



The
University
Of
Sheffield.

Effective Utilization of Molecular Genetic Screening of Patients with Sickle Cell Disease and Beta Thalassemia Major in Saudi Arabia

Rawabi Zahed

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
The Medical School
Department of Infection, Immunity and Cardiovascular Disease

May 2023

ABSTRACT

Hereditary blood diseases are prevalent in the Kingdom of Saudi Arabia. The majority of these blood disorders are sickle cell disease and β -thalassemia with variants located on the beta globin gene (*HBB*).

Aim: To determine the profile of novel or previously reported causative mutations in more than 150 transfusion dependent individuals using TaqMan genotyping and next-generation DNA sequencing. In addition, I explored the genomic variation in a family with transfusion dependency but without a definitive genetic diagnosis related to *HBB*. I also attempted to detect unknown genetic variations in functionally related genes and applied in-silico analysis of the detected variants to propose candidate genes that may contribute to the severe etiology of thalassemia within a family. **Methods:** To identify *HBB* variants, I conducted Taqman genotyping tests using SCD, c.92+5G>C, c.92+1G>A, c.93-21G>A, c.27dupG, and c.118C>T as the most frequently identified *HBB* variants within the Saudi population. After that, targeted next generation sequencing was performed on samples with either negative or only heterozygous results for these variants. The use of different molecular techniques including MLPA alpha thalassemia, whole exome sequencing, cytoscan HD array, and whole genome sequencing was undertaken on samples that needed further investigation. Implementation of different data filtering approaches and several in-silico techniques were utilized to investigate the detected variants. **Results:** After Taqman genotyping of the 154 DNA samples, 100 samples were either homozygous or compound heterozygous for the most frequently known *HBB* variants. The rest of these samples were sequenced using targeted NGS and 20 different common and rare *HBB* variants were identified. Three out of the 154 samples did not have any apparent *HBB* mutation and further investigation was applied using additional molecular techniques. This led to the identification of two gene candidates, SMC5 and TALDO1, with possible novel associations in increasing the severity of clinical manifestation in transfusion-dependent patients with heterozygous pathogenic variant of beta thalassemia. **Conclusion:** Beta thalassemia is a heterogenous disease with a wide range of clinical severity and the steps towards identification of the underlying genetic cause of the phenotype is different from case to case and may require a combination of several molecular techniques. Therefore, the interaction of illness-causing variations with the rest of an individual's genome is crucial to gaining a complete understanding of the condition. Excellent detection rates in less time may be achieved with a specialized filtering technique and strategy, making this an option for primary laboratory workflow.

ACKNOWLEDGMENTS

First, I owe a great debt of gratitude to my internal and external PhD advisors, Prof. Adel Abuzenadah from King Abdulaziz University and Dr. Martin Nicklin from University of Sheffield, for their unending guidance, encouragement, insightful feedback, and tolerance during my PhD period. Their vast knowledge and wealth of experience has been an inspiration to me throughout my studies. I would also like to thank Dr. Ashraf Dallol and Dr. Nouf Alghanmy for their constant help, support, and mentorship.

I would like to extend my gratitude to all members of the Joint Supervision Program, especially Dr. Ammar Ameen, for helping me succeed in finishing my research with all their continuous assistance from day one until the submission of my thesis.

A heartfelt "thank you" goes to my beloved husband for helping me in maintaining my sanity throughout my doctoral studies. Thank you for lifting my spirit when I was feeling down. Thank you for believing in me and for your comforting words and insightful advice over the years. I would also like to express my sincere thanks to my four children, Ali, Lian, Anmar, and Yara for their unconditional love, support, encouragement, and understanding. Thank you all for creating a great comfortable environment for me to thrive.

PUBLICATIONS

1. From chapter three

Alwazani, W., Zahid, R., Elaimi, A., Bajouh, O., Hindawi, S., Arab, B., Damanhour, G., Saka, M., Turki, R., Khan, J., Dallol, A., & Abuzenadah, A. (2016). Detection of beta-Thalassemia Mutations Using TaqMan Single Nucleotide Polymorphism Genotyping Assays. *Genet Test Mol Biomarkers*, 20(3), 154-157. <https://doi.org/10.1089/gtmb.2015.0222>

2. From chapter five

Zahed, R., & Sajer, B. (2023). Case report: Identification of a rare homozygous missense variant in the PKLR gene reported for the first time in transfusion-dependent Saudi Patient. *Journal of Contemporary Medical Sciences*, 9(3), 211–220. <https://doi.org/10.22317/jcms.v9i3.1354>

TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGMENTS	ii
PUBLICATIONS	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES	vii
LIST OF TABLES	xiii
ABBREVIATIONS	xv
CHAPTER ONE: GENERAL INTRODUCTION	1
1.1 Red Blood Cells (RBC)	2
1.2 RBC structure	2
1.3 RBC regulation and development.....	5
1.4 Hemoglobin structure	11
1.5 Beta globin and alpha globin gene clusters	15
1.5.1 Control of hemoglobin switch.....	18
1.5.2 Control of beta and alpha globin gene clusters	21
1.5.3 Properties of the different forms of hemoglobin tetramers	29
1.6 RBC disorders	32
1.7 Hemoglobinopathies	35
1.7.1 Sickle cell anemia	36
1.7.2 Genetics of hereditary persistent of fetal hemoglobin (HPFH)	42
1.7.3 Thalassemia.....	42
1.7.4 Alpha thalassemia	48
1.8 Clinical features of sickle cell anemia and thalassemia	51
1.9 Diagnosis of sickle cell disease and beta thalassemia.....	52
1.10 Treatment and outcomes.....	53
1.11 Prevalence of thalassemia in Saudi Arabia	56
1.12 Aim	60
CHAPTER TWO: EXPERIMENTAL PROCEDURES	61
2.1 Recruitment of thalassemia subjects and sample collection.....	62
2.2 Whole blood DNA extraction.....	63
2.3 DNA quantification	64
2.4 TaqMan assay genotyping.....	64
2.4.1 Principle	64

2.4.2	Assays.....	65
2.5	Next generation sequencing by Ion Torrent PGM™ semiconductor sequencing.....	67
2.5.1	Principle	68
2.5.2	Procedure.....	69
2.6	Whole exome sequencing by NovaSeq 6000 Sequencing system	79
2.6.1	Principle	79
2.6.2	Procedure.....	83
2.6.3	Analysis details.....	88
2.7	Sanger sequencing	90
2.7.1	Primers.....	91
2.7.2	Cycle sequencing.....	95
2.7.3	Product Denaturation	95
2.8	Multiplex ligation-dependent probe amplification (MLPA) of the alpha globin gene cluster ..	96
2.8.1	Principle	96
2.8.2	Procedure.....	97
2.9	Affymetrix CytoScan HD array	98
2.9.1	Protocol.....	99
2.10	Whole genome sequencing.....	102
2.10.1	Procedure.....	102
2.11	In silico variants analysis	106
2.11.1	FANCC.....	106
2.11.2	SMC5 and TALDO1	107
2.11.3	Molecular docking.....	108
CHAPTER THREE: IDENTIFICATION AND CHARACTERIZATION OF GENOTYPIC VARIATIONS IN PATIENTS WITH BETA THALASSEMIA MAJOR		109
3.1	Brief introduction	110
3.2	Use of TaqMan genotyping assays in identifying patients with the most frequently known β -thalassemia variants within Saudi population.....	111
3.3	Identification of beta globin gene variants in transfusion dependent thalassemia patients using targeted next generation sequencing by Ion Torrent PGM System	114
3.4	Confirmation of discovered β -thalassemia variants using Sanger sequencing	116
3.5	Discussion.....	117
3.5.1	Identification of genotype variations in transfusion-dependent beta-thalassemia patients.....	117
3.5.2	Genotype variation and phenotype correlation in transfusion-dependent patients.....	121
CHAPTER FOUR: A THERAPEUTIC APPROACH FOR AUTOLOGOUS GENE THERAPY OF SICKLE CELL AND BETA THALASSEMIA DISEASE.....		128
4.1	Brief Introduction	129
4.1.1	CRISPR/Cas9	130
4.1.2	Generation of Induced Pluripotent Stem Cells	133

4.2 Aim	134
4.3 Methods	134
4.3.1 Reprogramming of whole blood derived cells to iPSCs using ReprorNA-OKSGM vector by StemCell	134
4.3.2 Reprogramming of fibroblasts derived from whole blood to iPSCs using Epi5TM Episomal iPSC reprogramming kit from Invitrogen by Life Technologies	142
4.4 Limitations and delimitations	144
CHAPTER FIVE: CASE REPORT: IDENTIFICATION OF A RARE HOMOZYGOUS MISSENSE VARIANT IN THE PKLR GENE THAT CORRELATES WITH CHRONIC PYRUVATE KINASE DEFICIENCY ANEMIA	146
5.1 Brief Introduction	147
5.2 Diagnosis	149
5.3 Case Presentation	150
5.4 Clinical History	150
5.5 Results	152
5.6 Data Analysis	156
5.7 Conclusion	165
CHAPTER SIX: EXPLORING THE MOLECULAR HETEROGENEITY OF A FAMILY EXHIBITING THALASSEMIA MAJOR	166
6.1 Use of whole exome sequencing to elucidate the genetic bases of transfusion-dependent patients	167
6.2 Results and data analysis	167
6.3 Analysis of whole exome sequencing data	187
6.3.1 First scenario	187
6.3.2 Second scenario	190
6.3.3 Third scenario	193
6.3.4 Fourth scenario	202
6.4 Functional- related Candidate genes selection	208
6.4.1 Pathway analysis	210
6.4.2 Gene expression	211
6.4.3 Expression conservation in primate tissues	212
6.4.4 Knockout mouse model	212
6.4.5 Damage effect of the variant on the translated protein	214
6.4.6 Further in-silico analysis	218
6.4.7 Identification of deleterious SNPs using online in-silico prediction tools (Table 34).	228
6.5 General discussion	229
6.6 Conclusion	242
6.7 Future Recommendations	243
REFERENCES	245

LIST OF FIGURES

Figure 1: Different stages of red blood cell development.	5
Figure 2: Embden-Meyerhoff pathway of glycolysis.....	8
Figure 3: Hemoglobin structure.	12
Figure 4: The beta globin and the alpha globin gene clusters.	16
Figure 5: Schematic presentation of the globin chains synthesis during development.....	17
Figure 6: Definitive and primitive formation.	20
Figure 7: Chromatin looping in the activation of beta globin genes.....	23
Figure 8: Hemoglobin switch from HbF to HbA.	24
Figure 9: Function of fetal globin in γ and β -globin gene expression.	26
Figure 10: The molecular complex comprising the proximal γ -globin promoter and its essential transcription factors.....	29
Figure 11: The effect of HBB allele combinations on blood hemoglobin concentrations.....	36
Figure 12: Normal blood cells Vs. Sickled cells flow through capillaries.....	37
Figure 13: Sickle cell polymerization.	39
Figure 14: Samples of different nationalities of transfusion dependent patients with sickle cell and beta thalassemia.	63
Figure 15: Principle of TaqMan genotyping assay.	65
Figure 16: Mapping of the most frequent beta-thalassemia SNPs within Saudi population used in our study.	66
Figure 17: Next generation sequencing coverage.	69
Figure 18: DNA fragmentation and barcoding.....	70
Figure 19: Standard curve and amplification plot of standards (6.8 pM, 0.68 pM, 0.068 pM). Standards are used as known DNA concentrations to help in determining unknown DNA concentrations of samples.	73
Figure 20: qPCR quantification of samples before sample pooling.	73
Figure 21: Data analysis pipeline.	76
Figure 22: An example of FASTQ file format.	76
Figure 23: SAM file with annotated mandatory and optional fields.....	77
Figure 24: Bead Linked Transposome.....	79
Figure 25: Sequencing ready fragment with annotated Index primers.	80

Figure 26: Principle of bridge amplification.....	81
Figure 27: Electropherogram interpretation graph.....	85
Figure 28: Principle of Sanger sequencing.....	91
Figure 29: Primers' coverage of Sanger sequencing of HBB.	92
Figure 30: Sanger sequencing primer design for confirmation of PKLR c.1015G>A homozygous mutation in sample 183.	93
Figure 31: Sanger sequencing primer design for confirmation of FANCC: c.1015G>A heterozygous mutation.	93
Figure 32: Sanger sequencing primer design for confirmation of SMC5 gene (c.676G>A) heterozygous mutation.	94
Figure 33: Sanger sequencing primer design for confirmation of TALDO1 gene (c.604G>A) heterozygous mutation.	94
Figure 34: Genotyping result distribution	111
Figure 35: Example of genotyping of the rs33915217 SNP (c.92+5G>C).	112
Figure 36: Observed occurrence of homozygous HBB mutations among 154 transfusion dependent beta thalassemia patients.	113
Figure 37: Observed occurrence of compound heterozygous HBB mutations found in seven samples among 154 transfusion dependent beta thalassemia patients.	113
Figure 38: Observed occurrence of 20 single heterozygous HBB variants among 154 transfusion dependent patients.....	114
Figure 39: Location of mutations in HBB gene NP_000509.1. NM_000518.4 identified by Ion Torrent NGS.	116
Figure 40: A chromatogram showing a beta thalassemia case with homozygous c.315+1 G>A variant.	117
Figure 41: A chromatogram showing beta thalassemia case with compound heterozygous variations.	117
Figure 42: Workflow and molecular techniques used in the investigation of our cohort.....	119
Figure 43: Beta thalassemia variants located upstream of 5' UTR region of HBB gene.....	121
Figure 44: Beta thalassemia variants located in the coding regions of HBB gene found in patient's DNA in this study.....	122
Figure 45: Beta thalassemia variants located in the intronic regions of the HBB gene.	124
Figure 46: <i>CRISPRs are viral defense mechanisms in the bacterial genome.</i>	131

Figure 47: A chimeric single-guide RNA (sgRNA) facilitates Cas9 binding to the target DNA.	132
Figure 48: Reprogramming fibroblasts to iPSCs with ReproRNA-OKSGM timeline.....	141
Figure 49: Reprogramming erythroid progenitors expanded cells to iPSCs with ReproRNA-OKSGM timeline.....	142
Figure 50: Reprogramming fibroblasts derived from whole blood to iPSCs with Epi5™ Episomal iPSC vector timeline.....	144
Figure 51: Equation of pyruvate kinase catalytic reaction.....	147
Figure 52: Visual diagram of molecular techniques utilized in attempt to identify disease causing variations of sample 183.	152
Figure 53: MLPA reaction result showing no significant deletions.	153
Figure 54: c.1015G>A variation mapped on PKLR gene.....	155
Figure 55: Sanger validation of PKLR mutation.	155
Figure 56: Alignment of PKLR gene Sanger Sequence.	156
Figure 57: Pyruvate Kinase 3D structure.	157
Figure 58: Pyruvate kinase 3D structure.	157
Figure 59: Crystallography Ligand with 2-phosphoglycollic acid/metal interaction involving metal ion Mn ²⁺ in both chain A and chain B using LIGPLOT (Wallace et al., 1995).....	158
Figure 60: Pyruvate Kinase active site.	159
Figure 61: Amino Acid sequence of PKLR gene showing Asp339 residue conservation.	160
Figure 62: The amino acid, D (Asp339), conservation across different organisms.	160
Figure 63: A comparison in 2D structure of the active site chemical interactions.....	162
Figure 64: A comparison in H-bonds.	162
Figure 65: A comparison of protein's charges.	163
Figure 66 A comparison in hydrophobicity.....	163
Figure 67: Pedigree of a family with beta-thalassemia major phenotype showing pathogenic variants in two genes, HBB and G6PD.....	168
Figure 68: Comparison between A) the mother's sample (I-1) with the heterozygous splice region variation (c.92+5G>C) vs. B) the father's sample (I-2) with reference sequence.....	174
Figure 69: Chromatogram of G6PD c.653C>T variation with the chromosomal position of hg37: chrX:153762634.	174
Figure 70: Comparison of sequencing chromatograms.....	175

Figure 71: MLPA reaction result of sample II-9 (transfusion dependent female) indicating heterozygous deletion.....	180
Figure 72: A table that shows genomic locations of probe values for sample II-10 from MLPA reaction.	181
Figure 73: MLPA probes mapped on alpha globin cluster.	181
Figure 74: MLPA mapping sheet for comparison with common deletions that cause alpha thalassemia.	182
Figure 75: Pedigree of a family with beta-thalassemia major phenotype showing pathogenic variants in three genes, HBB, G6PD, and alpha globin gene.	182
Figure 76: Chromosomal image of sample I-2 from microarray.	183
Figure 77: Filtered nucleotide variants from exome data, based on chromosomal positions, which showed autosomal recessive inheritance model of homozygous state in non-consanguineous thalassemia disease family B.....	188
Figure 78: Filtered nucleotide variants from exome data, based on chromosomal positions, which showed autosomal recessive inheritance model of compound heterozygosity in non-consanguineous thalassemia disease family B.....	190
Figure 79: Variants filtration criteria.	193
Figure 80: Sanger validation of FANCC variant.....	196
Figure 81: A. The mutant FANCC Δ vibrational entropy energy can be visually represented by coloring the amino acids based on the vibrational entropy change caused by the mutation.	199
Figure 82: A-D RMSD and RMSF of FANCC wild and mutant models MD simulation for 100 ns.	200
Figure 83: SSE components of wild and mutant models for 100ns MD simulation period.....	200
Figure 84: Structural conformation of wild and mutant FANCC protein structure in between 1ns to 100ns C. Mutant residues localization during simulation period.	201
Figure 85: Filtered nucleotide variants from exome data, based on chromosomal positions, which showed shared heterozygous variants exclusively between the father, all the affected individuals, and I-19. Both mothers, II-11, II-12, II-13, and II-3 were excluded.	203
Figure 86: SMC5 protein 3D structure with zoomed in residue position 226 (Glu).	208
Figure 87: TALDO1 protein 3D structure with zoomed in Asp residue in position 202 (PDB ID: 1f05).	209
Figure 88: Sanger validation of TALDO1 variant.....	209

Figure 89: Variant is not present in I-1 (mother of family A) on the left. Heterozygous variant in father's sample I-2 on the right.	210
Figure 90: Transaldolase and metabolic stability along the pentose-phosphate pathway.	210
Figure 91: (A&B) Ensembl based transcriptomics expression data of different genes across different tissues and organs.	211
Figure 92: Phenotype overview of SMC5 homozygous knockout mouse (Blake et al., 2021).	213
Figure 93: Phenotype overview of TALDO1 homozygous knockout mouse.	213
Figure 94: SMC5 and TALDO1 protein model, Ramachandranplot and Prosa-Web plots.	214
Figure 95: Shows the vibrational entropy change caused by mutations in SMC5 and TALDO1.	215
Figure 96: MD analysis plots RMSD, RMSF, SSE of wildtype and mutant model of SMC5.	216
Figure 97: MD analysis plots RMSD, RMSF, SSE of Wildtype and mutant model of TALDO1.	217
Figure 98: SMC5 3D structure (UniProt ID: Q8IY18) with its predicted substrate (phosphothiophosphoric acid-adenylate ester) in A) Wildtype protein, and B) Mutated protein.	219
Figure 99: 2D chemical interaction in A) wildtype, and B) mutated proteins.	220
Figure 100: Comparison of hydrogen bonds between the wildtype (A) and the mutated (B) proteins.	221
Figure 101: Charges of the domain area of the docked complexes in both the wildtype (A) and mutated (B) with their predicted substrate.	221
Figure 102: The hydrophobicity of the docked complexes, in the wildtype and mutated domains with their predicted substrate.	222
Figure 103: Ionizability of the docked complexes, in the wildtype and mutated domains with their predicted substrate.	222
Figure 104: SAS interactions of the docked complexes, in the wildtype and mutated domains with their predicted substrate.	223
Figure 105: The 3-dimensional structure of normal TALDO1 protein vs mutated protein with their substrate fructose-6-phosphate predicted by I-TASSER.	224
Figure 106: Comparison of 2D chemical interactions of TALDO1 active site between wildtype and mutated proteins.	225
Figure 107: Hydrogen bonds of the docked complexes, in the wildtype and mutated domains with their predicted substrate.	225
Figure 108: The charges of the docked complexes in the wildtype and mutated domains with their predicted substrate.	226

Figure 109: The hydrophobicity of the docked complexes in the wildtype and mutated domains with their predicted substrate were compared. 226

Figure 110: The ionizability of the docked complexes in the wildtype and mutated domains with their predicted substrate..... 227

Figure 111: SAS interactions of the docked complexes in the wildtype and mutated domains with their predicted substrate..... 227

Figure 112: Approaches to whole exome analysis. 238

Figure 113: Most frequent possible outcomes of whole exome sequencing approach. 239

LIST OF TABLES

Table 1: Different types of normal human hemoglobins with their chain composition and synthesis stage.....	13
Table 2: Beta thalassemia genotype/ phenotype correlation with its hemoglobins content.....	47
Table 3: Alpha thalassemia genotype/ phenotype correlation.	49
Table 4: The most frequent beta-thalassemia SNPs within Saudi population used in our study.....	66
Table 5: TaqMan assays used on samples and their primer sequences.	67
Table 6: Workflow chart for WES run.....	87
Table 7: HBB forward and reverse sequencing primers.	92
Table 8: Forward and reverse primers for PKLR gene.	92
Table 9: Forward and reverse primers for FANCC gene.	93
Table 10: Forward and reverse primers for SMC5 gene.	94
Table 11: Forward and reverse primers for TALDO1 gene.	95
Table 12: Thermocycling used for amplifying PCR products.....	95
Table 13: Identified mutations in 54 samples using Next Generation Sequencing by Ion Torrent. ...	115
Table 14: 2d plot of all sample results.....	120
Table 15: Whole exome sequencing result of patient 183.	154
Table 16: QC metrics of thalassemia samples' results.....	167
Table 17: QC metrics of thalassemia samples' results continues.	167
Table 18: Variations found in each affected member of families A and B by initial investigation of WES data.....	169
Table 19: Variations found in each unaffected member of families A and B by initial investigation of WES data.....	170
Table 20: Clinical details of the severity in affected members.	173
Table 21: Complete blood count test results.	177
Table 22: Hb electrophoresis and G6PD deficiency blood test results.	177
Table 23: Copy number variation (CNV) results of sample I-2 from microarray.....	184
Table 24: List of two rare nucleotide variants from exome data which showed autosomal recessive inheritance model in family (A).....	189

Table 25: Filtered nucleotide variants from exome data, based on chromosomal positions, which showed autosomal recessive inheritance model of compound heterozygosity in non-consanguineous thalassemia disease family B.....	191
Table 26: Filtered nucleotide variants from exome data, based on chromosomal positions, which showed autosomal recessive inheritance model of compound heterozygosity in non-consanguineous thalassemia disease family A.....	192
Table 27: Filtering criteria for variants.	194
Table 28: Pathogenic variant found in a heterozygous state shared between the father and two of the affected children.....	196
Table 29: Pathogenic variant found in a heterozygous state shared between the father and all affected individuals of family B.	196
Table 30: Gene list of filtered nucleotide variants from exome data, based on chromosomal positions, representing scenario 4.....	203
Table 31: Two functionally related gene candidates from the scenario of compound heterozygosity between different genes.....	208
Table 32: Expression/ conservation of Taldo1 and SMC5 in primate tissues.....	212
Table 33: Phenotypes associated with SMC5 gene using mouse model ($Smc5^{tm1b(KOMP)Wtsj}$).....	213
Table 34: in-silico prediction tools	228
Table 35: The genetic make-up of both affected and non-affected members of our family cohort.	240

ABBREVIATIONS

2,3-DPG	Dephosphorelate 2,3-bisphosphoglycerate
ADP	Adenosine diphosphate
AHSP	Alpha-hemoglobin stabilizing protein
AMPK	AMP-activated kinase
ATP	Adenosine triphosphate
BFU-erythroid	Burst-forming-unit
CBC	Complete blood count
CD#	Cluster differentiation #
CDA	Congenital dyserythropoietic anemia
CFU-erythroid	Colony forming unit
CK1-alpha	Casein kinase 1-alpha
CPSF	The cleavage and polyadenylation specificity factor
CstF	The cleavage stimulation factor
DNA	Deoxyribonucleic Acid
DNase I	Deoxyribonuclease I
EBA	Erythrocyte binding antigen
EPO	Erythropoietin
F-actin	Actin filaments
FBP	Fructose-1,6 diphosphate
FOP	Friend of protein arginine methyltransferase I
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
Hb or Hgb	Hemoglobin
Hct	Hematocrit
HIF	High inducible transcription factor
HK	Hexokinase
HPFH	Hereditary persistent of fetal hemoglobin
HPLC	High-performance liquid chromatography
HS	Hypersensitive sites

KAHRP	Knob-associated histidine-rich protein
KCl	Potassium chloride
KLF1	Krüppel-like factor 1
LCR	Locus control region
MCHC	Mean Corpuscular Hemoglobin Concentration
MCV	Mean corpuscular volume
methHB	Methemoglobin
mRNA	Messenger Ribonucleic Acid
NAD⁺	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide phosphate
PEP	Phosphoenolpyruvate
PfEMP1	<i>P. falciparum</i> Erythrocyte Membrane Protein 1
PGK	Phosphoglycerate kinase
PK	Pyruvate kinase
PKC	Protein kinase C
PKD	Pyruvate kinase Deficiency
PRMT 1	Protein arginine methyltransferase 1
PS	Phosphatidylserine
RBC	Red Blood Cell
RDW	Red cell distribution width
RESA	Ring-infected erythrocyte surface antigen
ROS	Reactive oxygen species
SCA	Sickle cell anemia
SCD	Sickle cell disease
snRNA	Small nuclear RNA
Stat5	Signal transducer and activator of transcription 5
tRNA	Transfer RNA
TTD	Trichothiodystrophy
UTR	Untranslated Region
α -MRE	Alpha major regulatory element

CHAPTER ONE
GENERAL INTRODUCTION

1.1 Red Blood Cells (RBC)

More than 5 liters of blood are present in the normal adult human body. Blood transfers waste products from living cells and supplies them with nutrients and oxygen. Furthermore, it transports immune cells that control infections and has platelets that can block injured blood vessels to stop blood loss. The circulatory system allows blood to adjust to the body's requirements. Heart rate increases during exercise, and increased blood flow increases the available oxygen and glucose to heart, brain, and muscles. During an infection, more immune cells are transported by blood to the infection site, where they gather to generate and regulate an immune response against existing pathogens (Corrons et al., 2021a).

Blood is composed of plasma and various cell types that include red blood cells (RBC or erythrocytes), white blood cells (WBC), and platelets. The main function of RBC is to deliver oxygen to our body's tissues to aid in mitochondrial oxidative phosphorylation. Oxygen is the principal terminal oxidizing agent for catabolism that converts organic materials to carbon dioxide and water and part of the released free energy is turned into that of high energy phosphate bonds. The hematopoietic lineage accounts numerically about 90% of human cells (Sender et al., 2016), with RBC accounting about 40% of blood by volume.

1.2 RBC structure

For the erythrocyte to be able to sustain traveling through the blood circulation, it requires a membrane that is both deformable and elastic. This membrane, which is composed of a lipid bilayer, transmembrane proteins, and membrane skeleton proteins, fulfills both of these requirements. The contemporary model of the RBC membrane skeleton was initially postulated in 1979 (Lux, 1979), and since then, various new proteins have been found, and their structures and interactions have been studied and described. RBC can bend and stretch and can bind to chemokines and pathogens (Anderson et al., 2018) (Minton, 2021). RBC have biconcave shape, which enlarges the total surface of the cell (Piety et al., 2016). Energy plays a crucial role in maintaining the biconcave resting form of RBC. As ATP levels drop in the cell over time, the shape of the cell progressively become more spherical and stiffer. RBC are microscopic, with a diameter of only 6-7 μm , which makes it easier for them to alter shape and squeeze through capillaries and thus expose surface directly to the endothelial cell membrane. RBC shape and size enables them to circulate through the body within 60 seconds. The RBC membrane has a variety of effects, and their composition has a significant impact on RBC's

functions. The lipid bilayer contains numerous transmembrane proteins in addition to its basic lipidic components. Since it determines several physical characteristics including membrane permeability and fluidity, the lipid composition is significant. The membrane skeleton is a structural network of proteins positioned on the inner surface of the lipid bilayer. Proteins make up half of the dry membrane mass in RBC for most mammals including humans, while lipids, specifically cholesterol and phospholipids, make up the other half (Pretini et al., 2019).

Membrane lipids

The lipids of the cell membrane are found in an asymmetrical uneven distribution throughout the bilayer membrane with equal proportions of cholesterol and phospholipids. This distribution reflects a steady state that involves a continual exchange of phospholipids between the two hemileaflets of the bilayer (Fasano & Chou, 2016). The outer layer of the phospholipid bilayer consists of choline containing phospholipid (phosphatidylcholine and sphingomyelin) while the inner layer consists of amine containing phospholipid (phosphatidylethanolamine and phosphatidylserine). In pathological state such as in sickle cell disease and thalassemia, the asymmetry of the phospholipids is lost, and the PS translocate from the inner to the external half of the bilayer marking RBC to be engulfed by macrophages. Phosphatidylserine is involved in apoptotic cell identification and functions as a molecule that bridges several ligands (Kay & Fairn, 2019).

Interaction between hemoglobin and membrane proteins

Two different types of proteins, integral (within lipid bilayer as transporter molecules) and peripheral (scaffold anchoring and cytoskeletal proteins), make up the erythrocyte's membrane. Integral proteins, which span the entire cell membrane's thickness, are more in molecules number (70%) than peripheral proteins (30%). Integral proteins are intrinsic proteins with hydrophobic or hydrophilic parts that are permanently embedded in the plasma membrane and binds to the phospholipid bilayer electrostatically or by non-covalent interactions, with high interaction with the bilayer hydrophobic core such as glycoporphins (heavily glycosylated), transporter proteins (band 3), and NADPH dehydrogenase. Band 3, named after its gel electrophoresis migration, maintains and preserves protein-protein interactions, serves as an ion transporter, and provides binding sites for hemoglobin, skeleton proteins (ankyrin), and glycolytic enzymes (hexokinase, phosphofructokinase, pyruvate kinase). NADPH dehydrogenase protects against free radicals and oxidative damage, maintains

glutathione (GSH) in a reduced state, and reduces biliverdin in liver to bilirubin (Lux, 2016) (Moon et al., 2020).

Membranes are selectively permeable to bicarbonate. The transport of bicarbonate through membranes is essential for CO₂ metabolism and pH homeostasis in cells. Transmembrane bicarbonate flow is facilitated by the ubiquitous expression of bicarbonate transport proteins in mammalian cells (Sterling & Casey, 2002). A major part of the red blood cell membrane is taken up by proteins that act as anion exchangers (such as bicarbonate) which enable RBC to maintain the right balance between the water and solute content of the extracellular and intracellular spaces, which is necessary in order to maintain the normal physiological functionality and homeostasis of RBC (Azouzi et al., 2018). Most of the negative surface charge that red blood cells need to repel each other and the vascular wall is provided by glycoporphins (Aoki, 2017). Glycoporphins also carry *Plasmodium* receptors, a variety of viruses and bacteria, and the blood group (Jaskiewicz et al., 2019).

Peripheral proteins are extrinsic proteins with hydrophilic parts found on the inner or outer surface of the phospholipid bilayer, reversibly attached to the plasma membrane without penetrating the bilayer and extending into the cytoplasm, interacting with integral proteins and lipids. The structural protein spectrin is the major peripheral protein of the membrane skeleton. It is made up of alpha and beta, which are structurally similar to one another. A peculiar trait of spectrin is that it is very flexible and may take different conformations. The α -chain and the β -chain are aligned in antiparallel fashion to create a dimer. These dimers join together to create a tetramer, which is characterized as being lengthy and wormlike in appearance. Spectrin proteins form two junctions with integral proteins. The first junction forms with glycoporphins by the non-covalent attachment of the two ends of the spectrin tetramer to short actin filaments in a horizontal interaction. Band 4.1 protein and adducin are two other proteins that play a role in facilitating this interaction. Actin filaments (F-actin) create the nodes of a two-dimensional meshwork of fibers network by joining multiple spectrin tetramers. The second junction forms through spectrin interactions with the cytoplasmic domain of band 3 via ankyrin protein (vertical interactions). In addition, the peripheral membrane protein 4.2, through its interactions with ankyrin and band 3, contributes to the process of connecting the skeleton to the lipid bilayer (Li & Lykotrafitis, 2014) (Paradkar & Gambhire, 2021). The development and breaking of junctional complexes, which allow cytoskeletal proteins to retain their fluctuation and reversible deformability is an energy dependent process. This uncommon property of assuming various conformations and reversible deformability is essential for the elasticity of membranes.

