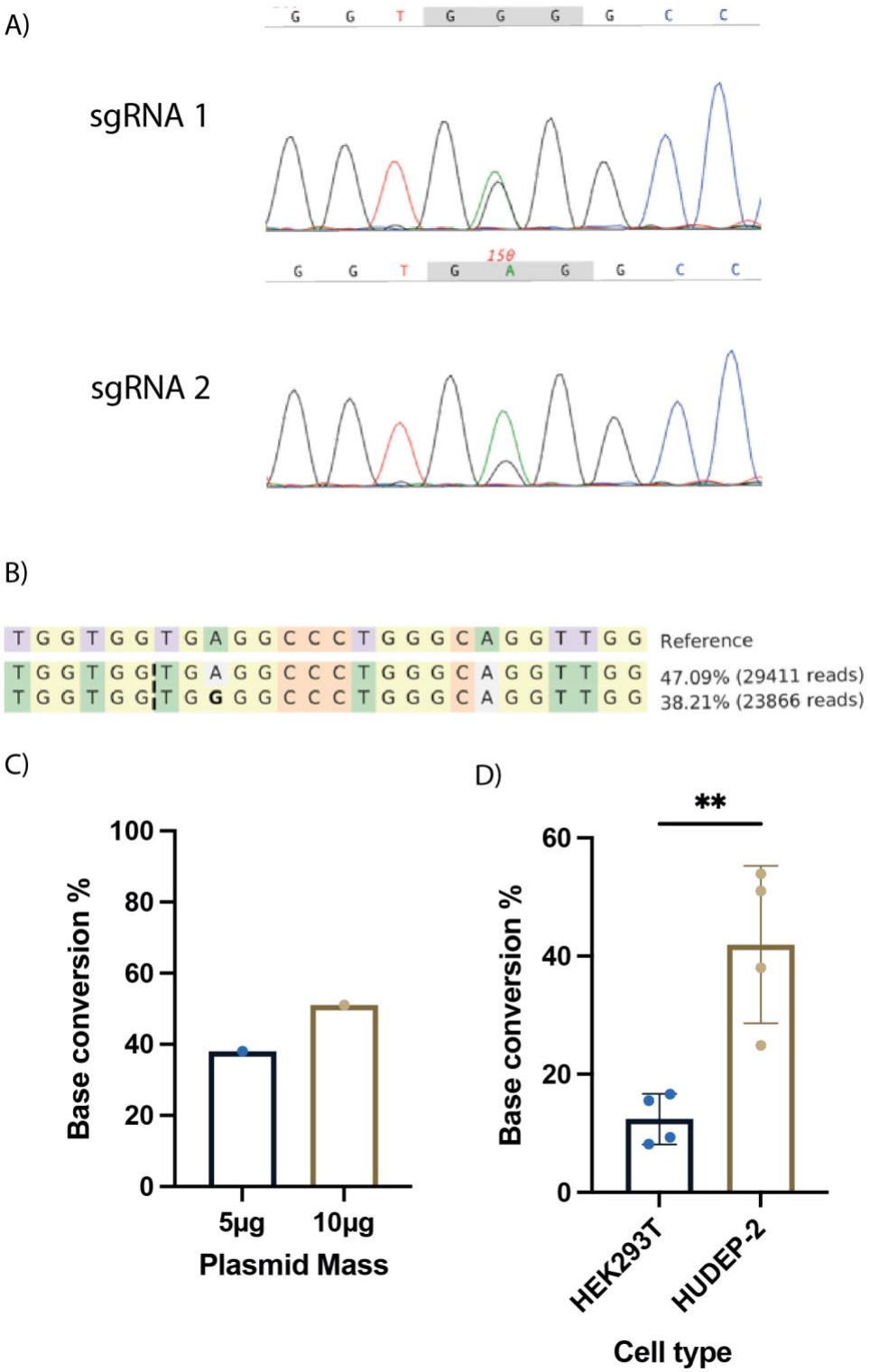


Figure 5.5 - Base editing in HUDEP-2 cells A) Representative Sanger sequencing data showing sgRNA 1 performs better than sgRNA2 in HUDEP-2 cells B) Editing results for sgRNA1 confirmed by NGS C) Transfecting 10µg of plasmid gives higher editing efficiency than 5µg D) Base editing A_6 is more efficient in HUDEP-2 cells as measured by NGS. Significance calculated by unpaired Mann-Whitney U-test.

5.2.3 Plasmid-based editing in wild type CD34+ cells

The ability of sgRNA1 to deaminate A_6 was tested in human haematopoietic stem and progenitor cells (HSPCs). CD34+ cells were transfected with xCas9-ABE7.10 and sorted for GFP positivity. After 3-5 days in culture, the cells were harvested and genotyped. Despite HUDEP-2 cells having similar transfection properties as CD34s, editing results in CD34+ cells were repeatedly poor (mean efficiency 0.81%). This was despite sequencing being performed in selected cells with good viability (Figure 5.6A), using an editing protocol previously shown to be effective when transfecting plasmids into CD34+ cells (Mettananda 2017).

Due to concerns about whether the vector was the limiting factor, transfection of mRNA coding for the ABE and sgRNA instead of plasmid was attempted. The ABE7.10 plasmid was linearised and a full-length PolyA-capped mRNA molecule generated using an *in-vitro* transcription kit, and a synthetic sgRNA bought from a commercial vendor (Synthego). Despite multiple attempts, editing efficiencies were below the limit of detection. Codon usage was also considered as a possible reason for the editing results. No data existed at this point on base editing in CD34+ cells and so a recently developed novel base editor, ABEmax, was used to test this hypothesis. ABEmax was developed using 1) addition of a bipartite nuclear localisation signal to the N and C termini of the base editor and 2) different codon usages to develop base editors with improved cellular expression (Koblan et al., 2018).The report describing these new iterations of ABEs found that poor expression was the cause for reduced editing efficiencies in human cells. The use of ABEmax in human CD34+ cells to target *HBB* immediately confirmed that this was the case, as editing

efficiencies were much improved (Figure 5.6B). On average the A_6 was deaminated with 48.3% efficiency (SD±6.8%).

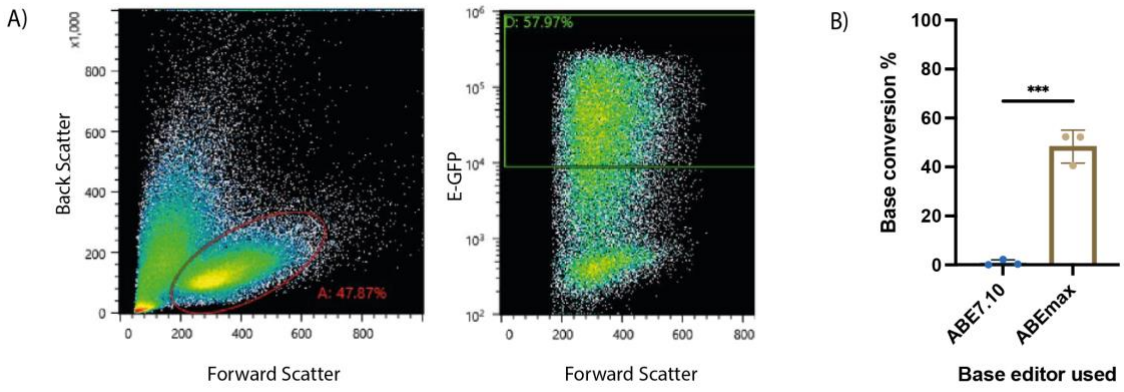


Figure 5.6 - Base editing in WT CD34+ cells *A) Representative flow cytometry data, showing that CD34+ cells edited with ABE7.10 had good viability and GFP positivity. The green box on the panel to the right shows the cells that were sorted by GFP positivity B) ABEmax was able to edit A_6 more efficiently than ABE7.10.*

5.2.4 Plasmid based editing in HUDEP-2 HbE cells

The work described thus far involved cells with a WT *HBB* gene, either in immortalised or primary cells. Base editing in cells with the HbE mutation was then tested. There were two unknown variables. The pathogenic mutation was at A_5, whereas the optimisation studies assessed deamination at A_6. Whilst A_5 was within the targeting window, it wasn't clear how well it would edit. The other open question was how the two adenines in the HbE codon would be edited relative to one another. If they were deaminated at similar levels, or if the A_5 was edited more than A_6, then the codon outcomes would mainly be to WT or Hb Aubenas, i.e a normal variant. This would be the optimal outcome. However, if the A_6 was

preferentially edited there may be a substantial number of codons converted to AGG which is an undescribed variant and potentially damaging.

A HUDEP-2 cell line was created which carried the HbE mutation in homozygosity (Figure 5.7A). This line was then transfected with ABE7.10 and ABEmax plasmid. The target adenine (A_5) was edited very efficiently, up to nearly 100% in many cases, and was consistently edited more efficiently than A_6 (Figure 5.7B). ABEs, unlike CBEs, are not known to be processive with identical adjacent bases, so this may reflect the more advantageous spatial position of A_5. As expected, ABEmax was a more efficient editor than ABE7.10. With the increased activity of ABEmax, when editing at A_5 nears 100% and these adenines have mostly been deaminated, this activity seems to be transferred to A_6. This is reflected in the codon outcomes as identified through NGS (Figure 5.7C). The lower activity of ABE7.10 doesn't edit A_5 and therefore A_6 as much, so the majority of the edited codons are left as WT (GAG), instead of Aubenas (GGG). When edited with the more active ABEmax, this ratio is reversed, with more reads left as Aubenas than as WT.

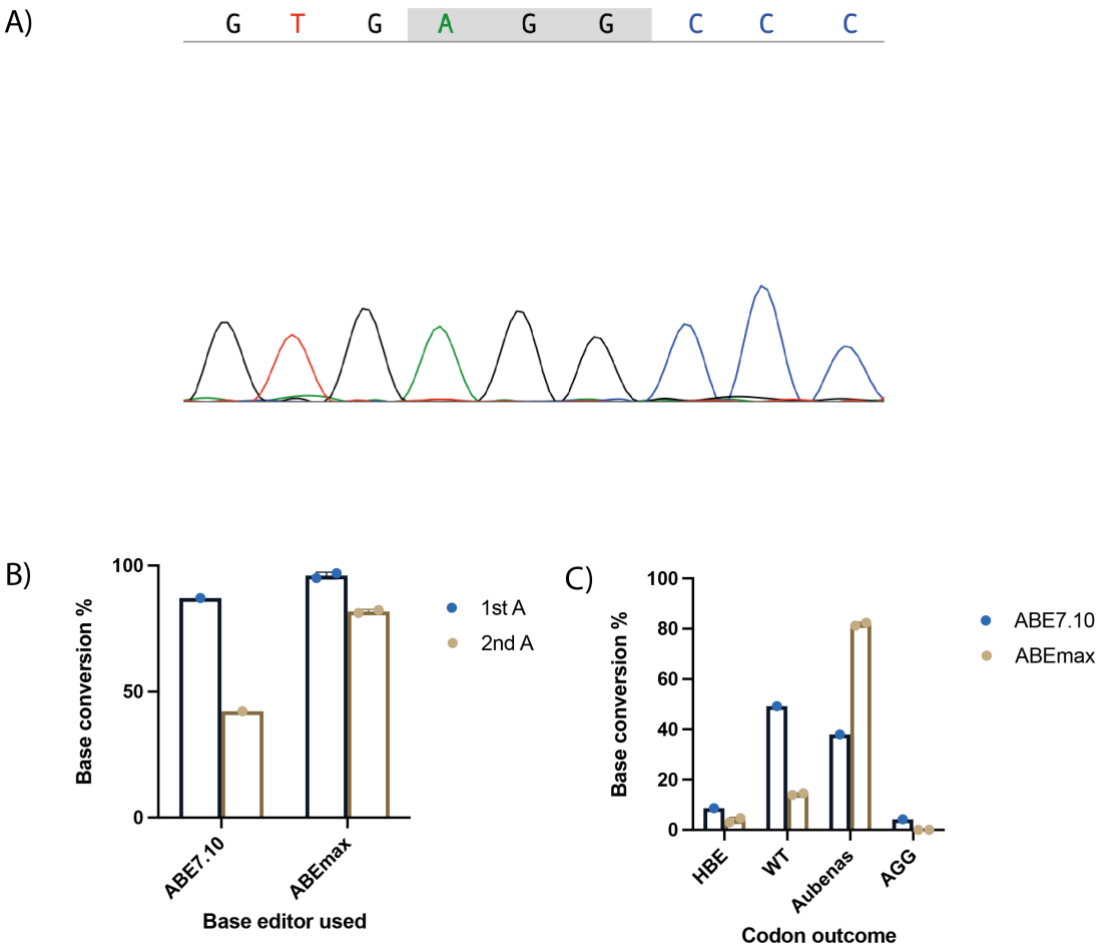


Figure 5.7 - Base editing a HUDEP-2 HbE cell-line *A) Sanger sequencing showing the HUDEP-2 HbE line carrying the mutated codon 26 AAG in homozygosity B) ABE7.10 and ABEmax both have high on-target efficiency, and varying bystander editing efficiency of the second A in codon 26. C) Codon outcomes as identified by NGS when editing the HbE line with ABE7.10 and ABEmax. ABEmax has higher on-target and bystander editing, resulting in a higher number of reads with the Aubenas codon.*

5.2.5 Plasmid-based editing in HbE/β-Thalassaemia patient CD34+ cells

CD34+ cells from patients with different β^{E}/β^{0}-thalassaemia genotypes were derived from peripheral blood. These were transfected with an ABEmax plasmid and a custom-designed minimal plasmid containing a HbE-targeting sgRNA and GFP marker. Cells were selected 24h after transfection for GFP positivity and differentiated along the erythroid lineage.

As with previous work in the HUDEP-2 line, the A_5 was edited more efficiently (Figure 5.8A). NGS using 500bp reads allowed deconvolution of bulk samples to assess the exact breakdown of alleles after editing. As one would expect approximately 50% of the reads were comprised of the thalassaemic allele. The other 50% were split between unedited HbE codons, codons converted to WT, to the Hb Aubenas codon and the undescribed AGG codon (Figure 5.8B). A closer look at the editing outcomes of the HbE allele alone showed an average 76% (SD±12.07) conversion to a normal codon (Figure 5.8C).

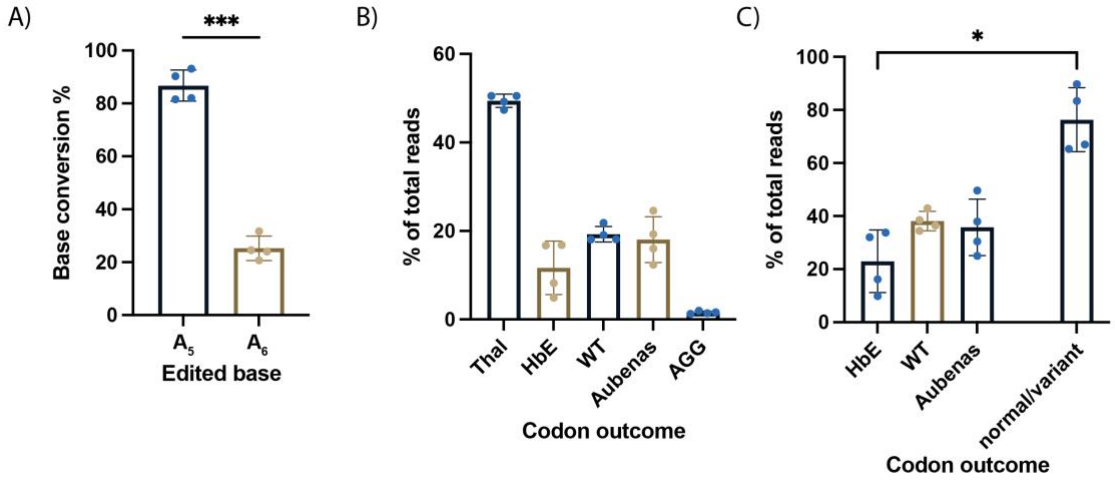


Figure 5.8 - Base editing in HbE/β-thalassaemia patient cells with plasmid A) Conversion of A_5 and A_6 in codon 26 to guanine B) Breakdown of codon outcomes. C) Codon outcomes on the HbE allele alone. The majority of codons were converted to normal or a normal variant.

The non-transfection control and edited cells were compared using microscopy and immunophenotyping to ascertain whether editing could result in perturbations in differentiation. There was no morphologic difference between edited and non-edited cells at day 10 of differentiation (Figure 5.9A). Similarly, there was no difference in expression of key erythroid cell surface markers including CD235, CD71, Band III and CD49D as assessed by flow cytometry (Figure 5.9B).

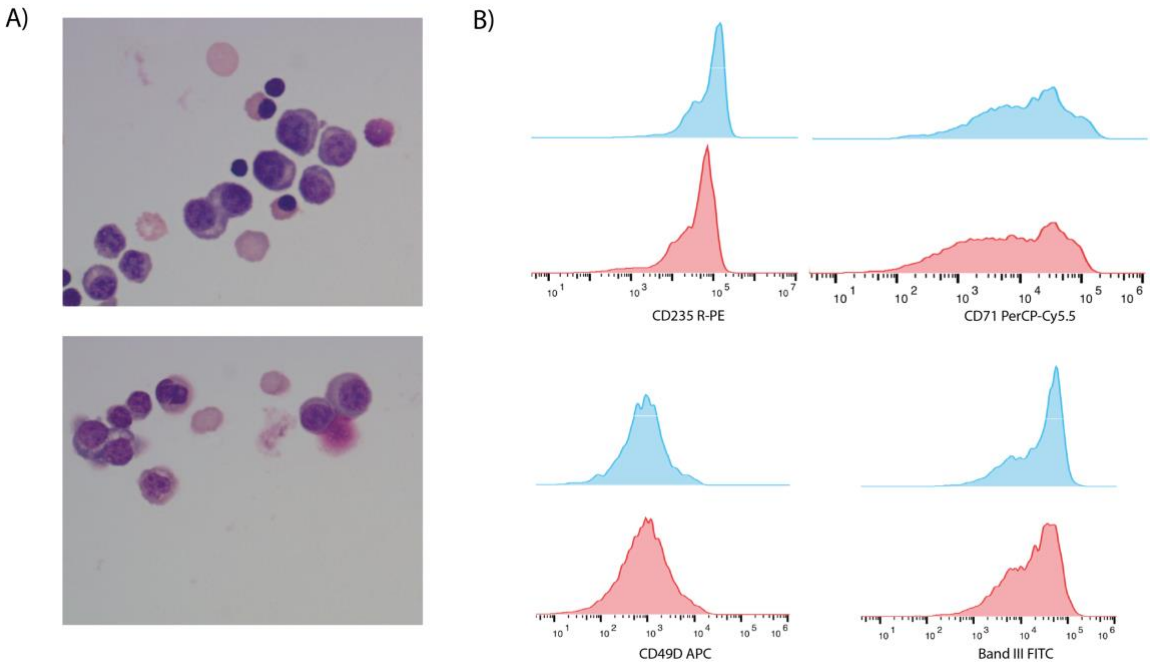


Figure 5.9 - Normal differentiation of HbE/β-thalassaemia patient cells edited by plasmid A) Representative cytospins at D10 of differentiation comparing WT (top) to edited patient sample (bottom) B) Flow cytometry data for key erythroid immunophenotypic markers. Data in red shows control, data in blue shows edited sample.

Gene expression in bulk edited cells was assessed by RT-qPCR. These patients were known to express high amounts of γ-globin, and so for clarity globin mRNA ratios are presented as α/β ratios. As β-thalassaemia is caused by an excess in α-globin chains, the desired therapeutic outcome would be to see a fall in the α/β-ratio. Indeed, there was a decrease in the α/β globin ratio in samples from patient 1 and 2, but not 3 (Figure 5.10A&B). To compare across samples, α/β-ratios in non-edited HbE/β-thalassaemia samples were normalised to 1, and so across all three samples the average reduction of the ratio was to 0.71. The largest reduction was in the sample with the most efficient editing (0.47).

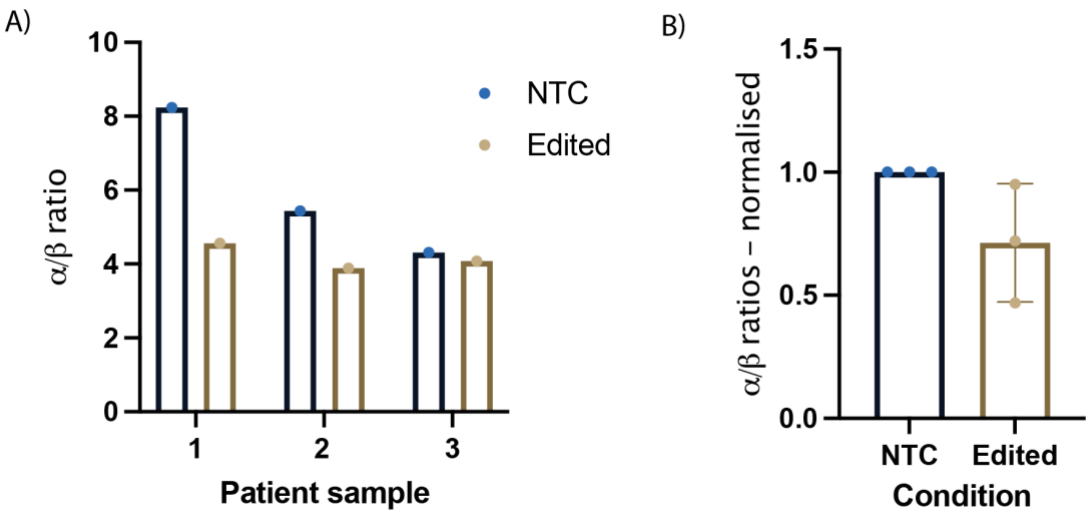


Figure 5.10 - Globin gene expression in plasmid-edited patient cells A) Raw RT-qPCR values for the three patients tested B) Normalised α/β ratios for all patients, showing a mean reduction in the ratio to 0.71. Non-edited patient cells normalised to 1.

As the pathophysiology of β-thalassaemia is caused by precipitation of excess alpha globin chains, protein analysis was conducted. Isoelectric focusing (IEF) - a semi quantitative method - showed that for the cells taken from patient FH, there was a noticeable decrease in the amount of HbE and HbF, and a concomitant increase in the amount of HbA and Hb Aubenas (Figure 5.11).

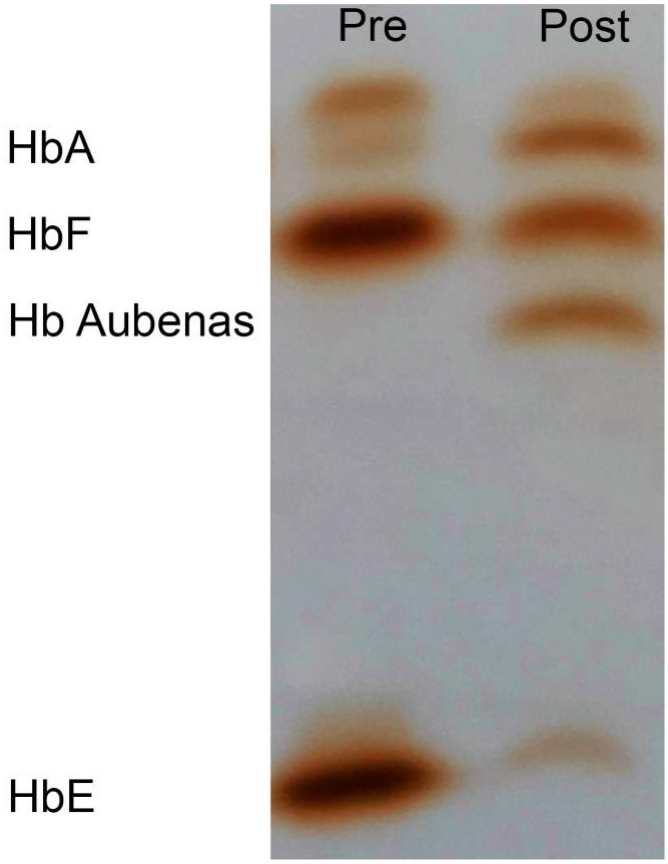


Figure 5.11 - Isoelectric focusing *Isoelectric focusing gel showing haemoglobins in the non-transfected control (pre) and edited (post) sample.*

Globin chain analysis was more quantitatively assessed using cation exchange high performance liquid chromatography (CE-HPLC). As expected, approximately half of the haemoglobin molecules in the unedited control were HbF. The remainder was detected in the HbE window, which is where HbA_2 also runs (Figure 5.12A&B). In the edited sample, the amount of HbF was largely similar, but there was an almost 10-fold increase in the amount of HbA and Hb Aubenas (the latter seems to run in the same window as HbA) and a 2-fold decrease in HbE (Figure 5.12C and D).

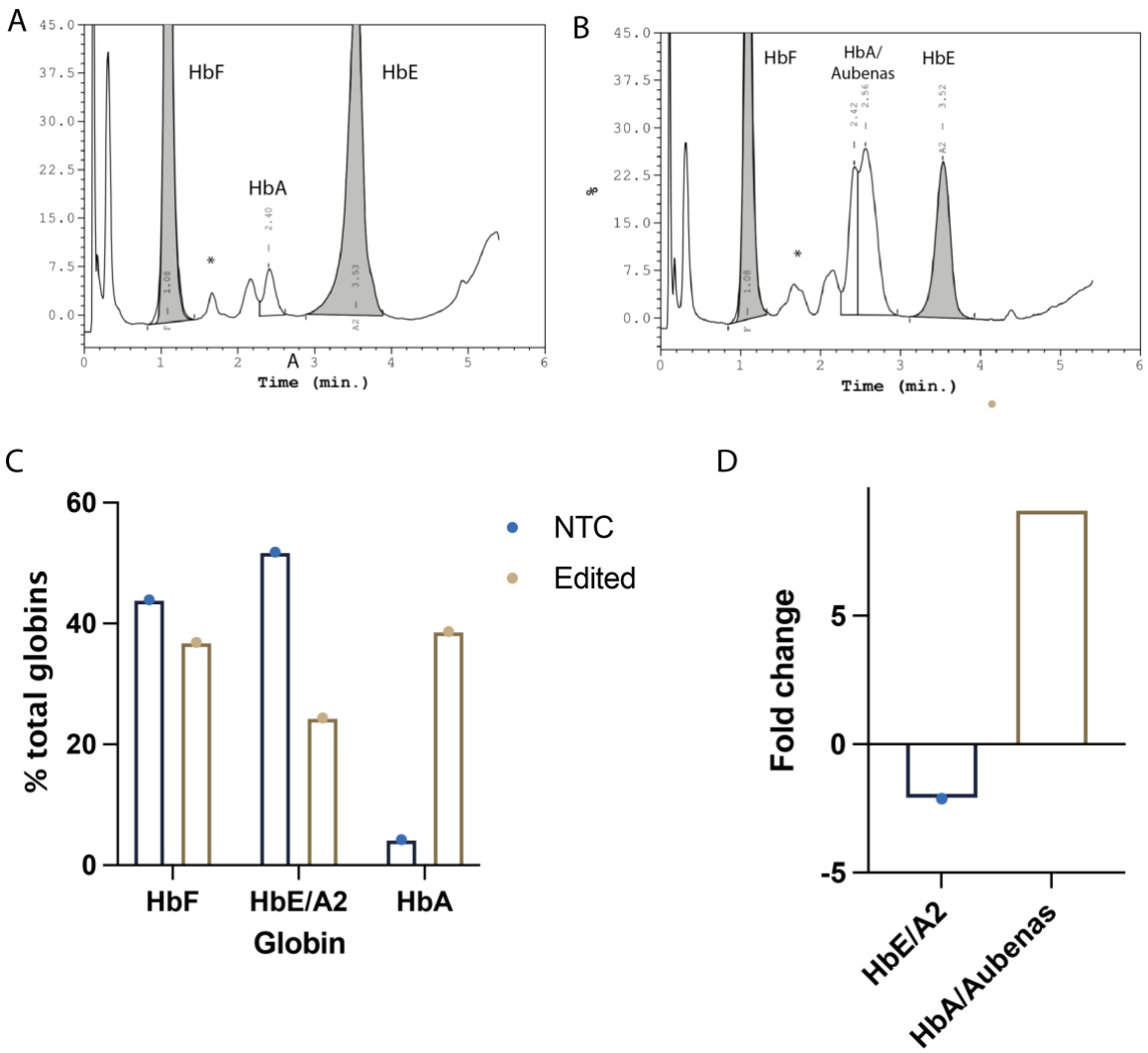


Figure 5.12 - CE-HPLC from plasmid-based editing *Quantitative globin chain analysis of unedited control (A) and edited (B) samples. * denotes a degradation peak due to freeze-thawing of samples. C) Haemoglobin molecules detected by CE-HPLC in control and edited samples, expressed as a percentage of total haemoglobin molecules. D) Relative fold changes in HbE/A_2 and normal haemoglobin molecules in edited samples compared to non-transfected control.*

5.2.6 mRNA-based editing in wild type CD34+ cells

These data provide compelling evidence that efficient editing of the HbE allele in HbE/β-thalassaemia patients is possible, with a concomitant improvement in normal haemoglobin expression. However, plasmid-based editing requiring selection via a fluorescent reporter is unlikely to be clinically appropriate. Unlike with WT Cas9, BE-RNP was not commercially

available. Due to the success in using cas9-RNPs to edit MCS-R2, attempts to produce base editor RNPs were made using a previously described protocol (Kim et al., 2019). Unfortunately, despite repeated attempts with a sgRNA known to cut efficiently *in-vitro* with WT-cas9 RNP, no base editing was detected with either purified CBEs or ABEs. To circumvent this problem, full length ABE mRNA was produced that could be co-transfected with a sgRNA. (Figure 5.13A and B).

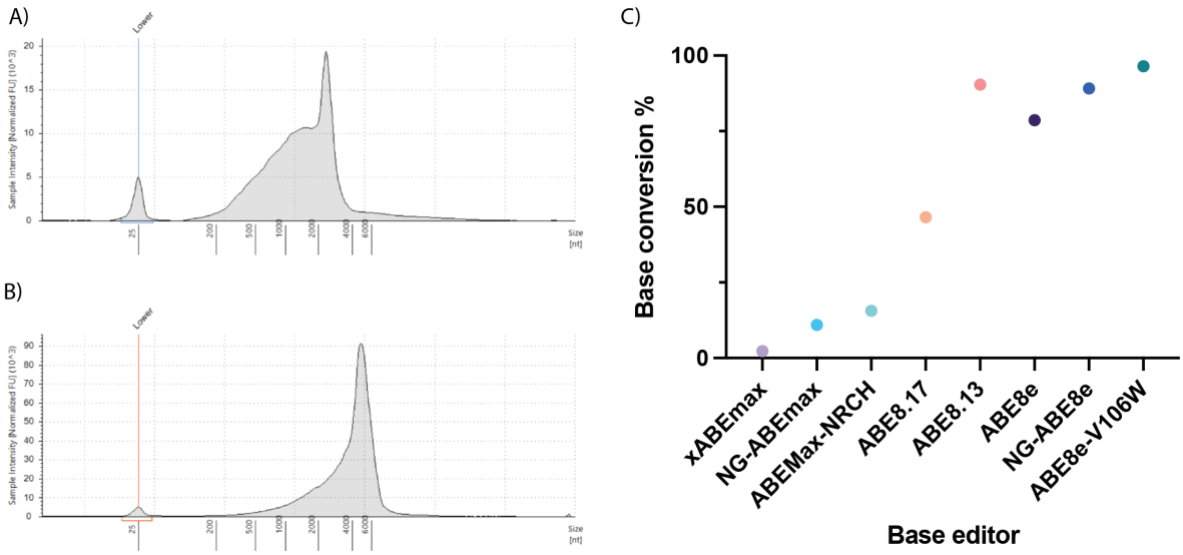


Figure 5.13 - in-vitro transcription of ABE mRNA *Representative Tapestation D5000 traces of ABE mRNA production A) prior to PolyA tail addition B) after PolyA tail addition. End products were the expected size. C) Editing efficiency of the 8th generation ABEs when delivered as mRNA. This optimisation experiment was performed with a single biological replicate but the results were consistent with later work, as shown in Figure 5.14A.*

The editors used for transcription were from a recently described suite of ABEs broadly described as 8th generation ABEs (Richter et al., 2020, Miller et al., 2020, Gaudelli et al., 2020). ABE8e has been shown to catalyse deamination approximately 1100-fold more efficiently than earlier ABEs (Lapinaite et al., 2020). WT CD34+ cells were edited with a variety of ABEs using a similar setup as with cas9-RNP transfection, and their ability to edit

A_6 was compared (Figure 5.13C). As these cells were WT, only deaminations at A_6 could be assessed. Most of the ABE8 editors worked well, with ABE8e (TadA-8e V106W) editing especially efficiently. V106W has a directed mutation in the TadA domain, which is reported to reduce its OT effects whilst maintaining high on-target efficiency (Richter et al., 2020). Due to this favourable profile, ABE8e (TadA-8e V106W) - referred to here as V106W - was selected for further work in patient cells.