1.3 RBC regulation and development

Erythropoiesis is a process that takes roughly seven days during which RBC (or erythrocytes) properly mature from committed stem cells and are then released into the bloodstream from the bone marrow where they are continually produced. It is a multistep process that may be broken down into the initial erythropoiesis stage (in bone marrow), terminal erythroid differentiation (in bone marrow), and reticulocyte maturation (in circulation) (Figure 1). Erythroid disorders may occur if defects were encountered during any of these steps.

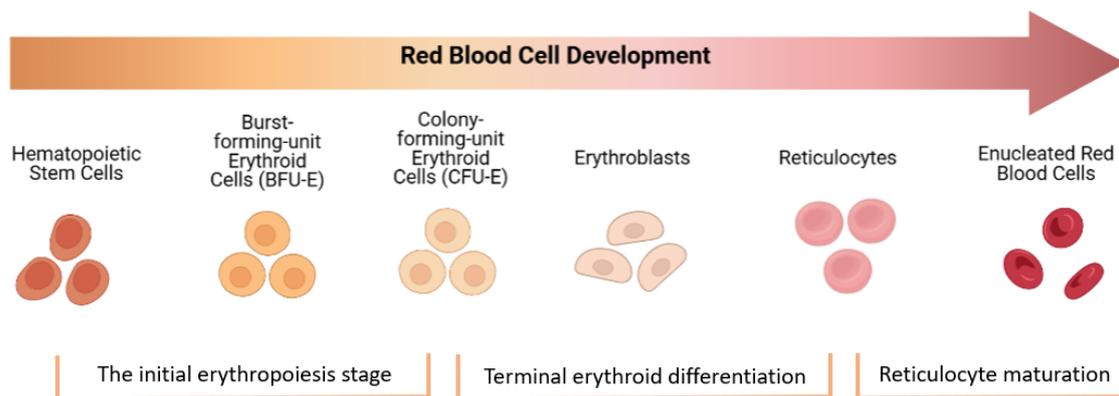


Figure 1: Different stages of red blood cell development.

The initial erythropoiesis stage starts in the bone marrow from a small multipotent population of HST developing the erythroid progenitors BFU-E which matures into the late erythroid progenitors CFU-E. Also in the bone marrow, these cells further differentiate into erythroblasts (terminal erythroid differentiation) which gradually get filled with hemoglobin and the nucleus get expelled. After that they are released into the bloodstream as reticulocytes and continue maturation. Detailed description is provided in the text below. Created with BioRender.com

In the bone marrow, the development of erythroid precursors (erythroblasts) generated from small population of multipotent stem cells (erythroid progenitors) in the mesenchyme to become mature erythrocytes go through several morphological changes during erythropoiesis (Dzierzak & Philipsen, 2013) (Corrons et al., 2021b). Starting from a slow proliferating burst forming unit (BFU-erythroid), the earliest progenitor committed to the erythroid lineage. BFU-erythroid show low expression of erythropoietin (EPO) receptors. EPO is kidney (and adult liver) derived cytokine receptor protein (glycoprotein hormone) that stimulates the production of erythrocytes. BFU-erythroid mature into rapid proliferating colony forming unit (CFU-E), a late erythroid progenitor that is highly dependent on erythropoietin (EPO) (Gregory & Eaves, 1977). Erythropoiesis also relies on fibronectin, a protein found in the extracellular matrix. CFU-E erythroid progenitor cells divide twice, increase transferrin receptor

expression, and start expressing Ter119 (erythroid specific marker) and hundreds of other genes essential to erythroid development. While EPO is necessary, fibronectin is not at this time. Two or three rounds of fast cell division occur on day two, with S and M phases being dominant and G1 and G2 phases are very brief or nonexistent. Even though EPO is not required for the final one or two cell divisions and promotion of enucleation, fibronectin adhesion is (Migliaccio, 2010) (Moras et al., 2017) (Rieu et al., 2000) (Ji et al., 2011). The cell then further differentiates into an erythroblast with decreased cell size, condensed chromatin, and fewer EPO receptors. The genes that produce globin proteins are activated as cells develop. The erythroblast is gradually filled with hemoglobin over 2-5 days, and the nucleus is expelled to produce enucleated cells (Granick & Levere, 1964). Mature erythrocytes are unable to produce proteins because they lack functioning ribosomes.

Macrophages play an essential role throughout the lifespan of RBC from the beginning in the bone marrow, in circulation, until the end in the spleen or liver. In 1958, using transmission electron micrographs of bone marrow sections, Marcel Bessis discovered the erythroblastic islands (EBI), which are anatomical niches where terminal erythroid development and hemoglobin synthesis occurs (Bessis, 1958). Typically, in EBI, macrophages are at the central of each island, surrounded by erythroid cells.

The interaction between erythroblasts and macrophages is mediated by adhesion molecules such as integrin-41 (which is found on erythroblasts) and Vcam1 (which is found on central macrophages) (Sadahira et al., 1995) (May & Forrester, 2020). These erythroblast cells are in various stages of red cell development anchored within the island by the macrophage.

These macrophages help the developing erythroblasts with heme synthesis by providing them with iron, as well as secrete cytokines to provide them with cellular connections that are required to stimulate erythroid differentiation and proliferation (Lee et al., 1988) (W. Li et al., 2019) (W Li et al., 2019). In addition, it has been demonstrated that the central macrophage is responsible for phagocytosing the enucleated erythroblast's expelled pyrenocytes (condensed nucleus with a little cytoplasm and plasma membrane) (Seki & Shirasawa, 1965) (Leimberg et al., 2008) (May & Forrester, 2020). EPO promotes erythropoiesis by acting on erythroid cells and erythroblastic island macrophages concurrently in the niche. The erythropoietin receptor EpoR was found expressed in mice and humans by both erythroblasts and EBI macrophages (Li et al., 2021).

Medullary macrophages are believed to be heterogenous, or at least not all at the same developmental stage, and their roles are tied to where in the bone marrow they are found. Macrophages near

sinusoids are associated with erythroblasts in the late stages of differentiation, but macrophages dispersed throughout the bone marrow are more likely to be close to erythroblasts in early development stages. Whether these macrophages are separate cells, or they move towards sinusoids and mature as erythropoiesis progresses is still debated (May & Forrester, 2020).

After that, they are released from the bone marrow into the bloodstream where they continue to mature as circulating reticulocytes until they become fully matured erythrocytes (Hattangadi et al., 2011).

Reticulocytes retain cytoplasmic organelles which make them able to produce proteins and perform oxidative phosphorylation to generate ATP. Organelles and mRNA are lost after one to two days from their release into the bloodstream, and they mature into erythrocytes (Wijk & Solinge, 2005b). During this time, the erythrocyte undergoes substantial membrane remodeling, resulting in its characteristic biconcave shape. Both reticulocytes and erythrocytes are free of EPO receptors and are no longer responsive to EPO (Gifford et al., 2006).

RBC have no mitochondria; therefore, they use none of the oxygen they carry to produce ATP via Krebs cycle. Instead, for the mature erythrocyte to be able to perform metabolic processes, they mainly use and store ATP produced by anaerobic glycolysis of glucose by the Embden-Meyerhoff pathway which is consistent with the fact that red blood cells are responsible for carrying oxygen to where it is needed rather than utilizing the oxygen themselves. RBC cannot produce new proteins to replace a damaged or otherwise inactive enzyme in red blood cells. Glycolysis involves a redox and a series of phosphorylation reactions, and several steps of enzymatic catalysis of phosphorylated intermediates to yield a net gain of two moles of lactate or pyruvate, two moles of ATP, two NADH, two H₂O, and two H⁺ from one mole of glucose with no oxygen consumed nor CO₂ produced (Figure 2). The insulin-independent glucose transporter (GLUT1), a red cell membrane protein, facilitates the entrance of glucose into erythrocyte via diffusion. Most of the glycolytic intermediates involved in the glycolysis process are used in maintaining membrane electrical activity which is needed to expel bicarbonate in exchange for chloride.

Furthermore, the pentose-phosphate pathway is used to produce and dephosphorylate 2,3-bisphosphoglycerate (2,3-BPG), the Rapoport-Luebering Shunt (RLS), from the glycolytic intermediate 1,3-bisphosphoglycerate. In the RLS, bisphosphoglycerate mutase shifts the 1-acylphosphate into the low-energy 2-position, an irreversible reaction (and effectively wastes one ATP) (Figure 2). This is

important as 2,3-bisphosphoglycerate directly reduces the affinity of hemoglobin for oxygen through its high affinity for deoxygenated hemoglobin (Wijk & Solinge, 2005b).

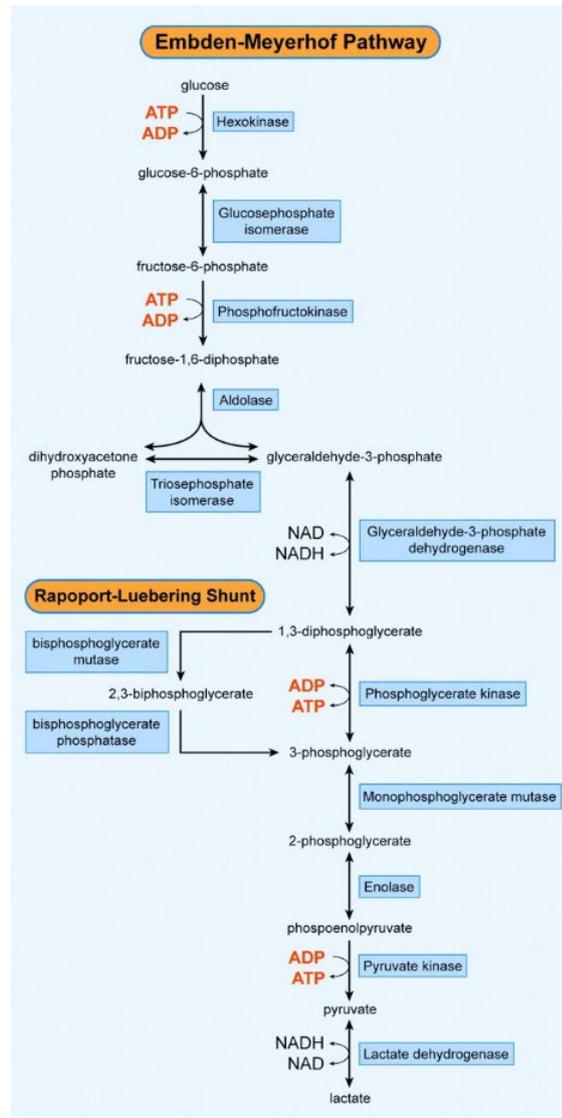


Figure 2: Embden-Meyerhoff pathway of glycolysis.

The energy-consumption phase, which occurs initially during glycolysis, involves the consumption of two ATP molecules to facilitate the processing and conversion of glucose into two 3-carbon sugar phosphates, namely dihydroxyacetone phosphate and glyceraldehyde-3-phosphate (G3P). The energy-release phase is initiated by the G3P, whereas its isomer, dihydroxyacetone phosphate, requires isomerase conversion to become G3P in order to advance to the energy-release phase. The second phase of glycolysis which is the energy-release phase, involves the production of energy-rich molecules (ATP and NADH). During this phase, each G3P molecule produces one NADH and two ATP molecules. The products and reactants of the processes occurring in this phase would be multiplied by two through the conversion of dihydroxyacetone phosphate to G3P. The energy-release phase can thus be anticipated to consist of two NADH molecules and four ATP molecules. Reproduced with permission from: (Wijk & Solinge, 2005a)

Growth factors, transcription factors, and cytokines in the bone marrow play a crucial role in regulating and controlling of the erythropoietic process. Both the local oxygen environment and the relative amount of hypoxia are responsible for controlling the pace at which the circulating hormone EPO gene transcription occurs. The production process of EPO by the interstitial fibroblasts of the kidney in response to low oxygen tension involves the participation of several transcription factors, including the hypoxia inducible factor and the GATA binding proteins. EPO mRNA expression is negatively controlled by the binding of GATA-1, 2, and 3 proteins to EPO promoter region.

Additionally, expression of GATA-2 is crucial for the regulation of committed erythroid progenitors while the expression of GATA-1 is crucial to erythroblast's terminal differentiation. Also, the expression of target genes required to induce the maturation of erythroid cells and ultimately globin genes expression is controlled by the relative proportional expression of GATA-1 and GATA-2 (Zivot et al., 2018). Many intracellular signal transduction pathways are activated in response to oxygen availability in the kidney and the binding of EPO to its receptor on the surface of erythroid progenitors. Erythropoietin serves as part of a feedback loop that ensures red blood cell synthesis is balanced with destruction and that there are enough red blood cells to maintain optimal tissue oxygen levels in healthy individuals.

These activated pathways include the signal transducer and activator of transcription 5 (Stat5), phosphoinositide-3 kinase/Akt, and Shc/Ras/mitogen-activated kinase (MAPK) pathways. Although blocking of MAPK pathway would only have a mild effect on terminal erythropoiesis (Zhang & Lodish, 2007), a blocking of Stat5 does lead to an early progenitor's termination or erythrocyte production will be drastically decreased (Hattangadi et al., 2011). There is a substantial rise in the number of early erythroblasts of hematopoietic tissues of Stat5a and 5b knocked out mice, that do not progress to differentiation and maturation. The severity of anemia (presented in these mice) is correlated with the increase in apoptosis of erythroblasts (Socolovsky et al., 2001).

In circulation, the macrophages in the liver and spleen continually survey erythrocytes based on interactions between RBC surface molecules and receptors on macrophages. They can identify and eliminate red blood cells that are damaged or have reached the end of their lifespan. Iron from these cells' hemoglobin molecules can be recycled by macrophages and reused to nourish nascent erythroblasts. (Crosby, 1957) (Corrons et al., 2021a). Recognition of senescent RBC depends mainly on the expression of specific molecules on the membranes of both RBC and macrophages that leads to either induction or inhibition to phagocytosis. The binding of cluster differentiation 47 (CD47)

molecule, a transmembrane surface protein, which is expressed on the surface of RBC, to the molecule CD172a (SIRP- α), expressed by macrophages, inhibits phagocytosis. On the other hand, aging cells gradually lose their expression of CD47 which makes them susceptible to degradation in the spleen by macrophages. Conversely, CD47 has been shown to undergo conformational change in aging RBC that leads to switching its signal from inhibition to activation to phagocytosis (de Back et al., 2014).

Normally, defective cells are eliminated by undergoing apoptosis that requires mitochondria and nuclei. Instead, because RBC have no nucleus or organelles, they undergo 'eryptosis' instead. The exposure of phosphatidylserine (PS) on the outer cellular membrane of senescent erythrocytes, membrane blebbing, and cell shrinkage are a key mark for macrophages recognition (stimulates phagocytoses). Stress factors such as hyperosmolality, oxidative stress, low energy levels, being exposed to xenobiotics, or Cl^- removal from the cytoplasm and the extracellular environment cause an increase in cytosolic calcium (Ca^{+2}) levels. Phospholipase A2 (PLA) and cyclo-oxygenase (COX) are activated in response to osmotic stress and Cl^- elimination, leading to prostaglandin E2 (PGE2) production. PGE2 raises intracellular Ca^{+2} levels by activating non-selective cation channels. Elevated cytoplasmic Ca^{+2} activates calpain (cysteine endopeptidase) which leads to breakdown of the anchorage of the cytoskeleton to the lipid bilayer of the erythrocyte, which in turn causes the membrane of the cell to bleb. In addition, osmotic stress induces the translocation of PS from the inner leaflet and expose it on the cell's surface by stimulation of ceramide production and contributes to the increase of cytosolic calcium ions as well. Upon Ca^{+2} elevated levels, the Ca^{+2} -sensitive K^+ channel is activated, causing the RBC to dehydrate and shrink as K^+ and Cl^- are released in synchrony (Lang et al., 2012). Macrophages recognize and bind to the exposed PS by their membrane-expressed scavenger receptors (TIM4, CD36, and TAM receptor family) which leads to phagocytosis of RBC (Schroit et al., 1985) (Zhang et al., 2022) (Repsold & Joubert, 2018).

EPO is considered an inhibitor to eryptosis by inhibiting the Ca^{+2} -permeable cation channels. Cell size reduction, membrane blebbing and scrambling leading to exposed PS are all hallmarks of apoptosis, which are also present in eryptosis. RBC eryptosis is regulated by several kinases. In contrast to AMP-activated kinase (AMPK) and cGMP-dependent protein kinase (cGKI) inhibition of eryptosis by inhibiting calcium entry into the cell, protein kinase C (PKC) and casein kinase 1-alpha (CK1-alpha) can trigger eryptosis through increasing calcium entry. If erythropoiesis is stimulated in conjunction with eryptosis, the total number of erythrocytes in circulation will stay stable. An elevated reticulocyte count is evidence for the accelerated erythrocyte turnover. If the production of new erythrocytes is

unable to keep up with the rate of destroyed ones, anemia will develop (Lang & Lang, 2015) (Corrons et al., 2021a). Enhanced eryptosis and higher levels of PS expression have been found in patients with sickle cell anemia, glucose-6-phosphate dehydrogenase deficiency, spherocytosis, and thalassemia, which cause the drastically shortened lifespan of RBC, a hallmark characteristic of these diseases (Wood et al., 1996) (Kuypers et al., 1998) (Lang & Lang, 2015).

In the embryo, primitive erythropoiesis starts in the yolk sac. After 8 weeks of gestational age, the liver serves as the primary site for definitive erythropoiesis. The site of erythropoiesis moves into the bone marrow around the gestational age of 32 weeks where it will continue throughout life (Corrons et al., 2021a).

1.4 Hemoglobin structure

The protein hemoglobin (Hb) found in red blood cells is in charge of transporting oxygen to the body's tissues. The crystal structure of hemoglobin in both its oxygenated and deoxygenated phases was solved by Perutz in 1970. These structural investigations revealed an abundance of information that provided insights into the remarkable molecular processes and precise explanation behind its physiological role as an oxygen transporter. Because of this, the clinical presentations of many of the "hemoglobinopathy" disorders can be understood in terms of the molecular structure underlying them, and hundreds of known mutations in the globin genes and proteins that can affect the hemoglobin may be explained.

Human hemoglobin research during the past 60 years has been crucial in the development of molecular medicine. The discovery of how hemoglobin is responsible for delivering oxygen to the areas of the body in which it is required the most was achieved by Reinhold and Benesch. They explained that the accumulation of carbon dioxide in RBC triggers the hemoglobin molecule that oxygen has been used up in the cell's metabolism of carbohydrates, and that the cell thus requires more oxygen. In 1967, they made the crucial finding that 2,3-bisphosphoglyceric acid in the hemoglobin molecule was required for oxygen release at low oxygen tension by loosening the molecule's tightly bound oxygen atoms. It binds to hemoglobin and hence stabilizes the structure of deoxyhemoglobin. It is present in RBC at a high concentration and favors the dissociation of oxygen (Benesch et al., 1968) (Dailey & Meissner, 2013).

The hemoglobin protein is a tetramer which consists of four protein subunits (globin chains). The synthesis of the globin proteins occurs in the cytosol. Each one of the globin subunits contains a heme

group that carries an iron atom (oxygen binding site). Heme synthesis occurs in the mitochondria and the cytosol, involving a series of steps and is composed of the prosthetic group, porphyrin ring IX (4 nitrogen ring), which forms a complex with ferrous cation iron (Fe^{2+}) (Dailey & Meissner, 2013). The alpha and beta subunits are formed of 7 helices, labeled A-H (the D motif is a loop). The E and F helices of each subunit act as a binding pocket for heme (covalently attached), where the iron atom is shielded from the aqueous environment (Figure 3) (Ahmed et al., 2020) (Marino et al., 2023).

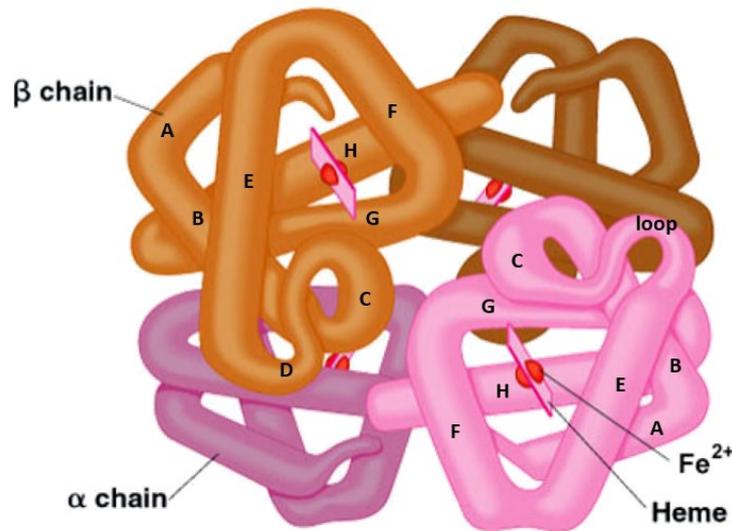


Figure 3: Hemoglobin structure.

Showing the tetramer of 2 β chains and 2 α chains with the iron-containing heme molecule between E and F helices in each subunit. Source: reproduced with permission from "themedicalbiochemistrypage, LLC"

During translation or soon afterward, heme attaches to the hydrophobic cleft of the globin subunit. The electrostatic interaction between a positively charged alpha-subunit and a negatively charged beta-subunit plays a key role in the production of the alpha beta-dimer (Bunn, 1987) (Mollan et al., 2010). The functional $\alpha_2\beta_2$ (HbA, major adult hemoglobin) tetramer is formed by the combination of two identical α -globins and two identical β -globins, with β -globin being longer than α -globin by five amino acid residues. Then, all four globins are folded together through interactions between oppositely charged amino acids to give the final hemoglobin quaternary structure, which is held together by hydrophobic interactions, hydrogen bonds, and salt bridges forming a tetramer that is nearly spherical (Ouellette & Rawn, 2015).

Depending on the developmental stage, normal hemoglobin can take on a few distinct forms combining different globin chains as shown in Table 1.

Table 1: Different types of normal human hemoglobins with their chain composition and synthesis stage.

Human hemoglobin	Chain subunit composition	Development state
Gower 1	2 zeta, 2 epsilon ($\zeta_2\epsilon_2$)	Embryonic
Gower 2	2 alpha, 2 epsilon ($\alpha_2\epsilon_2$)	Embryonic
Portland 1	2 zeta, 2 gamma ($\zeta_2\gamma_2$)	Embryonic
HbF	2 alpha, 2 gamma ($\alpha_2\gamma_2$)	Fetal
HbA2	2 alpha, 2 delta ($\alpha_2\delta_2$)	Adult
HbA	2 alpha, 2 beta ($\alpha_2\beta_2$)	Adult

Both gene clusters that control the production of different beta and alpha subunits exhibit the same, rigorously controlled pattern of expression during both whole-body development and individual erythrocyte differentiation, despite being positioned on different chromosomes and in very different chromosomal contexts. After translation of alpha and beta subunits, they form a dimer and insert a heme. Erythrocyte development relies on delicate balancing in the production of the two globin chains, since an excess of either chain and accumulation of free unbound beta or alpha subunits is extremely harmful to the survival of red blood cells and may result in cytotoxic precipitates. Multiple findings suggest that erythroid cells employ adaptive mechanisms to maintain free Hb and limit its toxicity.

Misfolded or toxic proteins are first identified and removed by proteolytic mechanisms. Therefore, erythroid cells have routes to digest excess free subunits, although the molecular processes were poorly understood (Shaeffer & Kania, 1995). In erythroid precursor cells, an effective adenosine triphosphate-dependent proteolytic system targets free subunits while leaving formed dimers and tetramers unharmed. Moreover, the presence of a chaperone called alpha-hemoglobin stabilizing protein (AHSP), highly produced in erythroid cells, binds selectively and firmly to heme-intact free alpha-globin subunits exclusively at the $\alpha_1\beta_1$ dimer interface with lower affinity than beta globin subunit, and protects the erythroid cells against the toxicity of oxidized (Fe^{3+}) heme until being reduced to Fe^{2+} by cytochrome b5 reductase. The alpha-globin subunit dissociates from AHSP when it comes into contact with a free heme-bound beta-globin subunit to form the stable alpha-beta dimer (Kihm et al., 2002) (Forget & Bunn, 2013) (Ahmed et al., 2020). However, when compared to free alpha subunits, which tends to self-associate into homo-tetramers that lend some stability, free beta subunits are more unstable because they reside mostly as monomers (F. Bunn & B. Forget, 1986) (Yu et al., 2007). Purified protein experiments in vitro and crystallographic studies revealed that AHSP

share the same binding on alpha globin as the beta globin, therefore, AHSP binds alpha globin subunits but not beta subunits (Feng et al., 2004) (Kihm et al., 2002). Energy-dependent proteases degrade irreparably damaged proteins, whereas molecular chaperones, which are present in high levels during erythropoiesis, support and stimulate normal protein folding, help fine-tune and improve the efficiency of protein synthesis, and prevent aggregation in immature erythroblasts and reticulocyte stage. Whether a protein is destroyed before it folds properly is dependent on the kinetics of partitioning between chaperones and proteases. When neither method of quality control is effective, aggregation and clumping of defective proteins occur (Wickner et al., 1999) (Bukau et al., 2006).

AHSP is not predicted to inhibit HbA ($\alpha_2\beta_2$), predominant adult hemoglobin, production or function at physiologically relevant concentrations of alpha and beta subunits because beta subunits bind more tightly to alpha subunits than AHSP. Generation of reactive oxygen species (ROS) generation by free alpha globin is decreased in vitro by using recombinant AHSP, and the protein's thermostability is improved and its tendency to precipitate is blocked. On the other hand, hemolytic anemia and elevated levels of reactive oxygen species (ROS) in erythroid cells and alpha subunits precipitates, seen in AHSP knocked down mice, show that AHSP plays a crucial role in hemoglobin homeostasis (Kong et al., 2004) (Yu et al., 2007). It is not known whether or if there are additional chaperones that aid in the maturation of the other globin subunits in erythroid cells (Ghosh et al., 2018). Nevertheless, since maturation of hemoglobin tetramers is dependent on heme integration into each globin, Ghosh and colleagues have examined the maturation of hemoglobin in cells that naturally express Hb- α with Hb- γ (K562 and HiDEP-1 cells), or with Hb- β (HUDEP-2) and found that in erythroid precursor cells, another protein, heat-shock protein 90 (hsp90), works in tandem with AHSP during erythropoiesis to generate functional hemoglobin tetramer by binding to beta globin subunits, beta (adult) and gamma (fetal) globin, and promote their heme insertion, and subsequently dissociate to allow binding of alpha subunit and formation of the heterotetramer (Ghosh et al., 2018).

The two most significant parts of hemoglobin in an erythrocyte are globin chains and iron-containing heme groups. When RBCs are old or damaged, they are phagocytosed by macrophages and the components are separated. In parallel with the iron being drawn from heme, the polypeptide globin chains are broken down into amino acids. Heme is converted into bilirubin while the iron is later transported to the bone marrow by transferrin proteins to be employed in fresh cycles of erythropoiesis. Bilirubin is eliminated through urine and feces after going through additional changes in the liver and intestines (Corrons et al., 2021a).

Each hemoglobin molecule has four iron atoms. Therefore, each hemoglobin molecule can carry four oxygen molecules reversibly and loosely in a positively cooperative process (Thom et al., 2013). The oxygen affinity of hemoglobin subunits rises as the molecule binds each oxygen. About 300 times as much energy is released in binding the fourth oxygen as the first. This cooperative process results in a sigmoidal oxygen dissociation curve (Thomas & Lumb, 2012). Hb is in equilibrium with oxygen between the relaxed R (oxy) and tense T (deoxy) tetrameric states. In order for oxygen to bind, there must be a certain amount of oxygen tension (local partial pressure) present. At saturating oxygen tensions all four sites will be occupied. The tetramer-dimer interface rearranges its subunit connections, and the location of the R/T equilibrium shifts as Hb either binds or releases oxygen. The binding of one oxygen molecule to an iron atom triggers and enhances the binding of the next oxygen at the other sites by the $\alpha\beta$ dimer rearrangement (15° rotation with respect to the other dimer), thus, changing the structural state of the hemoglobin in an allosteric effect. Likewise, oxygen dissociation is facilitated by low oxygen tension, once fully oxygenated hemoglobin loses one of its oxygen molecules, the rest of the molecules will readily be released. Additionally, 2,3-bisphosphoglycerate helps here too by binding and stabilizing the deoxygenated conformation of subunits. (Monod et al., 1965) (Berg JM, 2002; Chou, 1989) (Ahmed et al., 2020).

Alpha and beta chains make up the majority of the adult hemoglobin (around 97%), but there are other polypeptide chains (globin proteins) that are associated with the hemoglobin, γ_A , γ_G , δ , ϵ , ζ , θ , which are activated according to the sequence of development from embryonic to fetal to adult which occurs about 18 to 24 weeks after birth which will be discussed in details later on (Bunn & Forget, 1986).

1.5 Beta globin and alpha globin gene clusters

The human genome has two non-identical copies of the alpha globin locus, *HBA1* and *HBA2*, which encode identical alpha proteins, and a single beta globin gene *HBB*. Both *HBA1* and *HBA2* code for the synthesis of the alpha globin chain, therefore normally, a person has four alpha functional genes rather than just two (Hardison, 2012a). *HBA1* and *HBA2* are 4 kb long each, and located 3.4 kilobases apart, with identical coding exons and first intron sequences. These genes are slightly different in their introns and 5' untranslated regions, while significantly different in their 3' untranslated regions. Intron 2 of *HBA1* is 149 nt and *HBA2* is 142 nt. The difference is deletion of CTCGGCC and one base substitution. Previous studies have showed that *HBA2* gene has 2-3-fold higher mRNA production rate than *HBA1* (Hardison, 2012b; Liebhaber et al., 1980).

The beta-globin cluster has five functional genes and 1 pseudogene, and the order of those genes is 5'- ϵ (HBE), $G\gamma$ (HBG2), $A\gamma$ (HBG1), β_1 pseudogene (HBBP1), δ (HBD), β (HBB)- 3'on chromosome 11 (Efstratiadis et al., 1980; Lawn et al., 1980).

The globin gene clusters on chromosomes 11 and 16 contain pseudogenes in addition to the eight functional globin genes. Between $A\gamma$ and δ , only one beta globin pseudogene (*HBBP1*) has been discovered. The chromatin architecture that enables the interaction of the locus control region (LCR) with globin genes at different stages of development appears to be maintained in part by *HBBP1* (Ma et al., 2021). The alpha-globin cluster has three functional genes, three pseudogenes, and two genes of undetermined function (μ and $\theta 1$) in the following order 5'- ζ , $\psi\zeta 1$, $\psi\alpha 1$, μ , $\alpha 2$, $\psi\rho$, $\alpha 1$, $\theta 1$ -3' (Manning et al., 2007a) (Ju et al., 2020) (Albitar et al., 1989). The μ -globin gene was incorrectly labeled a pseudogene $\psi\alpha 2$ until recently. Microarray examination of adult reticulocytes and samples from cord blood allowed Goh et al. to show, however, that transcription of this gene occurs. Compared to α -globin expression, μ -globin expression was much lower. On the other hand, μ -globin expression was found to be higher than ζ and $\theta 1$ (Hsiao & Jim, 2008) (Goh et al., 2005).