5.2.7 mRNA-based editing in HbE/β-Thalassaemia patient derived CD34+ cells

CD34+ cells from three patients with varying β^E/β^0-thalassaemia genotypes were isolated from peripheral blood (Patients S, K and F). These were transfected with V106W mRNA and the HbE-mutation targeting sgRNA delivered as a synthetic sgRNA. Post-transfection, cells were not selected prior to differentiation along the erythroid lineage. For one of the donors, the best performing ABE8s were re-tested, to confirm V106W's performance. Once again, all the editors exhibited high on-target efficiency, and V106W had the highest base conversion detected (Figure 5.14A) The pattern of editing was the same as with the HbE HUDEP-2 line and plasmid based CD34+ editing, in that A5 was edited much more efficiently than A6. There was an efficient conversion of HbE codons to either WT or the Hb Aubenas codon (Figure 5.14B and C). As with plasmid editing, the majority of the codons were converted to Hb Aubenas rather than WT, but notably in one donor's cells this pattern was reversed (Figure 5.14C). Overall, there was a mean 86.17% editing efficiency on the HbE allele, with conversion efficiencies as high as 97.18% (Figure 5.14D). Despite deep targeted NGS of the on-target locus, indels were not detected in any sample and no A to non-G edits were found.

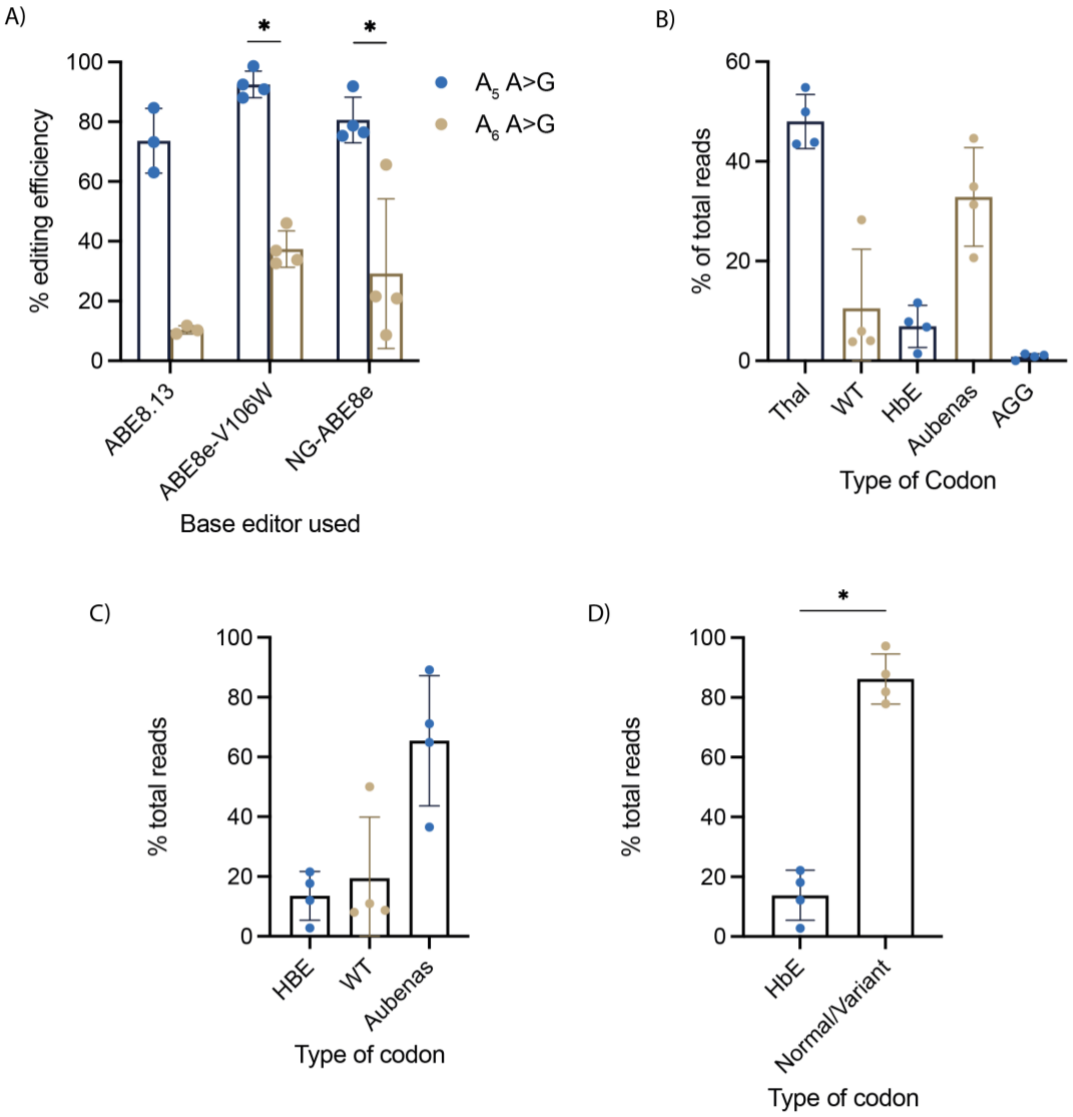


*Figure 5.14 - Base editing outcomes in HbE/β-thalassaemia patient cells with mRNA A) The most efficient editor was V106W, and A_5 was consistently edited more efficiently than A_6. B) Range of total codon outcomes C and D) Range of outcomes on the HbE allele. All data generated using NGS. Significance calculated using Mann Whitney U-test, P value * = <0.05*

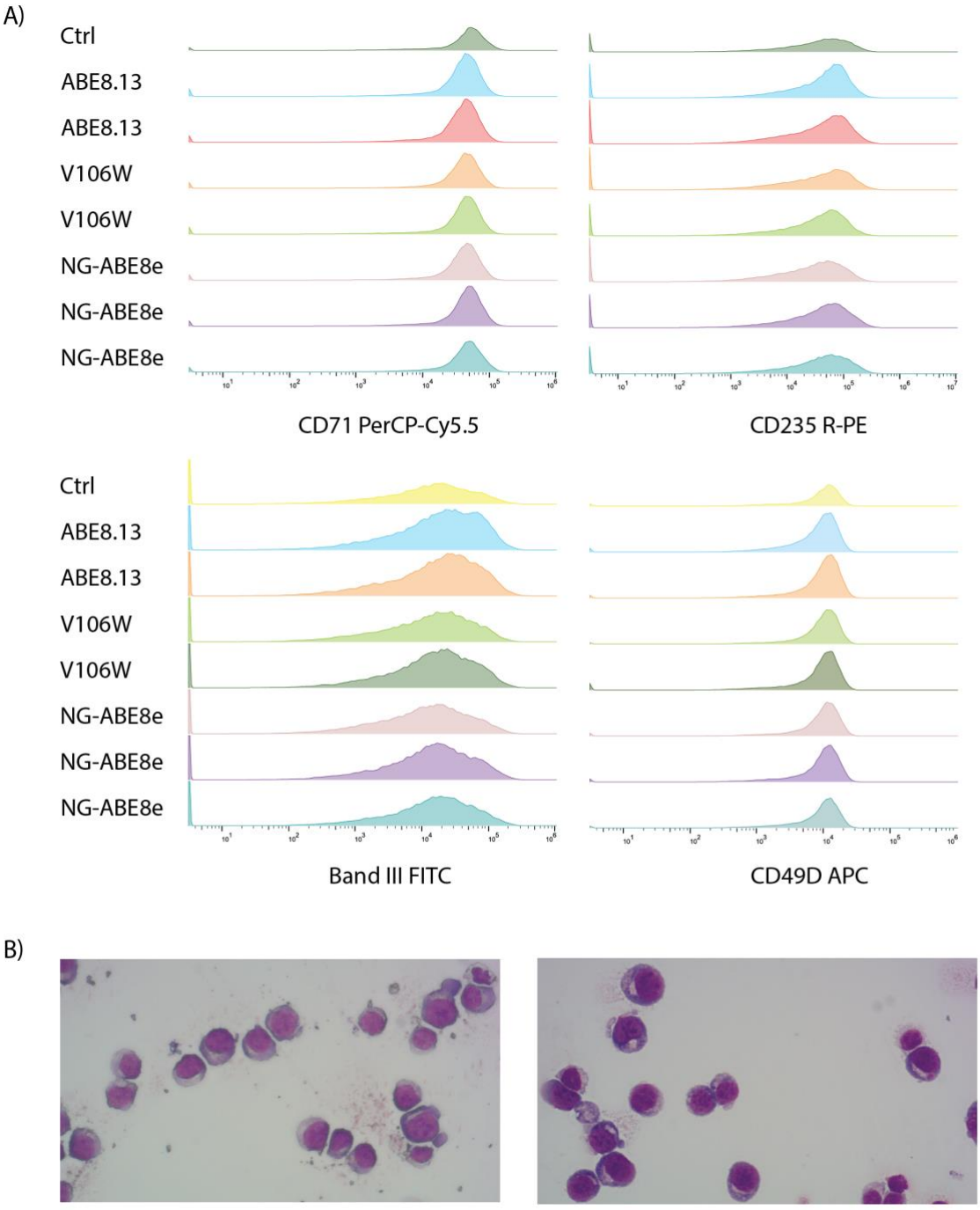


Figure 5.15 - Immunophenotyping of mRNA edited patient cells *Representative data from edited patient cells showing similar stages of differentiation at Day 10 in erythroid culture. A) Immunophenotyping data. Each track row represents a replicate, with the ABE used annotated on the left. B) Cytospins at 40X magnification*

There was no evidence of differentiation perturbation when comparing edited and non-edited samples as assessed by morphology on cytospins or immunophenotyping for erythroid cell surface markers on day 13 of differentiation (Figure 5.15).

Gene expression was assessed by RT-qPCR at D10 of erythroid differentiation. As observed with plasmid-based editing, there was a decrease in the α/β globin ratio in all samples edited, with a mean reduction to 0.632 (n=5, SD 0.08)(Figure 5.16A). The largest reduction, to 0.51, was found in the sample with the most efficient editing, showing that repairing the HbE allele directly caused an improvement in globin expression patterns in patient derived cells. CD34+ cells from a normal control were also edited from WT (GAG) with over 90% efficiency to the Hb Aubenas (GGG) codon, and RT-qPCR showed $\beta^{Aubenas}$ is transcribed at the same rate as β^{A} as assessed by the α/β and the α/β+γ ratios (Figure 5.16B&C). Protein based assays were performed in samples with efficient editing to look for evidence that there was a reduction in HbE and a concomitant rise in normal or normal β-globin variants that would produce HbA or Hb Aubenas. IEF gels and CE-HPLC showed this shift in protein expression; whilst HbF levels remained broadly the same, there was a significant fall in HbE and increase in HbA and especially Hb Aubenas (Figure 5.16D&E). Overall there was a mean 3-fold fall in HbE protein levels and a 13-fold increase in normal/normal variant haemoglobin expression (HbA/Hb Aubenas respectively)(Figure 5.16F).

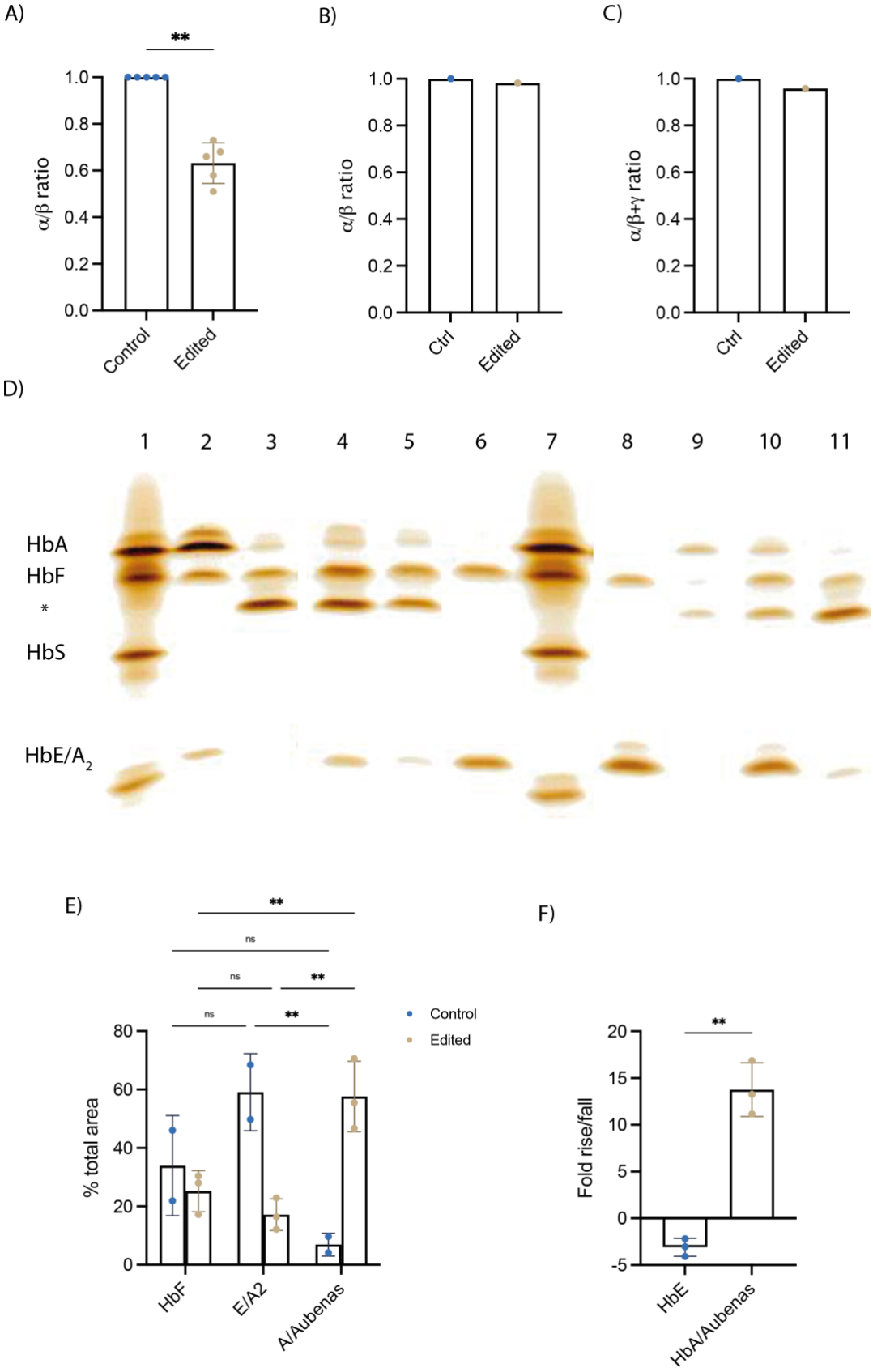

A)
**
α/β ratio
1.0
0.8
0.6
0.4
0.2
0.0
Control
Edited
B)
α/β ratio
Ctrl
Edited
C)
α/β+γ ratio
Ctrl
Edited
D)
1 2 3 4 5 6 7 8 9 10 11
HbA
HbF
*
HbS
HbE/A2
E)
**
ns
ns
**
ns
**
Control
Edited
% total area
80
60
40
20
0
HbF
E/A2
A/Aubenas
F)
**
Fold rise/fall
20
15
10
5
0
-5
HbE
HbA/Aubenas

Figure 5.16 - mRNA and protein expression in edited cells *A) RT-qPCR data showing a significant reduction in the α/β-ratio in edited cells. Ratios in patient control cells normalised to 1. B and C) Edited WT cells with the Hb Aubenas codon (GGG) shows normal β-globin transcription compared to WT control. D) IEF from control and edited cells. Lanes as follows: 1 and 7 - reference band, 2 - WT control, 3 - WT cells edited to have Hb Aubenas as in Figure 5.17B and C, 4 and 5 - edited cells from patient K using V106W, 6 - non-edited control cells from patient K, 8 - non-edited control cells from patient S, 9 - edited cells from patient S using V106W, 10 - edited cells from patient S using ABE8.13, 11 - edited cells from patient S using NG-ABE8e. E) Haemoglobin profile in cells edited with V106W using CE-HPLC. F) Fold changes in HbE and normal haemoglobins as measured by HPLC. Significance calculated in A) and F) by Mann-Whitney U test, in E) by two-way ANOVA. P values - * <0.05, ** <0.01.*

5.2.8 mRNA-based editing of long-term haematopoietic stem cells

In a pre-clinical *in-vitro* gene editing programme, it is important to make sure that long-term haematopoietic stem cells (LT-HSCs) are edited. In patients, a crucial issue is to ensure these edited cells are able to effectively reconstitute bone marrow and form stable grafts for decades. The current "gold standard" approach to test this is through murine serial xenotransplantation assays. To demonstrate that LT-HSCs are edited, input cells are injected into immunodeficient mice, allowed to engraft, extracted after 16-20 weeks, reinfused into a second mouse and then extracted after another 16-20 weeks. The presence of human CD45+ cells in the second xenotransplant indicates a positive result. The commonly used NOD-scid IL2R-gamma^{null} (NSG) mouse was used as the murine transplant recipient, and umbilical cord blood (UCB) CD34+ cells were used rather than adult peripheral blood CD34+ cells due to their relative enrichment in LT-HSCs (Ito et al., 2002, Shultz et al., 2005). The previously described optimisation experiments and those in patient-derived CD34+ cells showed that editing A_6 in the WT codon served as an excellent proxy for overall therapeutic editing of the HbE codon; editors that targeted A_6 efficiently edited A_5 even more efficiently in patient cells. Therefore, due to the relative scarcity of available patient-derived CD34+

cells, WT cells were used in the xenograft experiment, with the same strategy adopted as when optimising base editor usage in 5.2.3.

Three mice were injected with UCB CD34+ cells from three independent donors. All donors showed good engraftment in the primary mouse, as assessed by flow cytometry at 20 weeks (Figure 5.17A); a mean of 26.2% of live cells in murine bone marrow were positive for human CD45 (62% of all CD45+ cells in bone marrow). Human CD45+ cells were detected in bone marrow from all secondary xeno-transplanted mice (Figure 5.17); 0.25% of the total CD45+ cell compartment. As the threshold for positive secondary mouse engraftment in this model is 0.1% this represented robust human engraftment.

There was a variable level of editing efficiency in the engrafted human cells (Figure 5.17C). In the primary xenotransplants the first two donors showed increases in editing efficiency, from 55% to 68% and 44% to 72% respectively. In contrast the third donor showed a fall from 47% in the input cells to 23% in the primary mouse. In all three of the secondary xenotransplanted mice, human CD45+ cells showed editing of the target base, with an average 34.5% editing efficiency. The mean editing efficiency reduced from 46% in the input cells to 41% in the primary xenotransplanted mice, to 34.5% in the secondary xenotransplanted mice.

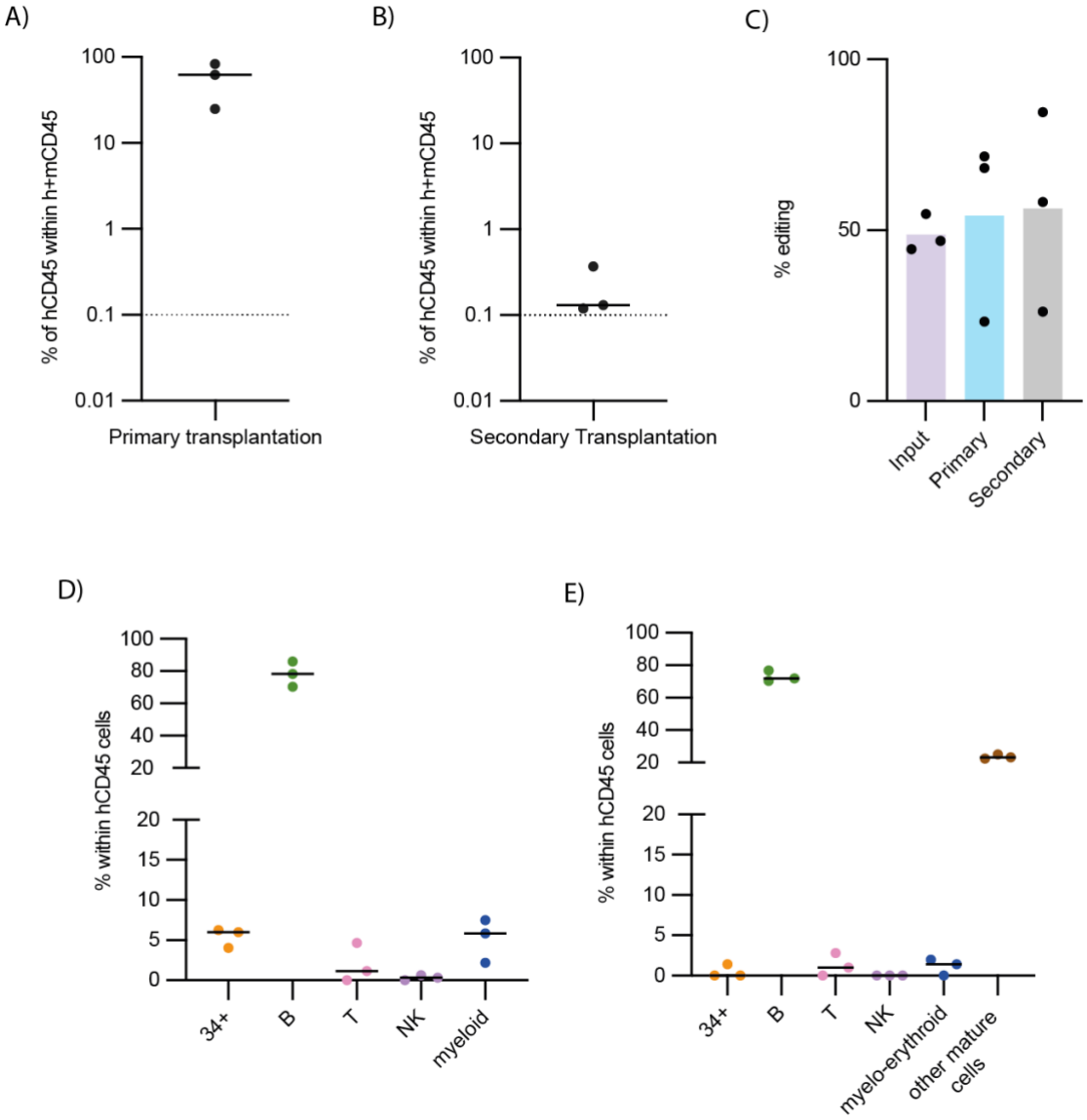


Figure 5.17 - Serial xenotransplantation data *Primary (A) and secondary (B) human CD45+ cell engraftment in murine bone marrow. C) Editing of A_6 in input, primary and secondary xenotransplants. Haematological differentiation in primary (D) and (E) xenotransplants in bone marrow.*

The NSG model allows for human lymphoid differentiation, but only an extremely limited amount of erythroid development. In the primary xenotransplanted mice, human cells were predominantly positive for B-cell markers and as expected showed low levels of T-cell, NK-cell and myeloid expression (Figure 5.17D). A broadly similar pattern was also found in murine spleen, though there was a higher proportion of T-cells in the human CD45+

compartment as compared to that in bone marrow. The same pattern of distribution between the haematopoietic lineages was found in bone marrow in the secondary xenotransplants, with a reduction in CD34+ expression as expected (Figure 5.17E).

5.2.9 Off-target effects - CIRCLE-seq of base-edited patient cells

To examine the potential OT editing landscape of base editors and the sgRNA used, *in-silico* and *in-vitro* assays were performed. The HbE editing sgRNA was initially profiled with Cas-OFFinder and CRISPOR (Bae et al., 2014, Concordet and Haeussler, 2018). These identified 1401 and 554 sites respectively. CIRCLE-seq was then performed on DNA extracted from the HUDEP-2 HbE cell line and two patients (patient F and patient S) with HbE/β-thalassaemia (Figure 5.18A). 227 sites were identified from the HUDEP-2 line, 979 from patient F and 351 from patient S. A total of 2830 potential off-target sites were identified between the three methods, but interestingly no single site was shared between all 3 CIRCLE-seq replicates and both *in-silico* methods. Only 51 sites were shared between each CIRCLE-seq replicate, CRISPOR and Cas-OFFinder, with the majority being found in only one method (Figure 5.18B). There was very little overlap found between the three OT profiling methods utilised (Figure 5.18C). Even between all CIRCLE-seq replicates only 154 (14%) sites were identified in all three replicates (Figure 5.18D). Potential OT editing sites were annotated using HOMER, showing that the majority of sites are in intergenic and intronic regions, with a small number in exonic or promoter/transcriptional start site (TSS) locations, although this profile may potentially change with in-vivo editing (CIRCLE-seq uses cell-free DNA as a substrate) (Figure 5.18E)(Heinz et al., 2010).

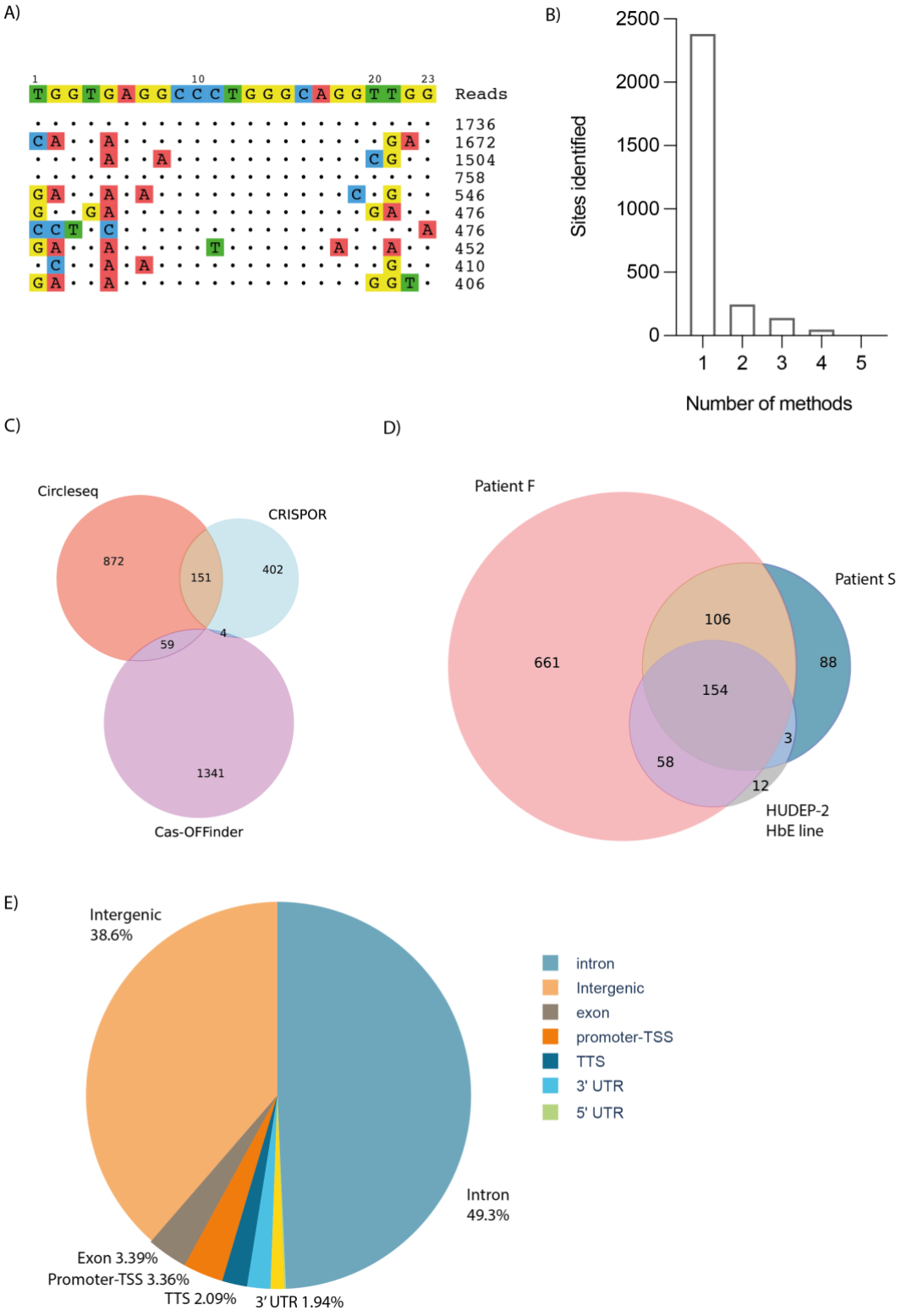


Figure 5.18 - The OT editing landscape *A) Representative CIRCLE-seq output showing the on-target site and the first nine OT sites. B) Number of sites identified between five methods/replicates (3 CIRCLE-seq replicates, Cas-OFFinder and CRISPOR). C) Venn diagram of candidate OT sites identified with three different methods D) Venn diagram of candidate OT sites identified by three CIRCLE-seq replicates (Patient F, S and HUDEP-2 HbE line) E) Predicted genomic features of candidate OT sites. TTS - transcriptional start site. TTS - transcriptional termination site, UTR - untranslated region.*

To see whether these potential sites corresponded to OT events in-vivo, the top 250 OT sites were assessed using targeted capture NGS of DNA from edited HbE/β-thalassaemia patient cells. This sequencing data is currently being analysed.

5.2.10 Off-target effects - assessment of chromatin accessibility in base-edited patient cells

To the assess the local and genome-wide effects of base editing on chromatin accessibility, ATAC-seq was performed on erythroid cells differentiated from patient CD34+ cells either transfected with a scrambled sgRNA or a sgRNA targeting the HbE mutation. There were no peaks that were significantly increased or decreased when assessed by DEseq2 (each analysis involved two technical replicates for each of the control and edited conditions). This was the case in the experiment targeting patient cells using plasmid and also with the same sgRNA targeting using mRNA with two different base editors. Even when the two mRNA experiments using different editors were combined, and all data including the two mRNA and plasmid experiments were combined, no significant differences were detected. Consistent with the data shown in chapter 4, the biggest difference between the datasets was between biological donors rather than between edited and non-edited samples.

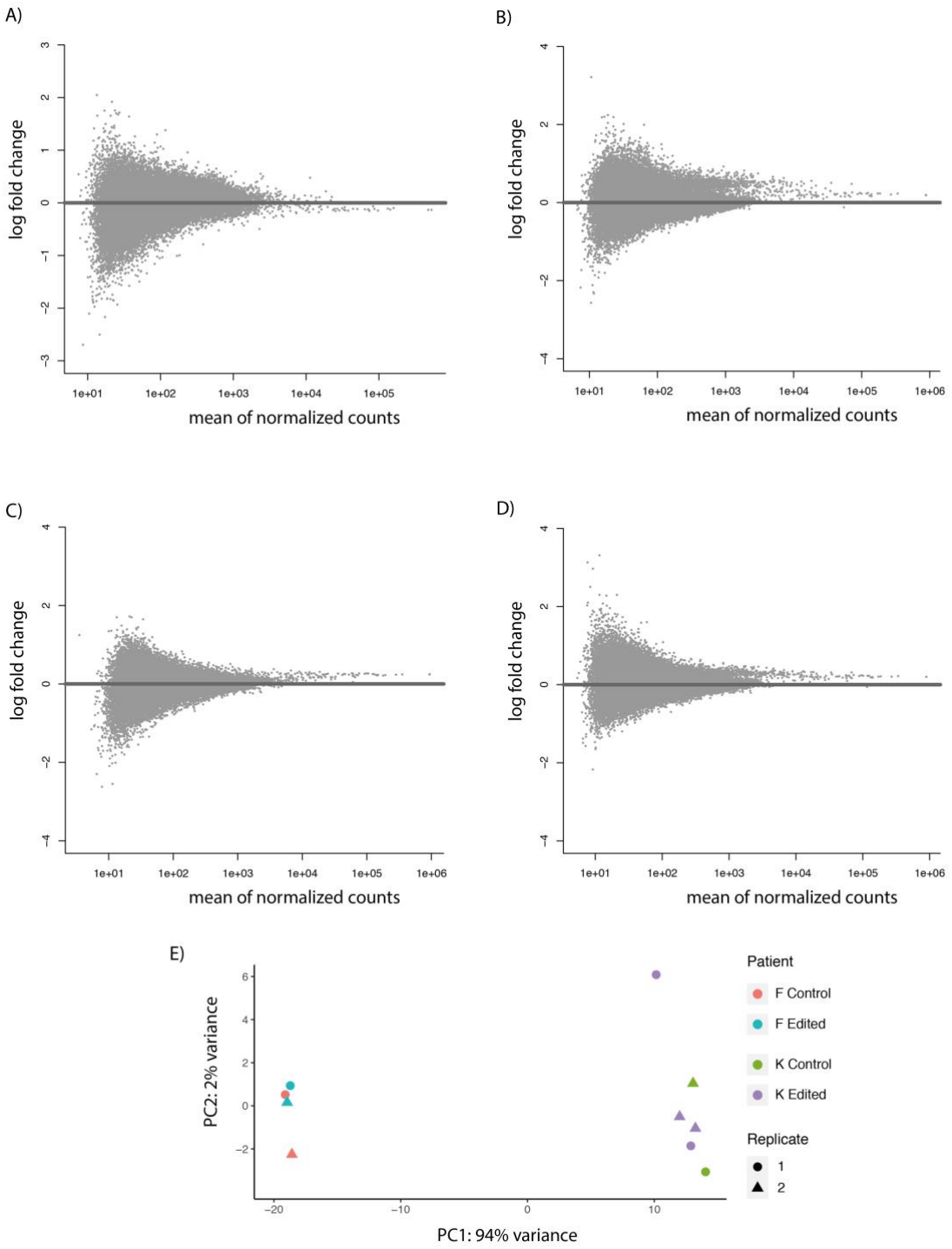


Figure 5.19 - Global chromatin accessibility profiles MA plots showing a lack of differentially accessible peaks between edited and non-edited patient samples A) plasmid based editing of cells from patient F B) mRNA based editing of cells from patient K using V106W and C) NG-ABE8e. D) Comparison combining all data from patient K E) PCA of dataset combining inputs from patient F and K, showing the majority of variation derives between biological replicates rather than between control and edited samples.

Whilst this data provides evidence that base-editing does not lead to genome-wide perturbations in chromatin accessibility in erythroid cells, other cells need to be profiled, as genotoxic events and malignant transformation are more likely to occur in HSPCs or other blood cell types such as T and B cells. As performing ATAC-seq in multiple cell types was not feasible, deepHaem, a deep convoluted neural network trained on 936 haematological chromatin-accessibility and epigenetic datasets, was used to profile 49 different blood cell types (Schwessinger et al., 2020).

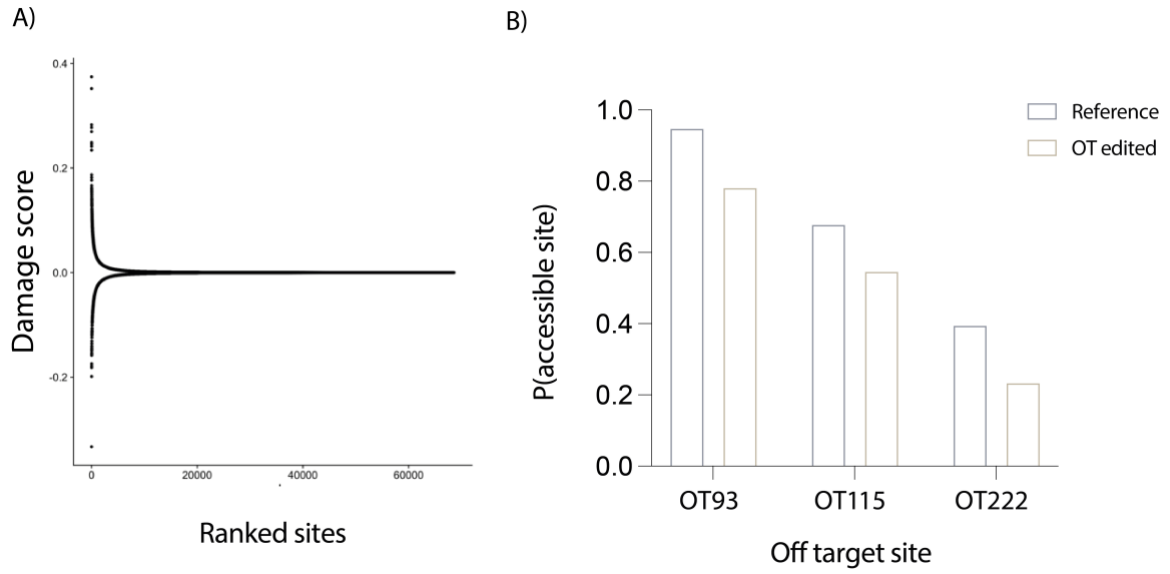


Figure 5.20 - DeepHaem results *A) Damage score for every possible OT edit across all blood cell types tested. The vast majority do not result in an accessibility change B) The three loci identified as possible significant OT sites. OT93 refers to the CHCHD6 site, OT115 to the PRR12 site and OT222 to an intergenic site.*

All putative sites identified in 5.2.9 that had an adenine in the editing window were tested (1400 sites). The vast majority of potential OT events were not predicted to result in any change in chromatin accessibility (**Figure 5.20**A). Of these, 173 sites were identified that were predicted to have an open chromatin signature in one of the cell types tested. 19 of these sites were predicted to have >10% reduction in chromatin accessibility upon editing.

Of note, only 3 sites were in the top 250 OT sites chosen for deep capture NGS; the other sites were only identified by one assay as noted in 5.2.9. None of these 3 potential OTs coincide with known SNPs causing disease (Figure 5.20B). One OT site is within the intron of *CHCHD6* and was predicted to have an effect in the MM.1S plasma-cell line. The second is 820bp away from the transcription termination site of *PRR12,* identified in K562 cells, and the third site is in intergenic region, predicted to have an effect in the GM12892 B-lymphocyte cell line. *CHCHD6* is involved in mitochondrial activity, and the function of *PRR12* is unknown, but it may play a role in embryonic neuronal development. Neither gene is expressed highly in haematological cells (Genotype-Tissue Expression [GTEx] Project), and the third intergenic site does not seem to be associated with a DNAseI accessible site in publicly available B-cell ENCODE data. Therefore, if these are bona fide OT sites, it is debateable whether they would have a deleterious effect in blood cells. As these sites are in the 250 chosen for deep capture NGS, it will be interesting to see if any editing occurs at these loci *in-vivo.*

5.3 Discussion

CRISPR/Cas9 has proved to be a major advance in basic scientific and therapeutic research, with several early-stage successful clinical trials in progress. Despite these advances, at least for curative applications, the WT CRISPR/Cas9 system is limited due to its reliance on creating DSBs which may have significant genotoxic effects. BEs offer an elegant solution to this problem. This chapter describes the use of ABEs to repair and cure one of the most common genetic disorders worldwide, the HbE/β-thalassaemias. This work is important for two broad reasons, its contribution to the collective experience of the field in using these relatively new editing technologies, as well as for its implications for the care of patients with HbE/β-thalassaemia.

The first part of this chapter involved the optimisation of the use of ABEs. A striking initial observation was that the ABEs were relatively effective at editing both non-haematopoietic and haematopoietic cells. In HEK293Ts they were able to modestly edit the HBB gene, with more efficient deaminations encountered in HUDEP-2 cells. As has been suggested elsewhere, this is likely to be specific to different cell-types, possibly due to chromatin accessibility (Arbab et al., 2020). It was surprising that the ABE7.10 editors did not exhibit any significant activity in CD34+ cells. In contrast, ABEmax was able to edit very efficiently in CD34+ cells, likely due to an altered codon usage of the ABE and a change in nuclear localisation signals (Koblan et al., 2018). Why this would improve the editing in comparison to that in the relatively similar HUDEP-2 cell line is not clear. Recent work targeting *HBB* showed base editing efficiencies of 80% in HSCs, with slightly higher levels in fibroblasts from patients with progeria (Koblan et al., 2021, Newby et al., 2021). The editing efficiencies

achieved in my work were generally in excess of 80%, with several of the later experiments achieving base conversions of almost 100%. This shows the efficiency of the approach used here, as well as the potential use of ABEs in other applications. The majority of the editing reported in this chapter involved patient cells, and so complete repair of the HbE allele still only represents deamination of one allele. However, these efficiencies are likely to be similar even if a homozygous mutation needed to be corrected: deamination of the homozygous HUDEP-2 HbE line and A_6 in WT CD34+ cells approached 100% as shown in Figure **5.7**B and discussed in 5.2.7 respectively. These findings are likely to be reproducible at other loci, as efficient multiplex editing was achieved at two TFBSs in MCS-R2 as shown in chapter four.

Early work using CBEs showed that C to non-T edits were observed (Rees and Liu, 2018). No A to non-G base conversions were noted in this work. Furthermore, while some studies showed a low level of indel formation distal to the target site induced by the nickase function of Cas9, no indels were detected in this project, despite multiple targeted NGS experiments. This specificity is important, particularly when using ABEs for precise mutation in exonic sequences. As this work involves developing a therapeutic strategy, an in-depth examination of OT events was conducted. Thus far OT work has involved performing targeted capture of sites identified by CIRCLE-seq or *in-silico* methods (Zeng et al., 2020, Newby et al., 2021, Koblan et al., 2021). Comprehensive OT work from different donors and cell types showed that no single experiment can be relied upon to draw general conclusions, particularly *in-silico* methods which did not show a large overlap with sites found through *in-vitro* work. CIRCLE-seq or equivalent methods are likely to be more reliable as they involve an assessment of the sites to which a particular sgRNA actually directs dCas9. However, there was still a degree of variability between the CIRCLE-seq replicates with nearly three times as many sites identified using Patient F's DNA compared to Patient S or the HUDEP-2 HbE

line. A degree of normal DNA sequence variation may account for some of these sites, but the majority remain unexplained. Experimental variability or oversampling remains a possibility, which would have to be controlled as much as possible when applying these techniques in the clinic. A caveat must be noted when using CIRCLE-seq to assess base-editing OT events; CIRCLE-seq relies upon Cas9's ability to bind DNA and catalyse DSBs. Whilst the former is ability is shared with base editing and WT CRISPR/Cas9 editing, the latter is notably not. That both forms of genome-editors have the same residence patterns on DNA, resulting in similar functional effects, be they DSBs or deaminations, is an assumption of the use of CIRCLE-seq to identify candidate base editing OT sites. Whilst this is a reasonable assumption to hold, there is the possibility that false positives and negatives may be present within the CIRCLE-seq results.

Unfortunately, these OT methods are potentially insufficient to detect sgRNA-independent OT base editing. This has been observed with CBEs, but not with ABEs, though the main study for the latter was in rice (Anzalone et al., 2020). The assessment of genome-wide single point mutations remains an incompletely resolved issue, given that the background mutation rate is approximately 15 mutations per cell division per diploid genome (Milholland et al., 2017). One approach to address this is to edit patient cells, sort them into single cells and then genotype clonal colonies of cells to compare the two groups. This work has been carried out in cells isolated from Patient F, but as the analysis is complex and on-going, the results have not been included here. Another form of sgRNA-independent base editing involves RNA editing; BEs were shown to catalyse transcriptome-wide edits with frequencies of 0.07-100% in up to 58% of expressed genes (Grunewald et al., 2019a). RNA-seq to characterise this in our work was not carried out for various reasons. Firstly, ABEs

have been shown to edit RNA far less than CBEs, and variants - like V106W - have been engineered to reduce this rate even further (Grunewald et al., 2019b, Rees et al., 2019, Richter et al., 2020). Secondly, by their nature, transcripts are short lived, and so unlike OTs in DNA any potential effect is likely to be transient. Rather than the risk of malignant transformation that is a concern with DNA OTs, a potential harmful effect of RNA OTs is the forcing of cells into different states through affecting the transcriptome. For example, LT-HSCs could potentially be forced to differentiate or become senescent, thereby compromising a therapeutic autograft that depends on these cells for robust haematopoietic activity over a lifetime. Whilst subtle changes cannot be ruled out, no abnormalities in differentiation were noted here in *in-vitro* assays or *in-vivo* mouse work.

As far as we are aware there is no work that has attempted to describe the functional consequence of any bona-fide OTs that do occur, particularly for the vast majority of events that will not involve previously annotated mutations associated with disease. Little, if any, work has been done to look at the genomic consequences of these changes in sequence; most work has focussed on looking at gross changes in cell differentiation and relatively short-term viability. As discussed in chapter 3, even single base-pair changes in sequence can result in potentially dramatic changes in chromatin accessibility, which may act as a proxy marker for altered gene expression leading to perturbed cell activity. The most striking manifestation of this would be the development of myelodysplastic syndrome or acute myeloid or lymphoid leukaemia. However, whilst not fully proven there is evidence to show that small mutations can lead to conditions that may not be immediately detectable, such as clonal haematopoeisis of indeterminate potential, which itself carries a risk of malignant transformation and is associated with significant disease in non-haematopoietic tissue (DiNardo and Cortes, 2016, Jaiswal and Libby, 2020, Silver et al., 2021). Analysis of ATAC-

seq offers an attractive option, as the assay is relatively easy and any functional changes in genome function may involve changes in chromatin accessibility. The fact that no editing events resulted in large-scale chromatin accessibility is reassuring. However, given that malignant events involving edited cells are likely to involve myeloid, lymphoid cells or more stem-like cells, profiling fully differentiated erythroid tissue will miss these. The machine learning model deepHaem offers a powerful solution to this problem. Reassuringly there were very few sites detected genome-wide that were predicted to produce altered chromatin accessibility. Whilst this technology is in its nascent stages it has been tested at several validated loci, and could become an important method to add to the OT analysis toolkit. As the number of potential functionally relevant OT targets is relatively low, these could then be functionally assessed through editing in relevant cell type models.

Generally speaking, managing OT risks involves navigating several important factors that were encountered in this study. The type of editor used is a major factor in OT events. The ABE8es are superior editors to previous ABEs, being able to catalyse deaminations 100 fold faster than their predecessors (Lapinaite et al., 2020). This may come with a downside, in that OTs may also be generated more efficiently. V106W, the editor used here was evolved to mitigate this risk, but it will be interesting to see results from detailed comparisons between this, other ABE8es and ABEmax, particularly as the ABE8es have been shown to catalyse increased RNA OT edits compared to ABEmax (Jeong et al., 2021). Another major consideration is the delivery method. RNPs have a favourable OT editing profile as they have immediate activity, but start to be degraded as soon as they enter the cell. In my work, significant efforts were made to synthesise ABE proteins without success (data not shown). mRNA transfection was adopted as the next most appropriate vector. Whilst clearly more

suitable than plasmid, mRNA and subsequent translated protein may persist in the cell for longer than RNP, and so the OT profile may be expanded. It is noteworthy however that in contrast to WT spCas9, mRNA seems to be more effective than RNPs for base editing as shown recently (Newby et al., 2021). A careful balance therefore needs to be made when choosing the type of editor and vector, particularly when therapeutic applications are involved.

The editing data generated here involving patient cells indicate that the repair of the HbE mutation is likely to result in a meaningful phenotypic improvement for patients with HbE/β-thalassaemia and could provide a functional cure. The HbF induction approach adopted in the BCL11a editing trials have been successful in the short-term, but the published data has not involved patients with HbE/β-thalassaemia. Patients with severe HbE/β-thalassaemia frequently have HbF levels at 40% and above, suggesting a level of pre-existing de-repression of gamma globin expression. Pre-clinical work editing BCL11a showed a rise in HbF levels to approximately 60%. The amount to which BCL11a downregulation imparts clinical benefit remains to be seen, particularly as studies in thalassaemia intermedia have shown that morbidity is observed until the HbF% exceeds 70-80% (Musallam et al., 2012a). The attractiveness of this current approach is it recapitulates the genotype of β-thalassaemia carriers, who have a completely normal clinical phenotype and comprise approximately 1.5% of the world's population (Colah et al., 2010). ABE editors, particularly ABE8e-V106W can efficiently edit the AAG HbE codon to WT or a normal variant. This normal variant (Hb Aubenas) has been described once in heterozygosity, in a patient who had completely normal haematological indices, in contrast to β-thalassaemia carriers, who have a reduced MCV and MCH. Therefore, it would be unexpected for an individual with a

$\beta^0/\beta^{Aubenas}$ to have a thalassaemic phenotype. To perform further studies, we attempted to contact the individual with Hb Aubenas who resides in the south of France, but she has been lost to follow-up. It was not possible to characterise Hb Aubenas further using HUDEP-2 HbE cells, as this line retains a fetal expression pattern and *HBB* transcription is minimal. It is notable however that Hb Aubenas seems to migrate in the HbA window on CE-HPLC, and has completely normal mRNA expression levels compared WT β-globin.

Repairing the HbE codon reproducibly and consistently reduces the α/β ratio to 0.6 (where the ratio in non-edited patient cells is normalised to 1). Whilst the mRNA ratios are significantly improved, it is also important that production of normal haemoglobin improves. This is shown by CE-HPLC, which shows a robust increase in HbA/Hb Aubenas and corresponding fall in HbE levels, although as HbA_2 is detected in the HbE window, the reduction in HbE Is likely to be greater. The switching of HbE for HbA/Aubenas has greater significance than the gross amount of haemoglobin protein; due to the reported instability of HbE and potentiating effect on the deleterious effects of excess unpaired α-globin chains, the presence of HbA/Aubenas will enable a more sustained and stable haemoglobin concentration, particularly in times of physiological stress (Hirsch et al., 2017). It is important to note the relatively high amount of HbF detected (~35%) in unedited cells. This is consistent with levels found in patients with HbE/β-thalassamia and is likely due to γ-globin expression from the thalassaemic allele (Fucharoen and Weatherall, 2012). While there seems to be a reduction in HbF on editing, this did not reach statistical significance.

The data in this chapter suggest that base editing occurs in LT-HSCs, given that edited human cells were found in all secondary transplanted mice, providing one of the first reports of this. The NSG mouse model was chosen due to its availability and local experience in using it for

serial xenotransplantation experiments. Good engraftment was noted in primary and secondary mice, with editing seen in LT-HSCs and expected levels of different blood cell types indicating that base editing doesn't interfere with normal differentiation of LT-HSCs.

A few caveats should be noted. Due to the time taken for these experiments (minimum 8 months) ABEmax was used in mRNA form, as it was the only editor available at the time. This meant that the editing efficiencies are lower than were seen in later WT and patient cell editing with ABE8e-V106W. Secondly, the NSG mouse model used, whilst excellent for assessing the presence of LT-HSCs is not a good system for observing human erythroid development. Given that the *in-vitro* differentiation experiments showed no abnormalities, the other haematopoietic lineages showed good reconstitution and LT-HSCs were edited, it is unlikely that the erythroid lineage would show a deficit in *in-vivo.* However if this needed to be tested the NOD B6.SCID *Il2rγ*$^{-/-KitW41/W41}$ (NBSGW) system has been shown to allow human erythroid engraftment (McIntosh et al., 2015).

In summary, this chapter shows that highly efficient repair of the HbE mutation that causes HbE/β-thalassaemia is feasible in primary human patient cells using adenine base editing. This editing recapitulates the naturally occurring β-thalassaemia carrier state, which is known to be asymptomatic, and is associated with significant improvements in globin mRNA expression and haemoglobin production. ABE editing is possible in LT-HSCs, and deamination events show persistence in secondary mouse xenotransplants. Together these data represent proof-of-principle pre-clinical work for a potentially effective therapeutic strategy to cure half the world's cases of severe β-thalassaemia.

6. Final discussion and conclusions

The globin loci and their associated genetic disorders have been described and studied for over a century. Through this work many of the principles underlying gene regulation have been and continue to be discovered. This knowledge has driven the development an ever-increasing variety of therapeutic strategies that offer the possibility of a cure for the haemoglobinopathies and genetic disorders more generally. The current work contributes to this scientific and medical endeavour by using the study of a rare patient mutation to assess how the enhancers control α-globin expression, as well as using knowledge gained about this control of α-globin expression and the pathophysiology of β-thalassaemia to develop two strategies for its cure. Whilst the globin loci have been worked on for decades, this study shows that new and valuable basic scientific and therapeutic insights can be gained by applying state-of-the-art technologies and careful clinical observation to well-known genes and their disorders.

6.1 MCS-R2 behaves according to a flexible enhancer model

Different models have been put forward to explain enhancer behaviour. On one end of the spectrum the 'enhanceosome' model is extremely stringent in terms of transcription factor binding site (TFBS) order, and spacing, and small perturbations in any of these determinants leads to a significant loss of function (Merika and Thanos, 2001). On the other side of the spectrum the 'billboard' model is more permissive; TFBS grammar is more flexible, and enhancer output is a function of how many TFs have bound to their respective sites (Spitz

and Furlong, 2012). In this model only a subset of sites need to be active at any time for the enhancer to work. Finally, in the 'TF collective' model, TFs bind their TFBSs as in the billboard model, but are also recruited through protein-protein interactions (Levo and Segal, 2014, Panigrahi and O'Malley, 2021). These higher-order binding events may require multiple TFs to be bound before protein-protein interactions can occur, meaning that unlike in the billboard model, enhancer output isn't simply additive in terms of how many TFs are bound to underlying DNA sequences. Individual TFBS motifs may vary, allowing for evolutionary change at these enhancers, as long as the TF complex can still form. Closer analysis of MCS-R2 suggests that it behaves according to a flexible enhancer architecture model, sharing features of the billboard and TF collective models. TFBSs were identified using a variety of methods, including new predictive *in-silico* tools, but no single site was predicted to be absolutely vital for the functioning of the enhancer. This was confirmed on functional testing when individual TFBSs were targeted in primary human CD34+ cells and differentiated into erythroid cells. Whilst α-globin transcription was reproducibly reduced, no single TFBS resulted in even a 50% reduction in expression and evidence suggest their contributions are additive which fits more with the billboard model. However certain central TFBSs were more important than others; namely the central NF-E2 and GATA1 motifs. These were predicted by a machine learning model to be more important than other sites for chromatin accessibility of the whole enhancer, which is consistent with the α-globin gene expression results. This may be because these sites co-operate with other bound TFs through protein-protein interactions once bound, as in the TF collective model. The erythroid pentameric complex, composed of GATA1, SCL/Tal1, Ldb1, LMO2 and E2a is known to be present at MCS-R2 (De Gobbi et al., 2007, Wadman et al., 1997). Of these only GATA1, E2A and SCL/Tal1 have DNA binding domains, with LMO2 and Ldb1 engaging in protein-protein

interactions with the DNA-bound TFs and one another. The GATA 5'2 site may be so important due to it being part of such a complex. Therefore, this region of the enhancer may exhibit TF collective like behaviour (though this remains to be formally tested), whilst other regions such as those containing the GATA1 5'1 and CACC 5' motifs function in a manner consistent with the billboard model

6.2 Continuing to develop genome engineering cures for red cell disorders

The promise of genome engineering strategies to cure genetic disease – particularly the haemoglobinopathies - has recently come to fruition with several phase I/II trials showing encouraging early results (Ribeil et al., 2017, Thompson et al., 2018, Esrick et al., 2020, Frangoul et al., 2021). Gene addition approaches using lentiviruses and BCL11a disruption methods using lentiviral shRNA and Cas9-mediated editing respectively can be described in many senses as a success. The data shown thus far shows an improvement in the clinical phenotype, but there are significant reasons to believe developing alternative strategies is a desirable. As mentioned elsewhere in this work, there are significant safety concerns regarding any therapy based on viral-mediated integration. Whilst the latest set-back to the lentiviral BlueBird Bio product has been cleared by the company, it remains a fact that the cargo was found to have integrated at a gene, *VAMP4*, though this is reportedly not known to cause myeloid malignancies. Non-viral integrative approaches, such as experimental work using retro-transposons could potential cause similar issues (Wang et al., 2019, Li et al., 2021). CRISPR/Cas9 based methods on the other hand are far more precise, efficient and easier to use. The HbF induction therapies have delivered good short-term results so far, but whether the levels induced will be satisfactory in the long-term is an open question.

Previous data has shown that the amount of HbF needed to convert a patient with a severe β-thalassaemia genotype to mild non-transfusion phenotype is approximately 60-70%. Levels below this result in a thalassaemia intermedia (TI) phenotype, which is still associated with significant long-term morbidity, particularly from iron overload (Musallam et al., 2012a).

The first strategy developed relies on recapitulating naturally occurring mutations in asymptomatic individuals, through knocking down α-globin expression by editing its main enhancer to reduce the excess pool of α-globin chains. The best candidate sites identified showed a 25-30% reduction in the α/β globin ratio. This approximates to the knockdown that might be observed in individuals with a single α-globin deletion. As individuals with a severe β-thalassaemia genotype and 1-2 α-globin deletions had mild symptoms or were asymptomatic, it is reasonable to assume that this approach may lead to a similar phenotype, or at worst non-transfusion dependent thalassaemia-intermedia (Chaisue et al., 2007, Sripichai et al., 2008b). In HbE/β-thalassaemia, the thalassaemia intermedia phenotype tends to appear at lower concentrations, at 60-70g/L, which is not seen in single or double gene deletion cohorts, and so it is possible that this level of knockdown of α-globin expression would be enough to avoid the complications of TI, and deliver meaningful clinical benefit to patients (Taher et al., 2013). It could be argued that a greater level of α-globin knockdown would be ideal. Using a single DSB this was not possible, but initial efforts were made to use multiplex base editing to disrupt multiple sites, with promising results. This could be done comprehensively across the enhancer to give even finer control of the α-globin expression. This approach could either be used as an alternative to HbF induction or as an adjunct, in case the latter approach is not effective enough in the long-term. It would

also circumvent some of the safety concerns with Cas9-cleavagae based therapies. DSB formation can be cytotoxic, but also lead to undesirable cell-type specific events. It has been shown that the majority of bone marrow derived CD34+ cells in patients with sickle cell disease and β-thalassaemia, the same cells used for gene editing grafts, are B-lymphoid progenitors, with a potentially more active IgH locus (Hua et al., 2019). Given that recombination events occur frequently at this locus, explaining why many lymphomas arise from translocations involving chromosome 14, it is not implausible that translocations could occur between this and the on-target DSB site, which in the case of BCL11a editing HbF-induction approaches is a strong enhancer of a gene known to be associated with lymphoid malignancy. Base editing on the other hand by and large does not result in DSBs, is far more programmable and was therefore used in the second approach.

The mutation causing HbE was highly efficiently repaired using base editing in primary patient-derived CD34+ cells. This was shown to cause up to a 50% reduction in the α/β-globin mRNA ratio and a robust increase in normal haemoglobin, HbA or Hb Aubenas. This editing was performed in long-term haematopoietic stem cells as assessed by serial xenotransplantation assays. The data shown here provide strong evidence that this approach would transform a patient with severe disease to an individual with an asymptomatic β-thalassaemia trait phenotype. This could be performed in isolation, or alongside the α-globin reduction approach using MCS-R2 base editing if needed. This base editing approach is one of the first described for haematological diseases, and amongst the most efficient. A previously described strategy using base editing to edit the BCL11a erythroid enhancer required serial transfections to efficiently edit the target site, which is not likely to be appropriate in the clinic (Zeng et al., 2020).

Because of the lack of DSB formation, base editing is highly clinically applicable, and is very likely to form the basis of the next generation of targeted gene-editing therapies. There is reason to believe improvements will be made to make its use even more attractive. Targeted evolution has improved the ability of base editors to cause A-G and C-T transitions, which have been used during this project. In our work this led to editing efficiencies in excess of 90%, approaching that of WT Cas9 cleavage. In addition to this efficiency concern, the flexibility has also been improved; whilst previously thought energetically unfavourable, editors that can catalyse C-G transversions have recently been developed. This opened up the window of pathogenic bases that can be edited (Kurt et al., 2021). As manufacturing pipelines improve and become more efficient, it is foreseeable that custom sgRNA and base editor combinations could be used to completely repair patient mutations. In HbE/β-thalassaemia this would potentially involve not only repairing the HbE mutation, but also the mutation on the other allele to completely revert the genotype back to normal. Of the 40 most common β-thalassaemia mutations as documented on the Ithagenes database, 18 (45%) can be corrected by base editing (Kountouris et al., 2014). This approach can be extended to any other condition presenting with compound heterozygous point mutations.

6.3 New ways of assessing off-targets in the clinic

As with WT Cas9, the use of base editors still comes with an off-target concern. It has been shown that adenine base editors have a favourable sgRNA-dependent OT profile (Jeong et al., 2020, Antoniou et al., 2021). However, sgRNA-independent editing can occur, both in RNA and DNA (Antoniou et al., 2021). Whole genome sequencing has been performed on clonal control and edited colonies grown from single CD34 cells isolated from a patient with

HbE\β-thalassaemia. Analysis of this large dataset is difficult, as the background presence and rate of naturally occurring mutations needs to be quantified before being able to assess the number of excess mutations potentially caused by the base editor. ATAC-seq has also been performed on these clonal colonies, to look for genome-wide chromatin accessibility changes. Finally, targeted capture NGS of the top 250 sites identified by CIRCLE-seq has been performed to see if any low-level OT editing is present. The analysis of these experiments is ongoing, but OT work on this scale has not yet been performed in published work using patient cells and so promises to be informative and instructive for the field in general. Considering OT analysis more generally for therapies, developing robust pipelines with rigorous variant calling algorithms will be crucial to rolling out gene editing treatments to patients outside of small-scale trials. It is likely that each individual graft will need an OT assay performed, possibly at multiple timepoints to look for clonal growth if OT edits are present.

Once OT events have been identified, it is not clear what the functional consequence of their presence might be, especially if they lie within non-coding regions. DeepHaem, a deep convoluted neural network trained on 936 haematological chromatin-accessibility and epigenetic datasets is a powerful tool to study chromatin accessibility, and was used to predict the effects of potential OT-sites identified by CIRCLE-seq. There were a very limited number of potential edits predicted to change chromatin accessibility over 49 different blood cell types. These reassuring findings were supported by the ATAC-seq studies performed in edited patient erythroid cells, which showed no significant changes in chromatin accessibility between control and edited cells.

The OT assessment in the present work has attempted to use multiple orthogonal approaches involving *in-silico*, *in-vitro* and *in-vivo* assays. It is likely that in the future, as genome engineering develops in the clinic, a similar multi-faceted approach will be needed to conduct a full profiling of the potential and actual OT landscape encountered in patient cells.

6.4 The future of genome editing - logistical and technical considerations

The work carried out in this project touches on some key issues that may guide the future of developing new treatments for the haemoglobinopathies and gene editing in general. Firstly, it is clear that whilst cutting edge therapies are mainly developed and utilised in highly economically-developed societies, the vast majority of patients who need them either cannot afford them if they rely on health-insurance based healthcare systems, or live in low-to-middle income countries. Gene editing may partially address this problem. Gene-editing is likely to be significantly cheaper than viral gene therapy programs due to relative simplicity of the products involved, especially if mRNA based editing is used. In this case, the price of a gene editing treatment will mainly be generated by the cost of performing an autologous stem-cell transplant, which is already routinely available in many middle-income countries (Aljurf et al., 2020). It may not be necessary for each country to have cell manufacturing facilities; in many viral and non-viral gene therapy trials, the cellular product is manufactured at a central site, often on another continent. This would lend itself well to lower-resource settings, so long as adequate cell collection facilities are present. How low and middle income countries will circumnavigate the closely guarded patents for

CRISPR/Cas9 editing and base-editing is an open question, and likely to require a significant amount of negotiation which could alter the price of therapeutics considerably.

Two alternative approaches for the haemoglobinopathies may make functionally curative treatments available even in settings where autologous HSCT is not available. The first is to develop small molecules that are able to ameliorate the symptoms of β-thalassaemia and sickle cell disease. The most obvious pathways to target would be those repressing γ-globin expression. A number of drugs are known to result in HbF upregulation in patients with β-thalassaemia, most notably hydroxyurea, but these either have limited efficacy or are non-specific in their action, resulting in intolerable side-effects (Musallam et al., 2013). An obvious strategy is to try and develop a drug that targets the transcription factors involved in γ-globin repression. This represents a holy grail for the field, as TFs have long been held to be undruggable. However recent technological strides have been made in conducting massively high-throughput compound screens, and the development of rationally designed micromolecules and drugs (Henley and Koehler, 2021). There is there hope therefore that an affordable and effective oral medication may one day be available. The second approach which would represent a major advance for individuals in any healthcare setting is developing *in-vivo* editing for red cell disorders. These are currently already available for disorders such as haemophilia B, whereby a hepatotropic virus is directed to the liver and induces Factor IX production in a targeted manner (Nathwani et al., 2014, George et al., 2017). Avoiding chemotherapy conditioning would remove the associated side effects, which include the risk of secondary MDS/AML, secondary solid cancers, endocrine dysfunction including infertility, relative immune suppression, as well as the immediate risks of an autologous transplant (Carreras et al., 2019). Using a virus to deliver editing machinery

is problematic as – along with the other issues related to viral transduction discussed at length previously - there are a number of immune responses associated with viruses that would lead to systemic symptoms but also viral neutralisation and poor efficacy (Shirley et al., 2020). An ideal *in-vivo* therapy would be non-immunogenic and able to target LT-HSCs, most of which reside in specific bone marrow niches. A potential approach would be to use nanoparticles containing the editing machinery with antibodies or other targeting moieties on their extrinsic surface that are specific for LT-HSCs (Mitchell et al., 2021). Whilst CD34 is used to identify haematopoietic stem cells, it is a non-specific marker in blood lineages but is also found on many other tissues such as the endothelium. CD49f and CD133 have been used as a selective marker for LT-HSCs, the latter in CD34-negative cells, and so could present candidate ligand targets for antibodies to promote internalisation of editing machinery (Notta et al., 2011, Sumide et al., 2018).

For decades the globin loci have been the focus of intense research which has significantly advanced our knowledge of gene regulation and helped establish many of the basic mechanisms that cause genetic disease. This work is a small contribution to this endeavour, and demonstrates that there is much yet to be gained from the detailed study of the α-globin locus that can deepen our understanding of cellular and molecular processes, but also drive forward therapeutic strategies for the world's most common genetic diseases.