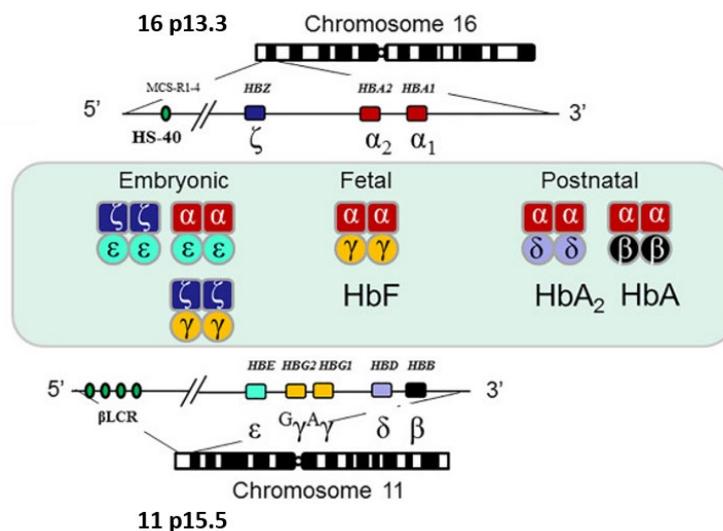


Figure 4: The beta globin and the alpha globin gene clusters.

The single locus control region (LCR) controls the expression of the beta gene cluster, and the multi-species conserved sequence regions (MSC-R) controls the alpha gene cluster and have the DNase 1 hypersensitive sites within them. The earliest hemoglobin is the embryonically expressed globin which consists of 2 ζ +2 ϵ (Gower1), and during embryogenesis the ζ globin becomes replaced by α globin (Gower 2). Also found in the embryonic stage are small traces of Portland hemoglobin that consists of 2 ζ +2 γ . In the fetal stage, 2 α +2 γ subunits form hemoglobin F (fetal hemoglobin) and at birth, the γ genes are downregulated and the β genes are upregulated so that β replaces γ , forming the adult $\alpha 2\beta 2$ tetramer (HbA) which makes up 95-98% of normal adult hemoglobin. The rest of adult hemoglobin is 2-1% of 2 α +2 δ (HbA₂) and >1% of fetal hemoglobin (Maniatis et al., 1980). (Levings & Bungert, 2002). Figure source (Farashi & Hartevelde, 2018). Permission under the terms of the Creative Commons Attribution-NonCommercial-No Derivatives License (CC BY NC ND).

All globin functional genes have the same exon-intron structure, which contains three functional exons and two introns. Both alpha globin proteins contain 142 amino acids and their genes (*HBA1* and *HBA2*) are located at position 13.3 of the short arm of chromosome 16 (Figure 4) (Marengo-Rowe, 2006). On the other hand, *HBB* encodes 146 amino acids.

The five beta globin-like genes are located specifically at band p15.5 of the short arm of chromosome 11 (Figure 4). All of these genes are restricted to the erythroid lineage by only being expressed in the progenitors of erythroid cells precursors (Dzierzak & Philipsen, 2013).

Hemoglobin ontogeny involves two significant transitions: from embryonic to fetal hemoglobin then from fetal to adult hemoglobin. The stage-specific activation of globin genes during development is in sequence moving from 5' to 3' on the chromosome (in the centromere to telomere direction), as seen in Figure 4, which is thought to be the result of competition for the enhancer element at the start of the cluster forming sequentially more distant interactions (Wilber et al., 2011).

The specific enhancer and promoter elements of each gene in the region can regulate each gene's expression in the red blood cell lineage, with the epsilon globin gene, for instance, being expressed in the embryo (Figure 5).

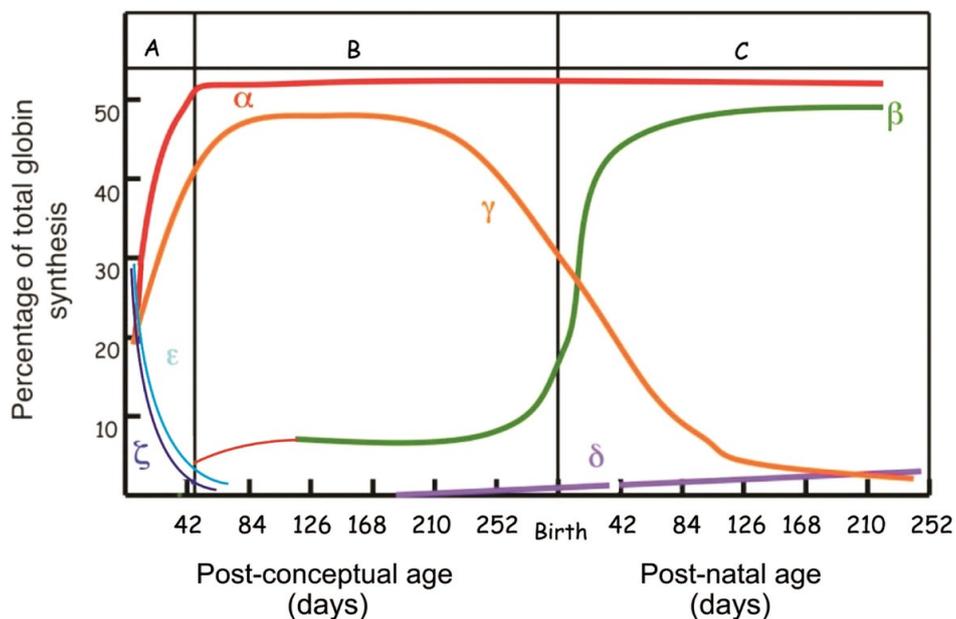


Figure 5: Schematic presentation of the globin chains synthesis during development. Alpha globin synthesis is continuous throughout life. The epsilon globin is synthesized during the embryonic stage, then declines to be replaced by the gamma globin after around 40 days post conception. After birth, gamma globin begins to decline and is replaced with beta globin (higher percentage) and delta globin (lower percentage). Gamma globin is still produced after birth but at a low rate. Reproduced with open access permission from Springer Nature (Kwaifa et al., 2020).

1.5.1 Control of hemoglobin switch

The construction of hemoglobin molecules with varying physiological characteristics results from changes in the makeup of globin subunits that occur naturally during ontogeny in a clonal population. The yolk sac serves as the primary location of erythropoiesis from weeks three through eight of gestation. Early in gestation, the main embryonic hemoglobin generated contain the alpha-like subunit, zeta as Gower 1 ($\zeta_2\varepsilon_2$), Gower 2 ($\alpha_2\varepsilon_2$), and Portland ($\zeta_2\gamma_2$) (Cividalli et al., 1974). Red blood cell synthesis switches from the yolk sac to the fetal liver after the eighth gestational week, which contains CD34+ hematopoietic multipotent progenitors since day 30 of human development, and subsequently to the spleen in which fetal hemoglobin ($\alpha_2\gamma_2$) takes over as the main hemoglobin generated. There are two γ genes, HBG1 (A γ) and HBG2 (G γ), that are only different in one amino acid. A γ has an alanine at position 136 while G γ contains a glycine (Dzierzak & Philipsen, 2013) (Kazazian & Woodhead, 1973). Production of beta-like subunits molecules of hemoglobin undergo two developmental transitions that depend on time rather than site. These varying progenies, however, display a range of expression levels for fetal and adult hemoglobin, are all descended from the same set of progenitors. The first genetic switch in the developmental transitions is from the primitive embryonic hemoglobin in the yolk sac to the definitive fetal hemoglobin in the fetal liver at the beta globin locus. A similar switch occurs on the alpha globin cluster as well, but from primitive embryonic hemoglobin (zeta-globin) to adult hemoglobin (alpha-globin). Almost 160 years ago, Korber and von Kruger were able to distinguish fetal hemoglobin from adult hemoglobin using their resistance property to alkaline denaturation (Korber, 1866) (Krüger, 1888). Later, in 1934, differentiating fetal and adult hemoglobin by their differential reaction kinetics (rate of hemoglobin denaturation in alkaline solution) was documented experimentally by Brinkman et al. They used alkaline denaturation, as determined by a spectrophotometer. Since the heme moiety between umbilical cord and adult blood was known, the authors reasoned that the difference in alkaline resistance must be attributed to the "globin" component of the molecule (Brinkman et al., 1934). Now, researchers are making significant headway in understanding the molecular processes that control the expression of globin genes throughout different stages of development using genetic methods and experimental work made possible by mouse transgenic, knockout, and knockdown technologies.

Assembly of γ -globin with adult alpha-globin subunits yields the fetal hemoglobin (HbF) tetramer ($\alpha_2\gamma_2$). The second switch is when the *HBG1* and *HBG2* (γ -globin) genes of the fetal cell begin to be replaced by the adult *HBB* (β -globin) gene as the newborn period draws closer. This changeover occurs

throughout the later stages of pregnancy and often lasts until around 6 months of age (Sankaran et al., 2010). Many hemoglobin switching models were proposed as a result, which includes gene competition, gene silencing, and chromatin looping. The first factor that determines stage-specific expression was proposed to be the position of the genes in relation to the LCR (Tanimoto et al., 1999). Using RNA fluorescence in situ hybridization, Wijgerde et al. demonstrated that, in day 12 mouse fetal erythroid cells, the transgenic human γ - and β -globin genes were mostly transcribed alternately, rather than concurrently which supports the idea that LCR may activate each gene independently through a direct connection. The transcription machinery was most often linked to the β -globin gene, which is in line with β -globin being the main messenger RNA (Wijgerde et al., 1995). A further study supported the idea of chromosomal looping by confirming direct contact of LCR sequences with a particular globin gene promoter with using chromosome confirmation capture technique (Tolhuis et al., 2002). Accordingly, these numerous proposed models do not compete with one another but rather enhance one another.

Epidemiological and clinical reports have showed the positive outcome from elevated HbF levels in ameliorating beta thalassemia and sickle cell diseases clinical complications (Steinberg, 2020). As a result, a significant amount of work has been focused on gaining an understanding of the molecular processes that are responsible for the developmental switch from fetal to adult hemoglobin. With the current progress in understanding the underlying molecular mechanisms involved in hemoglobin switching, a significant attention is directed towards stimulating human fetal hemoglobin production and elucidating the regulation of the developing hemoglobin switches. This approach is motivated by the principle that it would reduce the severity of β -hemoglobin diseases not only by reducing the levels of precipitation of excess free alpha chains and rescuing RBC from the toxic effect of globin subunits imbalance and ROS generation, but also by increasing hemoglobin levels and RBC survival (Perrine, 2005) (Lim et al., 2015) (Mukherjee et al., 2022) (Carrocini et al., 2011). Researchers in this field are concluding that to treat patients effectively, a potent and therapeutic response from multiple agents acting at different molecular levels will be required.

Moreover, several research teams, utilizing both targeted and genome-wide association analyses, in the study of genetic variations that led to differences in HbF levels have found three chromosomal regions containing common variations that alter HbF levels as potential regulators for HbF expression that are not linked to the beta-globin locus (Lettre et al., 2008; Menzel et al., 2007; Thein et al., 2007).

These regions include B-cell lymphoma/leukemia 11A (*BCL11A*) region on chromosome 2, and the intergenic region between the *HBS1L* and *MYB* genes on chromosome 6 (Sankaran & Orkin, 2013).

There are two different temporarily overlapping lineages of erythroid cells in mammals as described above. The yolk sac is the site of genesis for the primitive erythroid lineage, which produces nucleated erythrocytes. These mature to completion (including enucleation) in the embryonic circulation. Eventually, the definitive erythroid lineage takes over, giving rise to the smaller enucleated erythrocytes that are responsible for the bulk of the body's oxygen transport during fetal and the postnatal period (Palis, 2014). Hematopoietic stem cells (HSC) are responsible for definitive erythropoiesis, but primitive erythropoiesis only takes place once from mesoderm cells (right after gastrulation) in the developing embryo. Lineage-committed progenitors give rise to both lineages of erythroid cells, each with distinguishable precursors that enucleate to create mature RBCs. Both lineages are defined by the sequential passage of cells through a series of three compartments. These compartments are referred to as progenitors BFU-E, CFU-E (EPO-dependent), and EryP-CFC, erythroblast precursors, and red blood cells (reticulocytes and mature RBC), respectively. Rodent models have been extensively studied for primitive erythropoiesis due to the inaccessibility of the early human embryo (Figure 6).

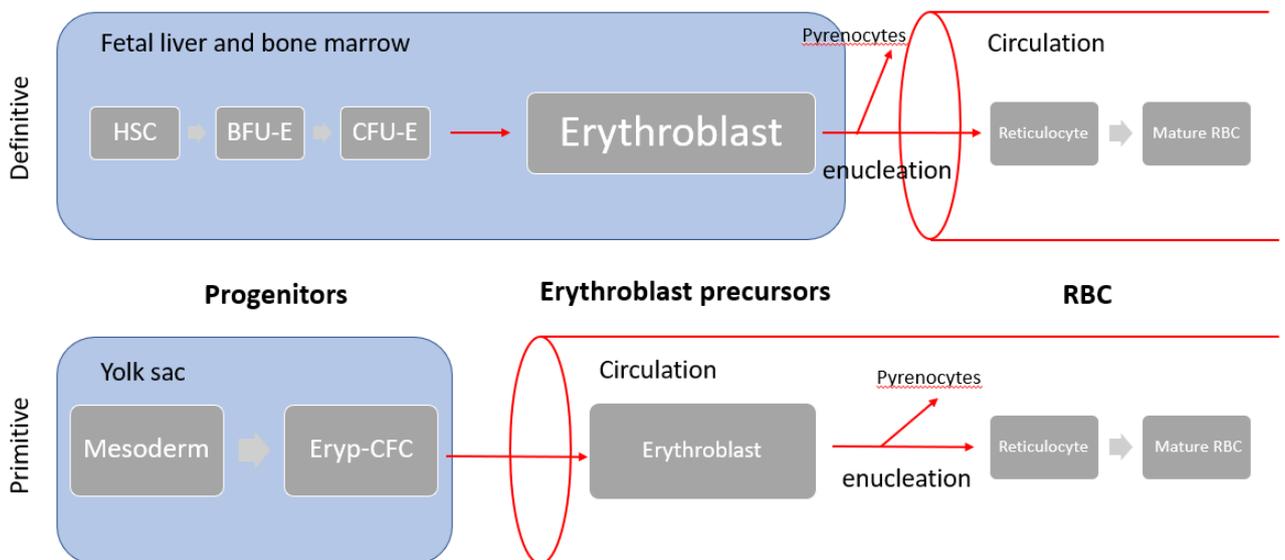


Figure 6: Definitive and primitive formation.

Primitive erythrocytes have distinctive cellular and molecular features that set them apart from mature erythrocytes. During the process of fetal development, as primitive cells are gradually replaced

by definitive erythroid cells, there is a changeover and a switch from the expression of embryonic globin genes to the definitive globin genes. Gata1 and Klf1 are both critical transcriptional regulators in both primitive and definitive erythropoiesis, although other regulators, including Myb, Sox6, and Bcl11A, are differentially expressed and play diverse roles in the two processes. Hematopoietic stem cells and their differentiated progeny continually form the definitive erythroid lineage. Variations in globin gene expression are one way that the definitive and primitive lineages may be told apart from one another. Primitive erythrocytes are the only cells that express primitive embryonic globin genes and have been found to undergo maturation and hemoglobin switch from Gower1 ($\zeta_2\epsilon_2$) to Gower2 ($\alpha_2\epsilon_2$) with dependency on EPO for their complete maturation (Qiu et al., 2008) (Malik et al., 2013). Within the primitive erythroid lineage, expression of erythroid globin genes has been found; however, this expression is at very low levels (Kingsley et al., 2006).

1.5.2 Control of beta and alpha globin gene clusters

Expression of the globin gene clusters is tightly regulated by a complex network of trans-acting proteins, such as transcription factors, and cis-acting regions that operate as promoters, enhancers, and silencers of gene activity. Through specific protein-protein interactions, transcription factors regulate gene transcription by recruiting RNA polymerase II to build DNA-protein complexes with sequences in regulatory regions of the globin gene cluster. Also, modifications to chromatin structure, which affect gene expression across whole globin gene clusters, are mediated by transcription factors (Chen & Pugh, 2021). Numerous transcription factor proteins, including GATA1, NFE2, and KLF1, have been found in the nucleus of erythroid cells. GATA factors are considered as 'master' transcription factors crucial for erythropoiesis (Gutiérrez et al., 2020). GATA1 knocked out mice develop embryonically fatal anemia (Migliaccio, 2020) (Katsumura & Bresnick, 2017).

Since GATA1 identifies and binds the DNA sequence motif (T/A) GATA (A/G), it was given that name. GATA1 is a DNA-binding protein that has two critical zinc fingers, the first is for DNA binding on the terminal C finger, and the other is the N finger for the physical binding of a zinc finger protein that acts as a cofactor named FOG1 (ZFPM1), friend of GATA1. The formation of GATA1/FOG1 heterodimer is required for GATA1 activity (Yu et al., 2002) (Fox et al., 1999) (Gutiérrez et al., 2020). The DNA sequence motif (T/C) GCT GA (C/G) TCA (T/C) is recognized by the protein NFE2 which functions as a transcriptional activator through cis-elements interaction with the locus control region of the beta

globin genes and belongs to the B-zip (basic leucine zipper) family (Lecine et al., 1998). Some researchers speculate that the erythroid Kruppel-like factor EKLF, also known as KLF1, is the most specific of all the erythroid transcription factors that have been found so far. It binds to the cis regulatory sequence CCACACCCT. The beta-globin gene promoter is a specific target of EKLF interaction (CAC box), suggesting that this protein may modulate the delta-beta switch. Lack of the *EKLF* gene is incompatible with life in mice due to ineffective fetal liver erythropoiesis (Nuez et al., 1995) (Perkins et al., 1995).

Another transcription factor KLF1 which has erythroid lineage specificity binds to the CACCC box and initiates beta globin gene transcription and may be involved in the last stages of human HbF silencing since it favorably binds to the gene promoter of the beta-globin over the γ globin. Mutations in the highly conserved CACCC motif of KLF1, located in the LCR region of beta globin cluster (HS1-HS3) can cause human beta-thalassemia by altering chromatin structure at this site (Miller & Bieker, 1993).

Mammalian β -globin loci were among the first cloned and sequenced gene loci, and as such have served as a valuable model system for the investigation of gene regulatory mechanisms (Fritsch et al., 1980). Since their cloning, researchers have learned a great deal about the mechanisms that regulate β -globin loci gene expression. The locus control region (LCR) was discovered to be a potent upstream enhancer of the β -globin loci, whose role is crucial for optimal gene expression.

To turn genes on and off in a stage-specific manner, intricate interactions between promoters and enhancers closely control the expression and synthesis of globin chains. The first 500 base pairs (5') of any globin gene include all of the proximal regulatory sequences necessary for effective transcription initiation. In addition, optimal expression requires participation from distal elements as well. Although all globin gene promoters exhibit striking similarity, each also reveals unique sequences that may be important for communicating the promoter's developmental stage specificity.

The TATA box, the CAAT box, and the CACCC box are three main regulatory elements that Myers and Maniatis found to be necessary for complete globin gene expression in their structural and genetic investigation of the beta globin promoter in mouse (Myers et al., 1986). All globin promoters have these components, but with slight sequence differences. On the other hand, the erythroid-specific DNase I hypersensitive sites (HS), which are the main distal regulatory element sequences that play an important role in both alpha and beta globin expression activity, are located 40 kb upstream of the cap

site of the zeta-globin mRNA of the alpha globin cluster (HS-40) (Chen et al., 1997), whereas for the beta-globin gene are located on a distant regulatory element region known as the locus control region LCR up to 50 kb upstream of the beta globin gene, and both show high sensitivity to nucleases in erythroid cells. In addition to DNA-protein interactions, HS-40 is also harboring active chromatin signatures such as H3K4me3 and H3K27me3 which promote chromatin structure modification and gene activation. H3K4me3, trimethylation of histone H3 at lysine 4, and H3K27me3, trimethylation of histone H3 at lysine 27, are a DNA packaging protein epigenetic alteration that commonly plays a role in gene expression control; it signifies tri-methylation at histone H3's fourth lysine residue and 27th lysine residue respectively which recognized as identifying the locations where active genes' transcription starts. Depending on the developmental stage and the globin needed, chromatin in specific locations remodel by unwinding into relaxed euchromatin (open) structure to activate transcription of which increase the accessibility of regulators and transcription factors (Bulger et al., 1999) (Kwaifa et al., 2020). In most cases, an HS will consist of a collection of DNA motifs for a set of transcription factors that interact with one another such as GATA1, KLF1, and NFE2 as main erythroid transcription factors (Lowrey et al., 1992) (Cao & Moi, 2002) (Jackson et al., 2003). The LCR forms a loop that allows them to interact with the promoters of certain globin genes, leading to the formation of a complex known as the active chromatin hub (Figure 7) (Kim & Dean, 2012) (Cavazzana et al., 2017) (Holwerda & De Laat, 2012) (Popay & Dixon, 2022).

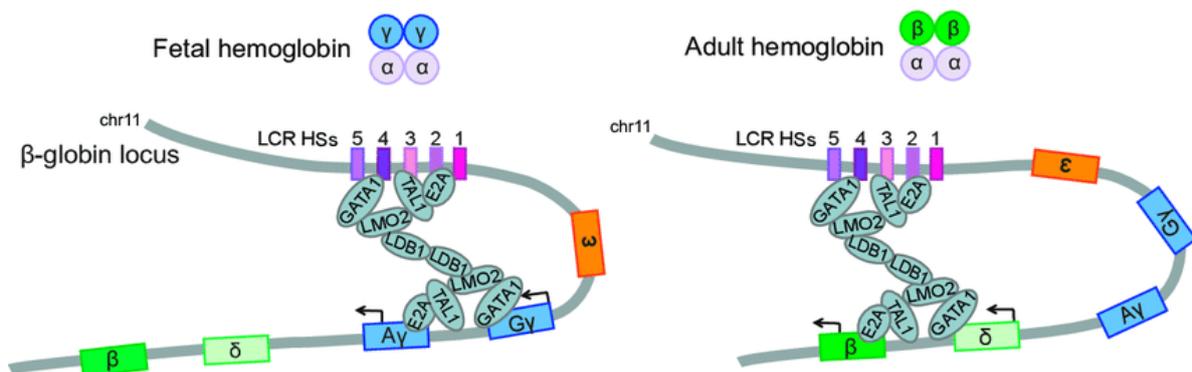


Figure 7: Chromatin looping in the activation of beta globin genes.

The LCR direct interaction with beta globin genes' promoters to regulate their expression with the involvement of several transcription factors, cofactors, and DNA motifs. Source: (Cavazzana et al., 2017). Permission under the terms of the Creative Commons Attribution-NonCommercial-No Derivatives License (CC BY NC ND).

Additionally, genome wide association studies of variations that involve HbF expression related traits have been very helpful in identifying transcription factors that are involved in fetal hemoglobin silencing (Uda et al., 2008). There are repressor transcriptional factors, such as BCL11A and LRF (also known as ZBTB7A), which independently decrease the expression of fetal hemoglobin in adult tissues by repressing the expression of γ -globin subunits (Figure 8) (Wang & Thein, 2018).

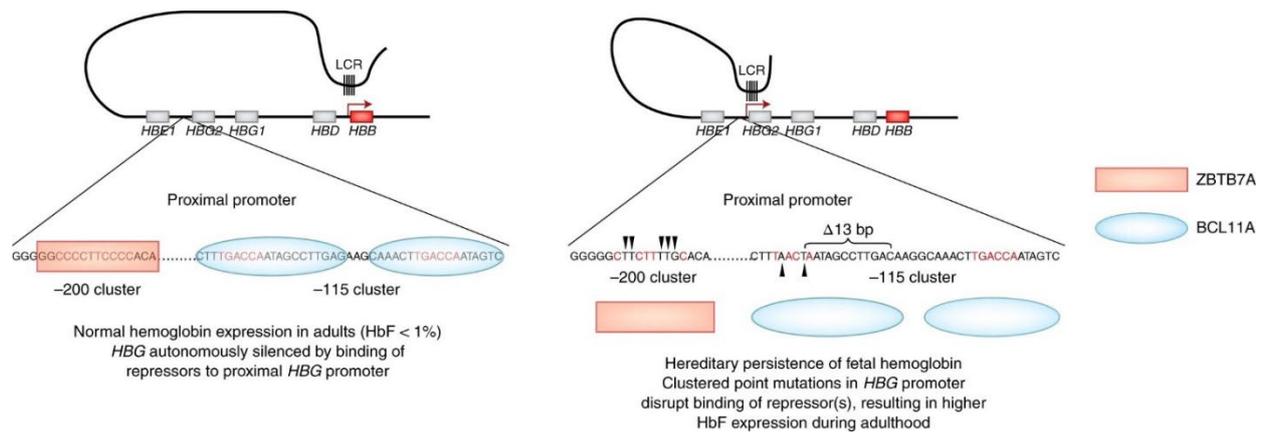


Figure 8: Hemoglobin switch from HbF to HbA.

HBG normal silencing by the binding of repressor transcription factors to proximal promoter of *HBG*. HbF induction can be achieved by disrupting the binding of these repressors. Source: reproduced with permission (Wang & Thein, 2018).

Recently, research using a functional assay for single-cell genome editing that allows for the independent and combined recapitulation of different mutations, have shed light on the ways in which numerous mutation-harboring functional elements (BCL11A and LRF) collectively contribute to HbF expression (Shen et al., 2021). BCL11A, which is a zinc finger transcription factor, has been found in shorter variant forms expressed in fetal cells with high levels of γ -globin expression. γ -globin expression is repressed by BCL11A mRNA expression (Zhou et al., 2010). In addition, erythroid cells in adulthood are the only ones where full-length versions of BCL11A are expressed at high levels (Sankaran et al., 2008). Additionally, BCL11A have been linked to HbF silencing by interacting physically with the chromatin at the beta-globin locus, as a part of a silencing complex that includes FOG1 and GATA1 transcription factors, and the nucleosome chromatin remodeling complex NuRD gene repressor (histone deacetylase 1 and 2) (Bradner et al., 2010) (Cao et al., 2004). Also, the transcription factor SOX6 (key players in directing cell fate and differentiation along certain lineages may be required for binding the proximal promoter of the γ -globin genes with the cooperation of BCL11A to assist in silencing these genes (Xu et al., 2010). Using gel filtration chromatography to look at BCL11A and SOX6

movement in nuclear extracts from human erythroid cells, Xu et al showed that BCL11A had significant peak movement. They also found that BCL11A robustly interacts with NuRD complexes in human erythroid cells, as evidenced by its elution pattern's high overlap with the Mi-2/NuRD components MBD3, HDAC1, and HDAC2. Most significantly, SOX6 also showed strong comigration with BCL11A and NuRD components. Investigators have also shown that the expression of BCL11A appears to be directly positively regulated by the erythroid-specific transcription factor KLF1. Correlation exists between mutations in KLF1 that contribute to a silencing impact of KLF1 on BCL11A and an increase in HbF levels (Gallienne et al., 2012). In a recent study, CRISPR-CAS 9 genetic screen was used to identify NFI-A and NFI-x as repressor transcription factors of HBG1 and HBG2 genes through their ability to stimulate BCL11A expression (Qin et al., 2022).

Mutations on any of these repressor transcription factors genes or their sequence targets that cause their inhibition led to substantial increase in the HbF (fetal hemoglobin) expression, which identifies it as a therapeutic target for patients with thalassemia or sickle cell disease to attenuate their phenotype. Masuda and colleagues created LRF Knock-out mice with a yeast artificial chromosome (YAC) transgene containing the human beta-globin gene cluster. LRF-deficient erythroblasts showed a considerable induction of human γ -globin transcripts, which accounted for 6–12% of total human beta-like globin in peripheral blood (Masuda et al., 2016).

The binding of FOP to the stable chromatin-interacting protein arginine methyltransferase 1 (PRMT1) may also play a part in hemoglobin switching process. Fop, friend of protein arginine methyltransferase I, is a PRMT1-methylated chromatin-associated protein that represses the gene expression of γ -globin. Increased expression of fetal globin genes was observed after knockdown of FOP. Although BCL11A levels were the same, a reduction in SOX6 protein levels was observed suggesting that FOP influence may be through SOX6 reduction (Figure 9) (van Dijk et al., 2010).

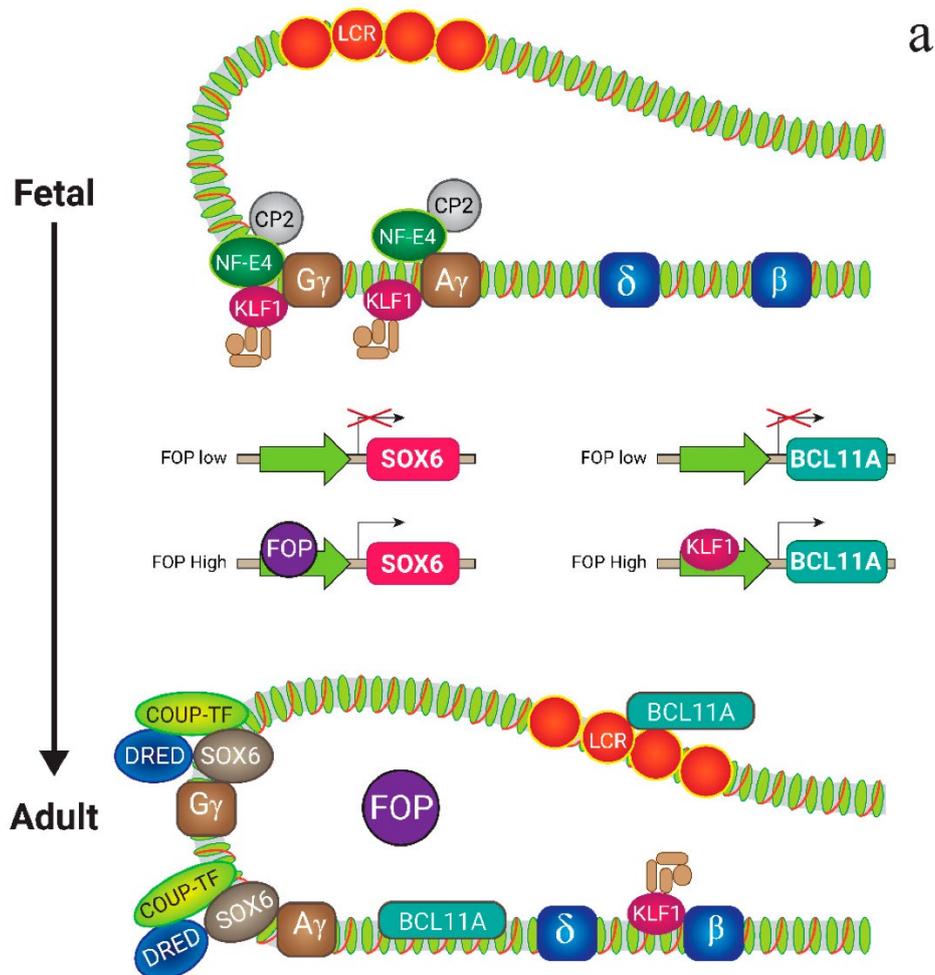


Figure 9: Function of fetal globin in γ and β -globin gene expression.

Expression of FOP is minimal in the developing fetus. Fetal hemoglobin is produced when transcription factors like NF-E4 bind to the gene's coding sequence, meaning that fetal globin repressors like BCL11A, SOX6, and KLF1 are inactive. On the other hand, adults have elevated levels of FOP expression, which activates fetal globin repressors, by binding to the gene's coding site, and cause erythroid progenitors to make β -globin. Source: (Rahimmanesh et al., 2022)

HBE is expressed during embryonic development. The progenitors that make up 6-week BFU-E are a novel lineage that has already transitioned from embryonic to fetal development (Peschle et al., 1984). The aneuploidy Trisomy 13 (Patau's syndrome) is the only known genetic disorder in which HBE is continuously expressed in newborns (with elevation in fetal hemoglobin) due to a delay in the normal globin switch from epsilon to gamma-globin synthesis, and from fetal to adult. Both in-vivo and in-vitro studies performed by Sankaran and colleagues have showed that the over expression of the candidates miR-15a and -16-1, two microRNA genes, located within the chromosomal band 13q14 play a role in embryonic and fetal elevation levels by acting via the transcription factor Myb (Sankaran et al., 2011)

The expression of *MYB*, which encodes c-Myb in erythroid progenitors, affects the amount of γ -globin generated by these cells. Several regulatory elements (three hypersensitive sites) with active chromatin located in the intergenic region between *HBS1L* and *MYB* on chromosome 6 play a significant role in this regulation (Tripathi et al., 2023). The researchers hypothesize that early differentiation is triggered because of suppression of cell division by a deficiency of Myb protein, which leads to production of red cells with greater amounts of HbF. More studies have shown that Studies of adults with partial trisomy of chromosome 13 have shown that *MYB* expression is linked with persistence of gamma and a delay in beta expression (Sankaran et al., 2011) (Wang et al., 2018). Homozygous mice for the c-Myb deletion (knockout of exon 6) died embryonically at day 15 due to severe anemia, which highlighted the significant role of c-Myb in definitive hematopoiesis (Lipsick, 2010).