Appendix 1 - Laboratory and radiological test results from patient NSE

Test	Patient Result	Normal Range
Red Blood Cell count, $x10^{12}$/L	5.85	4.2-5.5
Hemoglobin, g/L	69	120-155
Mean Corpuscular Volume, fL	40	80-96
Mean cell Hemoglobin, pg	12.7	27-33
Red Cell Distribution Width, %	26.5	12.2-16.1
Reticulocyte count, %	2.03	0.5-1.5
Haptoglobin, g/L	0.4	0.3-2.0
Total Bilirubin, mmol/L	21	1.71-20.50
Direct Bilirubin, mmol/L	5	<5.1
Lactate Dehydrogenase, U/L	573	100-190
Glucose-6-Phosphate Dehydrogenase, U/g Hb	41.2	5.5-20.5
Hexokinase, U/g Hb	5.55	0.8-1.5
Glucose-6-Phosphate Isomerase, U/g Hb	107	39.3-57.7
Erythropoietin, iU/L	14	3.7-36.0
Ferritin, mg/L	45	20-250
Zinc Protoporphyrin, mmol/mol heme	67	<100
Transferrin, g/L	2.01	1.7-3.7
Transferrin saturation, %	26.5	20-35
Soluble Transferrin Receptor, ug/mL	11.8	1.1-2.9
Erythroferrone, ng/mL	6.3	<1.0-2.4
MRI - Cardiac T2*, ms	38	>20
MRI - Liver Iron content T2*, µmol/g	40	36

Appendix 2 - Acknowledgement of previously published material

Section 3.3 and 3.4 of chapter 3 ‘Assessment of the *in-vivo* function of MCS-R2 and an analysis of its sequence’ describing and characterising α globin regulation in patient NSE has been previous published (Badat et al., 2021). I wrote the paper, was first author and produced all the figures.

Section 4.2.12 in chapter 4 ‘4.Development of a therapeutic strategy for ameliorating β-thalassaemia by editing MCS-R2 to reduce α-globin expression’ uses material that has been previously published (Hua et al., 2021). I was second author, jointly produced the figure shown and have only included material I was directly involved in and contributed to. I have permission from the corresponding author, James Davies, to reproduce the material here.

References

2021. Lentiviral vector cleared of causing blood cancer. *Nat Biotechnol,* 39**,** 398.

ACHOUR, A., KOOPMANN, T., CASTEL, R., SANTEN, G. W. E., DEN HOLLANDER, N., KNIJNENBURG, J., RUIVENKAMP, C. A. L., ARKESTEIJN, S. G. J., TER HUURNE, J., BISOEN, S., VERSCHUREN, M., VIJFHUIZEN, L., SCHAAP, R., GRIMBERGEN, A., SLOMP, J., TRAEGER-SYNODINOS, J., VRETTOU, C., PISSARD, S., GALACTEROS, F., BAAS, F. & HARTEVELD, C. L. 2020. A new gene associated with a beta-thalassemia phenotype: the observation of variants in SUPT5H. *Blood,* 136**,** 1789-1793.

ADIKUSUMA, F., PILTZ, S., CORBETT, M. A., TURVEY, M., MCCOLL, S. R., HELBIG, K. J., BEARD, M. R., HUGHES, J., POMERANTZ, R. T. & THOMAS, P. Q. 2018. Large deletions induced by Cas9 cleavage. *Nature,* 560**,** E8-E9.

ADLI, M. 2018. The CRISPR tool kit for genome editing and beyond. *Nat Commun,* 9**,** 1911.

ADVANI, R., RUBIN, E., MOHANDAS, N. & SCHRIER, S. L. 1992. Oxidative red blood cell membrane injury in the pathophysiology of severe mouse beta-thalassemia. *Blood,* 79**,** 1064-7.

ALJURF, M., WEISDORF, D., HASHMI, S. K., NASSAR, A., GLUCKMAN, E., MOHTY, M., RIZZO, D., PASQUINI, M., HAMADANI, M., SABER, W., HARI, P., KHARFAN-DABAJA, M., MAJHAIL, N., GERGES, U., ALI HAMIDIEH, A., HUSSAIN, F., ELHADDAD, A., MAHMOUD, H. K., TBAKHI, A., OTHMAN, T. B., HAMLADJI, R. M., BEKADJA, M. A., AHMED, P., BAZARBACHI, A., ADIL, S., ALKINDI, S., LADEB, S., DENNISON, D., PATEL, M., LU, P., QUESSAR, A. E., OKAMOTO, S., ATSUTA, Y., ALHEJAZI, A., AYAS, M., AHMED, S. O., NOVITZKY, N., SRIVASTAVA, A., SEBER, A., ELSOLH, H., GHAVAMZADEH, A., CONFER, D., KODERA, Y., GREINIX, H., SZER, J., HOROWITZ, M. & NIEDERWIESER, D. 2020. Worldwide Network for Blood and Marrow Transplantation (WBMT) recommendations for establishing a hematopoietic stem cell transplantation program in countries with limited resources (Part II): Clinical, technical and socio-economic considerations. *Hematol Oncol Stem Cell Ther,* 13**,** 7-16.

ALLEN, B. L. & TAATJES, D. J. 2015. The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol,* 16**,** 155-66.

ANDREANI, M., TESTI, M., BATTARRA, M., INDIGENO, P., GUAGNANO, A., POLCHI, P., FEDERICI, G. & LUCARELLI, G. 2008. Relationship between mixed chimerism and rejection after bone marrow transplantation in thalassaemia. *Blood Transfus,* 6**,** 143-9.

ANGELUCCI, E. 2010. Hematopoietic stem cell transplantation in thalassemia. *Hematology Am Soc Hematol Educ Program,* 2010**,** 456-62.

ANGELUCCI, E., MATTHES-MARTIN, S., BARONCIANI, D., BERNAUDIN, F., BONANOMI, S., CAPPELLINI, M. D., DALLE, J. H., DI BARTOLOMEO, P., DE HEREDIA, C. D., DICKERHOFF, R., GIARDINI, C., GLUCKMAN, E., HUSSEIN, A. A., KAMANI, N., MINKOV, M., LOCATELLI, F., ROCHA, V., SEDLACEK, P., SMIERS, F., THURET, I., YANIV, I., CAVAZZANA, M., PETERS, C., ERROR, E. I. & PARTIES, E. P. W. 2014. Hematopoietic stem cell transplantation in thalassemia major and sickle cell disease: indications and management recommendations from an international expert panel. *Haematologica,* 99**,** 811-20.

ANGUITA, E., HUGHES, J., HEYWORTH, C., BLOBEL, G. A., WOOD, W. G. & HIGGS, D. R. 2004. Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *EMBO J,* 23**,** 2841-52.

ANGUITA, E., SHARPE, J. A., SLOANE-STANLEY, J. A., TUFARELLI, C., HIGGS, D. R. & WOOD, W. G. 2002. Deletion of the mouse alpha-globin regulatory element (HS -26) has an unexpectedly mild phenotype. *Blood,* 100**,** 3450-6.

ANTONIOU, P., MICCIO, A. & BRUSSON, M. 2021. Base and Prime Editing Technologies for Blood Disorders. *Frontiers in Genome Editing,* 3.

ANZALONE, A. V., KOBLAN, L. W. & LIU, D. R. 2020. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol,* 38**,** 824-844.

ARBAB, M., SHEN, M. W., MOK, B., WILSON, C., MATUSZEK, Z., CASSA, C. A. & LIU, D. R. 2020. Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell,* 182**,** 463-480 e30.

ARLET, J. B., RIBEIL, J. A., GUILLEM, F., NEGRE, O., HAZOUME, A., MARCION, G., BEUZARD, Y., DUSSIOT, M., MOURA, I. C., DEMAREST, S., DE BEAUCHENE, I. C., BELAID-CHOUCAIR, Z., SEVIN, M., MACIEL, T. T., AUCLAIR, C., LEBOULCH, P., CHRETIEN, S., TCHERTANOV, L., BAUDIN-CREUZA, V., SEIGNEURIC, R., FONTENAY, M., GARRIDO, C., HERMINE, O. & COURTOIS, G. 2014. HSP70 sequestration by free alpha-globin promotes ineffective erythropoiesis in beta-thalassaemia. *Nature,* 514**,** 242-6.

AU, W. Y., CHEUNG, W. C., HU, W. H., CHAN, G. C., HA, S. Y., KHONG, P. L., MA, S. K. & LIANG, R. 2005. Hyperbilirubinemia and cholelithiasis in Chinese patients with hemoglobin H disease. *Ann Hematol,* 84**,** 671-4.

BABBS, C., BROWN, J., HORSLEY, S. W., SLATER, J., MAIFOSHIE, E., KUMAR, S., OOIJEVAAR, P., KRIEK, M., DIXON-MCIVER, A., HARTEVELD, C. L., TRAEGER-SYNODINOS, J., WILKIE, A. O. M., HIGGS, D. R. & BUCKLE, V. J. 2020. ATR-16 syndrome: mechanisms linking monosomy to phenotype. *J Med Genet,* 57**,** 414-421.

BACON, C. W. & D'ORSO, I. 2019. CDK9: a signaling hub for transcriptional control. *Transcription,* 10**,** 57-75.

BADAT, M., DAVIES, J. O. J., FISHER, C. A., DOWNES, D. J., ROSE, A., GLENTHOJ, A. B., VAN BEERS, E. J., HARTEVELD, C. L. & HIGGS, D. R. 2021. A remarkable case of HbH disease illustrates the relative contributions of the alpha-globin enhancers to gene expression. *Blood,* 137**,** 572-575.

BAE, S., PARK, J. & KIM, J. S. 2014. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics,* 30**,** 1473-5.

BAILEY, T. L. & ELKAN, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol,* 2**,** 28-36.

BAK, R. O., DEVER, D. P. & PORTEUS, M. H. 2018. CRISPR/Cas9 genome editing in human hematopoietic stem cells. *Nat Protoc,* 13**,** 358-376.

BANDYOPADHYAY, S., ROYCHOWDHURY, K., CHANDRA, S., DAS, M. & DASGUPTA, U. B. 2001. Variable severity of beta-thalassemia patients of eastern India: effect of alpha-thalassemia and xmnI polymorphism. *Clin Exp Med,* 1**,** 155-9.

BANERJI, J., RUSCONI, S. & SCHAFFNER, W. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell,* 27**,** 299-308.

BAO, E. L., CHENG, A. N. & SANKARAN, V. G. 2019. The genetics of human hematopoiesis and its disruption in disease. *EMBO Mol Med,* 11**,** e10316.

BARBOUR, V. M., TUFARELLI, C., SHARPE, J. A., SMITH, Z. E., AYYUB, H., HEINLEIN, C. A., SLOANE-STANLEY, J., INDRAK, K., WOOD, W. G. & HIGGS, D. R. 2000. alpha-thalassemia resulting from a negative chromosomal position effect. *Blood,* 96**,** 800-7.

BARONCIANI, D., ANGELUCCI, E., POTSCHGER, U., GAZIEV, J., YESILIPEK, A., ZECCA, M., OROFINO, M. G., GIARDINI, C., AL-AHMARI, A., MARKTEL, S., DE LA FUENTE, J., GHAVAMZADEH, A., HUSSEIN, A. A., TARGHETTA, C., PILO, F., LOCATELLI, F., DINI, G., BADER, P. & PETERS, C. 2016. Hemopoietic stem cell transplantation in thalassemia: a report from the European Society for Blood and Bone Marrow Transplantation Hemoglobinopathy Registry, 2000-2010. *Bone Marrow Transplant,* 51**,** 536-41.

BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science,* 315**,** 1709-12.

BASS, B. L. 2002. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem,* 71**,** 817-46.

BEAUDRY, M. A., FERGUSON, D. J., PEARSE, K., YANOFSKY, R. A., RUBIN, E. M. & KAN, Y. W. 1986. Survival of a hydropic infant with homozygous alpha-thalassemia-1. *J Pediatr,* 108**,** 713-6.

BENESCH, R. E., RANNEY, H. M., BENESCH, R. & SMITH, G. M. 1961. The chemistry of the Bohr effect. II. Some properties of hemoglobin H. *J Biol Chem,* 236**,** 2926-9.

BERNET, A., SABATIER, S., PICKETTS, D. J., OUAZANA, R., MORLE, F., HIGGS, D. R. & GODET, J. 1995. Targeted inactivation of the major positive regulatory element (HS-40) of the human alpha-globin gene locus. *Blood,* 86**,** 1202-11.

BIEKER, J. J. 1996. Isolation, genomic structure, and expression of human erythroid Kruppel-like factor (EKLF). *DNA Cell Biol,* 15**,** 347-52.

BLAESE, R. M., CULVER, K. W., CHANG, L., ANDERSON, W. F., MULLEN, C., NIENHUIS, A., CARTER, C., DUNBAR, C., LEITMAN, S., BERGER, M. & ET AL. 1993. Treatment of severe combined immunodeficiency disease (SCID) due to adenosine deaminase deficiency with CD34+ selected autologous peripheral blood cells transduced with a human ADA gene. Amendment to clinical research project, Project 90-C-195, January 10, 1992. *Hum Gene Ther,* 4**,** 521-7.

BLAESE, R. M., CULVER, K. W., MILLER, A. D., CARTER, C. S., FLEISHER, T., CLERICI, M., SHEARER, G., CHANG, L., CHIANG, Y., TOLSTOSHEV, P., GREENBLATT, J. J., ROSENBERG, S. A., KLEIN, H., BERGER, M., MULLEN, C. A., RAMSEY, W. J., MUUL, L., MORGAN, R. A. & ANDERSON, W. F. 1995. T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years. *Science,* 270**,** 475-80.

BOLUKBASI, M. F., GUPTA, A. & WOLFE, S. A. 2016. Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. *Nat Methods,* 13**,** 41-50.

BORG, J., PAPADOPOULOS, P., GEORGITSI, M., GUTIERREZ, L., GRECH, G., FANIS, P., PHYLACTIDES, M., VERKERK, A. J., VAN DER SPEK, P. J., SCERRI, C. A., CASSAR, W., GALDIES, R., VAN IJCKEN, W., OZGUR, Z., GILLEMANS, N., HOU, J., BUGEJA, M., GROSVELD, F. G., VON LINDERN, M., FELICE, A. E., PATRINOS, G. P. & PHILIPSEN, S. 2010. Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nat Genet,* 42**,** 801-5.

BOUTIN, J., ROSIER, J., CAPPELLEN, D., PRAT, F., TOUTAIN, J., PENNAMEN, P., BOURON, J., ROORYCK, C., MERLIO, J. P., LAMRISSI-GARCIA, I., CULLOT, G., AMINTAS, S., GUYONNET-DUPERAT, V., GED, C., BLOUIN, J. M., RICHARD, E., DABERNAT, S.,

MOREAU-GAUDRY, F. & BEDEL, A. 2021. CRISPR-Cas9 globin editing can induce megabase-scale copy-neutral losses of heterozygosity in hematopoietic cells. *Nat Commun,* 12**,** 4922.

BOYLE, A. P., DAVIS, S., SHULHA, H. P., MELTZER, P., MARGULIES, E. H., WENG, Z., FUREY, T. S. & CRAWFORD, G. E. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell,* 132**,** 311-22.

BOZHILOV, Y. K., DOWNES, D. J., TELENIUS, J., MARIEKE OUDELAAR, A., OLIVIER, E. N., MOUNTFORD, J. C., HUGHES, J. R., GIBBONS, R. J. & HIGGS, D. R. 2021. A gain-of-function single nucleotide variant creates a new promoter which acts as an orientation-dependent enhancer-blocker. *Nat Commun,* 12**,** 3806.

BRADAI, M., ABAD, M. T., PISSARD, S., LAMRAOUI, F., SKOPINSKI, L. & DE MONTALEMBERT, M. 2003. Hydroxyurea can eliminate transfusion requirements in children with severe beta-thalassemia. *Blood,* 102**,** 1529-30.

BRAUN, C. J., BOZTUG, K., PARUZYNSKI, A., WITZEL, M., SCHWARZER, A., ROTHE, M., MODLICH, U., BEIER, R., GOHRING, G., STEINEMANN, D., FRONZA, R., BALL, C. R., HAEMMERLE, R., NAUNDORF, S., KUHLCKE, K., ROSE, M., FRASER, C., MATHIAS, L., FERRARI, R., ABBOUD, M. R., AL-HERZ, W., KONDRATENKO, I., MARODI, L., GLIMM, H., SCHLEGELBERGER, B., SCHAMBACH, A., ALBERT, M. H., SCHMIDT, M., VON KALLE, C. & KLEIN, C. 2014. Gene therapy for Wiskott-Aldrich syndrome--long-term efficacy and genotoxicity. *Sci Transl Med,* 6**,** 227ra33.

BRENDEL, C., GUDA, S., RENELLA, R., BAUER, D. E., CANVER, M. C., KIM, Y. J., HEENEY, M. M., KLATT, D., FOGEL, J., MILSOM, M. D., ORKIN, S. H., GREGORY, R. I. & WILLIAMS, D. A. 2016. Lineage-specific BCL11A knockdown circumvents toxicities and reverses sickle phenotype. *J Clin Invest,* 126**,** 3868-3878.

BRITTENHAM, G. M., GRIFFITH, P. M., NIENHUIS, A. W., MCLAREN, C. E., YOUNG, N. S., TUCKER, E. E., ALLEN, C. J., FARRELL, D. E. & HARRIS, J. W. 1994. Efficacy of deferoxamine in preventing complications of iron overload in patients with thalassemia major. *N Engl J Med,* 331**,** 567-73.

BUENROSTRO, J. D., GIRESI, P. G., ZABA, L. C., CHANG, H. Y. & GREENLEAF, W. J. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods,* 10**,** 1213-8.

BURNS, M. B., LACKEY, L., CARPENTER, M. A., RATHORE, A., LAND, A. M., LEONARD, B., REFSLAND, E. W., KOTANDENIYA, D., TRETYAKOVA, N., NIKAS, J. B., YEE, D., TEMIZ, N. A., DONOHUE, D. E., MCDOUGLE, R. M., BROWN, W. L., LAW, E. K. & HARRIS, R. S. 2013. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature,* 494**,** 366-70.

CAMASCHELLA, C., MAZZA, U., ROETTO, A., GOTTARDI, E., PARZIALE, A., TRAVI, M., FATTORE, S., BACCHIEGA, D., FIORELLI, G. & CAPPELLINI, M. D. 1995. Genetic interactions in thalassemia intermedia: analysis of beta-mutations, alpha-genotype, gamma-promoters, and beta-LCR hypersensitive sites 2 and 4 in Italian patients. *Am J Hematol,* 48**,** 82-7.

CAMERON, P., FULLER, C. K., DONOHOUE, P. D., JONES, B. N., THOMPSON, M. S., CARTER, M. M., GRADIA, S., VIDAL, B., GARNER, E., SLORACH, E. M., LAU, E., BANH, L. M., LIED, A. M., EDWARDS, L. S., SETTLE, A. H., CAPURSO, D., LLACA, V., DESCHAMPS, S., CIGAN, M., YOUNG, J. K. & MAY, A. P. 2017. Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat Methods,* 14**,** 600-606.

CANVER, M. C., SMITH, E. C., SHER, F., PINELLO, L., SANJANA, N. E., SHALEM, O., CHEN, D. D., SCHUPP, P. G., VINJAMUR, D. S., GARCIA, S. P., LUC, S., KURITA, R., NAKAMURA, Y., FUJIWARA, Y., MAEDA, T., YUAN, G. C., ZHANG, F., ORKIN, S. H. & BAUER, D. E. 2015. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature,* 527**,** 192-7.

CAO, A. & GALANELLO, R. 2010. Beta-thalassemia. *Genet Med,* 12**,** 61-76.

CAPPELLINI, M. D., PORTER, J., ORIGA, R., FORNI, G. L., VOSKARIDOU, E., GALACTEROS, F., TAHER, A. T., ARLET, J. B., RIBEIL, J. A., GARBOWSKI, M., GRAZIADEI, G., BROUZES, C., SEMERARO, M., LAADEM, A., MITEVA, D., ZOU, J., SUNG, V., ZINGER, T., ATTIE, K. M. & HERMINE, O. 2019. Sotatercept, a novel transforming growth factor beta ligand trap, improves anemia in beta-thalassemia: a phase II, open-label, dose-finding study. *Haematologica,* 104**,** 477-484.

CAPPELLINI, M. D. & TAHER, A. T. 2021. The use of luspatercept for thalassemia in adults. *Blood Advances,* 5**,** 326-333.

CAPPELLINI, M. D., VIPRAKASIT, V., TAHER, A. T., GEORGIEV, P., KUO, K. H. M., COATES, T., VOSKARIDOU, E., LIEW, H. K., PAZGAL-KOBROWSKI, I., FORNI, G. L., PERROTTA, S., KHELIF, A., LAL, A., KATTAMIS, A., VLACHAKI, E., ORIGA, R., AYDINOK, Y., BEJAOUI, M., HO, P. J., CHEW, L. P., BEE, P. C., LIM, S. M., LU, M. Y., TANTIWORAWIT, A., GANEVA, P., GERCHEVA, L., SHAH, F., NEUFELD, E. J., THOMPSON, A., LAADEM, A., SHETTY, J. K., ZOU, J., ZHANG, J., MITEVA, D., ZINGER, T., LINDE, P. G., SHERMAN, M. L., HERMINE, O., PORTER, J., PIGA, A. & INVESTIGATORS, B. 2020. A Phase 3 Trial of Luspatercept in Patients with Transfusion-Dependent beta-Thalassemia. *N Engl J Med,* 382**,** 1219-1231.

CARRERAS, E., DUFOUR, C., MOHTY, M. & KRÖGER, N. 2019. The EBMT handbook : hematopoietic stem cell transplantation and cellular therapies. 7th ed. Cham, Switzerland: Springer,.

CAVAZZANA-CALVO, M., PAYEN, E., NEGRE, O., WANG, G., HEHIR, K., FUSIL, F., DOWN, J., DENARO, M., BRADY, T., WESTERMAN, K., CAVALLESCO, R., GILLET-LEGRAND, B., CACCAVELLI, L., SGARRA, R., MAOUCHE-CHRETIEN, L., BERNAUDIN, F., GIROT, R., DORAZIO, R., MULDER, G. J., POLACK, A., BANK, A., SOULIER, J., LARGHERO, J., KABBARA, N., DALLE, B., GOURMEL, B., SOCIE, G., CHRETIEN, S., CARTIER, N., AUBOURG, P., FISCHER, A., CORNETTA, K., GALACTEROS, F., BEUZARD, Y., GLUCKMAN, E., BUSHMAN, F., HACEIN-BEY-ABINA, S. & LEBOULCH, P. 2010. Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature,* 467**,** 318-22.

CENTIS, F., TABELLINI, L., LUCARELLI, G., BUFFI, O., TONUCCI, P., PERSINI, B., ANNIBALI, M., EMILIANI, R., ILIESCU, A., RAPA, S., ROSSI, R., MA, L., ANGELUCCI, E. & SCHRIER, S. L. 2000. The importance of erythroid expansion in determining the extent of apoptosis in erythroid precursors in patients with beta-thalassemia major. *Blood,* 96**,** 3624-9.

CHAIBUNRUANG, A., KARNPEAN, R., FUCHAROEN, G. & FUCHAROEN, S. 2014. Genetic heterogeneity of hemoglobin AEBart's disease: a large cohort data from a single referral center in northeast Thailand. *Blood Cells Mol Dis,* 52**,** 176-80.

CHAISUE, C., KITCHAROEN, S., WILAIRAT, P., JETSRISUPARB, A., FUCHAROEN, G. & FUCHAROEN, S. 2007. alpha/beta-Globin mRNA ratio determination by multiplex quantitative real-time reverse transcription-polymerase chain reaction as an indicator of globin gene function. *Clin Biochem,* 40**,** 1373-7.

CHAKRABARTI, A. M., HENSER-BROWNHILL, T., MONSERRAT, J., POETSCH, A. R., LUSCOMBE, N. M. & SCAFFIDI, P. 2019. Target-Specific Precision of CRISPR-Mediated Genome Editing. *Mol Cell,* 73**,** 699-713 e6.

CHAN, V., CHAN, V. W., TANG, M., LAU, K., TODD, D. & CHAN, T. K. 1997. Molecular defects in Hb H hydrops fetalis. *Br J Haematol,* 96**,** 224-8.

CHAROENKWAN, P., TEERACHAIMAHIT, P. & SANGUANSERMSRI, T. 2014. The correlation of alpha-globin gene mutations and the XmnI polymorphism with clinical severity of Hb E/beta-thalassemia. *Hemoglobin,* 38**,** 335-8.

CHAUDHARI, H. G., PENTERMAN, J., WHITTON, H. J., SPENCER, S. J., FLANAGAN, N., LEI ZHANG, M. C., HUANG, E., KHEDKAR, A. S., TOOMEY, J. M., SHEARER, C. A., NEEDHAM, A. W., HO, T. W., KULMAN, J. D., CRADICK, T. J. & KERNYTSKY, A. 2020. Evaluation of Homology-Independent CRISPR-Cas9 Off-Target Assessment Methods. *CRISPR J,* 3**,** 440-453.

CHECHIK, L., MARTIN, O. & SOUTOGLOU, E. 2020. Genome Editing Fidelity in the Context of DNA Sequence and Chromatin Structure. *Front Cell Dev Biol,* 8**,** 319.

CHEN, F. E., OOI, C., HA, S. Y., CHEUNG, B. M. Y., TODD, D., LIANG, R., CHAN, T. K. & CHAN, V. 2000. Genetic and Clinical Features of Hemoglobin H Disease in Chinese Patients. *New England Journal of Medicine,* 343**,** 544-550.

CHENG, Q., WEI, T., FARBIAK, L., JOHNSON, L. T., DILLIARD, S. A. & SIEGWART, D. J. 2020. Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR-Cas gene editing. *Nat Nanotechnol,* 15**,** 313-320.

CHERNOFF, A. I., MINNICH, V. & CHONGCHAREONSUK, S. 1954. Hemoglobin E, a hereditary abnormality of human hemoglobin. *Science,* 120**,** 605-6.

CLYNES, D., JELINSKA, C., XELLA, B., AYYUB, H., SCOTT, C., MITSON, M., TAYLOR, S., HIGGS, D. R. & GIBBONS, R. J. 2015. Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX. *Nat Commun,* 6**,** 7538.

COELHO, A., PICANCO, 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 Mol Dis,* 45**,** 147-53.

COHEN, D. M., LIM, H. W., WON, K. J. & STEGER, D. J. 2018. Shared nucleotide flanks confer transcriptional competency to bZip core motifs. *Nucleic Acids Res,* 46**,** 8371-8384.

COLAH, R., GORAKSHAKAR, A. & NADKARNI, A. 2010. Global burden, distribution and prevention of beta-thalassemias and hemoglobin E disorders. *Expert Rev Hematol,* 3**,** 103-17.

CONCORDET, J. P. & HAEUSSLER, M. 2018. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res,* 46**,** W242-W245.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

CONSORTIUM, E. P. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

CONTICELLO, S. G. 2008. The AID/APOBEC family of nucleic acid mutators. *Genome Biol,* 9**,** 229.

COX, D. B., PLATT, R. J. & ZHANG, F. 2015. Therapeutic genome editing: prospects and challenges. *Nat Med,* 21**,** 121-31.

CRADDOCK, C. F., VYAS, P., SHARPE, J. A., AYYUB, H., WOOD, W. G. & HIGGS, D. R. 1995. Contrasting effects of alpha and beta globin regulatory elements on chromatin structure may be related to their different chromosomal environments. *EMBO J,* 14**,** 1718-26.

CRAWFORD, G. E., DAVIS, S., SCACHERI, P. C., RENAUD, G., HALAWI, M. J., ERDOS, M. R., GREEN, R., MELTZER, P. S., WOLFSBERG, T. G. & COLLINS, F. S. 2006. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat Methods,* 3**,** 503-9.

CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M., SHARP, P. A., BOYER, L. A., YOUNG, R. A. & JAENISCH, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A,* 107**,** 21931-6.

CRISPINO, J. D. & HORWITZ, M. S. 2017. GATA factor mutations in hematologic disease. *Blood,* 129**,** 2103-2110.

CROMER, M. K., CAMARENA, J., MARTIN, R. M., LESCH, B. J., VAKULSKAS, C. A., BODE, N. M., KURGAN, G., COLLINGWOOD, M. A., RETTIG, G. R., BEHLKE, M. A., LEMGART, V. T., ZHANG, Y., GOYAL, A., ZHAO, F., PONCE, E., SRIFA, W., BAK, R. O., UCHIDA, N., MAJETI, R., SHEEHAN, V. A., TISDALE, J. F., DEVER, D. P. & PORTEUS, M. H. 2021. Gene replacement of alpha-globin with beta-globin restores hemoglobin balance in beta-thalassemia-derived hematopoietic stem and progenitor cells. *Nat Med,* 27**,** 677-687.

CULLOT, G., BOUTIN, J., TOUTAIN, J., PRAT, F., PENNAMEN, P., ROORYCK, C., TEICHMANN, M., ROUSSEAU, E., LAMRISSI-GARCIA, I., GUYONNET-DUPERAT, V., BIBEYRAN, A., LALANNE, M., PROUZET-MAULEON, V., TURCQ, B., GED, C., BLOUIN, J. M., RICHARD, E., DABERNAT, S., MOREAU-GAUDRY, F. & BEDEL, A. 2019. CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nat Commun,* 10**,** 1136.

DAVIES, J. O., TELENIUS, J. M., MCGOWAN, S. J., ROBERTS, N. A., TAYLOR, S., HIGGS, D. R. & HUGHES, J. R. 2016. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods,* 13**,** 74-80.

DE GOBBI, M., ANGUITA, E., HUGHES, J., SLOANE-STANLEY, J. A., SHARPE, J. A., KOCH, C. M., DUNHAM, I., GIBBONS, R. J., WOOD, W. G. & HIGGS, D. R. 2007. Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. *Blood,* 110**,** 4503-10.

DE GOBBI, M., GARRICK, D., LYNCH, M., VERNIMMEN, D., HUGHES, J. R., GOARDON, N., LUC, S., LOWER, K. M., SLOANE-STANLEY, J. A., PINA, C., SONEJI, S., RENELLA, R., ENVER, T., TAYLOR, S., JACOBSEN, S. E., VYAS, P., GIBBONS, R. J. & HIGGS, D. R. 2011. Generation of bivalent chromatin domains during cell fate decisions. *Epigenetics Chromatin,* 4**,** 9.

DE GOBBI, M., VIPRAKASIT, V., HUGHES, J. R., FISHER, C., BUCKLE, V. J., AYYUB, H., GIBBONS, R. J., VERNIMMEN, D., YOSHINAGA, Y., DE JONG, P., CHENG, J. F., RUBIN, E. M., WOOD, W. G., BOWDEN, D. & HIGGS, D. R. 2006. A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science,* 312**,** 1215-7.

DELTCHEVA, E., CHYLINSKI, K., SHARMA, C. M., GONZALES, K., CHAO, Y., PIRZADA, Z. A., ECKERT, M. R., VOGEL, J. & CHARPENTIER, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature,* 471**,** 602-7.

DENG, W., LEE, J., WANG, H., MILLER, J., REIK, A., GREGORY, P. D., DEAN, A. & BLOBEL, G. A. 2012. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell,* 149**,** 1233-44.

DEVER, D. P., BAK, R. O., REINISCH, A., CAMARENA, J., WASHINGTON, G., NICOLAS, C. E., PAVEL-DINU, M., SAXENA, N., WILKENS, A. B., MANTRI, S., UCHIDA, N., HENDEL, A., NARLA, A., MAJETI, R., WEINBERG, K. I. & PORTEUS, M. H. 2016. CRISPR/Cas9 beta-globin gene targeting in human haematopoietic stem cells. *Nature,* 539**,** 384-389.

DEWITT, M. A., MAGIS, W., BRAY, N. L., WANG, T., BERMAN, J. R., URBINATI, F., HEO, S. J., MITROS, T., MUNOZ, D. P., BOFFELLI, D., KOHN, D. B., WALTERS, M. C., CARROLL, D., MARTIN, D. I. & CORN, J. E. 2016. Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci Transl Med,* 8**,** 360ra134.

DHANYA, R., SEDAI, A., ANKITA, K., PARMAR, L., AGARWAL, R. K., HEGDE, S., RAMASWAMI, G., GOWDA, A., GIRIJA, S., GUJJAL, P., PUSHPA, H., RAMAIAH, J. D., KARRI, C., JALI, S., TALLUR, N. R., SHENOY, U. V., PINTO, D., RAMPRAKASH, S., RAGHURAM, C. P., TRIVEDI, D., CAO, X. & FAULKNER, L. 2020. Life expectancy and risk factors for early death in patients with severe thalassemia syndromes in South India. *Blood Adv,* 4**,** 1448-1457.

DINARDO, C. D. & CORTES, J. E. 2016. Mutations in AML: prognostic and therapeutic implications. *Hematology Am Soc Hematol Educ Program,* 2016**,** 348-355.

DOULATOV, S., NOTTA, F., EPPERT, K., NGUYEN, L. T., OHASHI, P. S. & DICK, J. E. 2010. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat Immunol,* 11**,** 585-93.

DOWNES, D. J., SCHWESSINGER, R., HILL, S. J., NUSSBAUM, L., SCOTT, C., GOSDEN, M. E., HIRSCHFELD, P. P., TELENIUS, J. M., EIJSBOUTS, C. Q., MCGOWAN, S. J., CUTLER, A. J., KERRY, J., DAVIES, J. L., DENDROU, C. A., INSHAW, J. R. J., LARKE, M. S. C., OUDELAAR, A. M., BOZHILOV, Y., KING, A. J., BROWN, R. C., SUCIU, M. C., DAVIES, J. O. J., HUBLITZ, P., FISHER, C., KURITA, R., NAKAMURA, Y., LUNTER, G., TAYLOR, S., BUCKLE, V. J., TODD, J. A., HIGGS, D. R. & HUGHES, J. R. 2019. An integrated platform to systematically identify causal variants and genes for polygenic human traits. *bioRxiv***,** 813618.