The δ -globin gene is first actively expressed in erythroid progenitors. No δ -globin mRNA production is detectable by the reticulocyte stage. When compared to the β -globin gene promoter, the δ -globin promoter is substantially weaker (due of a silencer element that is unique to the globin promoter and is found upstream of the δ -globin gene) and the *HBD* gene produces far less mRNA with lower stability than the *HBB* gene which accounts for the significantly lower levels of hemoglobin A₂ ($\alpha_2\delta_2$) compared to those of hemoglobin A ($\alpha_2\beta_2$).

On the other hand, HbA₂ is just as effective as HbF at blocking the polymerization of sickle hemoglobin and ameliorating beta thalassemia symptoms. As a proof of principle, Manchinu et al have demonstrated that HbA₂ may be raised pharmacologically which may pave the way for related research in the future (Manchinu et al., 2020). Additionally, unlike fetal hemoglobin, which is expressed heterocellularly, HbA₂ is present in all erythrocytes. Induced levels of HbA₂ might have clinical significance (Steinberg & Rodgers, 2015). To assess the therapeutic feasibility of increasing the delta globin gene's expression in living organisms, Porcu et al have crossed a mouse model of sickle cell disease with transgenic mice that had a copy of the delta-globin gene that had been genetically engineered to be expressed at a greater level. This cross has resulted in the steady synthesis of HbA₂ which ameliorated the SCD phenotype. These findings show that delta-globin has therapeutic potential for the first time in vivo, which might lead to new ways of treating sickle cell disease (Porcu et al., 2021).

HS-40 (alpha multispecies conserved regulator α -MSC-R) of the alpha globin contains six DNA sequence motifs which can be recognized by erythroid enriched factors (Katsumura et al., 2013) (Thein,

2013). The erythroid enriched factor Sp1 can recognize the GT motif. GATA-1 factor can recognize GATA-1-b, GATA-1-c, and GATA-1-d motifs. Both NF-E2/AP1 motifs (5-NA and 3-NA) can be recognized by several factors that include NF-E2, the ubiquitous AP1, Nrf proteins, Bach proteins, and the homodimers of the small Maf family. Albitar and colleagues suggested that human zeta and alpha-globin gene expression are regulated during development in a gene-autonomous manner, meaning that they are not affected by its proximity to another alpha-like gene or its position within the alpha-globin gene cluster, and that the two genes include all the information required for correct developmental silencing, located on their immediate flanking regions (Albitar et al., 1991). Contrary to this finding, this regulation could not be set up using the promoters of the alpha or zeta globin genes. In contrast, the zeta-globin gene's silence was an ensemble effort including the promoter, transcribed region, and 3'-flanking regions (Liebhaber et al., 1996) (King et al., 2021).

In both the embryonic and postnatal/adult phases, HS-40 plays a role in the regulation of the alpha-globin gene cluster developmental pattern control by binding several trans-acting factors. HS40-E (embryonic) and HS40-A (adult) are proposed to be two distinct complexes of HS-40 with distinct configurations and roles and different nuclear binding factors at 3-NA motif (Zhang et al., 2002).

A wide variety of transcription factors, cofactors, and regulatory complexes work together to ensure that beta globin genes are expressed in the appropriate tissues and stages of development. Erythroid cells may have a similar set of transcription factors that regulate globin gene expression and guide erythroid development. There is a common set of ubiquitous TFs, like NF-Y, that binds all globin promoters, but with differing affinity, across embryonic/fetal and adult cells. The general transcription factors TATA and CCAAT boxes, which are situated in the promoter region, 30–100 bases upstream of all of the globin genes, are highly conserved motifs because of their functional role in regulating of both alpha and beta globin genes by cooperating with the developmentally specific regulatory factors (such as GATA1, KLF1, and NF-E2) and ensuring efficient transcription (Barbarani et al., 2021). In general, NF-Y is the major activating transcription factor that binds to globin CCAAT box motifs (Martyn et al., 2017). The TATA-like element and the initiator sequence, both of which have been demonstrated to interact with high affinity to TFII-D protein complex, make up the basal promoter of the human adult -globin gene (Figure 10) (Lewis & Orkin, 1995) (Duan et al., 2002) (Fan et al., 2015) (Zhu et al., 2012).

NF-Y(-115) **NF-Y(-88)**
GGGTTGGCCA GCCTTGCCTT GACCAATAGC CTTGACAAGG CAAACTTGAC CAATAGTCTT
CoupTFII, TR2/TR4
GATA(-73) **BCL11A(-56)** **TATA(-30)**
AGAGTATCCA GTGAGGCCAG GGGCCGCGG CTGGCTAGGG ATGAAGAATA AAAGGAAGCA
NF-E4/CP2

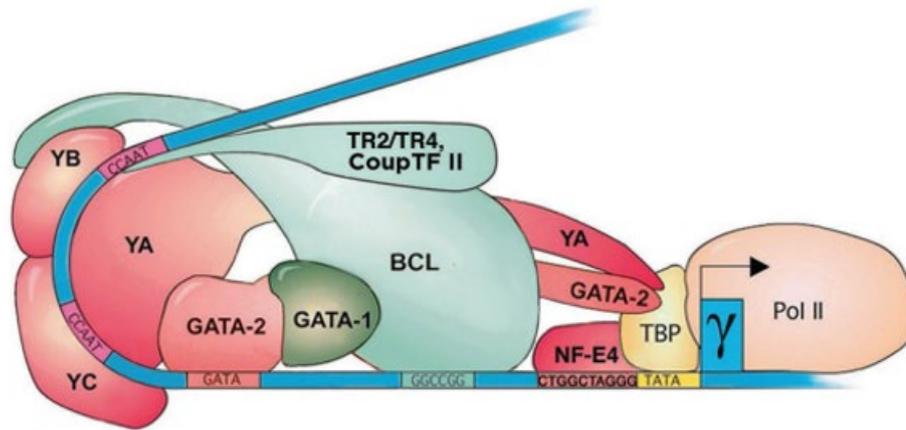


Figure 10: The molecular complex comprising the proximal γ -globin promoter and its essential transcription factors.

Red bars represent activating binding motifs while green bars represent repressor binding motifs. The black arrow represents transcription direction. In order to assemble the proximal promoter complex, NFY which consists of YA, YB, and YC binds to tandem CCAAT and twists the DNA by 70 degrees to create a pocket where other transcription factors may be recruited and interact. The transition from fetal to adult hemoglobin and suppression of gamma-globin gene expression in adult erythroblasts can be accomplished by BCL binding to SOX6, GATA1, FOG1, and the NuRD repressor complex. within the β -like globin gene cluster, the BCL11A protein occupies crucial sites at the cost of the HBG promoter, stimulating long-range physical connections between the promoter of HBB gene and the LCR. When BCL11A is not present, the LCR (upstream enhancers) is found in close proximity to the transcriptionally active HBG (Zhu et al., 2012) (Paikari & Sheehan, 2018).

Permission under the terms of the Creative Commons Attribution-NonCommercial-4.0 International.

1.5.3 Properties of the different forms of hemoglobin tetramers

Although different tetramers share similar architecture, the tetramer-dimer interface-strength and dissociation properties vary between the different types of hemoglobins (Manning et al., 2007b).

The heme-bound globin subunits subsequently pair up to form heterodimers, which are kept together by a dimer-monomer interface that, in the case of adult hemoglobin A, has an equilibrium that greatly favors the production of alpha-beta heterodimers ($\alpha\beta$). The combination of two identical heterodimers results in a single ($\alpha_2\beta_2$) tetramer with a flexible interface between the two dimers. This interface allows the tetramer to rearrange into one of two conformations, oxy (R) which binds oxygen, or deoxy (T) which releases oxygen.

Normal adult red cells abundantly produce $\alpha_2\beta_2$ (HbA), one of two human hemoglobins containing beta-subunits. On the other hand, $\zeta_2\beta_2$ (Hb Portland-2) is a very uncommon Hb that is only seen in alpha-thalassemias in which α -subunit production is absent (The normal partner of zeta being the beta-like protein, epsilon, in embryonic RBC). In order to provide an explanation for it, in 1984, Randhawa et al. suggested that α and β -subunits have a greater affinity for each other than ζ with β -subunits (i.e., a stronger monomer–monomer heterodimer interface); hence, $\zeta_2\beta_2$ is only observed when α -subunits are not present and the zeta gene continues to be expressed (Randhawa et al., 1984). In 2001, He and Russell group confirmed this notion and demonstrated that the affinities of the subunits of all embryonic, fetal, and adult hemoglobins span a wide range of values by comparing them (He & Russell, 2001). The work of several research groups have found that the Hill coefficients of $\zeta_2\beta_2$ (Hb Portland-2) and $\zeta_2\epsilon_2$ (Hb Gower-1), both of which have Ac-Ser at the N-terminus of their ζ -subunits, are lower (they give a weaker sigmoid oxygen saturation curve) than those of $\alpha_2\beta_2$ (HbA) and $\alpha_2\epsilon_2$ (Hb Gower-2), in which the alpha subunit replaces the zeta subunits (He & Russell, 2001) (Brittain, 2004) (Ashiuchi et al., 2005) (Manning et al., 2007b).

Embryonic, fetal, and adult hemoglobins have very different oxygen binding affinities that have evolved to allow each to do so optimally at a particular developmental stage. In comparison to adult hemoglobin, fetal and embryonic Hb bind oxygen more strongly which is helpful since the oxygen levels in their environment are significantly lower than in adult tissues. Different amino acid side chain changes have distinct effects on the electrostatics of each subunit structure. These differences might be a factor in the greater overall dissociation of subunit interfaces that include gamma subunits as opposed to beta subunits. The interface between dimers in the tetramer of HbF ($\alpha_2\gamma_2$) has an enhanced strength by 70-fold compared to HbA due to a combination of amino acid differences between gamma and alpha subunits sequences of the N terminal A helix. This A helix (in gamma) is not a part of the tetramer dimer interface suggesting that the long-distance strength enhancement of this helix comes from its deeper protrusion into the central cavity in the case of HbF (Manning et al., 1998) (Dumoulin et al., 1998) (Yagami et al., 2002). In erythrocytes, 2,3-bisphosphoglycerate (2,3-BPG), which is a highly anionic molecule (with 5 negative charges) formed by rearrangement of the glycolysis intermediate 1,3-bisphosphoglycerate, helps release oxygen from HbA and it is the main allosteric effector for hemoglobin. It binds reversibly to deoxyhemoglobin tetramer's central compartment (positively charged pocket) in a cleft between the beta chains dimer (lys82, his143, his2) and shifts its affinity to oxygen binding from high to low state by stabilizing its T tetrameric conformation and stimulates

oxygen dissociation and release at tissue sites. Fetal hemoglobin has higher oxygen affinity because it binds to 2,3-BPG less tightly than adult hemoglobin. Since the beta subunit in HbA contains more positive charges than the gamma subunit in HbF, the negatively charged 2,3-BPG (which promotes oxygen release) has higher preference towards HbA. Therefore, oxygen binds to HbF with higher affinity than HbA allowing HbF to carry oxygen to regions of lower oxygen tension than could HbA (MacDonald, 1977) (Kaufman et al., 2023).

Monomer to monomer interaction is stronger between alpha and beta subunits in HbA compared to fetal and embryonic hemoglobin. Manning and his group have showed that when alpha, gamma, and beta subunits are combined, HbA is generated but not HbF, suggesting that HbA is the more stable tetramer of the two with lower free energy (Manning et al., 2009).

Additionally, the formation of hybrid hemoglobin variations is caused by the extremely unusual process of globin gene fusion. In most cases, the misaligned globin genes cause uneven crossing over, which results in hemoglobins with structural abnormalities. When hybrid hemoglobins interact with other hemoglobin variations, the clinical presentation of carriers might range from healthy to thalassemic. Many hybrid hemoglobins made of β -like globins have been identified and studied (Kim et al., 2010). Two α -globin and two $\delta\beta$ -globin chains make up the hybrid hemoglobin variant known as hemoglobin Lepore. This fused globin chain is the product of a hybrid gene in which the delta and beta globin genes have not crossed over evenly due to a misalignment of homologous chromosomes during meiosis, which has resulted in a deletion between the two genes (Bhusal et al., 2022). Hb Lepore-Hollandia, Hb Lepore-Baltimore, and Hb Lepore-Washington-Boston have all been identified according to their deletion breakpoints: $\delta 22\text{Ala}/\beta 50\text{Thr}$, $\delta 68\text{Leu}/\beta 84\text{Thr}$, or $\delta 59\text{Lys}/\beta 86\text{Ala}$, and $\delta 87\text{Gln}/\psi\text{IVSII-8}$ or $\delta 87\text{Gln}/\beta 116\text{His}$, respectively. Heterozygote individuals may exhibit symptoms of thalassemia minor, although a wide range of clinical manifestations might result from interactions with other hemoglobinopathies (Chaibunruang et al., 2010). Another hybrid hemoglobin is $\text{G}\gamma\text{-}\beta$ Ulsan, identified by Kim et al, showed reduced oxygen affinity and originated from the deletion of 27,707 bp spanning from *HBG2* to *HBB* that resulted in the fusion of 13 N-terminal $\text{G}\gamma$ globin amino acids with 128 β globin amino acids (Kim et al., 2010).

1.6 RBC disorders

Red Blood Cells (RBC) disorders come in a wide variety of forms, including illnesses that impair the production rate, structure, longevity, and transport functions of RBCs. RBC production disorders include anemias, which are disorders of reduced RBC mass, and hyperemia, which is a disorder of excessive RBC mass (erythrocytosis or polycythemia). RBC dysfunction can present with a wide range of symptoms based on their nature, severity, and cell-related effects. However, certain symptoms of these conditions may overlap because they impact how RBCs operate (Iolascon et al., 2019). Red blood disorders can be categorized into hemoglobin disorders, enzymopathies, and cytoskeletal defects.

Anemias

When a person's RBC count is too low, or RBC does not work properly, anemia develops. Anemia may develop for many different reasons and comes in close to 100 various forms, each of which may be identified by the underlying cause as well as the size and hemoglobin concentration of the aberrant cells. Examples of anemia include iron-deficiency anemia which can be caused by a low iron intake from food, blood loss, improper iron absorption, and aplastic anemia that develops when insufficient new blood cells are not produced because of bone marrow damage (Rodrigo, 2019) (Auerbach & Adamson, 2016).

Cytoskeletal defects of red blood cells, also called membranopathies, are defects that alter the RBC's membranes' composition or permeability such as, spherocytosis and elliptocytosis (Da Costa et al., 2013). Spherocytosis, which is also a hereditary condition that gives erythrocytes less biconcave shape (rounder), which makes RBC more fragile. Hereditary spherocytosis is caused by mutations in the sequence of genes that are responsible for the production of proteins located on the membranes of erythrocytes and are responsible for maintaining the structure of cells (Zamora & Schaefer, 2022). These genes include ANK1 (codes for Ankyrin-1 protein), EPB42 (codes for protein 4.2), SLC4A1 (codes for band 3 protein), SPTA1 (codes for spectrin α -chain), SPTB genes (codes for spectrin β -chain protein), and EBP41 (codes for protein 4.1) (Ramasamy, 2020).

RBC enzymopathies are autosomal recessive genetic disorders that cause hereditary non-spherocytic hemolytic anemia, and are characterized by an insufficiency in enzymes involved in glycolysis and metabolism of erythrocytes leading to early destruction of erythrocytes (Sankaran & Weiss, 2015) (Ford, 2013) (Bogdanova et al., 2020) such as, glucose-6-phosphate dehydrogenase (Pimpakan et al., 2022) (Tavazzi et al., 2008), pyruvate kinase deficiency resulting in decreased ATP utilization (Bianchi

et al., 2020) (Koralkova et al., 2014), and triose phosphate isomerase deficiency which causes hemolytic anemia and affects young children (Myers et al., 2022).

Resistance to malaria

Malaria is caused by Plasmodium species, with *P. falciparum* being the most dangerous. The parasite infects mosquitoes that feed on humans and attacks their red blood cells which causes alterations in its shape, stiffens their membranes, increase its permeability to a wider range of ions, and increases its adherence properties, especially to endothelial surfaces. These alterations cause severe anemia that kills children and pregnant women more effectively. Mutations in genes affecting erythrocyte proteins may give partial resistance to highly pathogenic *Plasmodium falciparum* infection (Mohandas & An, 2012). The heterozygous advantage of these mutations outweighs the homozygous disadvantage in terms that the heterozygote's offspring survive better on average (Williams, 2006).

Heterozygotes with sickle cell anemia have a lower risk of contracting malaria, as they experience fewer incidences of malaria or less severe disease. The term "sickle cell trait" was coined in 1954 to describe the protective properties of sickle cell trait against malaria infection (Allison, 1954).

People with sickle cell trait have higher fetal hemoglobin concentrations in red cells, which has been linked to a reduction in malaria prevalence by hindering the development of the malaria parasite and conferring resistance. In 1998, Shear and her investigators group proposed a novel mechanism for the protection of malaria by increased expression of red blood cells containing HbF. HbF is resistant to digestion by the malaria parasite's protease, making it difficult to dissociate into monomer subunits. This suggests that the ability of the malaria parasite to develop into an adult in normal cells depends on the existence of protease-susceptible subunits generated by the dissociation of HbA in these cells.

As for people with sickle cell trait, because of induced phagocytosis activity, premature hemolysis, and activated immunity of the host, sickled erythrocytes harboring the plasmodium parasite are rather toxic to the parasite (Cholera et al., 2008).

Archer et al (Archer et al., 2018), proposed that the major mechanism for HbAS protective effect is the polymerization of HbAS under low oxygen conditions. Since the oxygen content differs between peripheral circulation and organs. By mapping the growth of the parasite in relation to oxygen concentration, they found that parasite growth in HbAS RBC was directly correlated with the amount of oxygen in the blood and the organs where they sequester.

In 1976, Pasvol et al have proposed that normal newborns and infants with beta-thalassemia have been shown to benefit from higher fetal hemoglobin concentrations in red cells, and this benefit has been linked to a reduction in malaria prevalence by hindering the development of the malaria parasite and confers resistance by influencing the parasite's growth rather than their actual invasion (Pasvol et al., 1976).

Hemoglobinopathies

Hemoglobinopathies are illnesses that affect the hemoglobin protein found inside of RBCs, and can be divided into two groups, defects that affect either the structure of the hemoglobin or its production rate (thalassemia). Hemoglobinopathies are the most common genetic defect in humans. There appears to have been natural selection for these deleterious mutations in regions where *Plasmodium falciparum*, the protozoan parasite that causes severe malaria is endemic (Taylor et al., 2012).

The majority of traits are inherited in a recessive manner, except for several reported cases with dominant variations. Heterozygotes or carriers of a single recessive globin gene pathogenic variations are asymptomatic or have minor clinical presentations. There are more than 1200 variants (intronic and exonic) of the Hb genes, with the majority being benign and non-pathogenic, that have been identified using various molecular genetics techniques. Around 400 variants are associated with beta thalassemia (Barrera-Reyes & Tejero, 2019). Variants are listed on the HbVar globin database (<http://globin.bx.psu.edu/hbvar/menu.html>).

Polycythemia

Also known as erythrocytosis, causes the body to produce more RBC than usual and is characterized by high levels of red blood cells. Polycythemia is caused by somatic, acquired mutations in the bone marrow. The excess blood cells may increase the viscosity of the blood the blood and interfere with blood flow, increasing the risk of serious health problems. Polycythemia can be primary, which is a form of chronic myeloid cancer and called polycythemia vera, or secondary which is caused by an underlying condition that triggers the kidney to produce high levels of erythropoietin that leads to increased production of RBC. The most common causes include tumors, obesity, and sleep apnea (Mallik et al., 2021).

1.7 Hemoglobinopathies

Many of the inherited hemoglobin diseases cause death in the first few years of life if untreated. It has only recently been clear how they affect disease burden, following an epidemiological shift brought on by advancements in nutrition, cleanliness, infection control, and genetic screening that have reduced childhood mortality. Now, infants with serious hemoglobin abnormalities can endure long enough to seek diagnosis and care.

To form normal functioning hemoglobin, alpha globin and beta globin molecules must be produced in balance and equal quantities. Failure of this balance leads to an abnormal buildup of uncombined subunits, leading to their premature denaturation (Smith, 1980) (Kaur Gharial, 2022).

The production of RBC and hemoglobin is affected in thalassemia. This often results in a person having fewer healthy RBCs. Thalassemias, are caused by mutations in regulatory genes or within alpha or beta globin genes that may lead to underproduction or no production of one of the two subunits of the globin molecules (Harteveld et al., 2022).

On the other hand, most recorded structural variations are caused by a single amino acid change in the alpha or beta chains. Depending on the nature of the mutation, they are usually harmless, but occasionally they can have devastating consequences that may change the stability or functional characteristics of the hemoglobin and result in a clinical illness. They have been identified by alphabetic letters or by the names of the locations where the condition was originally identified. Despite the fact that hundreds of structural hemoglobin variants have been found, only three, sickle cell HbS (Glu7Val), HbC (Glu7Lys), and HbE (Glu26Lys) are common because they appear to have been subject to positive heterozygous selection, presumably because they give resistance to severe malaria (Kohne, 2011).

Thalassemia is genetically and clinically heterogenous with a wide range of symptoms and severity between different affected individuals depending on the combination of alleles inherited from both parents, as shown in the diagram below (Figure 11). Alleles for mild beta-thalassemia result in the decreased production of beta-globin proteins accompanied with impaired function. On the other hand, severe alleles result in the production of either nonfunctional or nonexistent proteins (Harteveld et al., 2022).

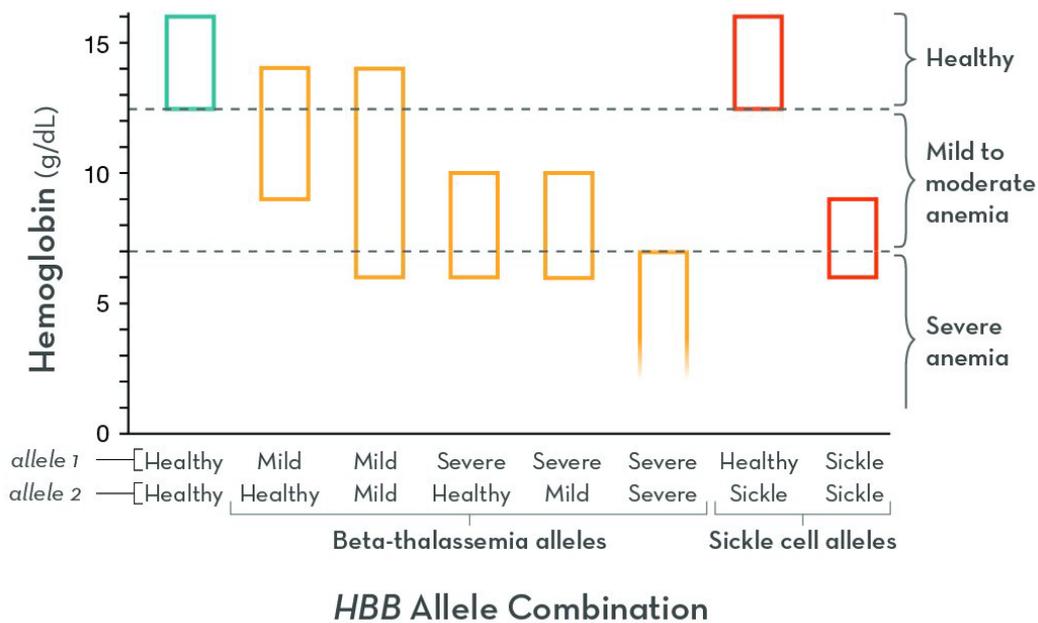


Figure 11: The effect of HBB allele combinations on blood hemoglobin concentrations.

Green triangle represents healthy individuals. Yellow triangles represent beta thalassemia alleles. Red triangles represent sickle cell alleles. The x axis shows different HBB allele combinations, while the y axis represents hemoglobin concentration in a blood sample. Source: Genetic Science Learning Center. "Hemoglobin Disorders." Learn.Genetics. June 10, 2019. Accessed with permission June 29, 2022. <https://learn.genetics.utah.edu/content/genetics/hemoglobin>.

Nevertheless, even family members who share the same HBB allele can experience diverse effects. While some people have ongoing medical needs, others never experience any symptoms. Changes in other genes, modifier genes, are suspected to be responsible for some of these discrepancies.

For instance, fetal globin-coding genes are normally repressed soon after birth. However, some people continue to produce fetal hemoglobin far into adulthood which, as a result, makes their symptoms less severe. In many populations, inheriting both the structural hemoglobin variation and the thalassemia gene is relatively frequent; these complicated interactions result in a very broad spectrum of clinical symptoms that collectively make up the thalassemia syndromes.

1.7.1 Sickle cell anemia

Sickle Cell Disease (SCD) is an autosomal recessive blood disorder caused by a point mutation in the HBB gene that leads to a structural abnormality in the beta chain of hemoglobin characterized by the presence of HbS (crescent shaped RBCs). It is a qualitative blood abnormality. When HbS RBC are under hypoxic stress, they collapse into crescent structures and aggregate with normal RBC by clumping

together into fibers leading to their precipitation. These sickling cells are rigid, with decreased flexibility and elasticity which cause capillaries blockage leading to trouble in transporting oxygen. Additionally, the cells become porous, fragile, and adhesive. As a result, the body's tissues do not receive adequate blood flow (Figure 12). Serious issues including episodes of pain, stroke, organ damage, frequent bacterial infections, and delayed growth may result from this (Bunn, 1997).

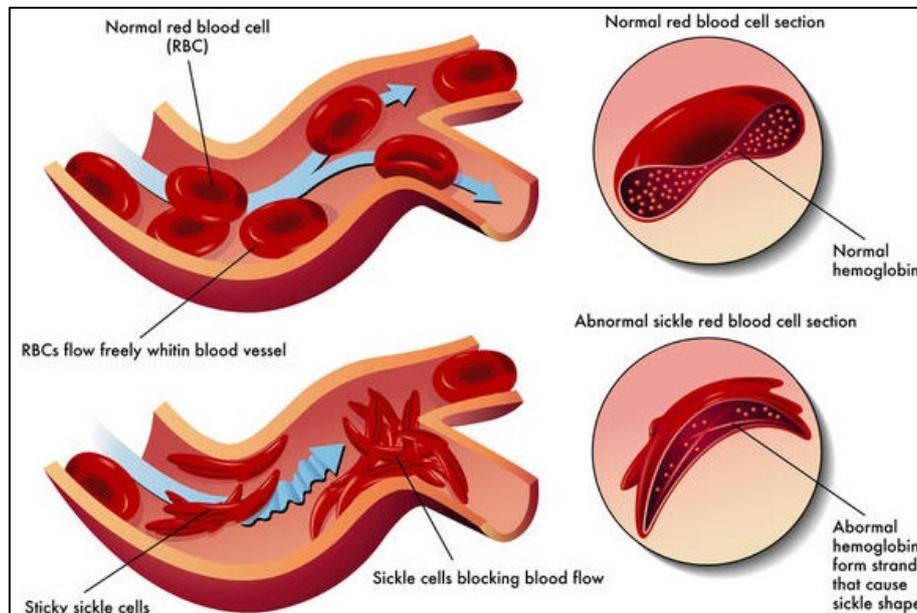


Figure 12: Normal blood cells Vs. Sickled cells flow through capillaries.

Normal biconcave shaped RBC flow normally and freely within capillaries while sickle shaped cells block the blood flow by clumping together forming sticky fibers.

Source: Reproduced with permission from (Montellano, 2018). Copyright (Amegbor) American Chemical Society.

A systemic literature review has showed that the countries of sub-Saharan Africa (500–2000/100,000) and South America and the Caribbean Islands (20–1000/100,000) had the highest birth prevalence of sickle cell disease (SCD) among children of one year old and younger, as determined primarily through newborn screening. The United States and European countries exhibited a birth prevalence of less than 500/100,000. In accordance with the findings regarding regional prevalence, the Middle East, sub-Saharan and North-East Africa, and India are "hotspots" for SCD on a global scale. Nevertheless, the availability of prevalence data was insufficient to generate consistent prevalence figures by region via meta-analysis (Colombatti et al., 2022).

In 1910, Herrick (Herrick, 1910), in the first medical study of sickle cell anemia (SCA) described the presence of sickle -shaped and crescent shaped red blood cells in a smear of blood taken from a

critically anemic dental student from Grenada. Several similar cases were observed during the following years (Sydenstricker, 1924) (Savitt, 2014). In 1927, By showing that saturating a cell solution with carbon dioxide may trigger shape alterations, Hahn and Gillespie hypothesized that hypoxia was responsible for red blood cell sickling (Hahn & Gillespie, 1927). This concept was proved experimentally in vivo in 1930 by Scriver and Waugh by employing a rubber band to cause venous stasis in a finger. They demonstrated that the percentage of sickle-shaped cells grew considerably, from around 15% to over 95%, due to stasis-induced hypoxia (Scriver & Waugh, 1930).

In 1945, Pauling took notice of this crucial research and proposed that the erythrocyte defect resided in the hemoglobin molecule itself because it precipitated when deoxygenated (Eaton, 2020). He validated his hypothesis by showing that sickle and normal hemoglobin migrate differently in a gel electrophoresis experiment, and explained how it occurs as a result of allelic change in a single gene that leads to the production of more positively charged hemoglobin than normal. He was the first to identify the molecular cause of a genetic defect in 1949 (Pauling et al., 1949). His discovery included additional conclusions stating that since aggregation can occur, the mutation location must be near the molecular surface. Moreover, electrophoretic patterns were shared by both parents and a 50/50 combination of hemoglobins A and S proved that the illness was inherited in an autosomal recessive manner, rather than autosomal dominant (Pauling et al., 1950) (Eaton, 2020).

In 1956, Ingram, employed by Perutz, specializing in protein chemistry discovered by peptide digestion, electrophoresis, and peptide chemistry that sickle cell disease is caused by single gene mutation that leads to single amino acid change from glutamic acid to valine in a tryptic peptide of the hemoglobin protein (Ingram, 1956) (Ingram, 1957).

In the late 70s and early 80s, sickled red blood cells were shown to obstruct blood flow by binding to the inner of blood vessels. Many health organizations and research groups took an active interest in SCD in the 1970s, which boosted public understanding of the disease, raised awareness, and provided vital research funds.

Although this disease is inherited in an autosomal recessive manner, the sickling is at least partially a deoxygenation artefact, such that HbAS heterozygote erythrocytes will sickle when completely deoxygenated. People with sickle cell trait have normal mortality and better quality of life. However, complications arising from the patient's trait status have been reported (Ashorobi et al., 2022) (Ashley-Koch et al., 2000; Neel, 1951; Taliaferro & Huck, 1923).

The change of normal hemoglobin to sickle cell hemoglobin occurs as a result of a single nucleotide substitution in the coding sequence of the beta globin chain. A single nucleotide change from GAG to GTG at codon 7 of the open reading frame of the beta globin gene (*HBB*) results in the substitution of the translated codon from glutamic to valine (p.E7V) (Ingram, 1956) (Bunn, 1997).

Glutamic acid is negatively charged hydrophilic amino acid located on the surface of the protein (originally shown by Perutz) while valine, in the HbS mutant, is an uncharged hydrophobic residue. Each hemoglobin beta molecule has the side chains of Phe86 and Leu89 exposed, to which the third hydrophobic side chain, Val7 is added in HbS. This mutation causes a dramatic change in the hemoglobin structure and severe malfunction under conditions of deoxygenation. Under deoxygenated conditions, Val7 interacts with Phe86 and Leu89, in an adjacent Hb tetramer which will stick together and polymerize causing the globin to aggregate, forming long chains of sickled filamentous precipitate that carries less oxygen than normal and cause the cell to deform into a long point sickle that are rigid and do not bend and have difficulty in passing through the smallest capillaries within tissues (Figure 13). This deoxygenation effect does not occur in normal HbA cells (Huang et al., 2003).

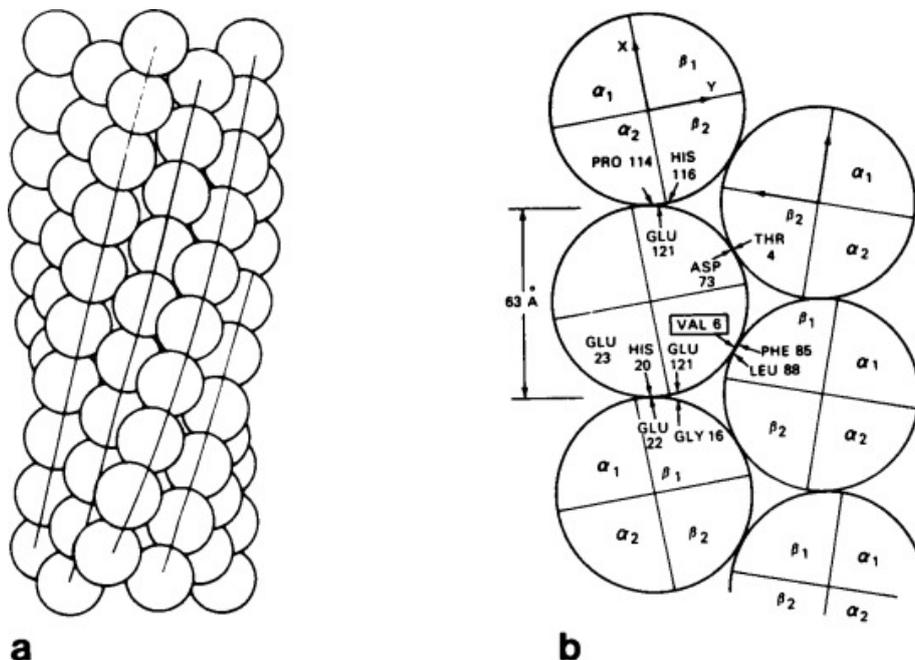


Figure 13: Sickle cell polymerization.

Electron microscopy view of Hemoglobin S molecules arrangement in deoxygenated state in a, and by x-ray diffraction in b. In a, it shows the formation of helical array by hemoglobin molecules while it shows in b the amino acid residues involved in their arrangement. Hydrophobic residues tend to clump together forming sickle shaped globin. Source: reproduced with permission (Noguchi & Schechter, 1981).

It has been shown that cells become sickled under low oxygen states and when the oxygen is restored, the cell can go back to its original shape. However repetition of sickling/unsickling cycle damages the cell membrane and makes it unable to return to normal (irreversibly sickled cells) and shortens its life which causes premature destruction of the cell (Shen et al., 1949) (Hazelwood, 2001). In 1967, Jensen and his group have noted that after four or five cycles of sickling and unsickling in single-cell preparations, cells lost their ability to return to their normal state (Jensen et al., 1967).

Besides blocking the blood vessels, the severity of the disease comes from the short life of the sickled cells. Sickled cells survive for only 10-20 days, in comparison to normal RBCs which last for 90-120 days, and cannot be replaced by new RBCs fast enough to compensate for this loss, leading to anemia (Kumar & Robbins, 2007). In addition, a major contributor to the disease pathophysiology is HbS cellular dehydration (as a consequence of potassium loss). The K-Cl cotransport (Mg dependent), the Gardos channel (calcium-dependent potassium efflux), and Psickle (polymerization-induced membrane permeability) are the three principal ion transport mechanisms implicated in sickle cell dehydration. PIEZO, a mechanosensitive cation channel is proposed as the protein mediating Psickle. This flux-mediated dehydration leads to elevated intracellular hemoglobin concentration, which in turn perpetuates polymerization, and eventually, irreversibly deforms erythrocytes (Vandorpe et al., 2022) (Brugnara, 2018).

Microvascular occlusion is also triggered by the interaction between sickled cells and activated neutrophils and their adherence to the endothelium of the vessels, which in turn causes hypoxia, followed by dilation of the vessels and eventually its damage. Another contributor to RBC endothelial cell adhesion is the reduced nitric oxide bioavailability in the endothelial surface. Hemolysis resulting from brittle membrane structure of defective RBC leads to the constant release of cell-free hemoglobin at the endothelial surface, which reduces nitric oxide bioavailability because Hb has a high irreversible affinity for NO and reacts very rapidly (McMahon et al., 2019). Moreover, monocytes and platelets are also activated in SCD, and in return activate vascular endothelial cells and aggregate with erythrocytes during P-selectin dependent vasoocclusion crisis (Belcher et al., 2000) (Wongtong et al., 2015) (Slimane et al., 2020).

The heterogeneity in SCA clinical presentation is influenced by both environmental and genetic variables. Monozygotic (identical) twin studies demonstrate that despite the fact that twins often share identical height and weight growth curves and similar laboratory measurements, there is a large

discrepancy in the occurrence of painful crises and other consequences, which must be because of non-genetic environmental factors (Weatherall et al., 2005) (Laurentino et al., 2019).

Those with sickle cell anemia can be classified according to their haplotype of the beta globin gene cluster, which is a cis-acting determinant that indicates their ethnicity or place of origin. Among the haplotypes are those corresponding to Senegal (SEN), Benin (BEN), Bantu or Central African Republic (CAR), Cameroon (CAM), and Arab-Indian (ARAB), and other unusual haplotypes were identified (Loggetto, 2013). The occurrence of HbF in adults is linked to haplotype and corresponds with the clinical manifestation of sickle cell anemia. The SEN and ARAB haplotypes are associated with reduced rates of organ damage and fewer clinical signs of sickle cell anemia correlating with the generation of the greatest quantities of HbF. Both the BEN and CAM haplotypes have intermediate clinical presentations because of their intermediate levels of HbF. On the other hand, the CAR haplotype generates the lowest levels of HbF, which leads to the worst clinical severity (Lobie et al., 1985) (Month et al., 1990).

When looking at the disease severity at the population level instead of the person level, the β -globin genotype is the primary factor. In general, SCD is most severely affected when inherited in a homozygous state (HbSS) or in a compound heterozygous state with beta globin variation that have been identified to cause thalassemia major (HbS β^0), whereas coinheritance of HbS with HbC (HbSC) or HbS with β^+ tend to be less severe (Quinn, 2016).

Other genetic factors can potentially impact the symptoms of SCD. As an illustration, the extent to which postnatal red blood cells retain fetal hemoglobin, a factor that mitigates sickle cell disease severity by disrupting the polymerization of sickle hemoglobin, is mostly dictated by genetics. Variants in the extended β -like globin locus induce the coinherited hereditary persistence of fetal hemoglobin, which leads to exceptionally high levels of HbF and the elimination of several symptoms of sickle cell disease. Variants in genes that control fetal hemoglobin were found frequently. In *BCL11A*, *HBS1L-MYB*, and *HBG2*, 19 single-nucleotide variations were shown to be substantially linked to fetal hemoglobin. Understanding the genetic moderators of SCD might have major therapeutic consequences; this finding led to gene therapy efforts that attempt to reduce erythroid BCL11A expression. Some of these treatments are showing early signs of success in clinical trials.

On the other hand, deletions in genes that lead to α -thalassemia were found to be strongly connected with an increased hemoglobin level, a decreased risk of albuminuria and stroke. Also, genetic

variations and gene pathway investigations related to specific complications, growth factors, cytokine receptors, oxidative stress, cellular adhesion, and transcriptional factors, among others have been explored as potential explanations for the observed clinical variation in SCA (Kirkham et al., 2023).

Many single nucleotide variations in several genes have been identified to be affecting sickle cell anemia and its association with cerebrovascular illness in large blood vessels such as *TGFBR3*, *GOLGB1*, *ENPP1*, *TNF-alpha*, *TEK*, *PON1*, *ANXA2*, *ADCY9*, and *APOE*. Additionally, genetic variations in *APOL1* have been shown to influence sickle cell nephropathy. Also, Gilbert's syndrome which is a benign condition caused by variations in *UDGT1A* gene have been shown to increase the risk of gall stone formation in patients with sickle cell disease (Rees et al., 2022).

1.7.2 Genetics of hereditary persistent of fetal hemoglobin (HPFH)

HPFH is a benign condition that occurs when fetal hemoglobin is persistently produced in significant amount during adulthood due to mutations in alpha or beta gene clusters, or mutations in the promoter region of γ -globin (Sharma et al., 2020). There are two kinds of HPFH that have been identified: deletional and non-deletional. Mutations in proximal promoter mutations in Ag- and Gg-globin genes (such as *GATA* site) have been identified in the non-deletional HPFH (Peterson et al., 2008)

1.7.3 Thalassemia

Thalassemia is an autosomal recessive disease characterized by reduced hemoglobin in the RBC. There are many forms of thalassemia ranging from mild to deadly (Marengo-Rowe, 2007). In 1925, a physician from Detroit who examined Italian children with severe anemia and other symptoms was the first to describe thalassemia. Other consequences of thalassemia, such as deformities of the bones and an enlarged spleen, might arise as a result of erythropoiesis acceleration caused by rapid loss of functional RBC.

Thalassemia is classified into two categories, beta thalassemia and alpha thalassemia, according to the mutated globin chain involved. The defect affects synthesis rate of the mutated globin leading to reduced or no production of that chain. At the molecular level, the thalassemias are genetically heterogeneous; more than 400 pathogenic distinct beta globin gene mutations have been associated with beta-thalassemia (Barrera-Reyes & Tejero, 2019), and the alpha thalassemias have more than 120 reported mutations (Kwaifa et al., 2020). Every affected community in the world has a small number

of regionally specific common mutations as well as different proportions of rare mutations (Clarke & Higgins, 2000).

1.7.3.1 Beta thalassemia

Beta thalassemia is a quantitative abnormality. It is characterized by decreased production of the beta globin chain (β^+ thalassemia) that may lead to microcytic hypochromic (small pale RBC) anemia or no production of beta globin chains (β^0 thalassemia). Beta globin proteins are modified due to point mutations in *HBB* gene (distinct from the sickle allele), it affects its expression and can lead to the production of impaired beta-globin protein. Alleles of this kind can result in a variety of genetic diseases, depending on how the beta-globin protein is changed.

Carriers of beta-thalassemia account for 80 to 90 million individuals, or around 1.5% of the world's population. 60,000 people a year are born with symptoms. Northern Europe and North America have historically had the lowest prevalence of β -thalassemia, whereas the Mediterranean area, the Middle East, and Southeast Asia have had the highest. Western Europe and North America are seeing an increase in the prevalence of β -thalassemia as a result of migratory trends (Kattamis et al., 2020). Between 2006 and 2018, the prevalence of thalassemia rose from 0.74 to 2.76 per 100,000. Specifically, between 2016 and 2018, the incidence rate nearly doubled, going from 0.22 to 0.41/100,000. On the contrary, the yearly transfusion rate fell from 34.7% in 2006 to 20.6% in 2018, showing a continuous decline (Lee et al., 2022).

Mutations of *HBB* that cause beta thalassemia can be intronic, exonic, 5' or 3' UTRs (affect RNA stability), or even in the promoter regions with the majority being single nucleotide substitutions or small insertion/deletions that lead to a frameshift. Mild forms of thalassemia can be caused by mutations that affect regulatory regions such as (TATA, CCAAT, and CACCC boxes). RNA splicing mutations would affect the maturation of pre-mRNA and associate with different range of severity (expected to reduce beta production because there will be less mRNA).

Mutations can result in beta thalassemia major if they occur in the intron-exon junction which are highly disruptive, while they can result in mild beta thalassemia if they happen to be in the adjacent splicing consensus sequences. Nonsense mutations that occur in exon 1 or exon 2 lead to a premature termination codon due to non-sense mediated decay, which result in a severe form of thalassemia due to the production of nonviable mRNA. On the other hand, mutations that occur in exon 3 would produce a more stable mRNA but result in a truncated protein, which would render it unstable (non-

functional with a dominant negative effect). In addition, initiation codon (ATG) variants that cause β^0 thalassemia have also been discovered. (Grosso et al., 2012) (Lee et al., 2021).

The severity of the disease is dependent on the alpha and non-alpha globin chain production imbalance. This imbalance results in ineffective red blood cell synthesis due to the damage created to the bone marrow precursors by excess free unpaired alpha chains inclusion bodies (α_4 tetramers) (Cao & Galanello, 2010). In the absence of normal beta-globin synthesis, alpha-globin will be made at levels that are higher than the capability of the AHSP that is readily accessible to bind it. Unbound alpha-globin molecules are highly unstable, and the deprivation of its binding partner, beta globin, leads to the accumulation of alpha subunits forming of insoluble aggregates in erythrocytes and precursors (Voon & Vadolas, 2008). On the alpha or beta globin chain of the hemoglobin molecule, the ferrous iron of the heme is connected to the proximal (F8) histidine, and the porphyrin ring is held in place by a phenylalanine. When functional hemoglobin is being oxygenated or deoxygenated, oxygen may move freely in the heme pocket between a second, distal (E7) histidine and the ferrous iron. Methemoglobin (metHb) is formed when ferrous iron spontaneously attaches to the distal histidine, where it can be oxidized by O₂ to ferric iron, rendering it useless as an oxygen carrier. About one percent of all circulating hemoglobin undergoes this reaction daily, but there is a cascade of protective systems in place to revert the methemoglobin back to hemoglobin (Voon & Vadolas, 2008). Hemoglobin function, stabilization of the chains, and appropriate folding of its polypeptides depend on the correct binding and coordination of the heme (Komar et al., 1997). Oxidative stress in the erythrocyte is largely triggered by hemoglobin. Excess of free alpha globin molecules in circulation in patients with beta thalassemia, go through auto-oxidation that changes ferrous iron to ferric iron and generates metHb and superoxide radicals, which can cause an auto-catalytic oxidation within the cell. Ferric iron binds the globin loosely which subsequently leads to heme degradation. Iron released from the degraded heme catalyzes reactions of free radicals (If superoxide dismutase converts two superoxide to hydrogen peroxide then free ferric iron can convert that to hydroxyl radicals (HO•) which are extremely damaging) (Nagababu et al., 2008) (Voskou et al., 2015) (Bou-Fakhredin et al., 2022). The imbalance of redox can cause oxidative stress and high-level production of reactive oxygen species (ROS). Increased generation of ROS in RBCs has been associated to caspase-3 activation. Caspase-3, which cleaves GATA-1 and proteins involved in cytoskeletal and DNA integrity, is usually synthesized during erythropoiesis. The interaction of band-3 with cytosolic proteins and the cytoskeleton is altered as a result of ROS-activated caspase-3. Phosphatidylserine (PS) exposure can occur as a result of

membrane impairment. All these are signs of RBC eryptosis. To maintain the redox balance, ROS is used by cytosol antioxidants to produce hydrogen peroxide, which can react with heme and trigger its degradation and lead to the stimulation of apoptotic pathways in the bone marrow early in the development of the erythrocytes.

Multiple investigations have indicated that beta-thalassemia patients' bone marrow contains an abnormally high number of activated macrophages, which is suggestive of accelerated senescence and eryptosis of erythroid precursors. Erythroid progenitor cells may experience premature death due to the inability to overcome the overproduction of ROS, which results in ineffective erythropoiesis (Voskou et al., 2015) (Gwozdziński et al., 2021).

Wickramasinghe and Bush examined bone marrow samples from individuals with homozygous beta-thalassemia using electron microscope. They found that cells during different stages of erythroblast maturation contain precipitates of alpha globin subunits. They suggested that an overabundance of free alpha-chains, rather than alpha-chain precipitates, may be responsible for the G1 phase arrest of many of these cells (ineffective erythropoiesis) (Wickramasinghe & Bush, 1975).

Mathias et al studied the timing of erythroid lineage commitment in beta-thalassemia major under erythroid culture conditions using purified bone marrow CD34+ progenitor cells. They concluded that accelerated erythroid development, increased clonogenicity of BM progenitor cells, and enhanced expression of proteins characteristic of the late erythroid lineage are all hallmarks of thalassemia major. However, apoptosis during the reticulocytes stage leads to inefficient erythropoiesis despite the apparent enhanced erythroid commitment (Mathias et al., 2000).

The cell surface contacts of FAS with FAS-ligand, the death receptor pathway, appear to drive this response in thalassemic erythroid precursor cells in vitro. In beta thalassemia major with frequent blood transfusions, the erythrocytes circulating in the bloodstream bind more IgG than usual. Normal IgG binding to senescent cell antigens signals macrophage clearance; however, in beta-thalassemia, the accumulation of oxidative stress, followed by heme breakdown and membrane damage, appears to speed up this process (Schrier et al., 2003).

Moreover, because beta chain production is reduced or absent in beta-thalassemia patients, the free alpha subunits are available to bind with both delta globin chains and cause a slight increase in HbA₂, and to gamma globin chains and cause an increase in HbF.

Thalassemia major (Cooley's anemia)

Individuals with thalassemia major inherit two mutated *HBB* alleles (homozygous or compound heterozygous) leading to insufficient beta chain production (deficiency in HbA). The predominant hemoglobin at birth, HbF ($\alpha_2\gamma_2$), has no beta chains, protects the baby at birth. After a few months, when the predominant hemoglobin transitions from HbF to HbA (about six months), symptoms start to appear which include severe anemia and bone marrow expansion (Muncie & Campbell, 2009).

Rather than whole blood transfusions, affected individuals need cleansed red cell transfusions, and lifelong treatment with iron-chelating agents. The degree of anemia on several measures, the extent of inefficient erythropoiesis, and clinical factors like failure to thrive or bone abnormalities should all factor into the decision of initiating lifelong therapy of transfusion (Cappellini et al., 2008). Iron chelators are used to treat patients and lessen the harmful effects of iron excess. Iron chelators are small molecules that can cross cell membranes and remove free iron from the body in urine or bile by forming a complex with it. Iron chelation therapy is widely acknowledged for the treatment of severe chronic anemia in young patients, but its efficacy in older patients is questioned because of myelodysplastic syndromes. Iron overload can be treated using a number of different chelators that can be used in single or combination forms. The DFP, DFX, and DFO are all significant iron chelators, and each have their own set of benefits and drawbacks (Entezari et al., 2022) (Di Maggio & Maggio, 2017) (Mobarra et al., 2016).

With regular transfusions, generally at monthly intervals, the child's growth will progress normally until about the age of 10-11. By that age, the child would have accumulated iron overload which without proper iron chelation therapy may result in many severe complications including delayed growth and later on myocardial disease and heart, liver, endocrine gland and kidney problems (Muncie & Campbell, 2009). Iron accumulation comes from turnover of transfused RBC without adequate endogenous RBC production and recycling and the iron released from RBC early destruction (Ballas et al., 2018).

Thalassemia intermedia

In thalassemia intermedia, the disease severity ranges from mild to severe. Patients with thalassemia intermedia experience similar clinical manifestation as thalassemia major but with lessened severity. However, as for transfusion dependency, most individuals are transfusion independent, or transfusion is only occasionally required (Musallam et al., 2012) (Asadov et al., 2018) (Ben Salah et al., 2017).

1.7.3.2 Genetics of beta thalassemia

The genotype in thalassemia disorders does not always predict the phenotype as shown in Table 2. It is worth noting that different laboratories use slightly different reference ranges.

Table 2: Beta thalassemia genotype/ phenotype correlation with its hemoglobins content. (Colaco et al., 2022) (Baird et al., 2022) (Brancaleoni et al., 2016a) (Zakaria et al., 2021).

Thalassemia	Allele	Hemoglobins
Asymptomatic	Normal <i>HBB</i> + β^+/β	HbA 98%, HbA2 2-1%, HbF >1% HbA 94-98%, HbA2 4-2%, HbF >1% or <1%
Minor (trait)	β^+/β β^0/β	HbA 94-97%, HbA2 4-2%, HbF >1% or <1%
Intermedia	Mostly two thalassemia alleles β^+/β^+	Ranges in severity
Major	Either β^+/β^0 , or β^0/β^0	HbA 10-30%, HbF 70-90%, HbA2 2-5% HbA is absent, HbF 95-98%, HbA2 2-5%

The main factor outside of the beta globin cluster determining disease severity is the genetic variations that affect the production of fetal hemoglobin (*BCL11A*, *HBS1L-MYB*, γ -globin genes (*Xmnl G- γ*), and *KLF1*). The clinical presentation and therapy of beta thalassemia are influenced by the HbF modulation. Additionally, coinheritance of HPFH or alpha thalassemia have been shown to ameliorate beta thalassemia symptoms. Other factors may influence the severity of the disease by modifying the severity of Jaundice (Gilbert's syndrome), cardiovascular complications (*APOE* and some *HLA* haplotypes), bone density (*COL1A1*, *COL1A2*, *TGFB1*, and *VDR*), and iron overload (*HFE* and *GSTM1*) (Panigrahi & Agarwal, 2008) (Danjou et al., 2012) (Jaing et al., 2021) (Tripathi et al., 2023).

Genetic factors located outside the beta globin gene cluster may result in a beta-thalassemia-like phenotype on rare occasions. Viprakasit et al, provide evidence that the lower expression of the beta-globin genes in people with trichothiodystrophy (TTD) is caused by specific mutations in XPD helicase component of the transcription factor TFIIH. They have studied eleven TTD patients and identified mutations in the XPD gene that exhibit beta-thalassaemia characteristic features, including decreased beta-globin production and mRNA levels (Viprakasit et al., 2001). Also, the production of thrombocytopenia associated with thalassemia trait produced by the GATA-1 X-linked transcription factor mutation (Cao & Galanello, 2010). In another study, Channing and his colleagues demonstrated that a missense mutation in the GATA-1 N finger is responsible for X-linked thrombocytopenia with thalassemia in humans. Specifically, the mutation in question is Arg216Gln. In addition, this is the first documented case of beta-thalassemia in a human being that was caused by a mutation in an erythroid

transcription factor. Authors showed that although the interaction between Arg216Gln GATA-1 and FOG-1 is quite similar to that of wild-type GATA-1 and able to facilitate the maturation of GATA-1 erythroid cells into mature erythroid cells, nevertheless, it does so with less efficacy (Yu et al., 2002).

To find updates about hemoglobin variations, a collaborative scholarly effort has resulted in HbVar (<http://globin.cse.psu.edu/globin/hbvar/>), a relational database that contains current and high-quality data on the genomic sequence alterations that cause hemoglobin variations and all forms of thalassemia and hemoglobinopathies. Each variation is accompanied by a wealth of data, such as its sequence change, biochemical and hematological consequences, related pathologies, frequency in different populations, and citations. The database has two major features as well. The mutation frequencies for many different types of beta-thalassemia in high-risk groups have been compiled from the scientific literature and made available for user enquiry, and users may now integrate data on hemoglobin variations and thalassemia mutations with a wide range of genomic data, since HbVar has been connected to the GALA (Genome Alignment and Annotation database, available at <http://globin.cse.psu.edu/gala/>) database.

1.7.4 Alpha thalassemia

Beginning in early embryonic development, significant levels of expression are maintained in the alpha-globin genes. Therefore, the consequences of abnormalities in alpha-globin chains are felt, beginning in early development and continuing throughout adulthood, in contrast with beta-globin gene mutations, which do not show symptoms until around six months after birth. Almost all alpha thalassemia mutations are caused by deletions (around 90%), which can range from single base to the absence of a whole gene, and the rest are caused by point mutations. Alpha thalassemia is inherited in an autosomal recessive way. Two genes (four alleles), *HBA1* and *HBA2*, on the alpha globin cluster encode alpha globin chains. Gene expression for α_2 -globin is greater than that for α_1 -globin under normal conditions. A discrepancy in transcription levels between the two duplicated -globin genes was first shown by the research groups of Liebhaber and Molchanova in the 1980s and 1990s, respectively (Liebhaber et al., 1986) (Molchanova et al., 1994). At steady state, α_2 mRNA outnumbers α_1 mRNA in both fetuses and adults, with neither form changing its translation profile. Based on this, the *HBA2* locus was expected to play the leading role (Shakin & Liebhaber, 1986). As heterozygotes for naturally occurring structural mutations in either the *HBA2* or *HBA1* genes were studied, they showed a slightly higher average of stable variants from the *HBA2* mutations compared to the *HBA1* mutations indicating that the translation of *HBA2* is less efficient with approximately 60:40 contribution at the protein level

from both genes (Farashi & Hartevelde, 2018). Alpha thalassemia is characterized according to the number of alpha globin alleles affected by the mutation which may lead to reduced production of alpha chains α^+ , or to complete absence of alpha globin chains α^- . Alpha thalassemia is categorized into four syndromes as demonstrated in Table 3: alpha thalassemia trait, silent carrier, HbH and HbBart's (Marengo-Rowe, 2007).

Table 3: Alpha thalassemia genotype/ phenotype correlation.

Alpha thalassemia type	Number of alpha globin alleles lost	Alpha genotype
HbBart's (γ_4) hydrops fetalis	Four	-/- -/-
HbH disease (β_4)	Three	-/- -/ α
Alpha thalassemia trait	Two	-/ - α/α (cis form) or -/ α -/ α (trans form)
Silent carrier	One	-/ α α/α

Because alpha chains are also involved in making normal HbA₂ ($\alpha_2\delta_2$), a mutation to the alpha gene may decrease the production of HbA₂ as well (W. Chen et al., 2010; Karamzade et al., 2014; Origa & Moi, 1993; Wang et al., 2003; Xu et al., 2004).

Silent carriers with three normal alpha globin genes have not been observed to develop health complications. Individuals with alpha thalassemia trait (2 normal alpha globin genes) may develop mild anemia, with small and pale (microcytic, hypochromic) RBC. In addition, free beta and gamma chains make homotetramers. Hemoglobin H is produced with four beta chains (β_4) while four gamma chains (γ_4) make hemoglobin Bart. Individuals with slight elevation in the concentration of (β_4) and (γ_4) are considered as markers for alpha thalassemia trait. On the other hand, the loss of three normal alpha globin genes leads to the high production of hemoglobin H (β_4). When an individual has only inherited a single copy of a normal alpha gene, the generation of alpha globin is considerably hindered. This will result in an excess of gamma chains during the neonatal period, and an excess of beta-globin chains in adults. Individuals who carry only one viable alpha allele will show excess beta-family subunits as γ_4 at birth, transitioning to β_4 as adults. Both HbH (β_4) (γ_4) and has extremely high oxygen affinity and resistant to oxygen depletion even under physiological stress, therefore, is considered as an ineffective oxygen transporter with an atypical oxyhemoglobin dissociation curve, which means it can carry oxygen but cannot release it and deliver it properly to tissues (Joshi et al., 2004). HbH (β_4) is highly unstable and can cause a variety of clinical symptoms that is variable from patient to patient because

it is more prone to oxidant damage with shortened lifespan and increased hemolysis rate and can precipitate within the cell. Individuals with HbH (β_4) disease suffer moderate to severe symptoms which may lead to serious complications that may include microcytic hypochromic anemia, enlarged spleen, slower growth rate, infections, bone deformities, and jaundice (Vichinsky, 2013) (Origa & Moi, 1993) (Harewood & Azevedo, 2023).

The loss of all four alpha globin alleles and the inability to produce any chains of alpha globin is the most severe form of alpha thalassemia, with hemoglobin Bart's (γ_4) as the only available hemoglobin (some have (β_4) or Hb Portland $\zeta_2 \gamma_2$). The clinical manifestations of Hb Bart's (γ_4), also called hydrops fetalis (abnormal accumulation of interstitial fluid) in this case, begin around week 20 to week 26 of pregnancy, when the mother has developed hypertension due to polyhydramnios (abnormal excess of amniotic fluid). Fetal ultrasound reveals hydrops. Due to their strong oxygen affinity, hemoglobin Bart's make tissues hypoxic and leave them open to degradation, leading to anemia. To provide adequate tissue perfusion in the face of fetal anemia, the fetal body undergoes a hemodynamically compensatory process characterized by an increase in intravascular volume. Due to their abnormal form, fetal erythrocytes are eliminated (self-hemolysis). Therefore, fetal iron overload may ensue, with iron accumulation in the cardiomyocytes and other tissues of internal organs. Because of this, cardiomyopathy may develop (Jatavan et al., 2018). Fetuses with α^0 thalassemia die in utero or shortly after birth for lacking the ability to produce any functional hemoglobin (A, F, or A2), unless intrauterine blood transfusion is administered and early delivery is induced (with limited success) (Vichinsky, 2009) (Joshi et al., 2004).

The majority of mutations that cause alpha thalassemia are deletional mutations. Nonetheless, non-deletion mutations causing alpha thalassemia are reported worldwide. These point mutations or insertions interfere with mRNA processing or translation, and the stability of alpha globin results in more severe forms of alpha thalassemia than deletion mutations in most patients. Examples of non-deletional mutations of the termination codon of *HBA2* include Hb^{Constant Spring} (Hb^{CS}, c.427T>C, p.*143Gln). Other common non-deletion mutations are Hb^{EVORA} (c.106T>C, p.Ser36Pro), Hb^{Koya Dora} (c.428A>C, p.*143>Ser), and Hb^{Adana} (c.179G>A, p.Gly60Asp), among others (Farashi & Hartevelde, 2018). The HbVar website provides an updated overview of alpha thalassemia mutations and their phenotypes.

1.8 Clinical features of sickle cell anemia and thalassemia

Sickle cell disease

The manifestations of sickle cell disease may differ among individuals and are subject to temporal progression. The course of the disease's progression will dictate the manifestation of any symptoms that may be experienced. The typical age range for clinical presentation of sickle cell disease is 6–24 months. Symptoms of sickle cell disease include anemia, fatigue, pale skin, jaundice, dark urine, dactylitis, enlargement of organs such as spleen (splenomegaly) and liver (hepatomegaly), weakness, headache, pain crisis, acute chest pain, problems breathing, lack of energy, vision problems, increased risk of infection, delayed growth or puberty, swelling of hands and feet, irregular heartbeat, and stroke. Splenomegaly may result in an extended abdomen and can lead to several related problems such as hypersplenism, an overactive spleen, due to the accumulation of excessive number of RBCs within the spleen to be destroyed within the spleen. Several other complications may develop if the condition is left untreated. To compensate for early destruction of erythrocytes, the bone marrow produces more RBCs which may lead to bone marrow expansion that causes bone deformity with fracture risk (Cario, 2018) (Piel et al., 2017) (Serjeant, 2013b).

Beta thalassemia

Beta thalassemia major and intermedia clinical severity is widely heterogeneous between affected individuals depending on the specific pathogenic variants inherited from their parents and the complexity of gene interaction and may be influenced by modifier genes. The typical age range for clinical presentation of beta-thalassemia major is 6–24 months. Symptoms of beta thalassemia major include anemia, fatigue, weakness, delayed growth or puberty, facial bone deformities, jaundice and dark urine, abdominal swelling, poor appetite, risk of infections, recurrent fever, diarrhea, enlarged spleen and/or liver, bone abnormalities, and heart palpitations (Langer, 1993) (Needs et al., 2023). These symptoms tend to be milder in patients with beta-thalassemia intermedia, who often exhibit symptoms at a later age. children with moderate cases may be symptomatic with just mild anemia, however children with more severe cases may exhibit delayed development and stunted growth (Taher et al., 2013).

1.9 Diagnosis of sickle cell disease and beta thalassemia

A good diagnosis would rely on collective clinical assessment by taking family history and recognizing and connecting the various symptoms.

Sickle cell disease

Hemoglobin electrophoresis is performed by passing an electrical current through it. Because of this, distinct bands of hemoglobin are formed depending on the electrical charge carried by the globin protein and the amount of hemoglobin present. These different hemoglobins are distinguishable based on their migration through the gel. The test also gives the percentage of each hemoglobin type present in the sample. In addition, capillary electrophoresis or HPLC can also be performed.

Blood smear examination which may exhibit crescent shape red blood cells. Also, solubility of hemoglobin test can be performed as well.