DULMOVITS, B. M., APPIAH-KUBI, A. O., PAPOIN, J., HALE, J., HE, M., AL-ABED, Y., DIDIER, S., GOULD, M., HUSAIN-KRAUTTER, S., SINGH, S. A., CHAN, K. W., VLACHOS, A., ALLEN, S. L., TAYLOR, N., MARAMBAUD, P., AN, X., GALLAGHER, P. G., MOHANDAS, N., LIPTON, J. M., LIU, J. M. & BLANC, L. 2016. Pomalidomide reverses gamma-globin silencing through the transcriptional reprogramming of adult hematopoietic progenitors. *Blood,* 127**,** 1481-92.

DUSSIOT, M., MACIEL, T. T., FRICOT, A., CHARTIER, C., NEGRE, O., VEIGA, J., GRAPTON, D., PAUBELLE, E., PAYEN, E., BEUZARD, Y., LEBOULCH, P., RIBEIL, J. A., ARLET, J. B., COTE, F., COURTOIS, G., GINZBURG, Y. Z., DANIEL, T. O., CHOPRA, R., SUNG, V., HERMINE, O. & MOURA, I. C. 2014. An activin receptor IIA ligand trap corrects ineffective erythropoiesis in beta-thalassemia. *Nat Med,* 20**,** 398-407.

ELWOOD, N. J., ZOGOS, H., PEREIRA, D. S., DICK, J. E. & BEGLEY, C. G. 1998. Enhanced megakaryocyte and erythroid development from normal human CD34(+) cells: consequence of enforced expression of SCL. *Blood,* 91**,** 3756-65.

EMBURY, S. H., MILLER, J. A., DOZY, A. M., KAN, Y. W., CHAN, V. & TODD, D. 1980. Two different molecular organizations account for the single alpha-globin gene of the alpha-thalassemia-2 genotype. *J Clin Invest,* 66**,** 1319-25.

ENACHE, O. M., RENDO, V., ABDUSAMAD, M., LAM, D., DAVISON, D., PAL, S., CURRIMJEE, N., HESS, J., PANTEL, S., NAG, A., THORNER, A. R., DOENCH, J. G., VAZQUEZ, F., BEROUKHIM, R., GOLUB, T. R. & BEN-DAVID, U. 2020. Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nat Genet,* 52**,** 662-668.

ESRICK, E. B., LEHMANN, L. E., BIFFI, A., ACHEBE, M., BRENDEL, C., CIUCULESCU, M. F., DALEY, H., MACKINNON, B., MORRIS, E., FEDERICO, A., ABRISS, D., BOARDMAN, K., KHELLADI, R., SHAW, K., NEGRE, H., NEGRE, O., NIKIFOROW, S., RITZ, J., PAI, S. Y., LONDON, W. B., DANSEREAU, C., HEENEY, M. M., ARMANT, M., MANIS, J. P. & WILLIAMS, D. A. 2020. Post-Transcriptional Genetic Silencing of BCL11A to Treat Sickle Cell Disease. *N Engl J Med*.

FENG, L., ZHOU, S., GU, L., GELL, D. A., MACKAY, J. P., WEISS, M. J., GOW, A. J. & SHI, Y. 2005. Structure of oxidized alpha-haemoglobin bound to AHSP reveals a protective mechanism for haem. *Nature,* 435**,** 697-701.

FERRAO, J., SILVA, M., GONCALVES, L., GOMES, S., LOUREIRO, P., COELHO, A., MIRANDA, A., SEUANES, F., REIS, A. B., PINA, F., MAIA, R., KJOLLERSTROM, P., MONTEIRO, E., LACERDA, J. F., LAVINHA, J., GONCALVES, J. & FAUSTINO, P. 2017. Widening the spectrum of deletions and molecular mechanisms underlying alpha-thalassemia. *Ann Hematol,* 96**,** 1921-1929.

FERRARI, G., THRASHER, A. J. & AIUTI, A. 2021. Gene therapy using haematopoietic stem and progenitor cells. *Nat Rev Genet,* 22**,** 216-234.

FESSAS, P., LOUKOPOULOS, D. & KALTSOYA, A. 1966. Peptide analysis of the inclusions of erythroid cells in beta-thalassemia. *Biochim Biophys Acta,* 124**,** 430-2.

FIBACH, E., BURKE, L. P., SCHECHTER, A. N., NOGUCHI, C. T. & RODGERS, G. P. 1993. Hydroxyurea increases fetal hemoglobin in cultured erythroid cells derived from normal individuals and patients with sickle cell anemia or beta-thalassemia. *Blood,* 81**,** 1630-5.

FLINT, J., CRADDOCK, C. F., VILLEGAS, A., BENTLEY, D. P., WILLIAMS, H. J., GALANELLO, R., CAO, A., WOOD, W. G., AYYUB, H. & HIGGS, D. R. 1994. <Healing of broken human chromosomes by the addition of telomeric repeats>. *Am J Hum Genet,* 55**,** 505-12.

FLINT, J., HARDING, R. M., BOYCE, A. J. & CLEGG, J. B. 1998. The population genetics of the haemoglobinopathies. *Baillieres Clin Haematol,* 11**,** 1-51.

FLINT, J., ROCHETTE, J., CRADDOCK, C. F., DODE, C., VIGNES, B., HORSLEY, S. W., KEARNEY, L., BUCKLE, V. J., AYYUB, H. & HIGGS, D. R. 1996. Chromosomal stabilisation by a subtelomeric rearrangement involving two closely related Alu elements. *Hum Mol Genet,* 5**,** 1163-9.

FLINT, J., THOMAS, K., MICKLEM, G., RAYNHAM, H., CLARK, K., DOGGETT, N. A., KING, A. & HIGGS, D. R. 1997. The relationship between chromosome structure and function at a human telomeric region. *Nat Genet,* 15**,** 252-7.

FORGET, B. G. & HARDISON, R. C. 2009. The Normal Structure and Regulation of Human Globin Gene Clusters. *In:* FORGET, B. G., WEATHERALL, D. J., HIGGS, D. R. & STEINBERG, M. H. (eds.) *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management.* 2 ed. Cambridge: Cambridge University Press.

FORRESTER, W. C., NOVAK, U., GELINAS, R. & GROUDINE, M. 1989. Molecular analysis of the human beta-globin locus activation region. *Proc Natl Acad Sci U S A,* 86**,** 5439-43.

FRANGOUL, H., ALTSHULER, D., CAPPELLINI, M. D., CHEN, Y. S., DOMM, J., EUSTACE, B. K., FOELL, J., DE LA FUENTE, J., GRUPP, S., HANDGRETINGER, R., HO, T. W., KATTAMIS, A., KERNYTSKY, A., LEKSTROM-HIMES, J., LI, A. M., LOCATELLI, F., MAPARA, M. Y., DE MONTALEMBERT, M., RONDELLI, D., SHARMA, A., SHETH, S., SONI, S., STEINBERG, M. H., WALL, D., YEN, A. & CORBACIOGLU, S. 2021. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and beta-Thalassemia. *N Engl J Med,* 384**,** 252-260.

FRIEDMANN, T. & ROBLIN, R. 1972. Gene therapy for human genetic disease? *Science,* 175**,** 949-55.

FRISCHER, H. & BOWMAN, J. 1975. Hemoglobin E, an oxidatively unstable mutation. *J Lab Clin Med,* 85**,** 531-9.

FRONTELO, P., MANWANI, D., GALDASS, M., KARSUNKY, H., LOHMANN, F., GALLAGHER, P. G. & BIEKER, J. J. 2007. Novel role for EKLF in megakaryocyte lineage commitment. *Blood,* 110**,** 3871-80.

FU, Y., FODEN, J. A., KHAYTER, C., MAEDER, M. L., REYON, D., JOUNG, J. K. & SANDER, J. D. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol,* 31**,** 822-6.

FUCHAROEN, S. & WEATHERALL, D. J. 2012. The hemoglobin E thalassemias. *Cold Spring Harb Perspect Med,* 2.

FUJISAWA, T. & FILIPPAKOPOULOS, P. 2017. Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nat Rev Mol Cell Biol,* 18**,** 246-262.

FURBETTA, M., TUVERI, T., ROSATELLI, C., ANGIUS, A., FALCHI, A. M., COSSU, P., MELONI, A., GIAGU, N. & CAO, A. 1983. Molecular mechanism accounting for milder types of thalassemia major. *J Pediatr,* 103**,** 35-9.

GALANELLO, R., DESSI, E., MELIS, M. A., ADDIS, M., SANNA, M. A., ROSATELLI, C., ARGIOLU, F., GIAGU, N., TURCO, M. P., CACACE, E. & ET AL. 1989. Molecular analysis of beta zero-thalassemia intermedia in Sardinia. *Blood,* 74**,** 823-7.

GALANELLO, R., SANNA, S., PERSEU, L., SOLLAINO, M. C., SATTA, S., LAI, M. E., BARELLA, S., UDA, M., USALA, G., ABECASIS, G. R. & CAO, A. 2009. Amelioration of Sardinian beta0 thalassemia by genetic modifiers. *Blood,* 114**,** 3935-7.

GALAS, D. J. & SCHMITZ, A. 1978. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res,* 5**,** 3157-70.

GARRICK, D., DE GOBBI, M., SAMARA, V., RUGLESS, M., HOLLAND, M., AYYUB, H., LOWER, K., SLOANE-STANLEY, J., GRAY, N., KOCH, C., DUNHAM, I. & HIGGS, D. R. 2008. The role of the polycomb complex in silencing alpha-globin gene expression in nonerythroid cells. *Blood,* 112**,** 3889-99.

GASIUNAS, G., BARRANGOU, R., HORVATH, P. & SIKSNYS, V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A,* 109**,** E2579-86.

GAUDELLI, N. M., KOMOR, A. C., REES, H. A., PACKER, M. S., BADRAN, A. H., BRYSON, D. I. & LIU, D. R. 2017. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature,* 551**,** 464-471.

GAUDELLI, N. M., LAM, D. K., REES, H. A., SOLA-ESTEVES, N. M., BARRERA, L. A., BORN, D. A., EDWARDS, A., GEHRKE, J. M., LEE, S. J., LIQUORI, A. J., MURRAY, R., PACKER, M. S., RINALDI, C., SLAYMAKER, I. M., YEN, J., YOUNG, L. E. & CIARAMELLA, G. 2020. Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat Biotechnol,* 38**,** 892-900.

GEHRKE, J. M., CERVANTES, O., CLEMENT, M. K., WU, Y., ZENG, J., BAUER, D. E., PINELLO, L. & JOUNG, J. K. 2018. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol,* 36**,** 977-982.

GEORGE, L. A., SULLIVAN, S. K., GIERMASZ, A., RASKO, J. E. J., SAMELSON-JONES, B. J., DUCORE, J., CUKER, A., SULLIVAN, L. M., MAJUMDAR, S., TEITEL, J., MCGUINN, C. E., RAGNI, M. V., LUK, A. Y., HUI, D., WRIGHT, J. F., CHEN, Y., LIU, Y., WACHTEL, K., WINTERS, A., TIEFENBACHER, S., ARRUDA, V. R., VAN DER LOO, J. C. M., ZELENAIA, O., TAKEFMAN, D., CARR, M. E., COUTO, L. B., ANGUELA, X. M. & HIGH, K. A. 2017. Hemophilia B Gene Therapy with a High-Specific-Activity Factor IX Variant. *N Engl J Med,* 377**,** 2215-2227.

GIARDINE, B., BORG, J., VIENNAS, E., PAVLIDIS, C., MORADKHANI, K., JOLY, P., BARTSAKOULIA, M., RIEMER, C., MILLER, W., TZIMAS, G., WAJCMAN, H., HARDISON, R. C. & PATRINOS, G. P. 2014. Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. *Nucleic Acids Res,* 42**,** D1063-9.

GIBBONS, R. J. 2012. alpha-Thalassemia, mental retardation, and myelodysplastic syndrome. *Cold Spring Harb Perspect Med,* 2.

GIBBONS, R. J., MCDOWELL, T. L., RAMAN, S., O'ROURKE, D. M., GARRICK, D., AYYUB, H. & HIGGS, D. R. 2000. Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet,* 24**,** 368-71.

GIBBONS, R. J., PELLAGATTI, A., GARRICK, D., WOOD, W. G., MALIK, N., AYYUB, H., LANGFORD, C., BOULTWOOD, J., WAINSCOAT, J. S. & HIGGS, D. R. 2003. Identification of acquired somatic mutations in the gene encoding chromatin-remodeling factor ATRX in the alpha-thalassemia myelodysplasia syndrome (ATMDS). *Nat Genet,* 34**,** 446-9.

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 alpha-thalassemia (ATR-X syndrome). *Cell,* 80**,** 837-45.

GILLESPIE, M. A., PALII, C. G., SANCHEZ-TALTAVULL, D., SHANNON, P., LONGABAUGH, W. J. R., DOWNES, D. J., SIVARAMAN, K., ESPINOZA, H. M., HUGHES, J. R., PRICE, N. D., PERKINS, T. J., RANISH, J. A. & BRAND, M. 2020. Absolute Quantification of Transcription Factors Reveals Principles of Gene Regulation in Erythropoiesis. *Mol Cell,* 78**,** 960-974 e11.

GOEL, R., TOBIAN, A. A. R. & SHAZ, B. H. 2019. Noninfectious transfusion-associated adverse events and their mitigation strategies. *Blood,* 133**,** 1831-1839.

GOLDBERG, A. D., BANASZYNSKI, L. A., NOH, K. M., LEWIS, P. W., ELSAESSER, S. J., STADLER, S., DEWELL, S., LAW, M., GUO, X., LI, X., WEN, D., CHAPGIER, A., DEKELVER, R. C., MILLER, J. C., LEE, Y. L., BOYDSTON, E. A., HOLMES, M. C., GREGORY, P. D., GREALLY, J. M., RAFII, S., YANG, C., SCAMBLER, P. J., GARRICK, D., GIBBONS, R. J., HIGGS, D. R., CRISTEA, I. M., URNOV, F. D., ZHENG, D. & ALLIS, C. D. 2010. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell,* 140**,** 678-91.

GOOSSENS, M., DOZY, A. M., EMBURY, S. H., ZACHARIADES, Z., HADJIMINAS, M. G., STAMATOYANNOPOULOS, G. & KAN, Y. W. 1980. Triplicated alpha-globin loci in humans. *Proc Natl Acad Sci U S A,* 77**,** 518-21.

GOOSSENS, M., LEE, K. Y., LIEBHABER, S. A. & KAN, Y. W. 1982. Globin structural mutant alpha 125Leu leads to Pro is a novel cause of alpha-thalassaemia. *Nature,* 296**,** 864-5.

GOTEA, V., VISEL, A., WESTLUND, J. M., NOBREGA, M. A., PENNACCHIO, L. A. & OVCHARENKO, I. 2010. Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res,* 20**,** 565-77.

GOURDON, G., SHARPE, J. A., WELLS, D., WOOD, W. G. & HIGGS, D. R. 1993. Analysis of a 70 kb segment of DNA containing the human zeta and alpha-globin genes linked to their regulatory element (HS-40) in transgenic mice. *Nucleic Acids Res,* 22**,** 4139-47.

GRANT, C. E., BAILEY, T. L. & NOBLE, W. S. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics,* 27**,** 1017-8.

GREVET, J. D., LAN, X., HAMAGAMI, N., EDWARDS, C. R., SANKARANARAYANAN, L., JI, X., BHARDWAJ, S. K., FACE, C. J., POSOCCO, D. F., ABDULMALIK, O., KELLER, C. A., GIARDINE, B., SIDOLI, S., GARCIA, B. A., CHOU, S. T., LIEBHABER, S. A., HARDISON, R. C., SHI, J. & BLOBEL, G. A. 2018. Domain-focused CRISPR screen identifies HRI as a fetal hemoglobin regulator in human erythroid cells. *Science,* 361**,** 285-290.

GRIFFITHS, R. E., KUPZIG, S., COGAN, N., MANKELOW, T. J., BETIN, V. M., TRAKARNSANGA, K., MASSEY, E. J., LANE, J. D., PARSONS, S. F. & ANSTEE, D. J. 2012. Maturing reticulocytes internalize plasma membrane in glycophorin A-containing vesicles that fuse with autophagosomes before exocytosis. *Blood,* 119**,** 6296-306.

GROSVELD, F., VAN ASSENDELFT, G. B., GREAVES, D. R. & KOLLIAS, G. 1987. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell,* 51**,** 975-85.

GRUNEWALD, J., ZHOU, R., GARCIA, S. P., IYER, S., LAREAU, C. A., ARYEE, M. J. & JOUNG, J. K. 2019a. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature,* 569**,** 433-437.

GRUNEWALD, J., ZHOU, R., IYER, S., LAREAU, C. A., GARCIA, S. P., ARYEE, M. J. & JOUNG, J. K. 2019b. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat Biotechnol,* 37**,** 1041-1048.

GUERRA, A., OIKONOMIDOU, P. R., SINHA, S., ZHANG, J., LO PRESTI, V., HAMILTON, C. R., BREDA, L., CASU, C., LA, P., MARTINS, A. C., SENDAMARAI, A. K., FLEMING, M. & RIVELLA, S. 2019. Lack of Gdf11 does not improve anemia or prevent the activity of RAP-536 in a mouse model of beta-thalassemia. *Blood,* 134**,** 568-572.

GUO, Y., MAHONY, S. & GIFFORD, D. K. 2012. High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. *PLoS Comput Biol,* 8**,** e1002638.

GURUNATHAN, A., TARANGO, C., MCGANN, P. T., NISS, O. & QUINN, C. T. 2020. Non-transfusion-dependent beta-Thalassemia Because of a Single beta-Thalassemia Mutation and Coinherited alpha-Globin Gene Triplication: Need for Increased Awareness to Prevent Incorrect and Delayed Diagnosis. *J Pediatr Hematol Oncol,* 42**,** e494-e496.

HAAPANIEMI, E., BOTLA, S., PERSSON, J., SCHMIERER, B. & TAIPALE, J. 2018. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med,* 24**,** 927-930.

HAASE, V. H. 2013. Regulation of erythropoiesis by hypoxia-inducible factors. *Blood Rev,* 27**,** 41-53.

HACEIN-BEY-ABINA, S., GARRIGUE, A., WANG, G. P., SOULIER, J., LIM, A., MORILLON, E., CLAPPIER, E., CACCAVELLI, L., DELABESSE, E., BELDJORD, K., ASNAFI, V., MACINTYRE, E., DAL CORTIVO, L., RADFORD, I., BROUSSE, N., SIGAUX, F., MOSHOUS, D., HAUER, J., BORKHARDT, A., BELOHRADSKY, B. H., WINTERGERST, U., VELEZ, M. C., LEIVA, L.,

SORENSEN, R., WULFFRAAT, N., BLANCHE, S., BUSHMAN, F. D., FISCHER, A. & CAVAZZANA-CALVO, M. 2008. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest,* 118**,** 3132-42.

HACEIN-BEY-ABINA, S., VON KALLE, C., SCHMIDT, M., LE DEIST, F., WULFFRAAT, N., MCINTYRE, E., RADFORD, I., VILLEVAL, J. L., FRASER, C. C., CAVAZZANA-CALVO, M. & FISCHER, A. 2003. A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med,* 348**,** 255-6.

HALDANE, J. 1949. Disease and evolution, Ricercha. Sci.

HALL, M. A., CURTIS, D. J., METCALF, D., ELEFANTY, A. G., SOURRIS, K., ROBB, L., GOTHERT, J. R., JANE, S. M. & BEGLEY, C. G. 2003. The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. *Proc Natl Acad Sci U S A,* 100**,** 992-7.

HARRIS, R. S., PETERSEN-MAHRT, S. K. & NEUBERGER, M. S. 2002. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell,* 10**,** 1247-53.

HARTEVELD, C. L. & HIGGS, D. R. 2010. Alpha-thalassaemia. *Orphanet J Rare Dis,* 5**,** 13.

HATTON, C. S., WILKIE, A. O., DRYSDALE, H. C., WOOD, W. G., VICKERS, M. A., SHARPE, J., AYYUB, H., PRETORIUS, I. M., BUCKLE, V. J. & HIGGS, D. R. 1990. Alpha-thalassemia caused by a large (62 kb) deletion upstream of the human alpha globin gene cluster. *Blood,* 76**,** 221-7.

HAY, D., HUGHES, J. R., BABBS, C., DAVIES, J. O., GRAHAM, B. J., HANSSEN, L. L., KASSOUF, M. T., OUDELAAR, A. M., SHARPE, J. A., SUCIU, M. C., TELENIUS, J., WILLIAMS, R., RODE, C., LI, P. S., PENNACCHIO, L. A., SLOANE-STANLEY, J. A., AYYUB, H., BUTLER, S., SAUKA-SPENGLER, T., GIBBONS, R. J., SMITH, A. J., WOOD, W. G. & HIGGS, D. R. 2016. Genetic dissection of the alpha-globin super-enhancer in vivo. *Nat Genet,* 48**,** 895-903.

HEINTZMAN, N. D., STUART, R. K., HON, G., FU, Y., CHING, C. W., HAWKINS, R. D., BARRERA, L. O., VAN CALCAR, S., QU, C., CHING, K. A., WANG, W., WENG, Z., GREEN, R. D., CRAWFORD, G. E. & REN, B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet,* 39**,** 311-8.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell,* 38**,** 576-89.

HEINZ, S., ROMANOSKI, C. E., BENNER, C. & GLASS, C. K. 2015. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol,* 16**,** 144-54.

HENLEY, M. J. & KOEHLER, A. N. 2021. Advances in targeting 'undruggable' transcription factors with small molecules. *Nat Rev Drug Discov,* 20**,** 669-688.

HEYER, W. D., EHMSEN, K. T. & LIU, J. 2010. Regulation of homologous recombination in eukaryotes. *Annu Rev Genet,* 44**,** 113-39.

HIGGS, D. R. 2009a. The Molecular Basis of α Thalassemia. *In:* FORGET, B. G., WEATHERALL, D. J., HIGGS, D. R. & STEINBERG, M. H. (eds.) *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management.* 2 ed. Cambridge: Cambridge University Press.

HIGGS, D. R. 2009b. The Pathophysiology and Clinical Features of α Thalassaemia. *In:* FORGET, B. G., WEATHERALL, D. J., HIGGS, D. R. & STEINBERG, M. H. (eds.) *Disorders*

of Hemoglobin: Genetics, Pathophysiology, and Clinical Management. 2 ed. Cambridge: Cambridge University Press.

HIGGS, D. R. 2013. The molecular basis of alpha-thalassemia. *Cold Spring Harb Perspect Med,* 3**,** a011718.

HIGGS, D. R., ENGEL, J. D. & STAMATOYANNOPOULOS, G. 2012. Thalassaemia. *Lancet,* 379**,** 373-83.

HIGGS, D. R., GARRICK, D., ANGUITA, E., DE GOBBI, M., HUGHES, J., MUERS, M., VERNIMMEN, D., LOWER, K., LAW, M., ARGENTARO, A., DEVILLE, M. A. & GIBBONS, R. 2005. Understanding alpha-globin gene regulation: Aiming to improve the management of thalassemia. *Ann N Y Acad Sci,* 1054**,** 92-102.

HIGGS, D. R., WOOD, W. G., JARMAN, A. P., SHARPE, J., LIDA, J., PRETORIUS, I. M. & AYYUB, H. 1990. A major positive regulatory region located far upstream of the human alpha-globin gene locus. *Genes Dev,* 4**,** 1588-601.

HIRSCH, R. E., SIBMOOH, N., FUCHAROEN, S. & FRIEDMAN, J. M. 2017. HbE/beta-Thalassemia and Oxidative Stress: The Key to Pathophysiological Mechanisms and Novel Therapeutics. *Antioxid Redox Signal,* 26**,** 794-813.

HO, P. J., HALL, G. W., LUO, L. Y., WEATHERALL, D. J. & THEIN, S. L. 1998. Beta-thalassaemia intermedia: is it possible consistently to predict phenotype from genotype? *Br J Haematol,* 100**,** 70-8.

HODGKINS, A., FARNE, A., PERERA, S., GREGO, T., PARRY-SMITH, D. J., SKARNES, W. C. & IYER, V. 2015. WGE: a CRISPR database for genome engineering. *Bioinformatics,* 31**,** 3078-80.

HORNUNG, V. & LATZ, E. 2010. Intracellular DNA recognition. *Nat Rev Immunol,* 10**,** 123-30.

HSU, P. D., SCOTT, D. A., WEINSTEIN, J. A., RAN, F. A., KONERMANN, S., AGARWALA, V., LI, Y., FINE, E. J., WU, X., SHALEM, O., CRADICK, T. J., MARRAFFINI, L. A., BAO, G. & ZHANG, F. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol,* 31**,** 827-32.

HU, J., LIU, J., XUE, F., HALVERSON, G., REID, M., GUO, A., CHEN, L., RAZA, A., GALILI, N., JAFFRAY, J., LANE, J., CHASIS, J. A., TAYLOR, N., MOHANDAS, N. & AN, X. 2013. Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. *Blood,* 121**,** 3246-53.

HU, J. H., MILLER, S. M., GEURTS, M. H., TANG, W., CHEN, L., SUN, N., ZEINA, C. M., GAO, X., REES, H. A., LIN, Z. & LIU, D. R. 2018. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature,* 556**,** 57-63.

HUA, P., BADAT, M., HANSSEN, L. L. P., HENTGES, L. D., CRUMP, N., DOWNES, D. J., JEZIORSKA, D. M., OUDELAAR, A. M., SCHWESSINGER, R., TAYLOR, S., MILNE, T. A., HUGHES, J. R., HIGGS, D. R. & DAVIES, J. O. J. 2021. Defining genome architecture at base-pair resolution. *Nature*.

HUA, P., ROY, N., DE LA FUENTE, J., WANG, G., THONGJUEA, S., CLARK, K., ROY, A., PSAILA, B., ASHLEY, N., HARRINGTON, Y., NERLOV, C., WATT, S. M., ROBERTS, I. & DAVIES, J. O. J. 2019. Single-cell analysis of bone marrow-derived CD34+ cells from children with sickle cell disease and thalassemia. *Blood,* 134**,** 2111-2115.

HUANG, L. Y., YAN, J. M., ZHOU, J. Y., LI, J., XIE, X. M. & LI, D. Z. 2017. A Severe Case of Hemoglobin H Disease due to Compound Heterozygosity for Deletion of the Major alpha-Globin Regulatory Element (MCS-R2) and alpha0-Thalassemia. *Acta Haematol,* 138**,** 61-64.

HUANG, P., PESLAK, S. A., LAN, X., KHANDROS, E., YANO, J. A., SHARMA, M., KELLER, C. A., GIARDINE, B., QIN, K., ABDULMALIK, O., HARDISON, R. C., SHI, J. & BLOBEL, G. A. 2020. The HRI-regulated transcription factor ATF4 activates BCL11A transcription to silence fetal hemoglobin expression. *Blood,* 135**,** 2121-2132.

HUGHES, J. R., CHENG, J. F., VENTRESS, N., PRABHAKAR, S., CLARK, K., ANGUITA, E., DE GOBBI, M., DE JONG, P., RUBIN, E. & HIGGS, D. R. 2005. Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. *Proc Natl Acad Sci U S A,* 102**,** 9830-5.

HUMBERT, O., RADTKE, S., SAMUELSON, C., CARRILLO, R. R., PEREZ, A. M., REDDY, S. S., LUX, C., PATTABHI, S., SCHEFTER, L. E., NEGRE, O., LEE, C. M., BAO, G., ADAIR, J. E., PETERSON, C. W., RAWLINGS, D. J., SCHARENBERG, A. M. & KIEM, H. P. 2019. Therapeutically relevant engraftment of a CRISPR-Cas9-edited HSC-enriched population with HbF reactivation in nonhuman primates. *Sci Transl Med,* 11.

HUNT, D. M., HIGGS, D. R., CLEGG, J. B., WEATHERBALL, D. J. & MARSH, G. W. 1980. Determination of alpha thalassaemia phenotypes by messenger RNA analysis. *Br J Haematol,* 45**,** 53-64.

IHRY, R. J., WORRINGER, K. A., SALICK, M. R., FRIAS, E., HO, D., THERIAULT, K., KOMMINENI, S., CHEN, J., SONDEY, M., YE, C., RANDHAWA, R., KULKARNI, T., YANG, Z., MCALLISTER, G., RUSS, C., REECE-HOYES, J., FORRESTER, W., HOFFMAN, G. R., DOLMETSCH, R. & KAYKAS, A. 2018. p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat Med,* 24**,** 939-946.

ITANO, H. A., BERGREN, W. R. & STURGEON, P. 1954. IDENTIFICATION OF A FOURTH ABNORMAL HUMAN HEMOGLOBIN. *Journal of the American Chemical Society,* 76**,** 2278-2278.

ITO, M., HIRAMATSU, H., KOBAYASHI, K., SUZUE, K., KAWAHATA, M., HIOKI, K., UEYAMA, Y., KOYANAGI, Y., SUGAMURA, K., TSUJI, K., HEIKE, T. & NAKAHATA, T. 2002. NOD/SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells. *Blood,* 100**,** 3175-82.

JAISWAL, S. & LIBBY, P. 2020. Clonal haematopoiesis: connecting ageing and inflammation in cardiovascular disease. *Nat Rev Cardiol,* 17**,** 137-144.

JANG, H. K., SONG, B., HWANG, G. H. & BAE, S. 2020. Current trends in gene recovery mediated by the CRISPR-Cas system. *Exp Mol Med,* 52**,** 1016-1027.

JARMAN, A. P., WOOD, W. G., SHARPE, J. A., GOURDON, G., AYYUB, H. & HIGGS, D. R. 1991. Characterization of the major regulatory element upstream of the human alpha-globin gene cluster. *Mol Cell Biol,* 11**,** 4679-89.

JEONG, J., JAGER, A., DOMIZI, P., PAVEL-DINU, M., GOJENOLA, L., IWASAKI, M., WEI, M. C., PAN, F., ZEHNDER, J. L., PORTEUS, M. H., DAVIS, K. L. & CLEARY, M. L. 2019. High-efficiency CRISPR induction of t(9;11) chromosomal translocations and acute leukemias in human blood stem cells. *Blood Adv,* 3**,** 2825-2835.

JEONG, Y. K., LEE, S., HWANG, G. H., HONG, S. A., PARK, S. E., KIM, J. S., WOO, J. S. & BAE, S. 2021. Adenine base editor engineering reduces editing of bystander cytosines. *Nat Biotechnol*.

JEONG, Y. K., SONG, B. & BAE, S. 2020. Current Status and Challenges of DNA Base Editing Tools. *Mol Ther,* 28**,** 1938-1952.

JEZIORSKA, D. M., MURRAY, R. J. S., DE GOBBI, M., GAENTZSCH, R., GARRICK, D., AYYUB, H., CHEN, T., LI, E., TELENIUS, J., LYNCH, M., GRAHAM, B., SMITH, A. J. H., LUND, J. N., HUGHES, J. R., HIGGS, D. R. & TUFARELLI, C. 2017. DNA methylation of intragenic

CpG islands depends on their transcriptional activity during differentiation and disease. *Proc Natl Acad Sci U S A,* 114**,** E7526-E7535.

JIANG, F., TAYLOR, D. W., CHEN, J. S., KORNFELD, J. E., ZHOU, K., THOMPSON, A. J., NOGALES, E. & DOUDNA, J. A. 2016. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science,* 351**,** 867-71.

JIN, S., ZONG, Y., GAO, Q., ZHU, Z., WANG, Y., QIN, P., LIANG, C., WANG, D., QIU, J. L., ZHANG, F. & GAO, C. 2019. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science,* 364**,** 292-295.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337**,** 816-21.

JOLMA, A., YIN, Y., NITTA, K. R., DAVE, K., POPOV, A., TAIPALE, M., ENGE, M., KIVIOJA, T., MORGUNOVA, E. & TAIPALE, J. 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature,* 527**,** 384-8.

JORE, M. M., LUNDGREN, M., VAN DUIJN, E., BULTEMA, J. B., WESTRA, E. R., WAGHMARE, S. P., WIEDENHEFT, B., PUL, U., WURM, R., WAGNER, R., BEIJER, M. R., BARENDREGT, A., ZHOU, K., SNIJDERS, A. P., DICKMAN, M. J., DOUDNA, J. A., BOEKEMA, E. J., HECK, A. J., VAN DER OOST, J. & BROUNS, S. J. 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol,* 18**,** 529-36.

KAGEY, M. H., NEWMAN, J. J., BILODEAU, S., ZHAN, Y., ORLANDO, D. A., VAN BERKUM, N. L., EBMEIER, C. C., GOOSSENS, J., RAHL, P. B., LEVINE, S. S., TAATJES, D. J., DEKKER, J. & YOUNG, R. A. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature,* 467**,** 430-5.

KALANTRI, S. A., RAY, R., CHOUDHURI, S., ROY, S. & BHATTACHARYYA, M. 2020. Key Determinants of Phenotypic Heterogeneity of Hb E/beta Thalassemia: A Comparative Study from Eastern India. *Indian J Hematol Blood Transfus,* 36**,** 123-128.

KAN, Y. W. & NATHAN, D. G. 1970. Mild thalassemia: the result of interactions of alpha and beta thalassemia genes. *J Clin Invest,* 49**,** 635-42.

KANAVAKIS, E., TRAEGER-SYNODINOS, J., LAFIONIATIS, S., LAZAROPOULOU, C., LIAKOPOULOU, T., PALEOLOGOS, G., METAXOTOU-MAVROMMATI, A., STAMOULAKATOU, A. & PAPASSOTIRIOU, I. 2004. A rare example that coinheritance of a severe form of beta-thalassemia and alpha-thalassemia interact in a "synergistic" manner to balance the phenotype of classic thalassemic syndromes. *Blood Cells Mol Dis,* 32**,** 319-24.

KANG, B. C., YUN, J. Y., KIM, S. T., SHIN, Y., RYU, J., CHOI, M., WOO, J. W. & KIM, J. S. 2018. Precision genome engineering through adenine base editing in plants. *Nat Plants,* 4**,** 427-431.

KATSUMURA, K. R., DEVILBISS, A. W., POPE, N. J., JOHNSON, K. D. & BRESNICK, E. H. 2013. Transcriptional mechanisms underlying hemoglobin synthesis. *Cold Spring Harb Perspect Med,* 3**,** a015412.

KELLEY, D. R., SNOEK, J. & RINN, J. L. 2016. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res,* 26**,** 990-9.

KHAMPHIKHAM, P., NUALKAEW, T., PONGPAKSUPASIN, P., KAEWSAKULTHONG, W., SONGDEJ, D., PAIBOONSUKWONG, K., ENGEL, J. D., HONGENG, S., FUCHAROEN, S., SRIPICHAI, O. & JEARAWIRIYAPAISARN, N. 2020. High-level induction of fetal

haemoglobin by pomalidomide in beta-thalassaemia/HbE erythroid progenitor cells. *Br J Haematol,* 189**,** e240-e245.

KHANDROS, E., HUANG, P., PESLAK, S. A., SHARMA, M., ABDULMALIK, O., GIARDINE, B. M., ZHANG, Z., KELLER, C. A., HARDISON, R. C. & BLOBEL, G. A. 2020. Understanding heterogeneity of fetal hemoglobin induction through comparative analysis of F and A erythroblasts. *Blood,* 135**,** 1957-1968.

KIEM, H. P., SELLERS, S., THOMASSON, B., MORRIS, J. C., TISDALE, J. F., HORN, P. A., HEMATTI, P., ADLER, R., KURAMOTO, K., CALMELS, B., BONIFACINO, A., HU, J., VON KALLE, C., SCHMIDT, M., SORRENTINO, B., NIENHUIS, A., BLAU, C. A., ANDREWS, R. G., DONAHUE, R. E. & DUNBAR, C. E. 2004. Long-term clinical and molecular follow-up of large animals receiving retrovirally transduced stem and progenitor cells: no progression to clonal hematopoiesis or leukemia. *Mol Ther,* 9**,** 389-95.

KIHM, A. J., KONG, Y., HONG, W., RUSSELL, J. E., ROUDA, S., ADACHI, K., SIMON, M. C., BLOBEL, G. A. & WEISS, M. J. 2002. An abundant erythroid protein that stabilizes free alpha-haemoglobin. *Nature,* 417**,** 758-63.

KIM, D., BAE, S., PARK, J., KIM, E., KIM, S., YU, H. R., HWANG, J., KIM, J. I. & KIM, J. S. 2015. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods,* 12**,** 237-43, 1 p following 243.

KIM, D., KIM, D. E., LEE, G., CHO, S. I. & KIM, J. S. 2019. Genome-wide target specificity of CRISPR RNA-guided adenine base editors. *Nat Biotechnol,* 37**,** 430-435.

KIM, D., LIM, K., KIM, S. T., YOON, S. H., KIM, K., RYU, S. M. & KIM, J. S. 2017a. Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat Biotechnol,* 35**,** 475-480.

KIM, H. & KIM, J. S. 2014. A guide to genome engineering with programmable nucleases. *Nat Rev Genet,* 15**,** 321-34.

KIM, S., KIM, D., CHO, S. W., KIM, J. & KIM, J. S. 2014. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res,* 24**,** 1012-9.

KIM, Y. B., KOMOR, A. C., LEVY, J. M., PACKER, M. S., ZHAO, K. T. & LIU, D. R. 2017b. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol,* 35**,** 371-376.

KING, A. J. & HIGGS, D. R. 2018. Potential new approaches to the management of the Hb Bart's hydrops fetalis syndrome: the most severe form of alpha-thalassemia. *Hematology Am Soc Hematol Educ Program,* 2018**,** 353-360.

KOBLAN, L. W., DOMAN, J. L., WILSON, C., LEVY, J. M., TAY, T., NEWBY, G. A., MAIANTI, J. P., RAGURAM, A. & LIU, D. R. 2018. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol,* 36**,** 843-846.

KOBLAN, L. W., ERDOS, M. R., WILSON, C., CABRAL, W. A., LEVY, J. M., XIONG, Z. M., TAVAREZ, U. L., DAVISON, L. M., GETE, Y. G., MAO, X., NEWBY, G. A., DOHERTY, S. P., NARISU, N., SHENG, Q., KRILOW, C., LIN, C. Y., GORDON, L. B., CAO, K., COLLINS, F. S., BROWN, J. D. & LIU, D. R. 2021. In vivo base editing rescues Hutchinson-Gilford progeria syndrome in mice. *Nature,* 589**,** 608-614.

KOMOR, A. C., BADRAN, A. H. & LIU, D. R. 2017. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell,* 168**,** 20-36.

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**,** 420-4.

KOSICKI, M., TOMBERG, K. & BRADLEY, A. 2018. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol,* 36**,** 765-771.

KOUNTOURIS, P., LEDERER, C. W., FANIS, P., FELEKI, X., OLD, J. & KLEANTHOUS, M. 2014. IthaGenes: an interactive database for haemoglobin variations and epidemiology. *PLoS One,* 9**,** e103020.

KOWALSKI, P. S., RUDRA, A., MIAO, L. & ANDERSON, D. G. 2019. Delivering the Messenger: Advances in Technologies for Therapeutic mRNA Delivery. *Mol Ther,* 27**,** 710-728.

KRAUS, L. M. 1961. Formation of different haemoglobins in tissue culture of human bone marrow treated with human deoxyribonucleic acid. *Nature,* 192**,** 1055-7.

KRIVEGA, I., DALE, R. K. & DEAN, A. 2014. Role of LDB1 in the transition from chromatin looping to transcription activation. *Genes Dev,* 28**,** 1278-90.

KULESSA, H., FRAMPTON, J. & GRAF, T. 1995. GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. *Genes Dev,* 9**,** 1250-62.

KURITA, R., SUDA, N., SUDO, K., MIHARADA, K., HIROYAMA, T., MIYOSHI, H., TANI, K. & NAKAMURA, Y. 2013. Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PLoS One,* 8**,** e59890.

KURT, I. C., ZHOU, R., IYER, S., GARCIA, S. P., MILLER, B. R., LANGNER, L. M., GRUNEWALD, J. & JOUNG, J. K. 2021. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat Biotechnol,* 39**,** 41-46.

KUSCU, C., ARSLAN, S., SINGH, R., THORPE, J. & ADLI, M. 2014. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol,* 32**,** 677-83.

LA NASA, G., CAOCCI, G., EFFICACE, F., DESSI, C., VACCA, A., PIRAS, E., SANNA, M., MARCIAS, M., LITTERA, R., CARCASSI, C. & LUCARELLI, G. 2013. Long-term health-related quality of life evaluated more than 20 years after hematopoietic stem cell transplantation for thalassemia. *Blood,* 122**,** 2262-70.

LACAN, P., FRANCINA, A., PROME, D., DELAUNAY, J., GALACTEROS, F. & WAJCMAN, H. 1996. Hb Aubenas [beta 62(B8)Glu-->Gly]: a new variant normally synthesized, affecting the same codon as in Hb E. *Hemoglobin,* 20**,** 113-24.

LAN, X., REN, R., FENG, R., LY, L. C., LAN, Y., ZHANG, Z., ABOREDEN, N., QIN, K., HORTON, J. R., GREVET, J. D., MAYURANATHAN, T., ABDULMALIK, O., KELLER, C. A., GIARDINE, B., HARDISON, R. C., CROSSLEY, M., WEISS, M. J., CHENG, X., SHI, J. & BLOBEL, G. A. 2020. ZNF410 uniquely activates the NuRD component CHD4 to silence fetal hemoglobin expression. *bioRxiv***,** 2020.08.31.274324.

LANDRUM, M. J., LEE, J. M., BENSON, M., BROWN, G., CHAO, C., CHITIPIRALLA, S., GU, B., HART, J., HOFFMAN, D., HOOVER, J., JANG, W., KATZ, K., OVETSKY, M., RILEY, G., SETHI, A., TULLY, R., VILLAMARIN-SALOMON, R., RUBINSTEIN, W. & MAGLOTT, D. R. 2016. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res,* 44**,** D862-8.

LAPINAITE, A., KNOTT, G. J., PALUMBO, C. M., LIN-SHIAO, E., RICHTER, M. F., ZHAO, K. T., BEAL, P. A., LIU, D. R. & DOUDNA, J. A. 2020. DNA capture by a CRISPR-Cas9-guided adenine base editor. *Science,* 369**,** 566-571.

LAUER, J., SHEN, C. K. & MANIATIS, T. 1980. The chromosomal arrangement of human alpha-like globin genes: sequence homology and alpha-globin gene deletions. *Cell,* 20**,** 119-30.

LAURENTI, E. & GOTTGENS, B. 2018. From haematopoietic stem cells to complex differentiation landscapes. *Nature,* 553**,** 418-426.

LAW, M. J., LOWER, K. M., VOON, H. P., HUGHES, J. R., GARRICK, D., VIPRAKASIT, V., MITSON, M., DE GOBBI, M., MARRA, M., MORRIS, A., ABBOTT, A., WILDER, S. P., TAYLOR, S., SANTOS, G. M., CROSS, J., AYYUB, H., JONES, S., RAGOUSSIS, J., RHODES, D., DUNHAM, I., HIGGS, D. R. & GIBBONS, R. J. 2010. ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell,* 143**,** 367-78.

LAZZAROTTO, C. R., MALININ, N. L., LI, Y., ZHANG, R., YANG, Y., LEE, G., COWLEY, E., HE, Y., LAN, X., JIVIDEN, K., KATTA, V., KOLMAKOVA, N. G., PETERSEN, C. T., QI, Q., STRELCOV, E., MARAGH, S., KRENCIUTE, G., MA, J., CHENG, Y. & TSAI, S. Q. 2020. CHANGE-seq reveals genetic and epigenetic effects on CRISPR-Cas9 genome-wide activity. *Nat Biotechnol,* 38**,** 1317-1327.

LAZZAROTTO, C. R., NGUYEN, N. T., TANG, X., MALAGON-LOPEZ, J., GUO, J. A., ARYEE, M. J., JOUNG, J. K. & TSAI, S. Q. 2018. Defining CRISPR-Cas9 genome-wide nuclease activities with CIRCLE-seq. *Nat Protoc,* 13**,** 2615-2642.

LEE, J. K., JEONG, E., LEE, J., JUNG, M., SHIN, E., KIM, Y. H., LEE, K., JUNG, I., KIM, D., KIM, S. & KIM, J. S. 2018. Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun,* 9**,** 3048.

LEIBOWITZ, M. L., PAPATHANASIOU, S., DOERFLER, P. A., BLAINE, L. J., SUN, L., YAO, Y., ZHANG, C. Z., WEISS, M. J. & PELLMAN, D. 2021. Chromothripsis as an on-target consequence of CRISPR-Cas9 genome editing. *Nat Genet,* 53**,** 895-905.

LETTRE, G., SANKARAN, V. G., BEZERRA, M. A., ARAUJO, A. S., UDA, M., SANNA, S., CAO, A., SCHLESSINGER, D., COSTA, F. F., HIRSCHHORN, J. N. & ORKIN, S. H. 2008. DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc Natl Acad Sci U S A,* 105**,** 11869-74.

LEVO, M. & SEGAL, E. 2014. In pursuit of design principles of regulatory sequences. *Nat Rev Genet,* 15**,** 453-68.

LEVY, A. & NOLL, M. 1981. Chromatin fine structure of active and repressed genes. *Nature,* 289**,** 198-203.

LI, C., GEORGAKOPOULOU, A., MISHRA, A., GIL, S., HAWKINS, R. D., YANNAKI, E. & LIEBER, A. 2021. In vivo HSPC gene therapy with base editors allows for efficient reactivation of fetal gamma-globin in beta-YAC mice. *Blood Adv,* 5**,** 1122-1135.

LI, J., HALE, J., BHAGIA, P., XUE, F., CHEN, L., JAFFRAY, J., YAN, H., LANE, J., GALLAGHER, P. G., MOHANDAS, N., LIU, J. & AN, X. 2014. Isolation and transcriptome analyses of human erythroid progenitors: BFU-E and CFU-E. *Blood,* 124**,** 3636-45.

LI, Z. H., LIU, D. P., YIN, W. X., GUO, Z. C. & LIANG, C. C. 2001. Targeted correction of the point mutations of beta-thalassemia and targeted mutagenesis of the nucleotide associated with HPFH by RNA/DNA oligonucleotides: potential for beta-thalassemia gene therapy. *Blood Cells Mol Dis,* 27**,** 530-8.

LIANG, P., DING, C., SUN, H., XIE, X., XU, Y., ZHANG, X., SUN, Y., XIONG, Y., MA, W., LIU, Y., WANG, Y., FANG, J., LIU, D., SONGYANG, Z., ZHOU, C. & HUANG, J. 2017. Correction of beta-thalassemia mutant by base editor in human embryos. *Protein Cell,* 8**,** 811-822.

LIANG, P., XIE, X., ZHI, S., SUN, H., ZHANG, X., CHEN, Y., CHEN, Y., XIONG, Y., MA, W., LIU, D., HUANG, J. & SONGYANG, Z. 2019. Genome-wide profiling of adenine base editor specificity by EndoV-seq. *Nat Commun,* 10**,** 67.

LIANG, X., POTTER, J., KUMAR, S., ZOU, Y., QUINTANILLA, R., SRIDHARAN, M., CARTE, J., CHEN, W., ROARK, N., RANGANATHAN, S., RAVINDER, N. & CHESNUT, J. D. 2015. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol,* 208**,** 44-53.

LIBANI, I. V., GUY, E. C., MELCHIORI, L., SCHIRO, R., RAMOS, P., BREDA, L., SCHOLZEN, T., CHADBURN, A., LIU, Y., KERNBACH, M., BARON-LUHR, B., POROTTO, M., DE SOUSA, M., RACHMILEWITZ, E. A., HOOD, J. D., CAPPELLINI, M. D., GIARDINA, P. J., GRADY, R. W., GERDES, J. & RIVELLA, S. 2008. Decreased differentiation of erythroid cells exacerbates ineffective erythropoiesis in beta-thalassemia. *Blood,* 112**,** 875-85.

LIEBHABER, S. A., CASH, F. E. & MAIN, D. M. 1985. Compensatory increase in alpha 1-globin gene expression in individuals heterozygous for the alpha-thalassemia-2 deletion. *J Clin Invest,* 76**,** 1057-64.

LIEBHABER, S. A., GRIESE, E. U., WEISS, I., CASH, F. E., AYYUB, H., HIGGS, D. R. & HORST, J. 1990. Inactivation of human alpha-globin gene expression by a de novo deletion located upstream of the alpha-globin gene cluster. *Proc Natl Acad Sci U S A,* 87**,** 9431-5.

LIN, Q. X. X., THIEFFRY, D., JHA, S. & BENOUKRAF, T. 2020. TFregulomeR reveals transcription factors' context-specific features and functions. *Nucleic Acids Res,* 48**,** e10.

LIN, S., STAAHL, B. T., ALLA, R. K. & DOUDNA, J. A. 2014. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife,* 3**,** e04766.

LINDAHL, T. 1993. Instability and decay of the primary structure of DNA. *Nature,* 362**,** 709-15.

LIU, N., HARGREAVES, V. V., ZHU, Q., KURLAND, J. V., HONG, J., KIM, W., SHER, F., MACIAS-TREVINO, C., ROGERS, J. M., KURITA, R., NAKAMURA, Y., YUAN, G. C., BAUER, D. E., XU, J., BULYK, M. L. & ORKIN, S. H. 2018a. Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. *Cell,* 173**,** 430-442 e17.

LIU, Z., LU, Z., YANG, G., HUANG, S., LI, G., FENG, S., LIU, Y., LI, J., YU, W., ZHANG, Y., CHEN, J., SUN, Q. & HUANG, X. 2018b. Efficient generation of mouse models of human diseases via ABE- and BE-mediated base editing. *Nat Commun,* 9**,** 2338.

LOCATELLI, F., WALTERS, M. C., KWIATKOWSKI, J. L., PORTER, J., SAUER, M. G., THURET, I., HONGENG, S., KULOZIK, A. E., LAL, A., THRASHER, A. J., YANNAKI, E., ELLIOT, H., TAO, G., ASMAL, M. & THOMPSON, A. A. 2018. Lentiglobin Gene Therapy for Patients with Transfusion-Dependent β-Thalassemia (TDT): Results from the Phase 3 Northstar-2 and Northstar-3 Studies. *Blood,* 132**,** 1025-1025.

LONG, H. K., PRESCOTT, S. L. & WYSOCKA, J. 2016. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell,* 167**,** 1170-1187.

LONKAR, P., KIM, K. H., KUAN, J. Y., CHIN, J. Y., ROGERS, F. A., KNAUERT, M. P., KOLE, R., NIELSEN, P. E. & GLAZER, P. M. 2009. Targeted correction of a thalassemia-associated beta-globin mutation induced by pseudo-complementary peptide nucleic acids. *Nucleic Acids Res,* 37**,** 3635-44.

LOUKOPOULOS, D., LOUTRADI, A. & FESSAS, P. 1978. A unique thalassaemic syndrome: homozygous alpha-thalassaemia + homozygous beta-thalassaemia. *Br J Haematol,* 39**,** 377-89.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

LUO, H. Y. & CHUI, D. H. 2016. Diverse hematological phenotypes of beta-thalassemia carriers. *Ann N Y Acad Sci,* 1368**,** 49-55.

LYNCH, M. D., SMITH, A. J., DE GOBBI, M., FLENLEY, M., HUGHES, J. R., VERNIMMEN, D., AYYUB, H., SHARPE, J. A., SLOANE-STANLEY, J. A., SUTHERLAND, L., MEEK, S., BURDON, T., GIBBONS, R. J., GARRICK, D. & HIGGS, D. R. 2012. An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J,* 31**,** 317-29.

MA, N., LIAO, B., ZHANG, H., WANG, L., SHAN, Y., XUE, Y., HUANG, K., CHEN, S., ZHOU, X., CHEN, Y., PEI, D. & PAN, G. 2013. Transcription activator-like effector nuclease (TALEN)-mediated gene correction in integration-free beta-thalassemia induced pluripotent stem cells. *J Biol Chem,* 288**,** 34671-9.

MANKHEMTHONG, K., PHUSUA, A., SUANTA, S., SRISITTIPOJ, P., CHAROENKWAN, P. & SANGUANSERMSRI, T. 2019. Molecular characteristics of thalassemia and hemoglobin variants in prenatal diagnosis program in northern Thailand. *Int J Hematol,* 110**,** 474-481.

MANN, R., MULLIGAN, R. C. & BALTIMORE, D. 1983. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell,* 33**,** 153-9.

MARKTEL, S., SCARAMUZZA, S., CICALESE, M. P., GIGLIO, F., GALIMBERTI, S., LIDONNICI, M. R., CALBI, V., ASSANELLI, A., BERNARDO, M. E., ROSSI, C., CALABRIA, A., MILANI, R., GATTILLO, S., BENEDICENTI, F., SPINOZZI, G., APRILE, A., BERGAMI, A., CASIRAGHI, M., CONSIGLIERI, G., MASERA, N., D'ANGELO, E., MIRRA, N., ORIGA, R., TARTAGLIONE, I., PERROTTA, S., WINTER, R., COPPOLA, M., VIARENGO, G., SANTOLERI, L., GRAZIADEI, G., GABALDO, M., VALSECCHI, M. G., MONTINI, E., NALDINI, L., CAPPELLINI, M. D., CICERI, F., AIUTI, A. & FERRARI, G. 2019. Intrabone hematopoietic stem cell gene therapy for adult and pediatric patients affected by transfusion-dependent ss-thalassemia. *Nat Med,* 25**,** 234-241.

MARTYN, G. E., WIENERT, B., YANG, L., SHAH, M., NORTON, L. J., BURDACH, J., KURITA, R., NAKAMURA, Y., PEARSON, R. C. M., FUNNELL, A. P. W., QUINLAN, K. G. R. & CROSSLEY, M. 2018. Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. *Nat Genet,* 50**,** 498-503.

MATHIAS, L. A., FISHER, T. C., ZENG, L., MEISELMAN, H. J., WEINBERG, K. I., HITI, A. L. & MALIK, P. 2000. Ineffective erythropoiesis in beta-thalassemia major is due to apoptosis at the polychromatophilic normoblast stage. *Exp Hematol,* 28**,** 1343-53.

MCGOWAN, S. J., HUGHES, J. R., HAN, Z. P. & TAYLOR, S. 2013. MIG: Multi-Image Genome viewer. *Bioinformatics,* 29**,** 2477-8.

MCINTOSH, B. E., BROWN, M. E., DUFFIN, B. M., MAUFORT, J. P., VEREIDE, D. T., SLUKVIN, II & THOMSON, J. A. 2015. Nonirradiated NOD,B6.SCID Il2rgamma-/- Kit(W41/W41) (NBSGW) mice support multilineage engraftment of human hematopoietic cells. *Stem Cell Reports,* 4**,** 171-80.

MCVEY, M. & LEE, S. E. 2008. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet,* 24**,** 529-38.

MELIS, M. A., GALANELLO, R. & CAO, A. 1983. Alpha globin gene analysis in a Sardinian family with interacting alpha and beta thalassaemia genes. *Br J Haematol,* 53**,** 667-71.

MENZEL, S., GARNER, C., GUT, I., MATSUDA, F., YAMAGUCHI, M., HEATH, S., FOGLIO, M., ZELENIKA, D., BOLAND, A., ROOKS, H., BEST, S., SPECTOR, T. D., FARRALL, M.,

LATHROP, M. & THEIN, S. L. 2007. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat Genet,* 39**,** 1197-9.

MERCOLA, K. E. & CLINE, M. J. 1980. Sounding boards. The potentials of inserting new genetic information. *N Engl J Med,* 303**,** 1297-300.

MERIKA, M. & THANOS, D. 2001. Enhanceosomes. *Curr Opin Genet Dev,* 11**,** 205-8.

MERRYWEATHER-CLARKE, A. T., ATZBERGER, A., SONEJI, S., GRAY, N., CLARK, K., WAUGH, C., MCGOWAN, S. J., TAYLOR, S., NANDI, A. K., WOOD, W. G., ROBERTS, D. J., HIGGS, D. R., BUCKLE, V. J. & ROBSON, K. J. 2011. Global gene expression analysis of human erythroid progenitors. *Blood,* 117**,** e96-108.

METAIS, J. Y., DOERFLER, P. A., MAYURANATHAN, T., BAUER, D. E., FOWLER, S. C., HSIEH, M. M., KATTA, V., KERIWALA, S., LAZZAROTTO, C. R., LUK, K., NEEL, M. D., PERRY, S. S., PETERS, S. T., PORTER, S. N., RYU, B. Y., SHARMA, A., SHEA, D., TISDALE, J. F., UCHIDA, N., WOLFE, S. A., WOODARD, K. J., WU, Y., YAO, Y., ZENG, J., PRUETT-MILLER, S., TSAI, S. Q. & WEISS, M. J. 2019. Genome editing of HBG1 and HBG2 to induce fetal hemoglobin. *Blood Adv,* 3**,** 3379-3392.

METTANANDA, S., FISHER, C. A., HAY, D., BADAT, M., QUEK, L., CLARK, K., HUBLITZ, P., DOWNES, D., KERRY, J., GOSDEN, M., TELENIUS, J., SLOANE-STANLEY, J. A., FAUSTINO, P., COELHO, A., DOONDEEA, J., USUKHBAYAR, B., SOPP, P., SHARPE, J. A., HUGHES, J. R., VYAS, P., GIBBONS, R. J. & HIGGS, D. R. 2017a. Editing an alpha-globin enhancer in primary human hematopoietic stem cells as a treatment for beta-thalassemia. *Nat Commun,* 8**,** 424.

METTANANDA, S., FISHER, C. A., SLOANE-STANLEY, J. A., TAYLOR, S., OPPERMANN, U., GIBBONS, R. J. & HIGGS, D. R. 2017b. Selective silencing of alpha-globin by the histone demethylase inhibitor IOX1: a potentially new pathway for treatment of beta-thalassemia. *Haematologica,* 102**,** e80-e84.

METTANANDA, S., GIBBONS, R. J. & HIGGS, D. R. 2015. alpha-Globin as a molecular target in the treatment of beta-thalassemia. *Blood,* 125**,** 3694-701.

METTANANDA, S., GIBBONS, R. J. & HIGGS, D. R. 2016. Understanding alpha-globin gene regulation and implications for the treatment of beta-thalassemia. *Ann N Y Acad Sci,* 1368**,** 16-24.

METTANANDA, S., YASARA, N., FISHER, C. A., TAYLOR, S., GIBBONS, R. & HIGGS, D. 2019. Synergistic silencing of alpha-globin and induction of gamma-globin by histone deacetylase inhibitor, vorinostat as a potential therapy for beta-thalassaemia. *Sci Rep,* 9**,** 11649.

MICHLITSCH, J., AZIMI, M., HOPPE, C., WALTERS, M. C., LUBIN, B., LOREY, F. & VICHINSKY, E. 2009. Newborn screening for hemoglobinopathies in California. *Pediatr Blood Cancer,* 52**,** 486-90.

MILHOLLAND, B., DONG, X., ZHANG, L., HAO, X., SUH, Y. & VIJG, J. 2017. Differences between germline and somatic mutation rates in humans and mice. *Nat Commun,* 8**,** 15183.

MILLER, I. J. & BIEKER, J. J. 1993. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol Cell Biol,* 13**,** 2776-86.

MILLER, S. M., WANG, T., RANDOLPH, P. B., ARBAB, M., SHEN, M. W., HUANG, T. P., MATUSZEK, Z., NEWBY, G. A., REES, H. A. & LIU, D. R. 2020. Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat Biotechnol,* 38**,** 471-481.

MINNICH, V., NA-NAKORN, S., CHONG-CHAREONSUK, S. & KOCHASENI, S. 1954. Mediterranean anemia; a study of thirty-two cases in Thailand. *Blood,* 9**,** 1-23.

MITCHELL, M. J., BILLINGSLEY, M. M., HALEY, R. M., WECHSLER, M. E., PEPPAS, N. A. & LANGER, R. 2021. Engineering precision nanoparticles for drug delivery. *Nat Rev Drug Discov,* 20**,** 101-124.

MIYATA, M., GILLEMANS, N., HOCKMAN, D., DEMMERS, J. A. A., CHENG, J. F., HOU, J., SALMINEN, M., FISHER, C. A., TAYLOR, S., GIBBONS, R. J., GANIS, J. J., ZON, L. I., GROSVELD, F., MULUGETA, E., SAUKA-SPENGLER, T., HIGGS, D. R. & PHILIPSEN, S. 2020. An evolutionarily ancient mechanism for regulation of hemoglobin expression in vertebrate red cells. *Blood,* 136**,** 269-278.

MODELL, B. & DARLISON, M. 2008. Global epidemiology of haemoglobin disorders and derived service indicators. *Bull World Health Organ,* 86**,** 480-7.

MOJICA, F. J., JUEZ, G. & RODRIGUEZ-VALERA, F. 1993. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol Microbiol,* 9**,** 613-21.

MOUT, R., RAY, M., LEE, Y. W., SCALETTI, F. & ROTELLO, V. M. 2017. In Vivo Delivery of CRISPR/Cas9 for Therapeutic Gene Editing: Progress and Challenges. *Bioconjug Chem,* 28**,** 880-884.

MUSALLAM, K. M., SANKARAN, V. G., CAPPELLINI, M. D., DUCA, L., NATHAN, D. G. & TAHER, A. T. 2012a. Fetal hemoglobin levels and morbidity in untransfused patients with beta-thalassemia intermedia. *Blood,* 119**,** 364-7.

MUSALLAM, K. M., TAHER, A. T., CAPPELLINI, M. D. & SANKARAN, V. G. 2013. Clinical experience with fetal hemoglobin induction therapy in patients with beta-thalassemia. *Blood,* 121**,** 2199-212; quiz 2372.

MUSALLAM, K. M., TAHER, A. T. & RACHMILEWITZ, E. A. 2012b. beta-thalassemia intermedia: a clinical perspective. *Cold Spring Harb Perspect Med,* 2**,** a013482.

MUSUNURU, K., CHADWICK, A. C., MIZOGUCHI, T., GARCIA, S. P., DENIZIO, J. E., REISS, C. W., WANG, K., IYER, S., DUTTA, C., CLENDANIEL, V., AMAONYE, M., BEACH, A., BERTH, K., BISWAS, S., BRAUN, M. C., CHEN, H. M., COLACE, T. V., GANEY, J. D., GANGOPADHYAY, S. A., GARRITY, R., KASIEWICZ, L. N., LAVOIE, J., MADSEN, J. A., MATSUMOTO, Y., MAZZOLA, A. M., NASRULLAH, Y. S., NNEJI, J., REN, H., SANJEEV, A., SHAY, M., STAHLEY, M. R., FAN, S. H. Y., TAM, Y. K., GAUDELLI, N. M., CIARAMELLA, G., STOLZ, L. E., MALYALA, P., CHENG, C. J., RAJEEV, K. G., ROHDE, E., BELLINGER, A. M. & KATHIRESAN, S. 2021. In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature,* 593**,** 429-434.

NAGABABU, E., FABRY, M. E., NAGEL, R. L. & RIFKIND, J. M. 2008. Heme degradation and oxidative stress in murine models for hemoglobinopathies: thalassemia, sickle cell disease and hemoglobin C disease. *Blood Cells Mol Dis,* 41**,** 60-6.

NALDINI, L. 2011. Ex vivo gene transfer and correction for cell-based therapies. *Nat Rev Genet,* 12**,** 301-15.

NANDAKUMAR, S. K., ULIRSCH, J. C. & SANKARAN, V. G. 2016. Advances in understanding erythropoiesis: evolving perspectives. *Br J Haematol,* 173**,** 206-18.

NATHAN, D. G. & GUNN, R. B. 1966. Thalassemia: the consequences of unbalanced hemoglobin synthesis. *Am J Med,* 41**,** 815-30.

NATHWANI, A. C., REISS, U. M., TUDDENHAM, E. G., ROSALES, C., CHOWDARY, P., MCINTOSH, J., DELLA PERUTA, M., LHERITEAU, E., PATEL, N., RAJ, D., RIDDELL, A., PIE, J., RANGARAJAN, S., BEVAN, D., RECHT, M., SHEN, Y. M., HALKA, K. G., BASNER-TSCHAKARJAN, E., MINGOZZI, F., HIGH, K. A., ALLAY, J., KAY, M. A., NG, C. Y., ZHOU, J., CANCIO, M., MORTON, C. L., GRAY, J. T., SRIVASTAVA, D., NIENHUIS, A. W. &

DAVIDOFF, A. M. 2014. Long-term safety and efficacy of factor IX gene therapy in hemophilia B. *N Engl J Med,* 371**,** 1994-2004.

NERLOV, C., QUERFURTH, E., KULESSA, H. & GRAF, T. 2000. GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription. *Blood,* 95**,** 2543-51.

NEUMANN, E., SCHAEFER-RIDDER, M., WANG, Y. & HOFSCHNEIDER, P. H. 1982. Gene transfer into mouse lyoma cells by electroporation in high electric fields. *The EMBO journal,* 1**,** 841-845.

NEWBY, G. A., YEN, J. S., WOODARD, K. J., MAYURANATHAN, T., LAZZAROTTO, C. R., LI, Y., SHEPPARD-TILLMAN, H., PORTER, S. N., YAO, Y., MAYBERRY, K., EVERETTE, K. A., JANG, Y., PODRACKY, C. J., THAMAN, E., LECHAUVE, C., SHARMA, A., HENDERSON, J. M., RICHTER, M. F., ZHAO, K. T., MILLER, S. M., WANG, T., KOBLAN, L. W., MCCAFFREY, A. P., TISDALE, J. F., KALFA, T. A., PRUETT-MILLER, S. M., TSAI, S. Q., WEISS, M. J. & LIU, D. R. 2021. Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature,* 595**,** 295-302.

NICHOLLS, R. D., FISCHEL-GHODSIAN, N. & HIGGS, D. R. 1987. Recombination at the human alpha-globin gene cluster: sequence features and topological constraints. *Cell,* 49**,** 369-78.