Different types of sickle cell disease can be identified by genetic testing. A person's sickle hemoglobin gene copy number can also be determined by genetic testing.

Beta thalassemia

The shape, size, and amount of RBC are typically measured during the initial thalassemia screening by performing standard complete blood count (CBC) which includes a variety of red blood cell indices such as RBC count, mean corpuscular volume (MCV) that measures the size of RBC, hemoglobin (Hb) content, hematocrit (Hct) that measures proportions of RBC in the blood which gives an indication of the carrying capacity of oxygen in the blood, mean corpuscular hemoglobin concentration (MCHC) that measures the hemoglobin concentration in the cell, and red cell distribution width (RDW) that gives the standard deviation of MCV (RBC size variation) (Corrons et al., 2021a). The level of iron in the blood, which is frequently higher in people with beta thalassemia, can also be examined (Musallam et al., 2013). Evidence points to a close relationship between iron overload and tissue hypoxia generated by anemia. Moreover, erythropoietin production is stimulated in response to hypoxia and anemia. (Mariani et al., 2009).

Biochemical tests such as high-performance liquid chromatography (HPLC), or carrier screening by hemoglobin electrophoresis are well established, reliable, and affordable methods to confirm the diagnosis of thalassemia by detecting the elevated proportions of HbA2 and HbF. Genetic sequencing

and DNA analysis using various available techniques is often used to confirm the diagnosis by identifying the exact variation a person has (Munkongdee et al., 2020a) (Lee et al., 2021).

Patients with red blood cell disorders often have erythrocytes with aberrant morphology in their peripheral blood (PB) smear. Carriers of beta-thalassemia minor (trait) often do not have any clinical symptoms but may exhibit lifelong microcytosis and hypochromia (Clarke & Higgins, 2000). Also, basophilic stippling, and target cells are commonly documented forms of beta thalassemia carriers (Harrington et al., 2008) (Körber et al., 2017). Thalassemia minor shares many characteristics with iron deficiency anemia on a peripheral blood smear. Iron study results and particular CBC (complete blood count) findings can be compared to distinguish between the two conditions. As for patients with beta-thalassemia intermedia, their blood smear shows similar findings as thalassemia minor, in addition to variable poikilocytosis and anisocytosis. Moreover, thalassemia major has additional morphological variations that include teardrop cells, elliptocytosis, spherocytosis, and Howell-Jolly bodies. Patients with beta-thalassemia who have undergone a splenectomy will mainly have target cells and burr cells, whereas those who have not had their spleens removed will have target cells and teardrop cells (Chaichompoo et al., 2019). Thalassemia intermedia and major show numerous erythroblasts in inverse proportion to the degree of inefficiency of erythropoiesis, which increases after splenectomy (Brancaleoni et al., 2016b).

1.10 Treatment and outcomes

For sickle cell disease and beta thalassemia patients, the only cure is a bone marrow transplant. A well-matched donor is essential to maximize a patient's chances of a successful transplant. When the donor and recipient are related and have the same human leukocyte antigen (HLA) profile, over 85 percent of children who have blood or bone marrow transplants recover. Transplants are not risk-free, even with this impressive success rate. Seizures, major infections, and other medical issues are all possible side effects. The fatality rate for recipients of these transplants is around 5% (Ndefo et al., 2008). It is possible for transplanted cells to cause organ rejection in some cases. This condition is known as graft-versus-host disease. Many of the issues may be prevented with medication, but they are still possible. Scientists are looking at genetic therapy as a potential solution to sickle cell and beta thalassemia diseases. Restoring a defective or missing gene or inserting a new gene that enhances cell function are both components of genetic therapy. Using genetic treatments, researchers change stem cells derived from patient blood or bone marrow in a laboratory setting. When a suitable donor cannot be found,

genetic treatments that alter a patient's own hematopoietic stem cells hold great promise as a treatment. After being put into the circulation, modified stem cells can make their way to the bone marrow where they can differentiate into healthy red blood cells.

Alternatively, treatment is offered to these patients to manage their symptoms and ease their pain. Ibuprofen and acetaminophen, both of which are over-the-counter pain relievers, can manage mild to moderate pain. Healthcare providers always recommend regular visits for evaluation, drink plenty of fluids, maintain an active lifestyle and consume a heart-healthy diet, ensure receiving every recommended vaccination, and avoid sudden change of temperature.

Medications

Morphine sulfate immediate-release (MSIR) and hydromorphone can be administered under healthcare provider supervision for extreme pain relief.

Adults and children older than 4 years old can be treated with Voxelotor for sickle cell disease. The sickle-shaped and clumping red blood cells are prevented from forming by this oral medication. This has the potential to enhance blood flow to organs and decrease the risk of anemia by reducing the breakdown of red blood cells. Headache, diarrhea, stomach discomfort, nausea, fever, and allergic reactions are some of the possible adverse effects.

Patients with sickle cell disease, whether they are adults or children aged 16 and above, are eligible to receive Crizanlizumab-TMCA. Administered via an intravenous (IV) line into a vein, the medication aids in preventing blood cell adhesions, reducing blocked vessels and pain crises. Nausea, sore joints, back discomfort, and high body temperature are all potential adverse effects.

Studies have demonstrated that the oral medication Hydroxyurea can lessen several sickle cell disease and beta thalassemia problems. Administered to both adults and children but not pregnant women, the frequency of pain crises and acute chest syndrome episodes was reduced with Hydroxyurea. It helped with anemia as well, which meant fewer hospitalizations and transfusions were needed. A decreased platelet or white blood cell count are potential adverse effects. Hydroxyurea can infrequently exacerbate anemia. Typically, these adverse effects resolve rapidly once the patient discontinues the medication.

L-glutamine was approved by the FDA for ages 5 and older to reduce the incidence of pain crises. Prescribed in the form of a powder, the medication is to be incorporated into foods or liquids. Among

the potential adverse effects are fatigue, chest pain, and musculoskeletal pain. In senior adults, L-glutamine has not yet been evaluated.

In 2005, Sildenafil (phosphodiesterase-5 inhibitor) was authorized by the FDA for the treatment of pulmonary hypertension (Ndefo et al., 2008).

L-arginine is an essential substrate for the synthesis of nitric acid. Suffering from SCD, adults may also exhibit markedly reduced arginine levels, in addition to a NO deficiency. The infusion of L-arginine has been demonstrated to decrease vascular resistance and enhance blood oxygenation in neonates with pulmonary hypertension.

Penicillin administered twice daily reduces the risk of significant bloodstream infections in infants.

Many medical professionals discontinue the prescription of penicillin for children older than five years. Certain healthcare providers advocate for the lifelong prescription of this antibiotic, especially in patients with hemoglobin SS or hemoglobin S β 0 thalassemia, due to the continued risk to individuals with sickle cell disease. Individuals who have undergone splenectomy, which is the surgical removal of the spleen, or who have had a previous pneumococcal infection, should continue to take penicillin for the duration of their lifetimes.

In certain instances, dietary supplements such as folic acid, which promotes erythrocyte synthesis, may be necessary to ameliorate anemia.

Transfusions may be prescribed by a medical professional to treat and prevent specific complications. Acute transfusions are utilized to treat significant anemia caused by complications. Transfusions may also be administered to patients who are experiencing multi-organ failure or acute stroke. Sickle cell disease patients are frequently transfused with blood prior to undergoing surgery. Red blood cell transfusions can increase the quantity of normal red blood cells. Continual or routine blood transfusions may reduce the risk of recurrent stroke in patients who have experienced an acute stroke. In addition, healthcare clinicians may advise transcranial Doppler ultrasound abnormalities in children to receive blood transfusions, as these transfusions have the potential to reduce the risk of initial stroke. Certain medical professionals employ this method to treat complications that are resistant to hydroxyurea. Also amenable to transfusions are patients who experience an excessive number of adverse effects from hydroxyurea. One potential complication is alloimmunization, which transpires when the recipient of the blood transfusion develops antibodies against the blood transfused. Such

antibodies can hinder the recipient's ability to locate a compatible unit of blood for subsequent transfusions. Also, possible complications are infection and iron overload.

Deferoxamine, Deferasirox, or Deferiprone are oral drugs used as iron chelation therapy to treat iron overload from frequent blood transfusions.

Life expectancy

The severity of the condition, the effectiveness of therapy, and the occurrence of complications all have a role in determining the expected lifespan of patients with sickle cell disease or beta thalassemia.

Sickle cell disease is associated with a decreased life expectancy. Lubeck et al showed that the matched non-SCD group had a life expectancy of 76 years, whereas the SCD cohort had a projected life expectancy of 54 years. After adjusting for quality of life, the difference narrowed to 33 years for SCD patients versus 67 years (Lubeck et al., 2019). A new report indicates that patients with sickle cell disease have a life expectancy that is over two decades inferior to that of the general population. Existing therapies for SCD are enhancing both life expectancy and quality of life. It is possible for individuals with sickle cell disease to live past the age of 50 with proper disease management (Jiao et al., 2023).

Similarly, beta thalassemia is associated with decreased life expectancy as well but prognosis for β -thalassemia patients has been steadily improving due to therapeutic advancements in the last several decades. The life expectancy of those who have the thalassemia trait is normal. From 1970 to 1980 and 1990, the average life expectancy in β -thalassemia increased from 17 to 27 and 37 years, respectively. In 2011, 89% of patients made it to age 40. Based on recent information, it appears that 63% of patients will make it to 50 years of age. The most common cause of mortality is due to iron excess and its effects on the heart (Farmakis et al., 2020).

1.11 Prevalence of thalassemia in Saudi Arabia

Middle Eastern countries have come a long way in regard to medical advances. They have decreased the prevalence of many infectious diseases and overcome many obstacles to elevate the health status in their populations. The fortunate use of modern medicine is having the effect of prolonging the lives of patients with hemoglobinopathies, therefore there is a need to develop better methods for treating them. Nevertheless, Saudi Arabia is no different from its neighboring adjacent Mediterranean and Asian countries in having high prevalence of sickle cell trait and thalassemia. Many environmental,

cultural, and religious factors contribute to the reason why hemoglobinopathies in these countries are present at high rates. The results of natural selection by historic infection with malaria are responsible for the high gene frequencies for hemoglobinopathies. Asymptomatic or minor heterozygotes for HbS, HbC, and possibly thalassemia and HbE, as well as those with mild forms of thalassemia, are more resistant to severe malarial infection than healthy individuals, even though severely affected homozygotes would have died young if medical interventions were not available. The “malaria hypothesis” suggests that in the ancestral populations, these alleles were under severe positive selection as a result of the resistance that they provide against prevalent, lethal outcomes to infection by endemic *Plasmodium falciparum*. The alleles remain present even where the threat from malaria has been largely eliminated. (Piel et al., 2010).

Different hemoglobin abnormalities may be co-inherited because of their high frequencies, leading to a complicated array of genotypes and clinical symptoms. In fact, structural Hb variations and thalassemic defects frequently coexist in various areas, and it is common for people from regions with a high prevalence of thalassemic defects to inherit more than one type of genetic thalassemic variation.

The most evident factor of the high prevalence rate of homozygotes is the tradition of consanguineous marriage (particularly between first cousins) which accounts for up to 60% of Saudi marriages and because hemoglobinopathies are inherited diseases, close family member marriages will increase the possibility of passing on the disease to the offspring. Also the large number of children per family elevate their chances for a family to have an affected child (since the disorders are inherited in an autosomal recessive pattern) (El-Hazmi et al., 2011) (Olwi et al., 2018).

In Saudi Arabia, the first sickle cell case was recognized and reported in the 1960s (Lehmann et al., 1963). On the other hand, research done outside of Saudi Arabia in 1986 by Boehm et al. was the earliest report of a beta-thalassemia mutation, Cd37 (G>A), being found in a Saudi patient (Boehm et al., 1986).

The majority of patients with hereditary blood disorders are carriers of sickle-cell disease and/or β -thalassemia. Over a six-year period, 99,968 individuals were evaluated, and the frequency of couples with sickle cell trait was 45.1 per 1000. 18.5 out of every 1000 people tested had beta-thalassemia in their families. For both sickle cell anemia and beta-thalassemia, the incidence was highest in the East, then the South and West, and finally the Midwest and the North (Memish et al., 2011). From February

2011 through December 2015, 1,230,582 individuals were included in a secondary data analysis of the premarital screening and genetic counseling program database. Carrier and disease states of beta thalassemia and sickle cell disease prevalence rates were estimated (per 1000 people). The total prevalence rate of beta thalassemia over the research period was 13.6 per 1000 people per year with 12.9 for the trait and 0.7 for the disease. The rate of SCD prevalence was 49.6 with 45.8 for the trait and 3.8 for the disease (Alsaeed et al., 2018b). The issue of consanguineous marriages and large family size may explain the high prevalence of sickle-cell disease and β -thalassemia in Saudi societies. The burden on health services is so great that the Saudi government initiated a premarital screening program in 2004 (AlHamdan et al., 2007) (Memish & Saeedi, 2011) and funded many research groups to study these disorders. Sickle cell trait is broadly prevalent all over Saudi Arabia with the highest prevalence in the eastern region (Serjeant, 2013a). Two clinical phenotypes have been identified, mild in the Eastern province with less complications (ARAB haplotype) and a severe phenotype in the Western region (African Benin haplotype) (Alabdulaali, 2007). What makes the ARAB haplotype unique is the way in which it affects the manifestation of sickle cell disease in homozygous carriers in unusually mild form. Clinically, the risk of avascular necrosis of the femoral head due to bone marrow hyperplasia and decreased blood flow to the bone is higher in the ARAB haplotype than the Benin haplotype, whereas the risk of stroke, acute chest syndrome, and dactylitis are more common in the Benin haplotype. Both haplotypes may get painful crisis and silent brain infarcts but with later onset in the ARAB haplotype (Jastaniah, 2011). The phenotypic variation degree correlates with differences in the fetal hemoglobin level in each haplotype. The mean level of HbF is much lower in the Benin haplotype than the Asian haplotype. This slight elevation in HbF helps protect against the disease severity by presumably preventing the abnormal hemoglobin from assembling into long polymers which decreases the painful crisis incidence (Gabriel, 2010).

Nowadays, a wide range of abnormal hemoglobin variants has been identified. In 2011, the highest number of documented variants (28 variations) found in a single research was carried out in Jeddah by Abuzenadah and colleagues (Abuzenadah et al., 2011) . An extensive screening program was conducted to help identify the different variants present amongst the Saudi population, some of which are found to be structural variants of the beta globin cluster, HbS (c.20A>T, p.Glu7Val, rs334), HbC (c.19G>A, p.Glu7Lys, rs33930165), Hb O-Arab (c.364G>A, p.Glu122Lys), Hb D-Punjab (c.364G>C, p.Glu122Gln, rs33946267), Hb Riyadh (c.363A>C, p.Lys120Asn) (El-Hazmi & Lehmann, 1976)) , Hb F-Dammam (c.238G>A, p.Asp79Asn) (Al-Awamy et al., 1985)), HbE (c.79G>A, p.Glu27Lys, rs33950507).

Besides, Hb-Handsworth, a missense mutation in the alpha globin gene c.55 G>C, p.Gly18Arg (Al Zadjali et al., 2014). Family migration from city to city within Middle East countries and other Arab countries looking for a better life has contributed to the dispersal of these variants (El-Hazmi et al., 2011) (Alsaeed et al., 2018a). In a review study published by AlAithan et al in 2018, they reported that there are several *HBB* mutations repeatedly reported across all studies of thalassemia prevalence in Saudi Arabia that include the Asian IVSI-5 (G>C) mutation on top of the list, followed by the Mediterranean Cd39 (C>T), and several others such as IVSII-1 (G>A), and IVSI-25bp mutations were consistently observed (Alaithan et al., 2018).

It is undeniable now that the time has come for immediate action towards the number of people affected with haemoglobinopathies. In a retrospective study conducted by Bin Zuair et al, that covers the period between 2016-2021, everyone admitted to the General Internal Medicine unit at King Abdulaziz University Hospital between 2016 and 2021 who had SCD was part of this retrospective observational research. In this investigation, 160 individuals were found to have SCD. It was southern Saudi Arabia from whence the majority came. About 19% of patients had three or more hospitalizations per year, and the average number of trips to the emergency department per year was four. The average duration of stay was six days. Around 7, 30, 60, and 90 days after discharge, the readmission rates were 8%, 24.5%, 13.6%, and 10.8%, correspondingly (Bin Zuair et al., 2023). The Saudi society bears a heavy economic burden and the need has increased to consider these conditions as a public health problem that needs to be prioritized in the context of prevention. The introduction of premarital screening and genetic counseling in Saudi Arabia, education and awareness of the complications and the chronic illness associated with the disease have essential roles for the prevention of the disease and can reduce rate of children born with hereditary hemoglobinopathies. As a result of thalassemia patient research, numerous mutations in the globin genes, the 3'UTR and 5'UTR, and the regulatory components governing the production of the alpha and beta globin gene families and thus the hemoglobin switch have been identified and more novel related pathogenic SNPs are still being identified. To learn more about disease mechanism and how expression is regulated, it is still crucial to study, investigate, and characterize naturally occurring deletions and point mutations in carriers and patients (Chen et al., 2018) (Sugiyama et al., 2019) (Luo et al., 2020) (Atroshi et al., 2021).

1.12 Aim

From severe transfusion-dependent thalassemia major to moderate non-transfusion dependent thalassemia intermedia to the asymptomatic or minor carrier state, the clinical symptoms of beta-thalassemia are exceedingly varied. Variations in beta globin genes are primarily responsible for the striking phenotypic diversity. β -thalassemia can be caused by many different genetic mutations affecting *HBB* gene, and there may be previously undiscovered beta-thalassemia alleles in the Saudi population. The goal of this present study is to utilize different molecular techniques approaches to establish the most frequent mutations responsible for the disorders that when analyzed will provide detection of up to 90% of the identified mutations. To determine the profile of novel or previously reported causative mutations in more than 150 transfusion dependent individuals using TaqMan genotyping and next-generation DNA sequencing. Early detection of specific *HBB* mutations can result in a more accurate prediction of the clinical presentation and severity of the disease. The study's findings will be used by public health campaigns at the national level to enhance genetic risk profiling.

In addition, I will explore the genomic variations in a family with transfusion dependency but without a definitive genetic diagnosis related to HBB. I will implement different data filtering approaches to detect unknown genetic variations in functionally related genes. Also, application of in-silico analysis of the detected variations to propose candidate genes that may have contributed to the severe etiology of thalassemia within this family.

CHAPTER TWO
EXPERIMENTAL PROCEDURES

Experimental procedures

All experiments were performed by me unless stated otherwise. They were conducted in the Center of Innovations in Personalized Medicine (CIPM) at King Fahd Medical Research Center at King Abdulaziz University, with the collaboration of the Genome Center in the same building.

As for the equipment and consumables, the thermal cycler, 0.1 MicroAmp optical 96-well PCR reaction plate, optical adhesive films, and 0.2 ml PCR reaction tubes were obtained from Life Technologies (California, USA). Pipettes, 1.5 ml microfuge tubes, and pipette tips were from Eppendorf (Hamburg, Germany). Gel electrophoresis Power Pac Basic was from Bio-Rad (California, USA). Vortex was from Thermo Scientific (Massachusetts, USA). The mini centrifuge was from VWR International (Pennsylvania, USA). The non-stick Dnase-Rnase free tubes were from Ambion (California, USA).

2.1 Recruitment of thalassemia subjects and sample collection

Work involving human participants was done according to the Unit of Biomedical Ethics Research Committee for the purpose of collecting blood samples from patients with thalassemia or SCD treated at the Day Care Unit at King Abdulaziz University Hospital. The approval was granted by the hospital (reference No 185-14).

Patients were selected based on their blood transfusion dependency status due to their thalassemia major or SCD diagnosis. Patients with SCD do not usually require blood transfusion, but depending on their hemoglobin level and if they had other specific indications such as previous stroke, recurrent acute chest syndrome, or pulmonary hypertension, the hematologist might recommend exchange transfusion (if hemoglobin >8 g/ dl) or simple top up transfusion (if hemoglobin <8 g/ dl).

Samples were collected from 154 transfusion-dependent beta thalassemia and SCD patients from DayCare Unit at King Abdulaziz University Hospital (KAUH) in collaboration with Dr Badr Arab. She obtained informed consent from all patients prior to sample collection. Patients were of 12 nationalities of either sex, and their ages covered a wide range from infants to over 60 years (Figure 14).

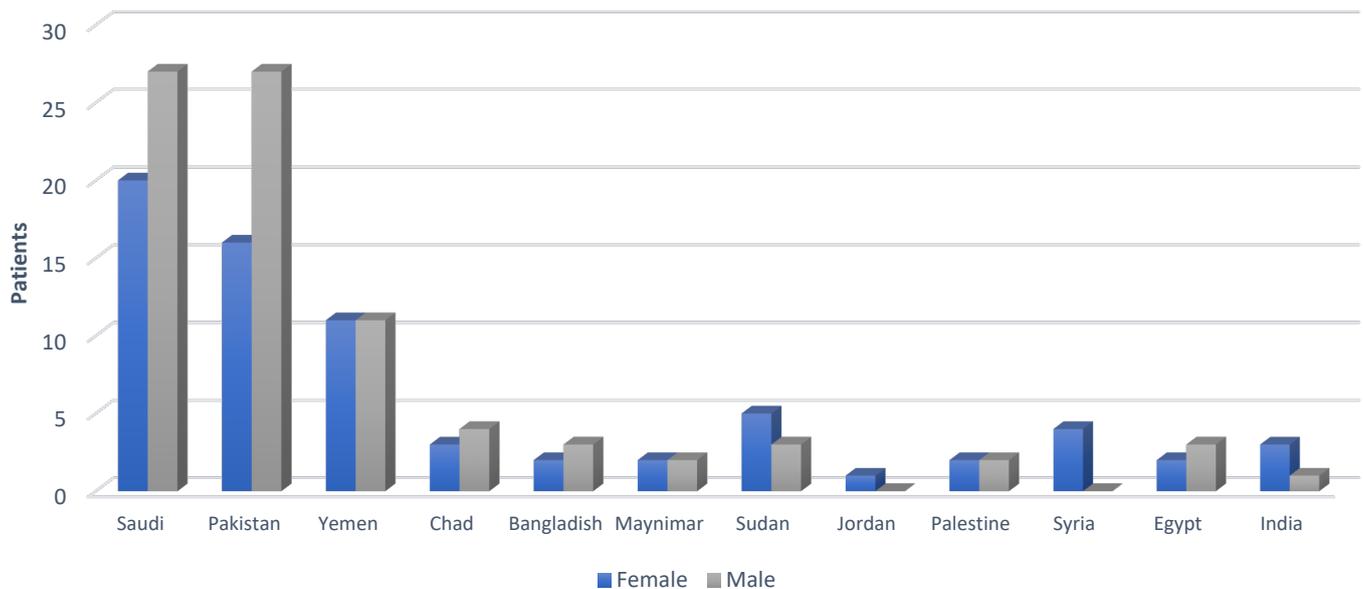


Figure 14: Samples of different nationalities of transfusion dependent patients with sickle cell and beta thalassemia.

Patients were recruited at their appointment for routine transfusion. To avoid admixture with donor cells, whole blood was drawn before transfusion. Collected samples (5 ml) were anticoagulated with EDTA and transferred to the Center of Innovations in Personalized Medicine (CIPM) located at King Fahd Research Center, King Abdulaziz University. Specimens, then, were logged at the receiving center and held at $-20\text{ }^{\circ}\text{C}$ for long-term storage.

2.2 Whole blood DNA extraction

I have used QIAamp Mini (CAT #51106) kits from Qiagen for DNA extraction. It is designed to purify an average of 6 μg of total DNA rapidly with the size of up to 50 kb in length that is free of proteins, nucleases and other contaminants. QIAamp spin-columns based extraction was used to eliminate sample-to-sample cross-contamination.

The procedure comprises four steps in which cells are lysed using lysis buffer and protease enzyme, then loaded into the column and bound to its membrane. Proteins and contaminants are not retained in the membrane because of the lysate buffering salt and pH conditions. After that, DNA is washed through two centrifugation steps then purified DNA is eluted in either elution buffer or water.

DNA concentration and purity was then measured using Thermo Scientific NanoDrop 2000c Spectrophotometer, then stored at $-20\text{ }^{\circ}\text{C}$ for later downstream applications.

2.3 DNA quantification

Qubit 2.0 Invitrogen (Life Technologies Corporations, California, USA): Fluorescence-based Qubit™ quantitation using molecular probe kits which include Qubit dsDNA Broad Range (BR) Dye Assay kit by Life Technologies.

Before using DNA for downstream applications, I measured DNA concentration using highly sensitive Qubit Fluorometer assays that would only fluoresce when dyes are bound to DNA. DNA quantitation is achieved with high levels of accuracy because it only measures the concentration of the molecule of interest, disregarding everything else.

2.4 TaqMan assay genotyping

2.4.1 Principle

TaqMan Assays (probe and primer sets) were used to detect specific single nucleotide polymorphism (SNV) in purified genomic DNA samples by real-time quantitative PCR using QuantStudio 12K Flex, Applied Biosystems by Life Technologies (USA).

The TaqMan SNP genotyping assay consists of two primers; forward and reverse, and two probes; one with the wild type reference allele and the other one with the variant allele (SNV). Each probe is covalently attached to a fluorescent reporter (VIC/reference and FAM/variant) at the 5' end and a quencher dye at the 3' end. The quencher absorbs the fluorophore emission from being excited via fluorescence resonance energy transfer (FRET) and inhibits any signal if the reporter and the quencher are in close proximity with each other. FRET is very sensitive to molecular level distance (close physical proximity). It transfers light energy without the use of radiation from a dipole fluorophore donor that being excited to another dipole fluorophore acceptor with matched resonance frequency (Hussain 2012). A minor groove binder (a dihydropyrrroloindole-carboxylate (CDPI3)) is added to each probe at the 3' end which increases the melting temperature of a given probe length and increases the probe and target binding affinity and produces robust allelic discrimination by stabilizing van der Waals forces. The probe anneals to the DNA during PCR only if the exact matched target sequence is present. It anneals downstream of the primer-binding site (anneals to single stranded DNA). The aim of the assay is to identify which nucleotide is present in each allele. When the probe anneals to the target sequence, the 5'-exonuclease activity of Taq polymerase enzyme cleaves the probe during the PCR extension phase. This cleavage separates the reporter and the quencher leading to fluorescence of the fluorophore. This happens each PCR cycle, increasing the fluorescence intensity proportionally to the

quantity of PCR products produced, and this increase is detected and measured during the exponential phase of PCR by TaqMan Genotype Software package provided with Applied Biosystems real-time PCR system (Figure 15) (Pub. No. 4448637). (Holland et al., 1991) (Shen et al., 2009) (Robledo et al., 2005)

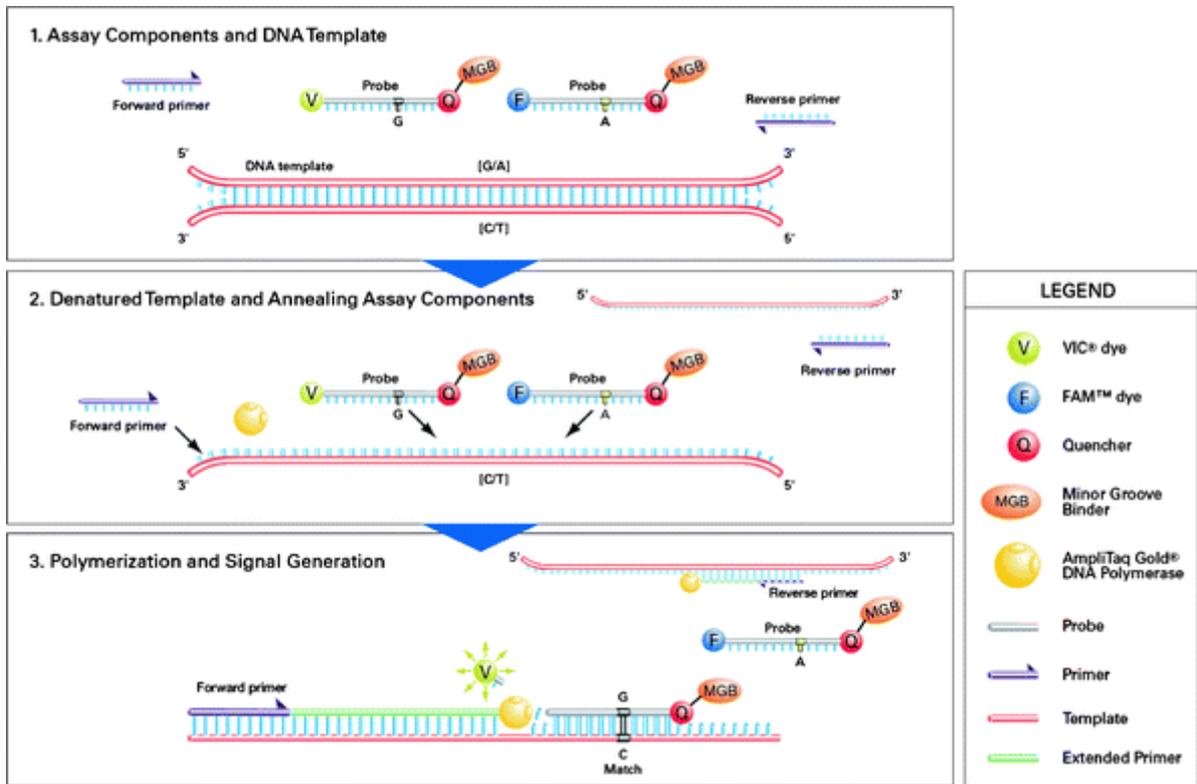


Figure 15: Principle of TaqMan genotyping assay.

Source: reproduced with permission (Schleinitz et al., 2011).

2.4.2 Assays

The Life Technologies TaqMan Assays Design Panel was used to design TaqMan SNV Genotyping Assays. We, myself included, have developed and custom designed the SNV Assays, probes and primers sequences, to detect specific single nucleotide mutations in *HBB* gene in purified DNA samples from thalassemia and SCD patients. A length of 100 bp flanking the variant from both sides was used as input for the design. These specific SNPs were chosen according to the most thalassemia mutations reported in Saudi population reported in the literature at the time we started the project. These variants of *HBB* gene are, c.20A>T, c.27dupG, c.92+1G>A, c.92+5G>C, c.93-21G>A, and c.118C>T (Abuzenadah et al., 2011). (Table 4) (Figure 16).

Table 4: The most frequent beta-thalassemia SNPs within Saudi population used in our study. Mutations are mapped to HBB transcript RefSeq NM_000518.4.

SNV (nomenclature)	Genomic Location (Hg19/GRCh37 chromosome 11)	Reference cluster ID (rs#)	Protein change	Molecular Consequence
c.20A>T	5248232	rs334	NP_000509.1: p. Glu7Val	Missense variant (Sickle Cell Disease)
c.27dupG	5248224-5248225	rs35699606	NP_000509.1: p. Ser10fs	Frameshift
c.92+1G>A	5248159	rs33971440	-	Splice donor variant
c.92+5G>C	5248155	rs33915217	-	Splice donor variant
c.93-21G>A	5248050	rs35004220	-	Intron variant
c.118C>T	5248004	rs11549407	NP_000509.1: p. Gln40Ter	Nonsense variant

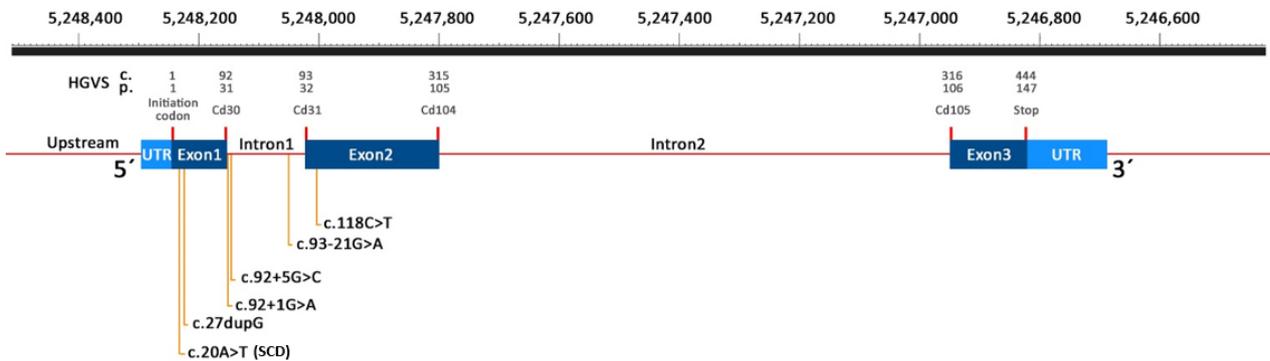


Figure 16: Mapping of the most frequent beta-thalassemia SNPs within Saudi population used in our study.