NIENHUIS, A. W., DUNBAR, C. E. & SORRENTINO, B. P. 2006. Genotoxicity of retroviral integration in hematopoietic cells. *Mol Ther,* 13**,** 1031-49.

NIENHUIS, A. W. & NATHAN, D. G. 2012. Pathophysiology and Clinical Manifestations of the beta-Thalassemias. *Cold Spring Harb Perspect Med,* 2**,** a011726.

NIENHUIS, A. W. & PERSONS, D. A. 2012. Development of gene therapy for thalassemia. *Cold Spring Harb Perspect Med,* 2.

NISHIDA, K., ARAZOE, T., YACHIE, N., BANNO, S., KAKIMOTO, M., TABATA, M., MOCHIZUKI, M., MIYABE, A., ARAKI, M., HARA, K. Y., SHIMATANI, Z. & KONDO, A. 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science,* 353.

NISHIKURA, K. 2010. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem,* 79**,** 321-49.

NORFOLK, D. R., MCCLELLAND, D. B. L. & UNITED KINGDOM BLOOD TRANSFUSION SERVICE 2013. *Handbook of transfusion medicine,* Norwich, Stationery Office.

NORTON, L. J., FUNNELL, A. P. W., BURDACH, J., WIENERT, B., KURITA, R., NAKAMURA, Y., PHILIPSEN, S., PEARSON, R. C. M., QUINLAN, K. G. R. & CROSSLEY, M. 2017. KLF1 directly activates expression of the novel fetal globin repressor ZBTB7A/LRF in erythroid cells. *Blood Adv,* 1**,** 685-692.

NOTTA, F., DOULATOV, S., LAURENTI, E., POEPPL, A., JURISICA, I. & DICK, J. E. 2011. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science,* 333**,** 218-21.

NUSSBAUM, L., TELENIUS, J. M., HILL, S., HIRSCHFELD, P. P., SUCIU, M. C., CONSORTIUM, W., DOWNES, D. J. & HUGHES, J. R. 2018. High-Throughput Genotyping of CRISPR/Cas Edited Cells in 96-Well Plates. *Methods Protoc,* 1.

O'SHAUGHNESSY, D. F., HILL, A. V., BOWDEN, D. K., WEATHERALL, D. J. & CLEGG, J. B. 1990. Globin genes in Micronesia: origins and affinities of Pacific Island peoples. *Am J Hum Genet,* 46**,** 144-55.

OGGIANO, L., DORE, F., PISTIDDA, P., GUISO, L., MANCA, L., MASALA, B., PIRASTU, M., ROSATELLI, C., CAO, A. & LONGINOTTI, M. 1988. Homozygous beta zero-39 mutation

with thalassemia intermedia in northern Sardinia: clinical, hematological and molecular analysis. *Hemoglobin,* 12**,** 673-80.

OKAZAKI, I. M., HIAI, H., KAKAZU, N., YAMADA, S., MURAMATSU, M., KINOSHITA, K. & HONJO, T. 2003. Constitutive expression of AID leads to tumorigenesis. *J Exp Med,* 197**,** 1173-81.

OLIVIERI, N. F., PAKBAZ, Z. & VICHINSKY, E. 2010. HbE/beta-thalassemia: basis of marked clinical diversity. *Hematol Oncol Clin North Am,* 24**,** 1055-70.

ORIGA, R., SOLLAINO, M. C., BORGNA-PIGNATTI, C., PIGA, A., FELIU TORRES, A., MASILE, V. & GALANELLO, R. 2014. alpha-globin gene quadruplication and heterozygous beta-thalassemia: a not so rare cause of thalassemia intermedia. *Acta Haematol,* 131**,** 162-4.

ORKIN, S. H., KAZAZIAN, H. H., JR., ANTONARAKIS, S. E., OSTRER, H., GOFF, S. C. & SEXTON, J. P. 1982. Abnormal RNA processing due to the exon mutation of beta E-globin gene. *Nature,* 300**,** 768-9.

OUDELAAR, A. M., BEAGRIE, R. A., GOSDEN, M., DE ORNELLAS, S., GEORGIADES, E., KERRY, J., HIDALGO, D., CARRELHA, J., SHIVALINGAM, A., EL-SAGHEER, A. H., TELENIUS, J. M., BROWN, T., BUCKLE, V. J., SOCOLOVSKY, M., HIGGS, D. R. & HUGHES, J. R. 2020. Dynamics of the 4D genome during in vivo lineage specification and differentiation. *Nat Commun,* 11**,** 2722.

OUDELAAR, A. M., BEAGRIE, R. A., KASSOUF, M. T. & HIGGS, D. R. 2021. The mouse alpha-globin cluster: a paradigm for studying genome regulation and organization. *Curr Opin Genet Dev,* 67**,** 18-24.

OWENS, D. D. G., CAULDER, A., FRONTERA, V., HARMAN, J. R., ALLAN, A. J., BUCAKCI, A., GREDER, L., CODNER, G. F., HUBLITZ, P., MCHUGH, P. J., TEBOUL, L. & DE BRUIJN, M. 2019. Microhomologies are prevalent at Cas9-induced larger deletions. *Nucleic Acids Res,* 47**,** 7402-7417.

PAGANI, A., NAI, A., SILVESTRI, L. & CAMASCHELLA, C. 2019. Hepcidin and Anemia: A Tight Relationship. *Front Physiol,* 10**,** 1294.

PALII, C. G., CHENG, Q., GILLESPIE, M. A., SHANNON, P., MAZURCZYK, M., NAPOLITANI, G., PRICE, N. D., RANISH, J. A., MORRISSEY, E., HIGGS, D. R. & BRAND, M. 2019. Single-Cell Proteomics Reveal that Quantitative Changes in Co-expressed Lineage-Specific Transcription Factors Determine Cell Fate. *Cell Stem Cell,* 24**,** 812-820 e5.

PANIGRAHI, A. & O'MALLEY, B. W. 2021. Mechanisms of enhancer action: the known and the unknown. *Genome Biol,* 22**,** 108.

PAPASSOTIRIOU, I., TRAEGER-SYNODINOS, J., KANAVAKIS, E., KARAGIORGA, M., STAMOULAKATOU, A. & KATTAMIS, C. 1998. Erythroid marrow activity and hemoglobin H levels in hemoglobin H disease. *J Pediatr Hematol Oncol,* 20**,** 539-44.

PAQUET, D., KWART, D., CHEN, A., SPROUL, A., JACOB, S., TEO, S., OLSEN, K. M., GREGG, A., NOGGLE, S. & TESSIER-LAVIGNE, M. 2016. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature,* 533**,** 125-9.

PARKER, S. C., STITZEL, M. L., TAYLOR, D. L., OROZCO, J. M., ERDOS, M. R., AKIYAMA, J. A., VAN BUEREN, K. L., CHINES, P. S., NARISU, N., PROGRAM, N. C. S., BLACK, B. L., VISEL, A., PENNACCHIO, L. A., COLLINS, F. S., NATIONAL INSTITUTES OF HEALTH INTRAMURAL SEQUENCING CENTER COMPARATIVE SEQUENCING PROGRAM, A. & AUTHORS, N. C. S. P. 2013. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A,* 110**,** 17921-6.

PASRICHA, S. R., MCHUGH, K. & DRAKESMITH, H. 2016. Regulation of Hepcidin by Erythropoiesis: The Story So Far. *Annu Rev Nutr,* 36**,** 417-34.

PATSALI, P., MUSSOLINO, C., LADAS, P., FLOGA, A., KOLNAGOU, A., CHRISTOU, S., SITAROU, M., ANTONIOU, M. N., CATHOMEN, T., LEDERER, C. W. & KLEANTHOUS, M. 2019. The Scope for Thalassemia Gene Therapy by Disruption of Aberrant Regulatory Elements. *J Clin Med,* 8.

PAVANI, G., FABIANO, A., LAURENT, M., AMOR, F., CANTELLI, E., CHALUMEAU, A., MAULE, G., TACHTSIDI, A., CONCORDET, J. P., CERESETO, A., MAVILIO, F., FERRARI, G., MICCIO, A. & AMENDOLA, M. 2021. Correction of beta-thalassemia by CRISPR/Cas9 editing of the alpha-globin locus in human hematopoietic stem cells. *Blood Adv,* 5**,** 1137-1153.

PAVEL-DINU, M., WIEBKING, V., DEJENE, B. T., SRIFA, W., MANTRI, S., NICOLAS, C. E., LEE, C., BAO, G., KILDEBECK, E. J., PUNJYA, N., SINDHU, C., INLAY, M. A., SAXENA, N., DERAVIN, S. S., MALECH, H., RONCAROLO, M. G., WEINBERG, K. I. & PORTEUS, M. H. 2019. Gene correction for SCID-X1 in long-term hematopoietic stem cells. *Nat Commun,* 10**,** 1634.

PENG, Q., ZHANG, Z., LI, S., CHENG, C., LI, W., RAO, C., ZHONG, B. & LU, X. 2020. Molecular epidemiological and hematological profile of thalassemia in the Dongguan Region of Guangdong Province, Southern China. *J Clin Lab Anal***,** e23596.

PERERA, S., ALLEN, A., SILVA, I., HAPUGODA, M., WICKRAMARATHNE, M. N., WIJESIRIWARDENA, I., ALLEN, S., REES, D., EFREMOV, D. G., FISHER, C. A., WEATHERALL, D. J. & PREMAWARDHENA, A. 2019. Genotype-phenotype association analysis identifies the role of alpha globin genes in modulating disease severity of beta thalassaemia intermedia in Sri Lanka. *Sci Rep,* 9**,** 10116.

PERKINS, A., XU, X., HIGGS, D. R., PATRINOS, G. P., ARNAUD, L., BIEKER, J. J., PHILIPSEN, S. & WORKGROUP, K. L. F. C. 2016. Kruppeling erythropoiesis: an unexpected broad spectrum of human red blood cell disorders due to KLF1 variants. *Blood,* 127**,** 1856-62.

PERSONS, D. A., ALLAY, E. R., SABATINO, D. E., KELLY, P., BODINE, D. M. & NIENHUIS, A. W. 2001. Functional requirements for phenotypic correction of murine beta-thalassemia: implications for human gene therapy. *Blood,* 97**,** 3275-82.

PESLAK, S. A., KHANDROS, E., HUANG, P., LAN, X., GERONIMO, C. L., GREVET, J. D., ABDULMALIK, O., ZHANG, Z., GIARDINE, B. M., KELLER, C. A., SHI, J., HARDISON, R. C. & BLOBEL, G. A. 2020. HRI depletion cooperates with pharmacologic inducers to elevate fetal hemoglobin and reduce sickle cell formation. *Blood Adv,* 4**,** 4560-4572.

PETLJAK, M., ALEXANDROV, L. B., BRAMMELD, J. S., PRICE, S., WEDGE, D. C., GROSSMANN, S., DAWSON, K. J., JU, Y. S., IORIO, F., TUBIO, J. M. C., KOH, C. C., GEORGAKOPOULOS-SOARES, I., RODRIGUEZ-MARTIN, B., OTLU, B., O'MEARA, S., BUTLER, A. P., MENZIES, A., BHOSLE, S. G., RAINE, K., JONES, D. R., TEAGUE, J. W., BEAL, K., LATIMER, C., O'NEILL, L., ZAMORA, J., ANDERSON, E., PATEL, N., MADDISON, M., NG, B. L., GRAHAM, J., GARNETT, M. J., MCDERMOTT, U., NIK-ZAINAL, S., CAMPBELL, P. J. & STRATTON, M. R. 2019. Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. *Cell,* 176**,** 1282-1294 e20.

PEVNY, L., SIMON, M. C., ROBERTSON, E., KLEIN, W. H., TSAI, S. F., D'AGATI, V., ORKIN, S. H. & COSTANTINI, F. 1991. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature,* 349**,** 257-60.

PHYLIPSEN, M., PRIOR, J. F., LIM, E., LINGAM, N., VOGELAAR, I. P., GIORDANO, P. C., FINLAYSON, J. & HARTEVELD, C. L. 2010. Thalassemia in Western Australia: 11 novel deletions characterized by Multiplex Ligation-dependent Probe Amplification. *Blood Cells Mol Dis,* 44**,** 146-51.

PIEL, F. B. & WEATHERALL, D. J. 2014. The alpha-thalassemias. *N Engl J Med,* 371**,** 1908-16.

PILLAI, A. S., CHANDLER, S. A., LIU, Y., SIGNORE, A. V., CORTEZ-ROMERO, C. R., BENESCH, J. L. P., LAGANOWSKY, A., STORZ, J. F., HOCHBERG, G. K. A. & THORNTON, J. W. 2020. Origin of complexity in haemoglobin evolution. *Nature,* 581**,** 480-485.

POOTRAKUL, P., SIRANKAPRACHA, P., HEMSORACH, S., MOUNGSUB, W., KUMBUNLUE, R., PIANGITJAGUM, A., WASI, P., MA, L. & SCHRIER, S. L. 2000. A correlation of erythrokinetics, ineffective erythropoiesis, and erythroid precursor apoptosis in thai patients with thalassemia. *Blood,* 96**,** 2606-12.

PREMAWARDHENA, A., ALLEN, A., PIEL, F., FISHER, C., PERERA, L., RODRIGO, R., GOONATHILAKA, G., RAMEES, L., PETO, T., OLIVIERI, N. & WEATHERALL, D. 2017. The evolutionary and clinical implications of the uneven distribution of the frequency of the inherited haemoglobin variants over short geographical distances. *Br J Haematol,* 176**,** 475-484.

PREMAWARDHENA, A., ARAMBEPOLA, M., KATUGAHA, N. & WEATHERALL, D. J. 2008. Is the beta thalassaemia trait of clinical importance? *Br J Haematol,* 141**,** 407-10.

PREMAWARDHENA, A., FISHER, C. A., OLIVIERI, N. F., DE SILVA, S., ARAMBEPOLA, M., PERERA, W., O'DONNELL, A., PETO, T. E., VIPRAKASIT, V., MERSON, L., MURACA, G. & WEATHERALL, D. J. 2005a. Haemoglobin E beta thalassaemia in Sri Lanka. *Lancet,* 366**,** 1467-70.

PREMAWARDHENA, A., FISHER, C. A., OLIVIERI, N. F., DE SILVA, S., SLOANE-STANLEY, J., WOOD, W. G. & WEATHERALL, D. J. 2005b. A novel molecular basis for beta thalassemia intermedia poses new questions about its pathophysiology. *Blood,* 106**,** 3251-5.

RACHMILEWITZ, E. A. & GIARDINA, P. J. 2011. How I treat thalassemia. *Blood,* 118**,** 3479-88.

RAHAL, I., GALAMBRUN, C., BERTRAND, Y., GARNIER, N., PAILLARD, C., FRANGE, P., PONDARRE, C., DALLE, J. H., DE LATOUR, R. P., MICHALLET, M., STESCHENKO, D., MOSHOUS, D., LUTZ, P., STEPHAN, J. L., ROHRLICH, P. S., YAKOUB-AGHA, I., BERNAUDIN, F., PIGUET, C., ALADJIDI, N., BADENS, C., BERGER, C., SOCIE, G., DUMESNIL, C., CASTEX, M. P., POIREE, M., LAMBILLIOTTE, A., THOMAS, C., SIMON, P., AUQUIER, P., MICHEL, G., LOUNDOU, A., AGOUTI, I. & THURET, I. 2018. Late effects after hematopoietic stem cell transplantation for beta-thalassemia major: the French national experience. *Haematologica,* 103**,** 1143-1149.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc,* 8**,** 2281-308.

RAY, R., KALANTRI, S. A., BHATTACHARJEE, S., BISWAS, A., SHAHAB, A., BISWAS, S. & BHATTACHARYYA, M. 2019. Association of alpha hemoglobin-stabilizing protein (AHSP) gene mutation and disease severity among HbE-beta thalassemia patients. *Ann Hematol,* 98**,** 1827-1834.

REES, D. C., CLEGG, J. B. & WEATHERALL, D. J. 1998. Is hemoglobin instability important in the interaction between hemoglobin E and beta thalassemia? *Blood,* 92**,** 2141-6.

REES, H. A., KOMOR, A. C., YEH, W. H., CAETANO-LOPES, J., WARMAN, M., EDGE, A. S. B. & LIU, D. R. 2017. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat Commun,* 8**,** 15790.

REES, H. A. & LIU, D. R. 2018. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet,* 19**,** 770-788.

REES, H. A., WILSON, C., DOMAN, J. L. & LIU, D. R. 2019. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci Adv,* 5**,** eaax5717.

REIMER, J., KNOSS, S., LABUHN, M., CHARPENTIER, E. M., GOHRING, G., SCHLEGELBERGER, B., KLUSMANN, J. H. & HECKL, D. 2017. CRISPR-Cas9-induced t(11;19)/MLL-ENL translocations initiate leukemia in human hematopoietic progenitor cells in vivo. *Haematologica,* 102**,** 1558-1566.

REKHTMAN, N., RADPARVAR, F., EVANS, T. & SKOULTCHI, A. I. 1999. Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes Dev,* 13**,** 1398-411.

RIBEIL, J. A., HACEIN-BEY-ABINA, S., PAYEN, E., MAGNANI, A., SEMERARO, M., MAGRIN, E., CACCAVELLI, L., NEVEN, B., BOURGET, P., EL NEMER, W., BARTOLUCCI, P., WEBER, L., PUY, H., MERITET, J. F., GREVENT, D., BEUZARD, Y., CHRETIEN, S., LEFEBVRE, T., ROSS, R. W., NEGRE, O., VERES, G., SANDLER, L., SONI, S., DE MONTALEMBERT, M., BLANCHE, S., LEBOULCH, P. & CAVAZZANA, M. 2017. Gene Therapy in a Patient with Sickle Cell Disease. *N Engl J Med,* 376**,** 848-855.

RICHTER, M. F., ZHAO, K. T., ETON, E., LAPINAITE, A., NEWBY, G. A., THURONYI, B. W., WILSON, C., KOBLAN, L. W., ZENG, J., BAUER, D. E., DOUDNA, J. A. & LIU, D. R. 2020. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol,* 38**,** 883-891.

RIFKIND, J. M., MOHANTY, J. G. & NAGABABU, E. 2014. The pathophysiology of extracellular hemoglobin associated with enhanced oxidative reactions. *Front Physiol,* 5**,** 500.

ROMAO, L., CASH, F., WEISS, I., LIEBHABER, S., PIRASTU, M., GALANELLO, R., LOI, A., PAGLIETTI, E., IOANNOU, P. & CAO, A. 1992. Human alpha-globin gene expression is silenced by terminal truncation of chromosome 16p beginning immediately 3' of the zeta-globin gene. *Hum Genet,* 89**,** 323-8.

ROMAO, L., OSORIO-ALMEIDA, L., HIGGS, D. R., LAVINHA, J. & LIEBHABER, S. A. 1991. Alpha-thalassemia resulting from deletion of regulatory sequences far upstream of the alpha-globin structural genes. *Blood,* 78**,** 1589-95.

ROPER, N., GAO, S., MAITY, T. K., BANDAY, A. R., ZHANG, X., VENUGOPALAN, A., CULTRARO, C. M., PATIDAR, R., SINDIRI, S., BROWN, A. L., GONCEARENCO, A., PANCHENKO, A. R., BISWAS, R., THOMAS, A., RAJAN, A., CARTER, C. A., KLEINER, D. E., HEWITT, S. M., KHAN, J., PROKUNINA-OLSSON, L. & GUHA, U. 2019. APOBEC Mutagenesis and Copy-Number Alterations Are Drivers of Proteogenomic Tumor Evolution and Heterogeneity in Metastatic Thoracic Tumors. *Cell Rep,* 26**,** 2651-2666 e6.

ROSATELLI, C., LEONI, G. B., TUVERI, T., SCALAS, M. T., MOSCA, A., GALANELLO, R., GASPERINI, D. & CAO, A. 1992. Heterozygous beta-thalassemia: relationship between the hematological phenotype and the type of beta-thalassemia mutation. *Am J Hematol,* 39**,** 1-4.

ROUET, P., SMIH, F. & JASIN, M. 1994. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol,* 14**,** 8096-106.

RUND, D. & RACHMILEWITZ, E. 2005. Beta-thalassemia. *N Engl J Med,* 353**,** 1135-46.

RYU, S. M., KOO, T., KIM, K., LIM, K., BAEK, G., KIM, S. T., KIM, H. S., KIM, D. E., LEE, H., CHUNG, E. & KIM, J. S. 2018. Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat Biotechnol,* 36**,** 536-539.

SALEH-GOHARI, N., KHADEMI BAMI, M., NIKBAKHT, R. & KARIMI-MALEH, H. 2015. Effects of alpha-thalassaemia mutations on the haematological parameters of beta-thalassaemia carriers. *J Clin Pathol,* 68**,** 562-6.

SANDELIN, A., ALKEMA, W., ENGSTROM, P., WASSERMAN, W. W. & LENHARD, B. 2004. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res,* 32**,** D91-4.

SANKARAN, V. G., MENNE, T. F., XU, J., AKIE, T. E., LETTRE, G., VAN HANDEL, B., MIKKOLA, H. K., HIRSCHHORN, J. N., CANTOR, A. B. & ORKIN, S. H. 2008. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science,* 322**,** 1839-42.

SANKARAN, V. G., XU, J., BYRON, R., GREISMAN, H. A., FISHER, C., WEATHERALL, D. J., SABATH, D. E., GROUDINE, M., ORKIN, S. H., PREMAWARDHENA, A. & BENDER, M. A. 2011. A functional element necessary for fetal hemoglobin silencing. *N Engl J Med,* 365**,** 807-14.

SANTARONE, S., PEPE, A., MELONI, A., NATALE, A., PISTOIA, L., OLIOSO, P., PAPALINETTI, G., CUCCIA, L., SPASIANO, A., LISI, R., DI IANNI, M., BONFINI, T., ACCORSI, P., SALVADORI, S., PAPOLA, F., ANGELINI, S. & DI BARTOLOMEO, P. 2018. Secondary solid cancer following hematopoietic cell transplantation in patients with thalassemia major. *Bone Marrow Transplant,* 53**,** 39-43.

SCHEP, R., BRINKMAN, E. K., LEEMANS, C., VERGARA, X., VAN DER WEIDE, R. H., MORRIS, B., VAN SCHAIK, T., MANZO, S. G., PERIC-HUPKES, D., VAN DEN BERG, J., BEIJERSBERGEN, R. L., MEDEMA, R. H. & VAN STEENSEL, B. 2021. Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol Cell,* 81**,** 2216-2230 e10.

SCHIROLI, G., CONTI, A., FERRARI, S., DELLA VOLPE, L., JACOB, A., ALBANO, L., BERETTA, S., CALABRIA, A., VAVASSORI, V., GASPARINI, P., SALATAJ, E., NDIAYE-LOBRY, D., BROMBIN, C., CHAUMEIL, J., MONTINI, E., MERELLI, I., GENOVESE, P., NALDINI, L. & DI MICCO, R. 2019. Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell,* 24**,** 551-565 e8.

SCHRIER, S. L., RACHMILEWITZ, E. & MOHANDAS, N. 1989. Cellular and membrane properties of alpha and beta thalassemic erythrocytes are different: implication for differences in clinical manifestations. *Blood,* 74**,** 2194-202.

SCHWESSINGER, R., GOSDEN, M., DOWNES, D., BROWN, R. C., OUDELAAR, A. M., TELENIUS, J., TEH, Y. W., LUNTER, G. & HUGHES, J. R. 2020. DeepC: predicting 3D genome folding using megabase-scale transfer learning. *Nat Methods,* 17**,** 1118-1124.

SCHWESSINGER, R., SUCIU, M. C., MCGOWAN, S. J., TELENIUS, J., TAYLOR, S., HIGGS, D. R. & HUGHES, J. R. 2017. Sasquatch: predicting the impact of regulatory SNPs on transcription factor binding from cell- and tissue-specific DNase footprints. *Genome Res,* 27**,** 1730-1742.

SCOTT, C., DOWNES, D. J., BROWN, J. M., BEAGRIE, R., OLIJNIK, A. A., GOSDEN, M., SCHWESSINGER, R., FISHER, C. A., ROSE, A., FERGUSON, D. J. P., JOHNSON, E., HILL, Q. A., OKOLI, S., RENELLA, R., RYAN, K., BRAND, M., HUGHES, J., ROY, N. B. A., HIGGS, D. R., BABBS, C. & BUCKLE, V. J. 2020. Recapitulation of erythropoiesis in congenital dyserythropoietic anaemia type I (CDA-I) identifies defects in differentiation and nucleolar abnormalities. *Haematologica,* Online ahead of print.

SHARMA, V., KUMAR, B., KUMAR, G. & SAXENA, R. 2009. Alpha globin gene numbers: an important modifier of HbE/beta thalassemia. *Hematology,* 14**,** 297-300.

SHARPE, J. A., CHAN-THOMAS, P. S., LIDA, J., AYYUB, H., WOOD, W. G. & HIGGS, D. R. 1992. Analysis of the human alpha globin upstream regulatory element (HS-40) in transgenic mice. *EMBO J,* 11**,** 4565-72.

SHARPE, J. A., SUMMERHILL, R. J., VYAS, P., GOURDON, G., HIGGS, D. R. & WOOD, W. G. 1993a. Role of upstream DNase I hypersensitive sites in the regulation of human alpha globin gene expression. *Blood,* 82**,** 1666-71.

SHARPE, J. A., WELLS, D. J., WHITELAW, E., VYAS, P., HIGGS, D. R. & WOOD, W. G. 1993b. Analysis of the human alpha-globin gene cluster in transgenic mice. *Proc Natl Acad Sci U S A,* 90**,** 11262-6.

SHEN, M. W., ARBAB, M., HSU, J. Y., WORSTELL, D., CULBERTSON, S. J., KRABBE, O., CASSA, C. A., LIU, D. R., GIFFORD, D. K. & SHERWOOD, R. I. 2018. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature*.

SHEN, Y., YUE, F., MCCLEARY, D. F., YE, Z., EDSALL, L., KUAN, S., WAGNER, U., DIXON, J., LEE, L., LOBANENKOV, V. V. & REN, B. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature,* 488**,** 116-20.

SHER, F., HOSSAIN, M., SERUGGIA, D., SCHOONENBERG, V. A. C., YAO, Q., CIFANI, P., DASSAMA, L. M. K., COLE, M. A., REN, C., VINJAMUR, D. S., MACIAS-TREVINO, C., LUK, K., MCGUCKIN, C., SCHUPP, P. G., CANVER, M. C., KURITA, R., NAKAMURA, Y., FUJIWARA, Y., WOLFE, S. A., PINELLO, L., MAEDA, T., KENTSIS, A., ORKIN, S. H. & BAUER, D. E. 2019. Rational targeting of a NuRD subcomplex guided by comprehensive in situ mutagenesis. *Nat Genet,* 51**,** 1149-1159.

SHI, L., CUI, S., ENGEL, J. D. & TANABE, O. 2013. Lysine-specific demethylase 1 is a therapeutic target for fetal hemoglobin induction. *Nat Med,* 19**,** 291-4.

SHIN, H. Y., WANG, C., LEE, H. K., YOO, K. H., ZENG, X., KUHNS, T., YANG, C. M., MOHR, T., LIU, C. & HENNIGHAUSEN, L. 2017. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat Commun,* 8**,** 15464.

SHIRLEY, J. L., DE JONG, Y. P., TERHORST, C. & HERZOG, R. W. 2020. Immune Responses to Viral Gene Therapy Vectors. *Mol Ther,* 28**,** 709-722.

SHULTZ, L. D., LYONS, B. L., BURZENSKI, L. M., GOTT, B., CHEN, X., CHALEFF, S., KOTB, M., GILLIES, S. D., KING, M., MANGADA, J., GREINER, D. L. & HANDGRETINGER, R. 2005. Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. *J Immunol,* 174**,** 6477-89.

SILVER, A. J., BICK, A. G. & SAVONA, M. R. 2021. Germline risk of clonal haematopoiesis. *Nat Rev Genet*.

SINGHA, K., TAWEENAN, W., FUCHAROEN, G. & FUCHAROEN, S. 2019. Erythrocyte indices in a large cohort of beta-thalassemia carrier: Implication for population screening in an area with high prevalence and heterogeneity of thalassemia. *Int J Lab Hematol,* 41**,** 513-518.

SIRIWORADECHKUL, S., JINDADAMRONGWECH, S., CHUNCHARUNEE, S. & AUPPARAKKITANON, S. 2014. Implication of globin gene expression, hemoglobin F and hemoglobin E levels on beta-thalassemia/Hb E disease severity. *Ann Clin Lab Sci,* 44**,** 437-42.

SOLLAINO, M. C., PAGLIETTI, M. E., LOI, D., CONGIU, R., PODDA, R. & GALANELLO, R. 2010. Homozygous deletion of the major alpha-globin regulatory element (MCS-R2) responsible for a severe case of hemoglobin H disease. *Blood,* 116**,** 2193-4.

SONG, B., FAN, Y., HE, W., ZHU, D., NIU, X., WANG, D., OU, Z., LUO, M. & SUN, X. 2015. Improved hematopoietic differentiation efficiency of gene-corrected beta-thalassemia induced pluripotent stem cells by CRISPR/Cas9 system. *Stem Cells Dev,* 24**,** 1053-65.

SONGDEJ, D., BABBS, C., HIGGS, D. R. & CONSORTIUM, B. I. 2017. An international registry of survivors with Hb Bart's hydrops fetalis syndrome. *Blood,* 129**,** 1251-1259.

SPITZ, F. & FURLONG, E. E. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet,* 13**,** 613-26.

SPLINTER, E. & DE LAAT, W. 2011. The complex transcription regulatory landscape of our genome: control in three dimensions. *EMBO J,* 30**,** 4345-55.

SRIPICHAI, O., MAKARASARA, W., MUNKONGDEE, T., KUMKHAEK, C., NUCHPRAYOON, I., CHUANSUMRIT, A., CHUNCHARUNEE, S., CHANTRAKOON, N., BOONMONGKOL, P., WINICHAGOON, P. & FUCHAROEN, S. 2008a. A scoring system for the classification of beta-thalassemia/Hb E disease severity. *Am J Hematol,* 83**,** 482-4.

SRIPICHAI, O., MUNKONGDEE, T., KUMKHAEK, C., SVASTI, S., WINICHAGOON, P. & FUCHAROEN, S. 2008b. Coinheritance of the different copy numbers of alpha-globin gene modifies severity of beta-thalassemia/Hb E disease. *Ann Hematol,* 87**,** 375-9.

STARICK, S. R., IBN-SALEM, J., JURK, M., HERNANDEZ, C., LOVE, M. I., CHUNG, H. R., VINGRON, M., THOMAS-CHOLLIER, M. & MEIJSING, S. H. 2015. ChIP-exo signal associated with DNA-binding motifs provides insight into the genomic binding of the glucocorticoid receptor and cooperating transcription factors. *Genome Res,* 25**,** 825-35.

STEIN, S., OTT, M. G., SCHULTZE-STRASSER, S., JAUCH, A., BURWINKEL, B., KINNER, A., SCHMIDT, M., KRAMER, A., SCHWABLE, J., GLIMM, H., KOEHL, U., PREISS, C., BALL, C., MARTIN, H., GOHRING, G., SCHWARZWAELDER, K., HOFMANN, W. K., KARAKAYA, K., TCHATCHOU, S., YANG, R., REINECKE, P., KUHLCKE, K., SCHLEGELBERGER, B., THRASHER, A. J., HOELZER, D., SEGER, R., VON KALLE, C. & GREZ, M. 2010. Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. *Nat Med,* 16**,** 198-204.

STEINBERG, M. H. 2009. *Disorders of hemoglobin : genetics, pathophysiology, and clinical management,* Cambridge, Cambridge University Press.

STEWART, M. P., SHAREI, A., DING, X., SAHAY, G., LANGER, R. & JENSEN, K. F. 2016. In vitro and ex vivo strategies for intracellular delivery. *Nature,* 538**,** 183-192.

STORZ, J. F. 2018. Hemoglobin : insights into protein structure, function, and evolution. First edition. ed. Oxford: Oxford University Press,.

STRADER, M. B., KASSA, T., MENG, F., WOOD, F. B., HIRSCH, R. E., FRIEDMAN, J. M. & ALAYASH, A. I. 2016. Oxidative instability of hemoglobin E (beta26 Glu-->Lys) is increased in the presence of free alpha subunits and reversed by alpha-hemoglobin stabilizing protein (AHSP): Relevance to HbE/beta-thalassemia. *Redox Biol,* 8**,** 363-74.

STRAUSS, E. C., ANDREWS, N. C., HIGGS, D. R. & ORKIN, S. H. 1992. In vivo footprinting of the human alpha-globin locus upstream regulatory element by guanine and adenine ligation-mediated polymerase chain reaction. *Mol Cell Biol,* 12**,** 2135-42.

STUART, M. J. & NAGEL, R. L. 2004. Sickle-cell disease. *Lancet,* 364**,** 1343-60.

STURGEON, P., ITANO, H. A. & BERGREN, W. R. 1955. Genetic and biochemical studies of intermediate types of Cooley's anaemia. *Br J Haematol,* 1**,** 264-77.