The exons shown are those of the RefSeq beta globin transcript NM_000518.4.

Life Technologies prepared the assays as a 40x concentrate, containing one VIC-labeled probe to detect allele 1 with reference sequence, one FAM-labeled probe to detect Allele 2 with mutation sequence, and sequence-specific forward and reverse primers. (Table 5)

Table 5: TaqMan assays used on samples and their primer sequences.

dbSNV rs# is the reference SNV cluster ID which is used to refer to a specific SNV. NM_ and NP_ are the nucleotide (cDNA) and protein reference sequences respectively.

SNV (nomenclature)	Forward Primer Sequence	Reverse Primer Sequence	Probe Sequence
rs33915217 NM_000518.4: c.92+5G>C	GGTGAACGTGGATGAAGTTGGT	GCCCAGTTTCTATTGGTCTCCTTAA	CAGGTTG(G/C)TATCAAGG
rs33971440 NM_000518.4: c.92+1G>A	GGTGAACGTGGATGAAGTTGGT	GCCCAGTTTCTATTGGTCTCCTTAA	CTGGGCAG(G/A)TTGGTAT
rs35004220 NM_000518.4: c.93-21G>A	GGGTTTCTGATAGGCACTGACT	GCAGCCTAAGGGTGGGAAA	CTCTGCCTATT(G/A)GTCTAT
rs11549407 NM_000518.4: c.118C>T	CTTAGGCTGCTGGTGGTCTAC	AGTGGACAGATCCCCAAAGGA	AAGAACCTCT(G/A)GGTCAA
rs334 NM_000518.4: c.20A>T	TCAAACAGACACCATGGTGCAT	CCCCACAGGGCAGTAACG	CTGACTCTG(A/T)GGAGAA
rs35699606 NM_000518.4: c.27dupG	CCTCTTATCTTCTCCCA	AAGCGAGCTTAGTGATAC	AGTAACGGCAGAC(T/C)TTCTC

I have personally performed the procedure using the same SNP genotyping conditions that were previously optimized by a former PhD student in the same laboratory (Alwazani et al., 2016). Each qPCR reaction contained 0.5 µl of diluted SNP assay mix (20X), 5 µl TaqMan Fast Universal PCR Master Mix (2X) by Applied Biosystems, 2.5 µl of nuclease-free water and 2 µl of purified genomic DNA sample (1-20 ng). PCR stages were used as indicated by the manufacturer. The default real-time PCR involved heating to 95.0°C for polymerase activation for 10 min, followed by 40 cycles of 15 s at 95.0°C (denaturation), decreasing by 2.42°C/s to 60°C for 1 min (annealing/extension). All other temperature ramps were 2.63°C/s. A final 20 s extension allowed a post-cycling fluorescence reading.

2.5 Next generation sequencing by Ion Torrent PGM™ semiconductor sequencing

I performed NGS using Ion Torrent PGM™ (Personal Genome Machine) on 54 thalassemia samples. I have sequenced the *HBB* gene to identify the mutations that corroborate their beta thalassemia diagnosis. All reagents, assays, kits, and equipment used in this procedure were manufactured by Life Technologies. The Ion PGM™ 200 Sequencing Kit (Cat. no. 4474004) was used that includes reagents

and materials for performing sequencing runs using the Ion 316™ Chip and Ion OneTouch™ 200 Template Preparation Kit v2 DL (Cat. no. 4480285).

2.5.1 Principle

This technique is based on label-free DNA sequencing (no fluorescence). In the first stage of the NGS workflow, 10 ng of gDNA is quantified. A library of DNA fragments flanked by ion torrent adaptors was obtained by amplification of the target, then DNA fragmentation to a uniform size with blunt ends of amplicons for adaptors ligation. Xpress™ Barcode Adaptors were used for sequencing. These adaptors are specifically designed to enable sample identification when numerous different library samples are pooled together to be sequenced on a single sequencing chip. Unamplified DNA was then cleaned up and purified. Library quantification using QuantStudio 12K Flex (must be 100 pM to proceed with pooling) was performed to ensure a good ratio of DNA library to Ion Sphere Particles in the next step. After quantification, barcoded samples were combined in one pool (Forth & Hoper, 2019). Then, using emulsion PCR, DNA pooled library is clonally amplified onto the Ion Spheres™ Particles (ISPs) (up to 200 base read libraries). ISPs are hydrogel beads coated with primer. These primer sequences are complementary to the adaptor sequences ligated to the amplicons during library preparation. In emPCR, single molecules that have been attached to individual beads are amplified and then brought into isolated aqueous droplets containing the general reagents for PCR, in an aqueous/oil emulsion. The pool library concentration was adjusted to optimize the loading of single amplicons to each bead. Amplicons are amplified so that this bead will contain thousands of copies of this amplicon conjugated to it. Beads containing more than one amplicon will be discarded during data analysis (unreadable). After the amplification, extraction with an organic solvent and centrifugation steps is then performed to isolate and break the emulsion and recover the beads. Then, using Ion One Touch Enrichment instrument, template positive ISPs are enriched and selected, then deposited by centrifugation into a chip that contains nano-sized wells with each well holding a different DNA template. The chip is then loaded on the PGM. The run was planned and created using the PGM touch screen and the sequencing program was initialized. The sequencing reaction works by using a semiconductor chip that contains millions of nanowells. pH-mediated sequencing uses four dNTPs that are added separately and sequentially and washed away from an amplified, primed template, with bound polymerase, in a nanowell in a metal–oxide–semiconductor field-effect transistor that captures chemical change from DNA sequencing and translate it into digital information. Each nanowell should contains a single ISP that is coated with thousands of copies of a single DNA fragment that can be assigned to an individual library by its barcode. Any well that contains two ISPs with different amplified fragments on them

cannot be interpreted later on (which cannot occur easily because of the relative size of the ISP to the well). The DNA polymerase will incorporate only the cognate complementary nucleotide from the individual dNTP at each position, resulting in the release of protons only when the cognate nucleotide is supplied using DNA polymerase and buffers. The released H⁺ ion decreases the pH of the solution and pH sensors detect that change. This biochemical change enables determination of which base was added. The excess dNTPs are then washed away, and the process is repeated with a different dNTP species in a cycling process (Rothberg et al., 2011) (Rusk, 2011). After sequencing was done, the generated data was automatically transferred to the required torrent server for analysis (Rothberg et al., 2011) (Pennisi, 2010).

2.5.2 Procedure

2.5.2.1 Preparation of the library

To create the library, the targeted sequence primers were custom designed and manufactured through the Life Technologies AmpliSeq portal (<http://ampliseq.com>) to cover 31.9 kb (chr11:5245611-5277531) of the genomic sequence of the beta globin cluster with a Human Reference Sequence GRCh37/hg19. Intergenic and intragenic regions were covered to find any gene alterations. After filtering low specificity regions within chr11:5245611-5277531, 30.55 kb was covered with 95.62% coverage and 125-275 bp amplicon range (Figure 17).

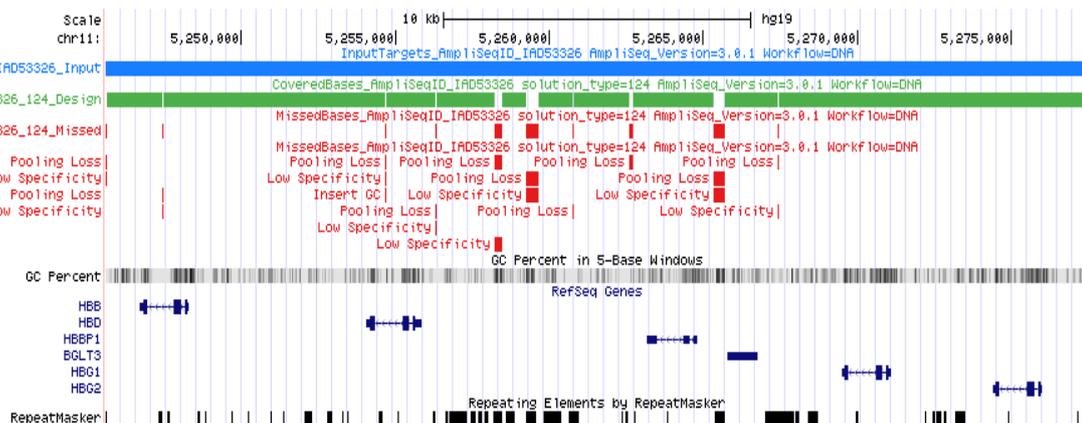


Figure 17: Next generation sequencing coverage.

For the preparation of the library, the The Ion AmpliSeq™ Library Kit 2.0 (Cat. no. 4475345) was used. Reagents and solutions used in this stage include: 5X Ion AmpliSeq™ HiFi Master Mix, FuPa Reagent (for primers digestion and amplicons phosphorelation), Switch Solution, DNA Ligase, Ion AmpliSeq™

Adapters, Platinum® PCR SuperMix HiFi, Library Amplification Primer Mix, Low TE, Ion Xpress™ P1 Adapter, and Ion Xpress™ Barcode (short identifier oligonucleotide sequence).

As shown in Figure 18, library construction involves amplifying the *HBB* cluster target and fragmenting the DNA of each sample separately to a uniform size using quantitative-PCR (QPCR) generally to 200-400 bp, then flanking these fragments with the ion sequencing adapters (barcodes) as a library-specific identifying sequences for each sample.

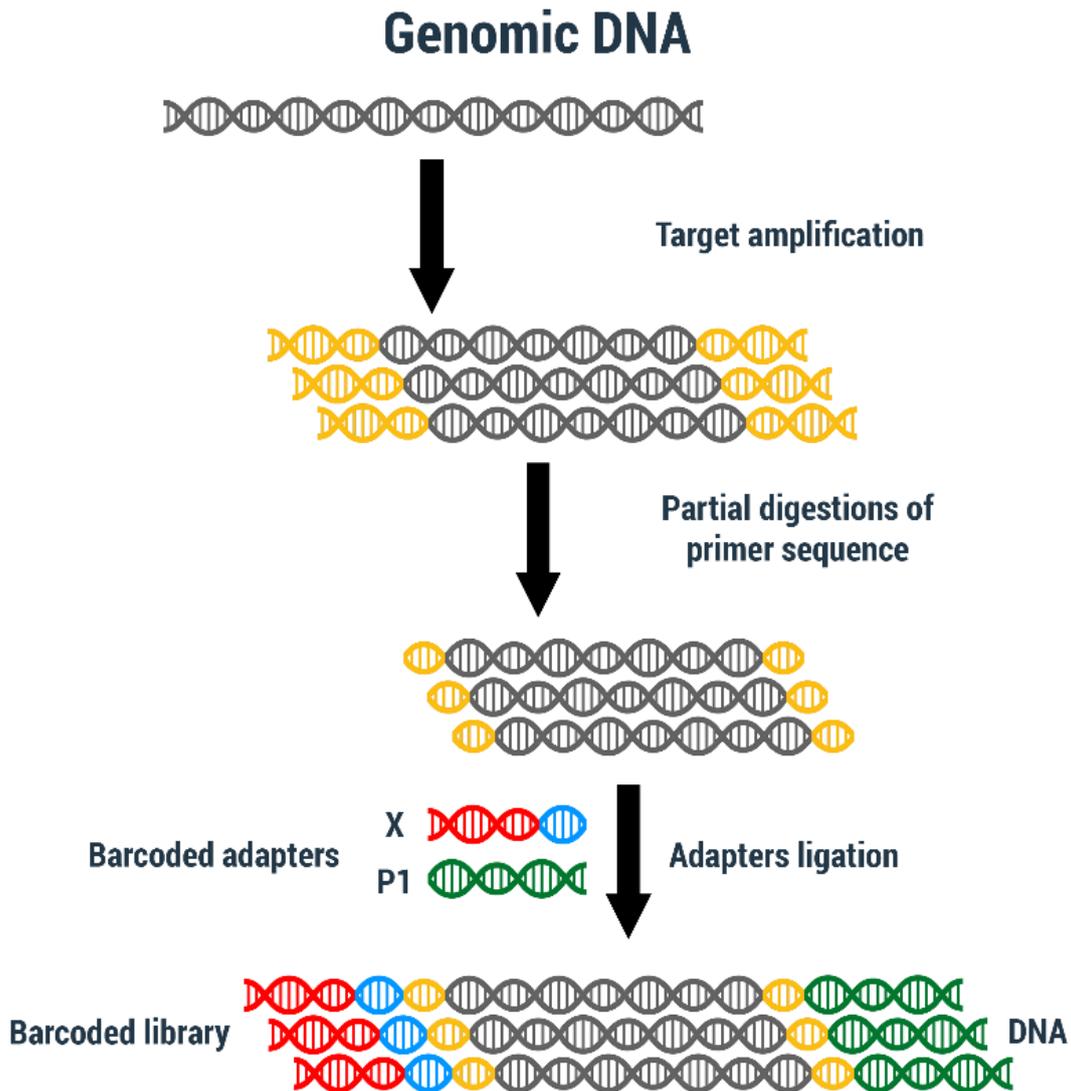


Figure 18: DNA fragmentation and barcoding.

In library preparation, the target sequence on the genomic DNA of the sample is amplified. After that, DNA is fragmented into a uniform size. Next, two synthesized known adapter sequences are ligated at the 3' and 5' ends of the DNA. The first is X in red with added short identifier oligonucleotide sequence in blue, and the second is P1 in green. P1 and X both contain universal primer. X adaptor contains a short unique oligonucleotide sequence (barcode) in blue to be assigned to each sample when sequencing more than one specimen at the same time which will be recognized during the analysis step. Different barcodes are assigned for each sample.

DNA quantification

We first measured the DNA sample concentration by Qubit 2.0. DNA concentration needed to be 10 ng/ μ l in a maximum of 3 μ l. All samples were diluted to 5 ng/ μ l.

Amplification of the target

For our custom-designed HB panel, we have 2 pools (2X each)/sample, 400 primer pairs/pool.

In a PCR tube, 2 μ l of 5X HiFi master mix (for cleaner amplification and increased coverage uniformity), 5 μ l of 2X primer pool, 2 μ l of Sample (conc. 5 ng/ μ l) and 1 μ l of Nuclease free (NF) water giving a total volume of 10 μ l were added for each pool (which means ending up with two tubes for each sample).

Amplification of the target DNA was achieved by PCR which runs for 2 min at 99°C, then 16 cycles (decided by the number of primer pairs per pool: 385-768 pairs) of 15 s at 99°C and 4 min at 60°C, then down to 10°C for up to one hour.

Partial digestion of primer sequences

The two pools for each sample were combined into one tube. After that, 2 μ l of FuPa reagent, a proprietary product for primer digestion, was added to each combined amplified sample and a PCR step was performed after that by loading the tube into the thermal cycler at 50°C for 10 min, 55°C for 10min, 60°C for 20 min, then down to 10°C for up to one hour.

Ligation of adapter sequences to 3' and 5' ends of double stranded amplicons

Two adaptors (each are double stranded 20-50 bp known sequences) are ligated to the amplicons. Barcodes are added to the adaptors (pre-prepared). One adaptor sequence contains the annealing site for the primer for sequencing, while the other adaptor is used to anchor the DNA amplicon to a surface for sequencing. Each sample was assigned to a specific barcode.

To set up and run the ligation reaction, the following 4 μ l of pipetted Switch solution (to facilitate the conditions changing between the PCR and ligation steps during library construction) and 2 μ l of diluted barcode adapter mix were added to each sample.

Then, 2 μ l of DNA ligase was added to each tube and the tube was loaded into a thermal cycler for 30 min at 22°C then 10min at 72°C and down to 10°C for up to one hour.

Purification of unamplified library

After the PCR step, the DNA library was purified by pull-down with adapter-specific magnetic beads (Agencourt AMPure XP by Beckman Coulter, USA). The beads will purify the unamplified libraries because they contain complementary sequences that would attach to the DNA with added barcoded adaptors.

DNA that did not ligate with adaptors will stay in the supernatant. Also, for DNA size selection, various concentrations of magnetic beads are used to isolate DNA fragments of interest. Low TE (Tris-EDTA with low EDTA) buffer is used to elute the amplified library from the beads.

qPCR quantification

Library quantification was performed to ensure the highest quality sequencing. This quantification step will dictate during template preparation whether the emPCR is clonal or polyclonal because of the ratio of DNA to Ion Sphere Particles. Polyclonal reads are discarded during the analysis process because beads that contain multiple sequences cannot be interpreted by the PGM instrument. Using a too high concentration of the library DNA increases the proportion of filled wells that yield no data. On the other hand, too little DNA will increase the number of untemplated beads which will lead to the reduction of the number of reads obtained for each sample. Data from qPCR of the library is used to optimize loading of DNA onto the ISPs.

Using the Ion Library Quantitation kit, a three 10-fold simple serial dilution of the Ion control library ready-to-use standards (68 pM) for the qPCR were prepared (6.8 pM, 0.68 pM, and 0.068 pM to be programmed in the qPCR instrument software to create the standard curve). Each sample was prepared by adding 1 μ l of eluted sample and 99 μ l nuclease free (NF) water (1:100 dilution). Additionally, a reaction mix was prepared for each sample and standard by adding 10 μ l of Fast TaqMan PCR master mix and 1 μ l dye (library TaqMan quantification assay 20X probe/primer). A total volume of 10 μ l in each well (5.5 μ l of reaction mix + 4.5 μ l of the standards and diluted samples). The plate was then loaded into the QuantStudio 12K Flex and the experiment was set up (Figure 19)(Figure 20). After the run, the concentration determined by qPCR for each sample was then multiplied by 100 (sample dilution factor) to calculate the average concentration of the undiluted library, and according to that result, the sample was then diluted to 100 pM using water.

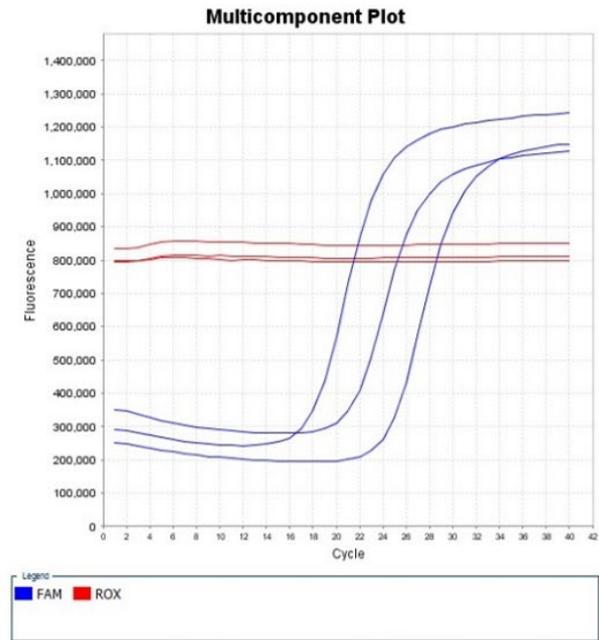
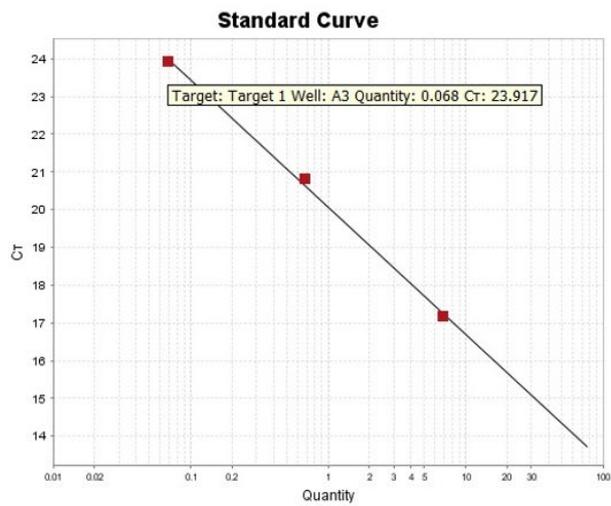


Figure 19: Standard curve and amplification plot of standards (6.8 pM, 0.68 pM, 0.068 pM). Standards are used as known DNA concentrations to help in determining unknown DNA concentrations of samples.

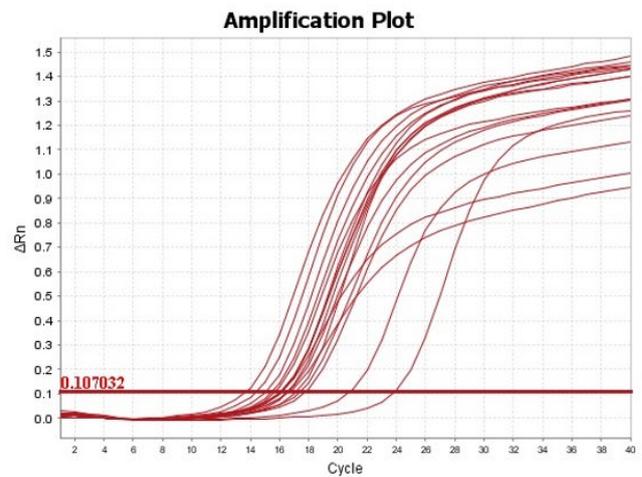
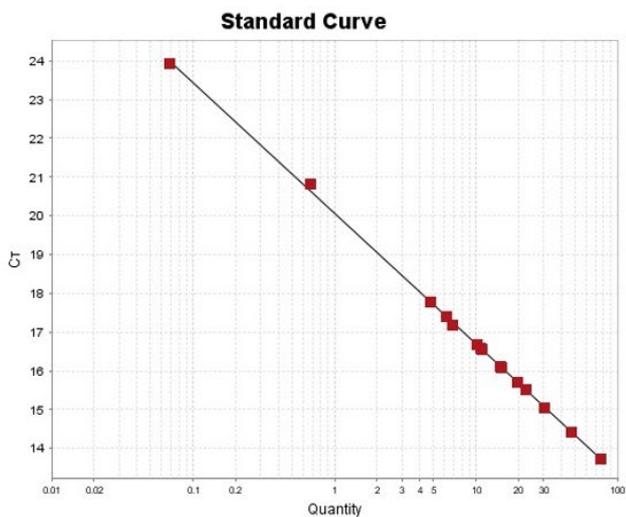


Figure 20: qPCR quantification of samples before sample pooling. Standard curve and amplification plot of samples.

Combination of barcoded libraries from different sample DNAs (pooling) was done by combining 5 μ l of each diluted sample (100 pM) together in one tube (HB pool library).

2.5.2.2 *Template preparation on ion sphere particles*

For the preparation of the template, The Ion OneTouch™ 200 Template Kit v2 DL (Part no. 4480285) was used. Reagents and solutions used in this stage include Ion OneTouch™ Enzyme Mix, Ion OneTouch™ 200 Ion Sphere™ Particles (with known attached sequences complementary to the adaptor sequences), Ion OneTouch™ 2X Reagent Mix, Ion OneTouch™ Oil, Ion OneTouch™ Reaction Oil, and Ion OneTouch™ Recovery Solution.

Principle of emulsion PCR

In EmPCR, the first step is the denaturation of amplicons into single stranded DNA. A biotin is attached on the 5' P1 primer adaptor. A primer coated bead is then ligated to the fragmented barcoded DNA, and then attached to a tagged primer, and finally PCR amplification is conducted within an aqueous emulsion in a continuous oil phase. Each aqueous droplet contains only one molecule of bead + DNA. This DNA is amplified so the bead would contain thousands of copies of clonally amplified DNA. The template single strand has been produced by extension of the covalently attached primers and therefore it remains attached to the bead under denaturing conditions while the complementary strand is removed (Kanagal-Shamanna, 2016) (Chai, 2019).

Clonal amplification and preparation of template positive ion sphere particles

A template amplification solution was prepared using Ion PGM Hi-Q OT2 kit by adding 25 µl of NF water, 50 µl of Ion PGM Hi-Q enzyme mix, 25 µl of diluted library and 100 µl of Ion PGM Hi-Q ISPs in order into the Ion PGM Hi-Q reagent mix tube (contains 800 µl) bringing the total volume to 1000 µl. The whole 1000 µl was then loaded into the emulsion PCR filter sample port, then 850 µl of reaction oil was added, and another 850 µl of reaction oil was added into the emPCR filter. The filter was then inverted, and the Ion One Touch 2 instrument was assembled as described in the manual.

2.5.2.3 *Enrichment of the amplified beads*

The purpose of enrichment was to separate template positive ISP from the ISP that do not have attached target. This is done with a streptavidin coated magnetic bead which that attaches to the biotin on the 5' P1 primer adaptor (biotin was incorporated to P1 during emPCR). The biotin-labelled P1 primers become incorporated into the successful products that are built on the P2 primers which are covalently attached to the ISP. Loaded ISPs, because of their biotin moieties, can thus be bound to streptavidin-coated magnetic beads. Template free (failed) ISPs will not bind. After washing the

magnetic beads to elute failed ISPs, template-loaded ISPs can be eluted in sodium hydroxide solution (melt-off solution), which leaves the covalent attachment of the template to the ISP intact but denatures streptavidin, releasing it from the biotin-loaded ISP. For the enrichment of the template, The Ion OneTouch™ 200 Template Kit v2 DL (Part no. 4480285) was used.

A fresh melt-off solution (denatures each dsDNA template attached to the ISPs to a single strand), a proprietary product, was prepared with the final composition of 0.1% Tween 20 solution (polyoxyethylene sorbitol which is a nonionic detergent) and 125 mM NaOH. A MyOne Streptavidin C1 beads was prepared as indicated in the user guide. The 8-well strip for the Ion One Touch ES was then filled starting from the square shaped well as follows: 100 µl template positive ISPs (my sample), 130 µL MyOne Beads, 300 µl Ion One Touch wash solution, 300 µl Ion One Touch wash solution, 300 µl Ion One Touch wash solution, empty, 300 µl freshly prepared melt-off solution and the last well was left empty. The well strip was then inserted into the enrichment instrument and 10 µl of neutralization solution was pipetted in a PCR tube (where the final sample will be collected) and inserted opened in the tube slot in the enrichment instrument which was assembled and run.

2.5.2.4 Loading the sequencing ion chip (316 chip V2)

Ion Chip 316 V2 has an output of 300 Mb- 1.0 Gb with 2-3 million reads. Each sample will cover 30.5 kb x 100 (coverage: fragments overlapping with variant) = 3050 kb. To calculate how many samples the Ion 316 Chip can hold simultaneously we have to divide 300,000/3050= 98.3. Therefore, the 316 chip can hold up to 98 samples simultaneously. Ion PGM Hi-Q sequencing kit (200-base read, 500 flows) was used in chip loading.

The ion sphere particles (ISPs) were applied to the 316 Ion Chip, one ISP per well, and then the chip was loaded onto the PGM. Maximum occupancy of the nanowells was achieved by loading the sample pool with constant speed in a linear manner with no interruptions while dispensing and no air bubbles.

2.5.2.5 Data output

The semiconductor membrane of the chip will detect the release of hydrogen ions when a nucleotide is incorporated which will be censored by an ion-sensitive layer beneath each nano-well. The ion-sensitive field-effect transistor (ISFET) sensor plate measures the solution's ion concentrations by measuring the current voltage change through the transistor which sends electrical pulses to the computer. Afterwards, the computer uses these pulses to translate them into DNA sequences.

Low quality makes our reads difficult to be mapped to the reference genome. This is due for many reasons such as accumulated errors as the run progresses. For that, sequence data went through trimming before alignment.

Trimming will make the sequence easier to align with reference genome by discarding the reads that fall below a pre-defined read quality average cut-off. For example, if we set the average read quality score to be 30 (precision of base call is 99.9%) for every 5 consecutive bases, the base that scores below 30 after that would be discarded. Afterwards, the reads in FASTQ files are mapped to a reference genome which is Human (*Homo sapiens*): hg19 in our case and result will be in .bam (stands for 'binary alignment map') file format which is usually compressed (into binary) form of .sam (stands for 'sequence alignment map) file (Figure 23), which can be recovered from it using SAMtool. The SAM file consists of a heading section and an alignment section. The heading section which begins with '@' symbol serves as an ID with name of the sequence and chromosome and their length. The alignment section describes in detail each sequence alignment with 11 mandatory fields (with optional tags) (figure).

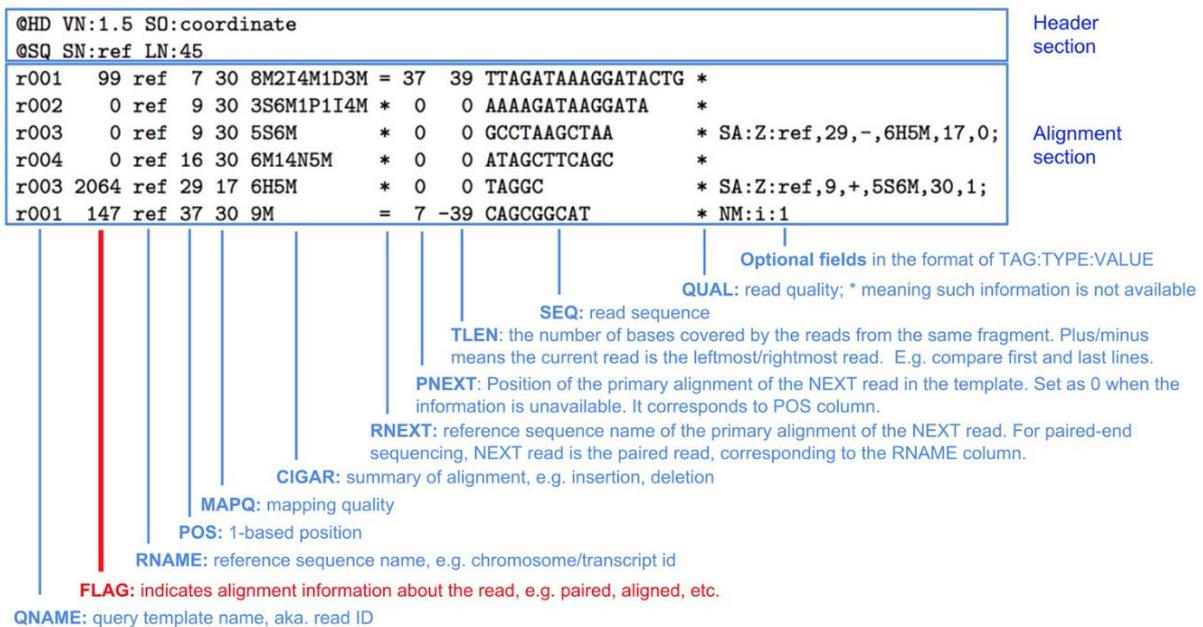


Figure 23: SAM file with annotated mandatory and optional fields.

Source: SAM format specification (Version: 1 Jun 2017)

After creating .bam files, our primary goal was to identify Single Nucleotide Polymorphisms (SNPs), and small insertions and deletions (indels) which is never a straightforward process because it depends on several factors such as base quality, mapping quality, number of reads, and overall coverage of position.

Even after identifying the variant, we it was necessary to annotate the variants to suggest which variant is benign and which is harmful or damaging in a disease context. To this end, the server accessed GRCh37.p13 Primary Assembly database to generate the annotation in the .vcf file. (variant call format). used another tool to filter the variants specifically to meet our patients' dataset to end up with a smaller .vcf file. The final result was downloaded as excel file for easier view and review.

Software parameters for variant calling

Alignment to The Reference Genome GRCh37/hg19 was supported with the Torrent Mapping Alignment Program (TMAP) map4 module (the default) which is based on the Burrows-Wheeler Alignment (BWA) algorithms that uses FASTQ format files to generate SAM files as the final alignment output.