SU, M. Y., STEINER, L. A., BOGARDUS, H., MISHRA, T., SCHULZ, V. P., HARDISON, R. C. & GALLAGHER, P. G. 2013. Identification of biologically relevant enhancers in human erythroid cells. *J Biol Chem,* 288**,** 8433-44.

SUMIDE, K., MATSUOKA, Y., KAWAMURA, H., NAKATSUKA, R., FUJIOKA, T., ASANO, H., TAKIHARA, Y. & SONODA, Y. 2018. A revised road map for the commitment of human cord blood CD34-negative hematopoietic stem cells. *Nat Commun,* 9**,** 2202.

SURESH, S., FISHER, C., AYYUB, H., PREMAWARDHENA, A., ALLEN, A., PERERA, A., BANDARA, D., OLIVIERI, N. & WEATHERALL, D. 2013. Alpha thalassaemia and extended alpha globin genes in Sri Lanka. *Blood Cells Mol Dis,* 50**,** 93-8.

TAHER, A., VICHINSKY, E., MUSALLAM, K., CAPPELLINI, M. D. & VIPRAKASIT, V. 2013. *In:* WEATHERALL, D. (ed.) *Guidelines for the Management of Non Transfusion Dependent Thalassaemia (NTDT).* Nicosia (Cyprus).

TAHER, A. T., KARAKAS, Z., CASSINERIO, E., SIRITANARATKUL, N., KATTAMIS, A., MAGGIO, A., RIVELLA, S., HOLLAENDER, N., MAHUZIER, B., GADBAW, B. & AYDINOK, Y. 2018. Efficacy and safety of ruxolitinib in regularly transfused patients with thalassemia: results from a phase 2a study. *Blood,* 131**,** 263-265.

TAHER, A. T., MUSALLAM, K. M. & CAPPELLINI, M. D. 2021. beta-Thalassemias. *N Engl J Med,* 384**,** 727-743.

TALBOT, D., PHILIPSEN, S., FRASER, P. & GROSVELD, F. 1990. Detailed analysis of the site 3 region of the human beta-globin dominant control region. *EMBO J,* 9**,** 2169-77.

TEBAS, P., STEIN, D., TANG, W. W., FRANK, I., WANG, S. Q., LEE, G., SPRATT, S. K., SUROSKY, R. T., GIEDLIN, M. A., NICHOL, G., HOLMES, M. C., GREGORY, P. D., ANDO, D. G., KALOS, M., COLLMAN, R. G., BINDER-SCHOLL, G., PLESA, G., HWANG, W. T., LEVINE, B. L. & JUNE, C. H. 2014. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med,* 370**,** 901-10.

TELENIUS, J. & 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***,** 393413.

THANOS, D. & MANIATIS, T. 1995. Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell,* 83**,** 1091-100.

THEIN, S. L. 2013. The molecular basis of beta-thalassemia. *Cold Spring Harb Perspect Med,* 3**,** a011700.

THEIN, S. L., MENZEL, S., PENG, X., BEST, S., JIANG, J., CLOSE, J., SILVER, N., GEROVASILLI, A., PING, C., YAMAGUCHI, M., WAHLBERG, K., ULUG, P., SPECTOR, T. D., GARNER, C., MATSUDA, F., FARRALL, M. & LATHROP, M. 2007. Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. *Proc Natl Acad Sci U S A,* 104**,** 11346-51.

THEIN, S. L. & REES, D. 2015. Haemoglobin and the Inherited Disorders of Globin Synthesis. *Postgraduate Haematology.*

THEIN, S. L. & WOOD, W. G. 2009. The Molecular Basis of β Thalassemia, δβ Thalassemia, and Hereditary Persistence of Fetal Hemoglobin. *In:* FORGET, B. G., WEATHERALL, D. J., HIGGS, D. R. & STEINBERG, M. H. (eds.) *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management.* 2 ed. Cambridge: Cambridge University Press.

THEODORIDOU, S., BALASSOPOULOU, A., BOUTOU, E., DELAKI, E. E., YFANTI, E., VYZANTIADIS, T. A., VETSIOU, E., VOSKARIDOU, E. & VLACHAKI, E. 2020. Coinheritance of Triplicated Alpha-Globin Gene and Beta-Thalassemia Mutations in Adulthood: Ten Years of Referrals in Northern Greece. *J Pediatr Hematol Oncol,* 42**,** e762-e764.

THOMAS, E. D., BUCKNER, C. D., SANDERS, J. E., PAPAYANNOPOULOU, T., BORGNA-PIGNATTI, C., DE STEFANO, P., SULLIVAN, K. M., CLIFT, R. A. & STORB, R. 1982. Marrow transplantation for thalassaemia. *Lancet,* 2**,** 227-9.

THOMPSON, A. A., WALTERS, M. C., KWIATKOWSKI, J., RASKO, J. E. J., RIBEIL, J. A., HONGENG, S., MAGRIN, E., SCHILLER, G. J., PAYEN, E., SEMERARO, M., MOSHOUS, D., LEFRERE, F., PUY, H., BOURGET, P., MAGNANI, A., CACCAVELLI, L., DIANA, J. S., SUAREZ, F., MONPOUX, F., BROUSSE, V., POIROT, C., BROUZES, C., MERITET, J. F., PONDARRE, C., BEUZARD, Y., CHRETIEN, S., LEFEBVRE, T., TEACHEY, D. T., ANURATHAPAN, U., HO, P. J., VON KALLE, C., KLETZEL, M., VICHINSKY, E., SONI, S., VERES, G., NEGRE, O., ROSS, R. W., DAVIDSON, D., PETRUSICH, A., SANDLER, L., ASMAL, M., HERMINE, O., DE MONTALEMBERT, M., HACEIN-BEY-ABINA, S., BLANCHE, S., LEBOULCH, P. & CAVAZZANA, M. 2018. Gene Therapy in Patients with Transfusion-Dependent beta-Thalassemia. *N Engl J Med,* 378**,** 1479-1493.

TRAEGER, J., WOOD, W. G., CLEGG, J. B. & WEATHERALL, D. J. 1980. Defective synthesis of HbE is due to reduced levels of beta E mRNA. *Nature,* 288**,** 497-9.

TRAEGER-SYNODINOS, J., HARTEVELD, C. L., KANAVAKIS, E., GIORDANO, P. C., KATTAMIS, C. & BERNINI, L. F. 1999. Hb Aghia Sophia [alpha62(E11)Val-->0 (alpha1)], an "in-frame" deletion causing alpha-thalassemia. *Hemoglobin,* 23**,** 317-24.

TRAXLER, E. A., YAO, Y., WANG, Y. D., WOODARD, K. J., KURITA, R., NAKAMURA, Y., HUGHES, J. R., HARDISON, R. C., BLOBEL, G. A., LI, C. & WEISS, M. J. 2016. A genome-editing strategy to treat beta-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *Nat Med,* 22**,** 987-90.

TRENT, R. J., HIGGS, D. R., CLEGG, J. B. & WEATHERALL, D. J. 1981. A new triplicated alpha-globin gene arrangement in man. *Br J Haematol,* 49**,** 149-52.

TROMPETER, S. & COHEN, A. 2014. *In:* RD, CAPPELLINI, M. D., COHEN, A., PORTER, J., TAHER, A. & VIPRAKASIT, V. (eds.) *Guidelines for the Management of Transfusion Dependent Thalassaemia (TDT).* Nicosia (CY).

TSAI, S. Q., NGUYEN, N. T., MALAGON-LOPEZ, J., TOPKAR, V. V., ARYEE, M. J. & JOUNG, J. K. 2017. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods,* 14**,** 607-614.

TSAI, S. Q., ZHENG, Z., NGUYEN, N. T., LIEBERS, M., TOPKAR, V. V., THAPAR, V., WYVEKENS, N., KHAYTER, C., IAFRATE, A. J., LE, L. P., ARYEE, M. J. & JOUNG, J. K. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol,* 33**,** 187-97.

TUFARELLI, C., STANLEY, J. A., GARRICK, D., SHARPE, J. A., AYYUB, H., WOOD, W. G. & HIGGS, D. R. 2003. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat Genet,* 34**,** 157-65.

TURCHIANO, G., ANDRIEUX, G., KLERMUND, J., BLATTNER, G., PENNUCCI, V., EL GAZ, M., MONACO, G., PODDAR, S., MUSSOLINO, C., CORNU, T. I., BOERRIES, M. & CATHOMEN, T. 2021. Quantitative evaluation of chromosomal rearrangements in gene-edited human stem cells by CAST-Seq. *Cell Stem Cell,* 28**,** 1136-1147 e5.

UDA, M., GALANELLO, R., SANNA, S., LETTRE, G., SANKARAN, V. G., CHEN, W., USALA, G., BUSONERO, F., MASCHIO, A., ALBAI, G., PIRAS, M. G., SESTU, N., LAI, S., DEI, M., MULAS, A., CRISPONI, L., NAITZA, S., ASUNIS, I., DEIANA, M., NAGARAJA, R., PERSEU, L., SATTA, S., CIPOLLINA, M. D., SOLLAINO, C., MOI, P., HIRSCHHORN, J. N., ORKIN, S. H., ABECASIS, G. R., SCHLESSINGER, D. & CAO, A. 2008. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci U S A,* 105**,** 1620-5.

VAKULSKAS, C. A., DEVER, D. P., RETTIG, G. R., TURK, R., JACOBI, A. M., COLLINGWOOD, M. A., BODE, N. M., MCNEILL, M. S., YAN, S., CAMARENA, J., LEE, C. M., PARK, S. H., WIEBKING, V., BAK, R. O., GOMEZ-OSPINA, N., PAVEL-DINU, M., SUN, W., BAO, G., PORTEUS, M. H. & BEHLKE, M. A. 2018. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med,* 24**,** 1216-1224.

VAN DEN BERG, J., A, G. M., KIELBASSA, K., FERINGA, F. M., FREIRE, R. & MEDEMA, R. H. 2018. A limited number of double-strand DNA breaks is sufficient to delay cell cycle progression. *Nucleic Acids Res*.

VAN OVERBEEK, M., CAPURSO, D., CARTER, M. M., THOMPSON, M. S., FRIAS, E., RUSS, C., REECE-HOYES, J. S., NYE, C., GRADIA, S., VIDAL, B., ZHENG, J., HOFFMAN, G. R., FULLER, C. K. & MAY, A. P. 2016. DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol Cell,* 63**,** 633-646.

VERNIMMEN, D., DE GOBBI, M., SLOANE-STANLEY, J. A., WOOD, W. G. & HIGGS, D. R. 2007. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J,* 26**,** 2041-51.

VERNIMMEN, D., LYNCH, M. D., DE GOBBI, M., GARRICK, D., SHARPE, J. A., SLOANE-STANLEY, J. A., SMITH, A. J. & HIGGS, D. R. 2011. Polycomb eviction as a new distant enhancer function. *Genes Dev,* 25**,** 1583-8.

VIPRAKASIT, V., GIBBONS, R. J., BROUGHTON, B. C., TOLMIE, J. L., BROWN, D., LUNT, P., WINTER, R. M., MARINONI, S., STEFANINI, M., BRUETON, L., LEHMANN, A. R. & HIGGS, D. R. 2001. Mutations in the general transcription factor TFIIH result in beta-thalassaemia in individuals with trichothiodystrophy. *Hum Mol Genet,* 10**,** 2797-802.

VIPRAKASIT, V., HARTEVELD, C. L., AYYUB, H., STANLEY, J. S., GIORDANO, P. C., WOOD, W. G. & HIGGS, D. R. 2006. A novel deletion causing alpha thalassemia clarifies the importance of the major human alpha globin regulatory element. *Blood,* 107**,** 3811-2.

VIPRAKASIT, V., KIDD, A. M., AYYUB, H., HORSLEY, S., HUGHES, J. & HIGGS, D. R. 2003. De novo deletion within the telomeric region flanking the human alpha globin locus as a cause of alpha thalassaemia. *Br J Haematol,* 120**,** 867-75.

VIPRAKASIT, V., TANPHAICHITR, V. S., CHINCHANG, W., SANGKLA, P., WEISS, M. J. & HIGGS, D. R. 2004. Evaluation of alpha hemoglobin stabilizing protein (AHSP) as a genetic modifier in patients with beta thalassemia. *Blood,* 103**,** 3296-9.

VITRANO, A., CALVARUSO, G., LAI, E., COLLETTA, G., QUOTA, A., GERARDI, C., CONCETTA RIGOLI, L., PITROLO, L., CUCCIA, L., GAGLIARDOTTO, F., FILOSA, A., CARUSO, V., ARGENTO, C., CAMPISI, S., RIZZO, M., PROSSOMARITI, L., FIDONE, C., FUSTANEO, M., DI MAGGIO, R. & MAGGIO, A. 2017. The era of comparable life expectancy between thalassaemia major and intermedia: Is it time to revisit the major-intermedia dichotomy? *Br J Haematol,* 176**,** 124-130.

VOON, H. P. & VADOLAS, J. 2008. Controlling alpha-globin: a review of alpha-globin expression and its impact on beta-thalassemia. *Haematologica,* 93**,** 1868-76.

VOON, H. P., WARDAN, H. & VADOLAS, J. 2007. Co-inheritance of alpha- and beta-thalassaemia in mice ameliorates thalassaemic phenotype. *Blood Cells Mol Dis,* 39**,** 184-8.

WADMAN, I. A., OSADA, H., GRUTZ, G. G., AGULNICK, A. D., WESTPHAL, H., FORSTER, A. & RABBITTS, T. H. 1997. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J,* 16**,** 3145-57.

WAINSCOAT, J. S., KANAVAKIS, E., WOOD, W. G., LETSKY, E. A., HUEHNS, E. R., MARSH, G. W., HIGGS, D. R., CLEGG, J. B. & WEATHERALL, D. J. 1983a. Thalassaemia intermedia in Cyprus: the interaction of alpha and beta thalassaemia. *Br J Haematol,* 53**,** 411-6.

WAINSCOAT, J. S., OLD, J. M., WEATHERALL, D. J. & ORKIN, S. H. 1983b. The molecular basis for the clinical diversity of beta thalassaemia in Cypriots. *Lancet,* 1**,** 1235-7.

WAJCMAN, H., TRAEGER-SYNODINOS, J., PAPASSOTIRIOU, I., GIORDANO, P. C., HARTEVELD, C. L., BAUDIN-CREUZA, V. & OLD, J. 2008. Unstable and thalassemic alpha chain hemoglobin variants: a cause of Hb H disease and thalassemia intermedia. *Hemoglobin,* 32**,** 327-49.

WALANI, S. R. & BIERMANN, J. 2017. March of Dimes Foundation: leading the way to birth defects prevention. *Public Health Rev,* 38**,** 12.

WALLACE, H. A., MARQUES-KRANC, F., RICHARDSON, M., LUNA-CRESPO, F., SHARPE, J. A., HUGHES, J., WOOD, W. G., HIGGS, D. R. & SMITH, A. J. 2007. Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. *Cell,* 128**,** 197-209.

WANG, H., GEORGAKOPOULOU, A., PSATHA, N., LI, C., CAPSALI, C., SAMAL, H. B., ANAGNOSTOPOULOS, A., EHRHARDT, A., IZSVAK, Z., PAPAYANNOPOULOU, T., YANNAKI, E. & LIEBER, A. 2019. In vivo hematopoietic stem cell gene therapy ameliorates murine thalassemia intermedia. *J Clin Invest,* 129**,** 598-615.

WANG, J., ZHANG, X., CHENG, L. & LUO, Y. 2020. An overview and metanalysis of machine and deep learning-based CRISPR gRNA design tools. *RNA Biol,* 17**,** 13-22.

WANG, X., LI, J., WANG, Y., YANG, B., WEI, J., WU, J., WANG, R., HUANG, X., CHEN, J. & YANG, L. 2018. Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat Biotechnol,* 36**,** 946-949.

WASI, P., SOOKANEK, M., POOTRAKUL, S., NA-NAKORN, S. & SUINGDUMRONG, A. 1967. Haemoglobin E and alpha-Thalassaemia. *Br Med J,* 4**,** 29-32.

WASSERMAN, W. W. & SANDELIN, A. 2004. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet,* 5**,** 276-87.

WATTANAPANITCH, M., DAMKHAM, N., POTIRAT, P., TRAKARNSANGA, K., JANAN, M., Y, U. P., KHEOLAMAI, P., KLINCUMHOM, N. & ISSARAGRISIL, S. 2018. One-step genetic correction of hemoglobin E/beta-thalassemia patient-derived iPSCs by the CRISPR/Cas9 system. *Stem Cell Res Ther,* 9**,** 46.

WEATHERALL, D. J. 2001. Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat Rev Genet,* 2**,** 245-55.

WEATHERALL, D. J. 2008. Genetic variation and susceptibility to infection: the red cell and malaria. *Br J Haematol,* 141**,** 276-86.

WEATHERALL, D. J. 2010a. The inherited diseases of hemoglobin are an emerging global health burden. *Blood,* 115**,** 4331-6.

WEATHERALL, D. J. 2010b. *Thalassaemia : the biography,* Oxford, Oxford University Press.
WEATHERALL, D. J. 2010c. Thalassemia as a global health problem: recent progress toward its control in the developing countries. *Ann N Y Acad Sci,* 1202**,** 17-23.
WEATHERALL, D. J. & CLEGG, J. B. 2001. *The thalassaemia syndromes,* Oxford ; Malden, MA, Blackwell Science.
WEATHERALL, D. J., CLEGG, J. B. & NAUGHTON, M. A. 1965. Globin synthesis in thalassaemia: an in vitro study. *Nature,* 208**,** 1061-5.
WEATHERALL, D. J. & OLIVIERI, N. F. 2009. Clinical Aspects of β Thalassemia and Related Disorders. *In:* FORGET, B. G., WEATHERALL, D. J., HIGGS, D. R. & STEINBERG, M. H. (eds.) *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management.* 2 ed. Cambridge: Cambridge University Press.
WEATHERALL, D. J., PRESSLEY, L., WOOD, W. G., HIGGS, D. R. & CLEGG, J. B. 1981. Molecular basis for mild forms of homozygous beta-thalassaemia. *Lancet,* 1**,** 527-9.
WEATHERALL, D. J., SCHECHTER, A. N. & NATHAN, D. G. 2013. *Hemoglobin and its diseases,* Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.
WEBER, L., FRATI, G., FELIX, T., HARDOUIN, G., CASINI, A., WOLLENSCHLAEGER, C., MENEGHINI, V., MASSON, C., DE CIAN, A., CHALUMEAU, A., MAVILIO, F., AMENDOLA, M., ANDRE-SCHMUTZ, I., CERESETO, A., EL NEMER, W., CONCORDET, J. P., GIOVANNANGELI, C., CAVAZZANA, M. & MICCIO, A. 2020. Editing a gamma-globin repressor binding site restores fetal hemoglobin synthesis and corrects the sickle cell disease phenotype. *Sci Adv,* 6.
WEIL, L. G., CHARLTON, M. R., COPPINGER, C., DANIEL, Y. & STREETLY, A. 2020. Sickle cell disease and thalassaemia antenatal screening programme in England over 10 years: a review from 2007/2008 to 2016/2017. *J Clin Pathol,* 73**,** 183-190.
WEINGARTEN-GABBAY, S. & SEGAL, E. 2014. The grammar of transcriptional regulation. *Hum Genet,* 133**,** 701-11.
WEINTRAUB, H. & GROUDINE, M. 1976. Chromosomal subunits in active genes have an altered conformation. *Science,* 193**,** 848-56.
WEISS, I. M. & LIEBHABER, S. A. 1994. Erythroid cell-specific determinants of alpha-globin mRNA stability. *Mol Cell Biol,* 14**,** 8123-32.
WEN, S. C., RODER, K., HU, K. Y., ROMBEL, I., GAVVA, N. R., DAFTARI, P., KUO, Y. Y., WANG, C. & SHEN, C. K. 2000. Loading of DNA-binding factors to an erythroid enhancer. *Mol Cell Biol,* 20**,** 1993-2003.
WHIPPLE, G. H. & BRADFORD, W. L. 1932. RACIAL OR FAMILIAL ANEMIA OF CHILDREN: ASSOCIATED WITH FUNDAMENTAL DISTURBANCES OF BONE AND PIGMENT METABOLISM (COOLEY-VON JAKSCH). *American Journal of Diseases of Children,* 44**,** 336-365.
WHYTE, W. A., ORLANDO, D. A., HNISZ, D., ABRAHAM, B. J., LIN, C. Y., KAGEY, M. H., RAHL, P. B., LEE, T. I. & YOUNG, R. A. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell,* 153**,** 307-19.
WICKRAMASINGHE, S. N., HUGHES, M., WASI, P., FUCHAROEN, S. & MODELL, B. 1981. Ineffective erythropoiesis in haemoglobin E beta -thalassaemia: an electron microscope study. *Br J Haematol,* 48**,** 451-7.
WIENERT, B., MARTYN, G. E., FUNNELL, A. P. W., QUINLAN, K. G. R. & CROSSLEY, M. 2018. Wake-up Sleepy Gene: Reactivating Fetal Globin for beta-Hemoglobinopathies. *Trends Genet,* 34**,** 927-940.

WIENERT, B., MARTYN, G. E., KURITA, R., NAKAMURA, Y., QUINLAN, K. G. R. & CROSSLEY, M. 2017. KLF1 drives the expression of fetal hemoglobin in British HPFH. *Blood,* 130**,** 803-807.

WIENERT, B., WYMAN, S. K., RICHARDSON, C. D., YEH, C. D., AKCAKAYA, P., PORRITT, M. J., MORLOCK, M., VU, J. T., KAZANE, K. R., WATRY, H. L., JUDGE, L. M., CONKLIN, B. R., MARESCA, M. & CORN, J. E. 2019. Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science,* 364**,** 286-289.

WILKIE, A. O., BUCKLE, V. J., HARRIS, P. C., LAMB, J., BARTON, N. J., REEDERS, S. T., LINDENBAUM, R. H., NICHOLLS, R. D., BARROW, M., BETHLENFALVAY, N. C. & ET AL. 1990a. Clinical features and molecular analysis of the alpha thalassemia/mental retardation syndromes. I. Cases due to deletions involving chromosome band 16p13.3. *Am J Hum Genet,* 46**,** 1112-26.

WILKIE, A. O., LAMB, J., HARRIS, P. C., FINNEY, R. D. & HIGGS, D. R. 1990b. A truncated human chromosome 16 associated with alpha thalassaemia is stabilized by addition of telomeric repeat (TTAGGG)n. *Nature,* 346**,** 868-71.

WILLIAMS, D. A., LEMISCHKA, I. R., NATHAN, D. G. & MULLIGAN, R. C. 1984. Introduction of new genetic material into pluripotent haematopoietic stem cells of the mouse. *Nature,* 310**,** 476-80.

WILLIAMS, T. N. & WEATHERALL, D. J. 2012. World distribution, population genetics, and health burden of the hemoglobinopathies. *Cold Spring Harb Perspect Med,* 2**,** a011692.

WINICHAGOON, P., FUCHAROEN, S., CHEN, P. & WASI, P. 2000. Genetic factors affecting clinical severity in beta-thalassemia syndromes. *J Pediatr Hematol Oncol,* 22**,** 573-80.

WINICHAGOON, P., FUCHAROEN, S., WEATHERALL, D. & WASI, P. 1985. Concomitant inheritance of alpha-thalassemia in beta 0- thalassemia/Hb E disease. *Am J Hematol,* 20**,** 217-22.

WU, M. Y., HE, Y., YAN, J. M. & LI, D. Z. 2017. A novel selective deletion of the major alpha-globin regulatory element (MCS-R2) causing alpha-thalassaemia. *Br J Haematol,* 176**,** 984-986.

WU, Y., ZENG, J., ROSCOE, B. P., LIU, P., YAO, Q., LAZZAROTTO, C. R., CLEMENT, K., COLE, M. A., LUK, K., BARICORDI, C., SHEN, A. H., REN, C., ESRICK, E. B., MANIS, J. P., DORFMAN, D. M., WILLIAMS, D. A., BIFFI, A., BRUGNARA, C., BIASCO, L., BRENDEL, C., PINELLO, L., TSAI, S. Q., WOLFE, S. A. & BAUER, D. E. 2019. Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat Med,* 25**,** 776-783.

XIE, F., YE, L., CHANG, J. C., BEYER, A. I., WANG, J., MUENCH, M. O. & KAN, Y. W. 2014. Seamless gene correction of beta-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Res,* 24**,** 1526-33.

XU, J., BAUER, D. E., KERENYI, M. A., VO, T. D., HOU, S., HSU, Y. J., YAO, H., TROWBRIDGE, J. J., MANDEL, G. & ORKIN, S. H. 2013. Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. *Proc Natl Acad Sci U S A,* 110**,** 6518-23.

XU, L., YANG, H., GAO, Y., CHEN, Z., XIE, L., LIU, Y., LIU, Y., WANG, X., LI, H., LAI, W., HE, Y., YAO, A., MA, L., SHAO, Y., ZHANG, B., WANG, C., CHEN, H. & DENG, H. 2017. CRISPR/Cas9-Mediated CCR5 Ablation in Human Hematopoietic Stem/Progenitor Cells Confers HIV-1 Resistance In Vivo. *Mol Ther,* 25**,** 1782-1789.

XU, P., TONG, Y., LIU, X. Z., WANG, T. T., CHENG, L., WANG, B. Y., LV, X., HUANG, Y. & LIU, D. P. 2015. Both TALENs and CRISPR/Cas9 directly target the HBB IVS2-654 (C > T) mutation in beta-thalassemia-derived iPSCs. *Sci Rep,* 5**,** 12065.

XU, S., LUK, K., YAO, Q., SHEN, A. H., ZENG, J., WU, Y., LUO, H. Y., BRENDEL, C., PINELLO, L., CHUI, D. H. K., WOLFE, S. A. & BAUER, D. E. 2019. Editing aberrant splice sites efficiently restores beta-globin expression in beta-thalassemia. *Blood,* 133**,** 2255-2262.

YAMSRI, S., PAKDEE, N., FUCHAROEN, G., SANCHAISURIYA, K. & FUCHAROEN, S. 2016. Molecular Understanding of Non-Transfusion-Dependent Thalassemia Associated with Hemoglobin E-beta-Thalassemia in Northeast Thailand. *Acta Haematol,* 136**,** 233-239.

YAMSRI, S., SANCHAISURIYA, K., FUCHAROEN, G., SAE-UNG, N. & FUCHAROEN, S. 2011. Genotype and phenotype characterizations in a large cohort of beta-thalassemia heterozygote with different forms of alpha-thalassemia in northeast Thailand. *Blood Cells Mol Dis,* 47**,** 120-4.

YIN, H., KANASTY, R. L., ELTOUKHY, A. A., VEGAS, A. J., DORKIN, J. R. & ANDERSON, D. G. 2014. Non-viral vectors for gene-based therapy. *Nat Rev Genet,* 15**,** 541-55.

YU, C., NIAKAN, K. K., MATSUSHITA, M., STAMATOYANNOPOULOS, G., ORKIN, S. H. & RASKIND, W. H. 2002. X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. *Blood,* 100**,** 2040-5.

YU, X., KONG, Y., DORE, L. C., ABDULMALIK, O., KATEIN, A. M., ZHOU, S., CHOI, J. K., GELL, D., MACKAY, J. P., GOW, A. J. & WEISS, M. J. 2007. An erythroid chaperone that facilitates folding of alpha-globin subunits for hemoglobin synthesis. *J Clin Invest,* 117**,** 1856-65.

YUAN, J., KANNAN, R., SHINAR, E., RACHMILEWITZ, E. A. & LOW, P. S. 1992. Isolation, characterization, and immunoprecipitation studies of immune complexes from membranes of beta-thalassemic erythrocytes. *Blood,* 79**,** 3007-13.

YUAN, J., RUBIN, E., ALJURF, M., MA, L. & SCHRIER, S. L. 1994. Defective assembly of membrane proteins in erythroid precursors of beta-thalassemic mice. *Blood,* 84**,** 632-7.

YUE, F., CHENG, Y., BRESCHI, A., VIERSTRA, J., WU, W., RYBA, T., SANDSTROM, R., MA, Z., DAVIS, C., POPE, B. D., SHEN, Y., PERVOUCHINE, D. D., DJEBALI, S., THURMAN, R. E., KAUL, R., RYNES, E., KIRILUSHA, A., MARINOV, G. K., WILLIAMS, B. A., TROUT, D., AMRHEIN, H., FISHER-AYLOR, K., ANTOSHECHKIN, I., DESALVO, G., SEE, L. H., FASTUCA, M., DRENKOW, J., ZALESKI, C., DOBIN, A., PRIETO, P., LAGARDE, J., BUSSOTTI, G., TANZER, A., DENAS, O., LI, K., BENDER, M. A., ZHANG, M., BYRON, R., GROUDINE, M. T., MCCLEARY, D., PHAM, L., YE, Z., KUAN, S., EDSALL, L., WU, Y. C., RASMUSSEN, M. D., BANSAL, M. S., KELLIS, M., KELLER, C. A., MORRISSEY, C. S., MISHRA, T., JAIN, D., DOGAN, N., HARRIS, R. S., CAYTING, P., KAWLI, T., BOYLE, A. P., EUSKIRCHEN, G., KUNDAJE, A., LIN, S., LIN, Y., JANSEN, C., MALLADI, V. S., CLINE, M. S., ERICKSON, D. T., KIRKUP, V. M., LEARNED, K., SLOAN, C. A., ROSENBLOOM, K. R., LACERDA DE SOUSA, B., BEAL, K., PIGNATELLI, M., FLICEK, P., LIAN, J., KAHVECI, T., LEE, D., KENT, W. J., RAMALHO SANTOS, M., HERRERO, J., NOTREDAME, C., JOHNSON, A., VONG, S., LEE, K., BATES, D., NERI, F., DIEGEL, M., CANFIELD, T., SABO, P. J., WILKEN, M. S., REH, T. A., GISTE, E., SHAFER, A., KUTYAVIN, T., HAUGEN, E., DUNN, D., REYNOLDS, A. P., NEPH, S., HUMBERT, R., HANSEN, R. S., DE BRUIJN, M.,

et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature,* 515**,** 355-64.

ZENG, J., WU, Y., REN, C., BONANNO, J., SHEN, A. H., SHEA, D., GEHRKE, J. M., CLEMENT, K., LUK, K., YAO, Q., KIM, R., WOLFE, S. A., MANIS, J. P., PINELLO, L., JOUNG, J. K. & BAUER, D. E. 2020. Therapeutic base editing of human hematopoietic stem cells. *Nat Med,* 26**,** 535-541.

ZHANG, L., JIA, R., PALANGE, N. J., SATHEKA, A. C., TOGO, J., AN, Y., HUMPHREY, M., BAN, L., JI, Y., JIN, H., FENG, X. & ZHENG, Y. 2015. Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. *PLoS One,* 10**,** e0120396.

ZHANG, Q., REDDY, P. M., YU, C. Y., BASTIANI, C., HIGGS, D., STAMATOYANNOPOULOS, G., PAPAYANNOPOULOU, T. & SHEN, C. K. 1993. Transcriptional activation of human zeta 2 globin promoter by the alpha globin regulatory element (HS-40)- functional role of specific nuclear factor-DNA complexes. *Mol Cell Biol,* 13**,** 2298-308.

ZHANG, Q., ROMBEL, I., REDDY, G. N., GANG, J. B. & SHEN, C. K. 1995. Functional roles of in vivo footprinted DNA motifs within an alpha-globin enhancer. Erythroid lineage and developmental stage specificities. *J Biol Chem,* 270**,** 8501-5.

ZHANG, Y., ZHAO, G., AHMED, F. Y. H., YI, T., HU, S., CAI, T. & LIAO, Q. 2020. In silico Method in CRISPR/Cas System: An Expedite and Powerful Booster. *Front Oncol,* 10**,** 584404.

ZHONG, L., GAN, X., XU, L., LIANG, C., XIE, Y., LIN, W., CHEN, P. & LIU, M. 2018. The phenomena of balanced effect between alpha-globin gene and of beta-globin gene. *BMC Med Genet,* 19**,** 145.

ZHOU, C., SUN, Y., YAN, R., LIU, Y., ZUO, E., GU, C., HAN, L., WEI, Y., HU, X., ZENG, R., LI, Y., ZHOU, H., GUO, F. & YANG, H. 2019. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature,* 571**,** 275-278.

ZHOU, J. & TROYANSKAYA, O. G. 2015. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods,* 12**,** 931-4.

ZHOU, S., OLSON, J. S., FABIAN, M., WEISS, M. J. & GOW, A. J. 2006. Biochemical fates of alpha hemoglobin bound to alpha hemoglobin-stabilizing protein AHSP. *J Biol Chem,* 281**,** 32611-8.

ZUCCARO, M. V., XU, J., MITCHELL, C., MARIN, D., ZIMMERMAN, R., RANA, B., WEINSTEIN, E., KING, R. T., PALMEROLA, K. L., SMITH, M. E., TSANG, S. H., GOLAND, R., JASIN, M., LOBO, R., TREFF, N. & EGLI, D. 2020. Allele-Specific Chromosome Removal after Cas9 Cleavage in Human Embryos. *Cell*.

ZUO, E., SUN, Y., WEI, W., YUAN, T., YING, W., SUN, H., YUAN, L., STEINMETZ, L. M., LI, Y. & YANG, H. 2019. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science,* 364**,** 289-292.