Annotation set settings

The annotation set used to annotate the variants was 5000exomes, canonical RefSeq transcripts, ClinVar, DGV, DrugBank, Gene Ontology, OMIM, Pfam, PhyloP scores, and dbSNV. For the annotation analysis.

Annotation statistics and reporting options

dbSNV hit level was set to 'overlap' (all annotations that have loci overlap with variant are matched), COSMIC hit level was set to 'locus' (annotations that have loci start at variant locus are matched), ClinVar hit level was set to 'allele' (annotations that have locus matches with variant with at least one allele in common are matched), variantDB hit level was set to 'locus' as well, and splice site size was set to be '2 bases' (the size in bases of the small intronic region immediately upstream of an exon).

Variant finding

Variant finding was set to allow INDELS, SNVs, multiple nucleotide variants MNVs. Reporting of complex variants was disabled. The minimum mapping quality value (QV) required for reads to be allowed was 4.

Bamstatistics

The maximum mapping quality value was 255 (the alignment was ignored if it had more than the specified value). The maximum mismatches allowed in the alignments were 500 (any alignment with more than 500 mismatches was ignored and a report of mismatches number was generated).

Quality score is $P \leq 0.05$.

2.6 Whole exome sequencing by NovaSeq 6000 Sequencing system

I performed WES on 3 thalassemia samples that had no clear genetic explanation linked for their severe symptoms. Since all samples have been collected from transfusion-dependent patients with confirmed diagnosis of thalassemia, this technique was done to examine whole exomes in addition to the *HBB* gene in search of genomic variations that would explain the disease.

Later, I have repeated the procedure on an additional 12 samples related to two of the previous three thalassemia samples that had no clear genetic explanation for their transfusion dependency.

2.6.1 Principle

The Illumina WES workflow of NovaSeq 6000 Sequencing System consists of four basic steps, sample library preparation, cluster generation, sequencing, and data analysis. Library preparation allows for sample identification and adherence to flow cell surface of the platform. In library preparation, bead-based 'tagmentation' method is used for DNA fragmentation, where the engineered magnetic bead-linked transposase (BLT) enzyme complex is pre-loaded with two adapter sequences. BLT performs several processes, which include DNA fragmentation, adapters insertion, and its saturation plays a critical role in DNA normalization since each BLT can be used only once (Figure 24).

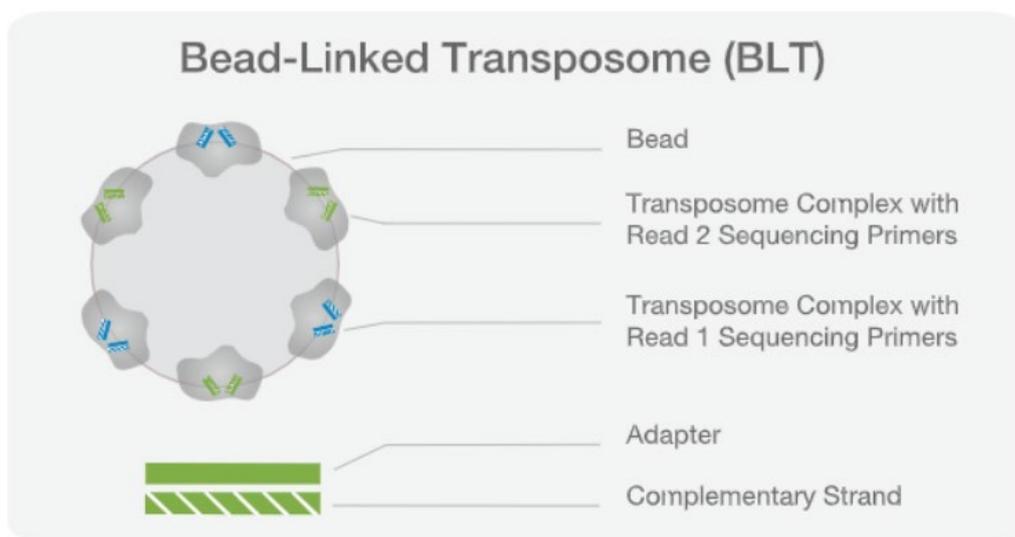


Figure 24: Bead Linked Transposome.

Transposomes with white-striped green and blue read sequences are complementary to the white-striped complementary strand. Source: Adapted from © 2017 Illumina, Inc. (Techniques/ Sequencing/ Library preparation: Tagmentation).

The first step in library preparation is the binding of unfragmented DNA to the BLT. After binding, BLT automatically generates DNA fragments that are mainly within a specified size range and inserts double stranded adapter sequences simultaneously to the dsDNA, which are used for DNA amplification by PCR later. Then, additional motifs are introduced through reduced cycle amplification that are designed to interact with the surface of Illumina’s NovaSeq flow cell, such as 10 base-pair indices at both ends of target fragments (index 1 (P7) and index 2 (P5)) with sequences complementary to the flow cell oligonucleotides required for cluster generation (two different sequences on each end to complement two types of oligonucleotides on the cell flow surface), and sequences for primer binding at each end for paired end sequencing (Figure 25).

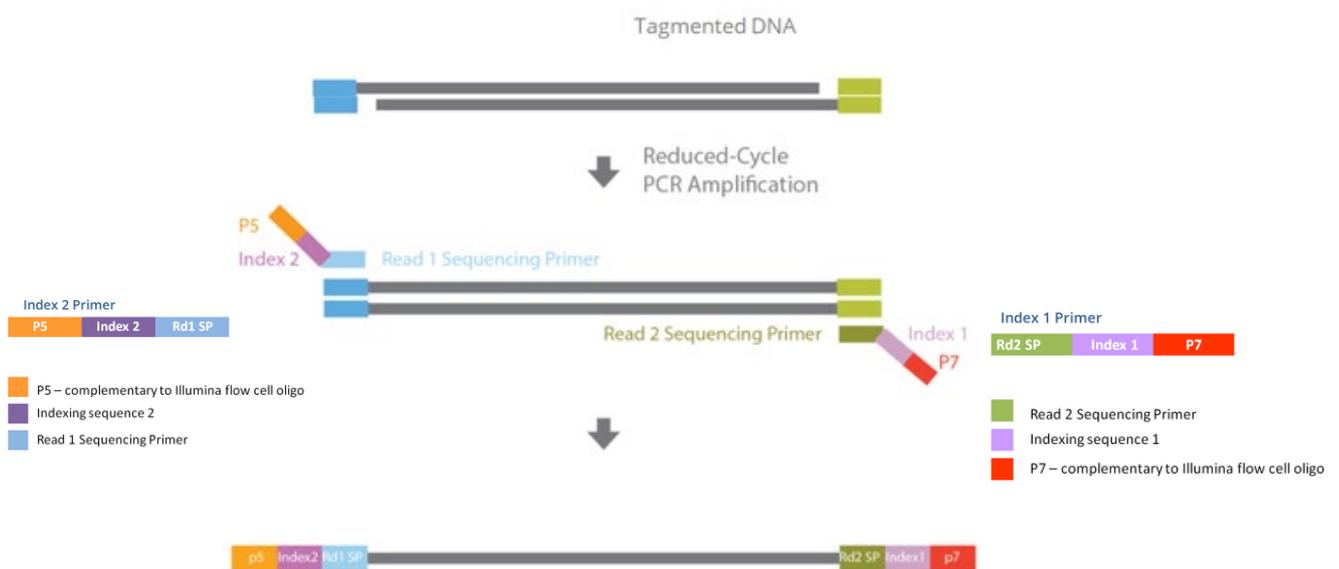


Figure 25: Sequencing ready fragment with annotated Index primers.

P5 and P7 binding regions are for clustering. Rd1 and Rd2 binding regions are for sequencing. Index 1 and Index 2 are for sample identification when different samples are pooled together.

Source: adapted from the Nextera XT DNA Library Prep Kit Reference Guide (© 2017 Illumina, Inc.)

Cluster generation occurred on the surface of a hydrogel-coated glass flow cell with lanes where each DNA fragment is amplified isothermally. Each lane on the flow cell is supplied pre-coated with two types of oligonucleotides. Each type of oligonucleotide is complementary to the 5’-sequence of one of the library adapters sequences. They hybridize and a complementary strand of the hybridized fragment is created by a polymerase then denatured and the original strand is washed away. The remaining complementary strand then folds over to hybridize with the other oligonucleotide on the cell at its adapter region on the other end of the fragment forming a bridge. A complementary strand

is generated by the polymerase forming a double stranded bridge which will denature as well, resulting in two single stranded molecules attached to the flow cell (reverse and forward strands). The process is isothermal, all fragments are clonally amplified by repeating this process of bridge amplification for millions of clusters simultaneously over and over. For this to be possible, the polymerase needed to be of the strand-displacing sort, producing (initially) two single strands. Each cluster starts with one DNA fragment and generates multiple copies by PCR. The density of loading of target DNA is critical as overloading gives overlap between clusters and a steep decline in the usable data.

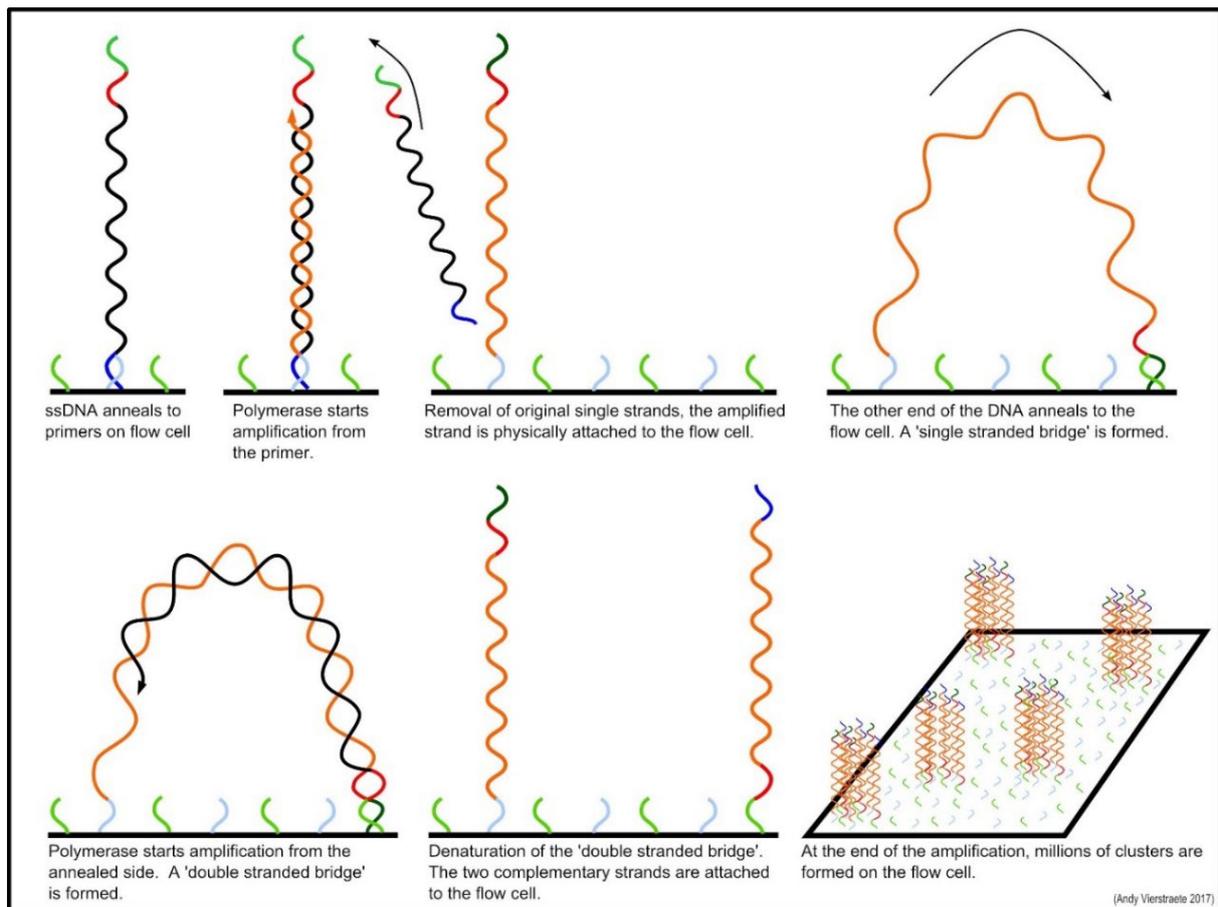


Figure 26: Principle of bridge amplification.

It starts with DNA fragmentation and addition of adaptors to both ends of the strand. Single stranded DNA fragments are added to the glass flow cell and hybridized to the oligonucleotide attached on the flow cell. A primer binds to the strand and DNA polymerase performs PCR extension. The double-strand DNA is denatured, and the original strand is washed away and the remaining strand folds over to hybridize using its other adaptor to a corresponding oligonucleotide on the flow cell forming a bridge. The bridge is denatured giving two single-stranded DNAs each attached to a different oligonucleotide on the flow cell with their different adaptors. The process is repeated until a cluster of strands is formed, then the sequencing process begins. Figure adapted with permission from Andy Vierstraete 2017.

As described in the above (Figure 26), for the first phase of sequencing, both strands are attached to the well at their 5'ends (they are complementary) but there are proprietary 'enzymes' that can distinguish between the two primers and can cleave one, removing that strand without removing the other. This means that at the start of sequencing each cluster, only one strand is present. All 3'-OH group is blocked so that the only possible site for incorporation of the fluorescently tagged nucleotide (next) will be the 3'-end of the read strand (preventing unwanted priming). Once the first strand is read, it is removed by denaturation and washing, and the complementary strand can be recovered from the first strand because the stub of the fragmented oligonucleotide is sufficient to act as a primer again for bridge amplification and the first template can be destroyed. The purpose of reading both strands, in paired end sequencing, is that it confirms the forward read and it extends the data obtained from the forward read, allowing longer overlaps with other reads (paired-end sequencing). The second strand read is generally of lower quality than the first. In the presence of a bound polymerase, analogues of all four nucleotides are added.

The growing nucleotides are identifiable by their fluorescence spectrum (only two colors that are either present or absent - G is not labelled). The nucleotides are also reversibly 3'-blocked and the fluorescent label is also removable, so that after detection of the nucleotide attached to a cluster, the fluorophore and the block are removed chemically. A laser detector (a confocal microscope with epifluorescent detection) excites the clusters after each nucleotide addition and emits a fluorescent signal (sequence by synthesis). The base is called depending on the wavelength emission and intensity. The read occurs simultaneously for all the strands within a cluster. Clusters are imaged by a camera that detects colored lights using bidirectional scanning and two-channel sequencing chemistry. Based on the colored signal ratios for each cluster, base calling is performed and repeated for each sequencing cycle and stored by the Real-Time Analysis (RTA) software on the instrument.

Forward and reverse reads are paired during data analysis, creating contiguous sequences (Mardis, 2013) (Pettersson et al., 2009). The optical system detects the position of each read, so forward and reverse reads can be put back together during data analysis. Sequence alignment with reference genome, variant calling, and annotation can be achieved using commercial or free bioinformatics analysis tools.

2.6.2 Procedure

Equipment and reagents used throughout the procedure are manufactured by Illumina. Illumina® DNA Prep with Enrichment, (S) Tagmentation (Cat. No. 20025524), Illumina Exome Panel – Enrichment Oligos Only (Cat. No. 20020183), IDT® for Illumina® DNA/RNA UD Indexes Set A, Tagmentation (96 Indexes, 96 Samples) (Cat. No. 20027213), and NovaSeq 6000 SP Reagent Kit v1.5 (200 cycles) (Cat. No. 20040719), and SP as a flow cell. DNA input is 10-1000 ng with insert size of 150-220 bp. Panel size is 45 Mb, probe size is 80 bp, and fragment length median is about 200 bp.

2.6.2.1 Preparation of DNA libraries for Whole Exome Sequencing with Illumina Technology

DNA normalization and quantification

We first measured the gDNA sample purity using NanoDrop UV spectrophotometry (ratio of 260/280) was in the range of 1.8-2 and 260/230 in the range of 2.0-2.2). Then, we measured its concentration by Qubit 2.0. gDNA was then normalized to 10 ng/μl in 10 mM Tris/HCl pH 8.5.

All samples were measured again by Qubit and then diluted to a final volume of 10 μl at 5 ng/μl (50 ng total).

DNA tagmentation

In this step, DNA is fragmented (this fragmentation helps create normalized libraries with a tight fragment size distribution using a variety of DNA input ranges) and tagged with adapter sequences in one step using Enrichment bead-linked transposomes. In a PCR plate, the following items were added in order: Tagment DNA Buffer (25 μl), Normalized gDNA is 50 ng in 10 μl (5 ng/ μl), and Tagment DNA Enzyme (15 μl) giving a total volume of 50 μl in each well. These items were mixed thoroughly using High-Speed Microplate Shaker by Illumina at 1600 rpm for 1 minute. Then, the plate was centrifuged at 280x g for 1 minute. The plate was then placed in the thermal cycler while the preheat lid option was set to 100°C (to inhibit sample evaporation to the lid and allows efficient heating of the sample) with the parameters set at 55°C for 5 min with a final hold at 10°C. After 10°C was reached (when transposome is still active), the plate was let to stand 2 min at room temperature before Stop Tagment Buffer (10 μl) was added to each well, and resuspended using micropipette, and incubated for 5 min at room temperature. Tagmented DNA was purified with magnetic Sample Purification Beads (SPB) from the transposomes to prevent them from binding to DNA ends and interfering with downstream processes, from unviable fragments (fragments with only one adapter, too long fragments, or too short fragments), and from adaptor dimers (ligation of adaptors without target insert) which can bind to the

flow cell and occupy valuable spaces without generating any useful data. The plate was placed on a magnetic stand until the liquid was clear then it was removed and discarded. Tagment washing buffer (TWB) was added to the beads slowly (to avoid foaming of TWB which may lead to incomplete mixing or aspiration of incorrect volumes). After resuspending the beads by pipetting, the plate was placed on the magnetic stand again until the liquid was cleared, and the wash was discarded. The washing step was repeated twice.

Cleaned up tagmented DNA was amplified by adding extension ligation mix (ELM) and index adapters (Index 1 (i7) adapters and Index 2 (i5) adapters) with a different i7 and i5 combination for each sample. Additionally, while using the adapters as primers, the PCR step added the sequences required for cluster generation using a limited-cycle PCR program which was set for 3 min at 72°C, then 3 min at 98°C, then 9 cycles of 98°C for 20 s, 60°C for 30 s, and 72°C for 3 min. Amplified DNA was purified by selecting the abundant insert size, and unwanted products were washed away using Illumina purification beads (double-sided beads for selecting inputs of ≥ 100 ng) by repeating the same steps mentioned for cleaning tagmented DNA. After the second wash, supernatant was discarded and while on the magnetic stand, 80% of freshly prepared ethanol was used to wash the beads two times. Purified library was eluted using resuspension buffer (RSB). The DNA library was quantified using Qubit 2.0 high sensitive dsDNA kit.

Quality control

Quantification of tagmented samples before pooling for enrichment is essential to check that our DNA is at useful length, and to know the concentration of it to maximize sequence output. For a quality check, libraries were run on an Advanced Fragment Analyzer Automated CE System by Advanced Analytical Technologies with the DNF-930 dsDNA kit 75 bp – 20,000 bp manufactured by Agilent (expected DNA fragments range from 200-500 bp). This system utilizes automated capillary electrophoresis. Each DNA library was diluted 1:10 with nuclease free water and run on the Analyzer (Figure 27). Reagents used in this step include intercalating dye, capillary storage solution, 5x capillary conditioning solution, 5x dsDNA inlet buffer, 0.25x TE rinse buffer, high sensitivity genomic DNA diluent marker, high sensitivity genomic DNA ladder, and blank solution. After all reagents are loaded in the instrument, screen instructions were followed to start the run, and the results were looked at by ProSize 2 software.

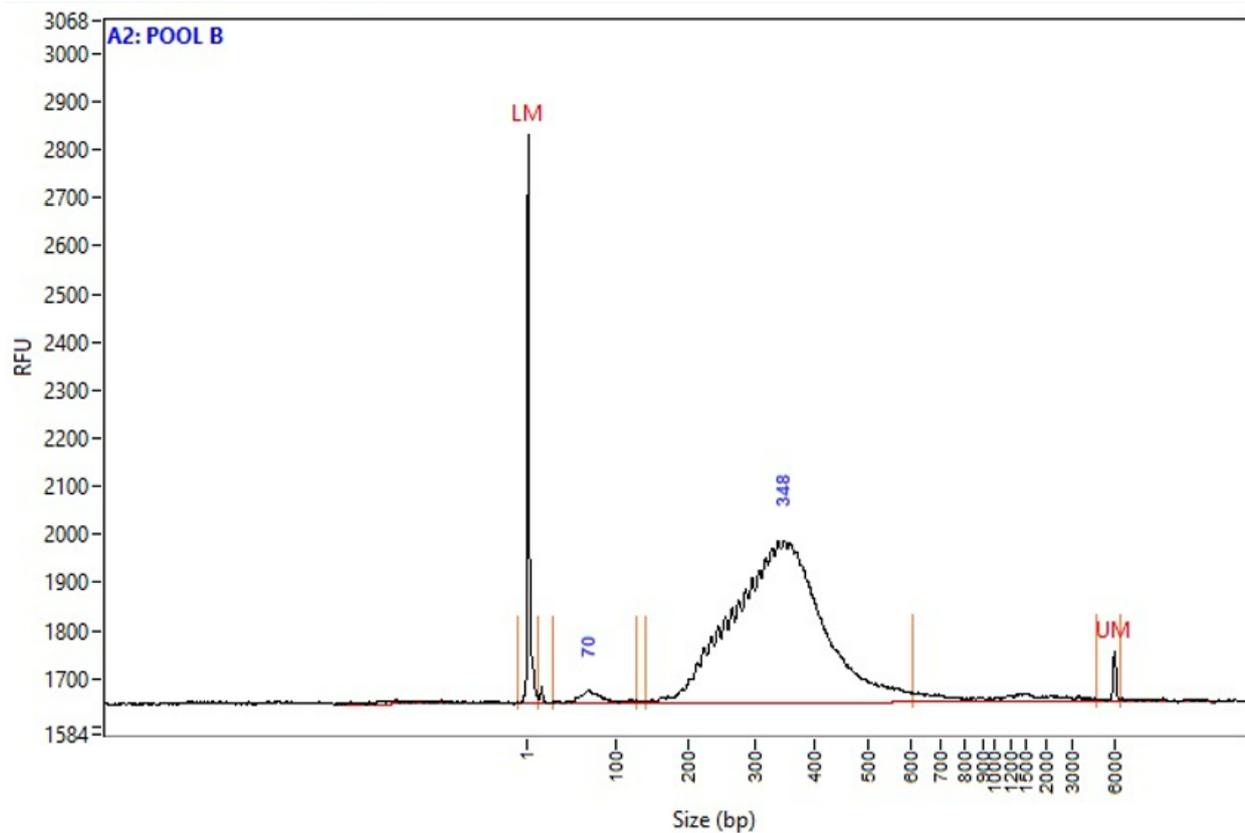


Figure 27: Electropherogram interpretation graph. Example of using Advanced Analytical Fragment Analyzer for DNA libraries quality check and proper sizing of samples. The lower marker (LM) is at 1 bp and upper marker (UM) is at 6000 bp. Good quality DNA fragments would range between 200-400 base pairs. RFU, relative fluorescence unit.

A concentration of 500 ng of each DNA library was pooled together in one new microcentrifuge tube (each has a unique Index identifier). The yield obtained from each library for pooling was quantified by calculating mass/ DNA library concentration.

2.6.2.2 Selection of whole exome DNA

Denatured double-stranded DNA libraries were hybridized to biotinylated oligonucleotide probes. Capture probes were used to hybridized with DNA target sequences by adding the following reagents in order: DNA pool (30 μ l), Blocker (synthetic oligonucleotide sequences, 50 μ l) to prevent target enrichment probes and adapter sequences from non-specific hybridization, and Enrichment Probe Panel (10 μ l), then Enrichment Hybridization Buffer (10 μ l) was added, resuspended, and incubated for 10 minutes at room temperature then centrifuged and placed on the thermal cycler for 5 min at 95°C, then 16 cycles/ 1 min each starting at 94°C with 2°C decrease with each cycle afterwards, then hold 30 min at 58°C.

The enrichment probes are biotin-labelled (340,427 probes with the size of 95mer). Each probe targets libraries of 300-350 bp. They are captured by adding Streptavidin Magnetic Beads (SMB) and incubated for 15 minutes on heat block at 62°C. Tube was then centrifuged and placed on a magnetic stand until liquid was clear, and supernatant was removed. Then, non-specific binding was removed by performing heated wash. Enhanced Enrichment Wash Solution (EEW) was added to the library pool, and the tube was incubated on the 62°C heat block for 5 minutes using Hybex microsample Incubator. The tube was immediately returned to the magnetic stand until liquid was clear and supernatant was removed. The washing step was repeated twice. The enriched library was eluted from by adding preheated EEW, then transfer to a new tube. The tube was incubated on the 62°C heat block for 5 minutes then placed on a magnetic stand until liquid was clear. Supernatant was then discarded. After removing the tube from magnetic stand, Elution mix was then added to the library and incubated for 2 min at room temperature then placed on magnetic stand until liquid was cleared. Supernatant was transferred to a new microcentrifuge tube and Elute target buffer was added to the library.

2.6.2.3 ***Amplification of enriched library***

PCR primer Cocktail and enhanced PCR mix were added to the enriched library tube. The tube (50 µl total volume) was placed on the preprogrammed thermal cycler that was set at 98°C for 30 s, then 10 cycles of 98°C for 10 s, 60°C for 30 s, and 72°C for 30 s and it finished at 72°C for 5 min. The 50 µl of amplified library was cleaned up using AMPure XP Beads, washed using 80% ethanol, then eluted with resuspension buffer (RSB).

The post-enriched library was quantified by using Qubit 2.0 to ensure optimum cluster densities on the flow cell (yield of ≥ 3 ng/ µl). The following formula was used to calculate the molarity value of the library and convert ng/ µL to nM:

$$\frac{(\text{Concentration in ng/ } \mu\text{L})}{(660 \frac{\text{g}}{\text{mol}} \times \text{average library size (bp)})} \times 10^6 = \text{concentration in nM}$$

Then, the DNA library was diluted to the starting concentration of 2 nM using resuspension buffer (RSB). DNA Library pool is denatured and diluted to a final concentration of 1.25 nM using High Output Reagent kit v2 by Illumina with maximum output of 60 Gb (150-cycle), and maximum 400 million clusters read per run.

Below, is a table chart used in each run to track the workflow steps (Table 6)

Table 6: Workflow chart for WES run.

Samples highlighted in yellow are our three transfusion dependent samples.

MiSeqV3/ NextSeq Flowcell Lot# Reagent Cartridge V2 Lot#	20368730		20367124			20367893		
Sample ID	1097	1256	1300	1287	148	123	15-46	183
Library prep start date:	30-12-19							
1ST DNA Conc. (ng/ul) by Qubit	49	75	61.2	60.8	84.2	92.4	41.3	104.7
Index i5	505	505	506	506	517	517	506	5017
Index i7	705	711	711	714	714	706	706	711
After Amplification of tagmented DNA: Libraries Conc. (ng/ul) by Qubit	77.1	57.3	81.4	90	53.1	51.1	51.2	39.3
Volume used in Pool libraries (ng/ul)	6.5	8.7	6.1	5.5	9.4	9.7	9.7	12.7
Total DNA Library Volume	68.3							
Library concentration After amplification of enriched library Qubit (ng/ul)	21.2				Library Average Size by Fragment Analysis:		350	
Library con. In nM Note: You can use the diluted one used for Analyzer (Make sure that it is already measured by Qubit)	<p>91.77 nM</p> <p>Volume to dilute the library to 4nM= $\frac{4\text{nM} \times 15 \text{ (final volume)}}{\text{Library conc.}}$ = ul</p> <p>$4 \times 15 / 91.77 = 0.65$ (from library) + 14.35 (NFW)</p>							
Run Date	2-1-2020				Run Status: pass			

2.6.2.4 Normalization of library for pooling

The recommended loading concentration for our flow cell type (SP) using PhiX as library type with insert size less than 450 bp is 250 pM with pooled loading concentration of 1.25 nM. Library (2 nM) was normalized to 1.25 nM using 10 mM Tris-HCl, pH 8.5.

2.6.2.5 Dilution of PhiX control to 2.5 nM

PhiX Control which is an adapter-ligated library used as a sequencing control and quality control for cluster generation was diluted using 10 mM Tris-HCl, pH 8.5 from 10 nM to 2.5 nM. The diluted PhiX control was added to the normalized library pool.

2.6.2.6 Denaturation of combined library with PhiX

Libraries denaturation for sequencing was prepared using freshly prepared 0.2 N NaOH which was added to the library with PhiX, then centrifuged at $280 \times g$ for 1 minute. Within 30 minutes, the denatured library was loaded to the library tube into the cluster cartridge then onto the NovaSeq 6000 instrument.

2.6.2.7 Library hybridization to flow cell and sequencing run

Setup steps using the NovaSeq Control Software (NVCS) were initiated and BaseSpace Sequence Hub setting was used. The prefilled reagent cartridge was prepared for use and libraries were loaded into reservoir labeled Load Library for sequencing. ExAmp master mix is added to the library before cluster generation onboard the instrument.

2.6.3 Analysis details

The reference genome used was (UCSC hg19). Base call files were automatically transferred to BaseSpace Sequence Hub by NovaSeq control software (NVCS) for data analysis.

When sequencing is completed, the run folder is copied from the NovaSeq 6000 machine to the University high performance computer (HPC). The in-house GenaTi NGS analysis pipeline reads the run folder. The main steps in the workflow of variant calling are as follows. The run folder holds the base calls in BCL files that require conversion to FASTQ format for use with user-developed or downstream analysis tools. This process is called BCL to FASTQ conversion or demultiplexing, the first step in the NGS analysis pipeline. In this process, the .BCL files are converted to the universally used FASTQ format using a bcl2fastq tool made by Illumina. This tool will assign a name to each read, which includes its location on the flow cell as well as the index specified, and the base calls that constitute each read with their accompanying quality scores.

FASTQ format is a text-based format for storing both a biological sequence (usually nucleotide sequence) and its corresponding quality scores. Both the base call and quality score are each encoded with a single ASCII character for brevity. For a single-read run, one FASTQ file is created for each sample per flow cell lane. For a paired-end run, two FASTQ files are created for each sample for each lane.

In order to achieve high-quality and high-confidence variations in downstream data analysis, quality control and pre-processing of FASTQ files are essential. Adapter contamination, biases in the data's base content, and overrepresented sequences are all potential issues. Inaccuracies in representing original nucleic acid sequences are inevitable throughout library preparation and sequencing stages.

The "FASTQ Quality Check" (FASTQC) program is a Java-based quality control check on sequence data that includes per-base and per-read quality profiling capabilities.

The pre-processing of the FASTQ files in this study removed adapter sequences "CTGTCTCTTATACACATCT", base calls with PHRED scores <20 and reads of <20 bp.

The FASTQ files are mapped to the reference genome. The Burrows-Wheeler Aligner (BWA) mapping (version 0.7.12) application was used to map reads to the human genome reference UCSC hg19 GRCh37 and to generate a technology-independent SAM file format. This SAM file generated is converted to BAM which is the compressed binary version of a SAM file.

Next, we sort and merge the BAM files generated from different lanes of the same sample into one BAM file using Samtools (version 1.2), and the duplicated fragments are marked and eliminated with Picard (version 2.2.1). The genome sequence files for *Homo sapiens* UCSC hg19 were downloaded from the UCSC Genome Browser website.

The Genome Analysis Tool Kit GATK-HaplotypeCaller was utilized for variant calling. To identify single nucleotide polymorphisms (SNPs) and small indels (ranging from 1 to 100 base pairs) within the target region, HaplotypeCaller is employed.

For filtration of variants, the in-house developed pipeline generates the VCF (variant calling file) with around 40,000 variants as expected in the WES data. The VCF file is a text file and can be converted into a MS Excel file that represents various variant calls such as SNPS, indels, and structural variants. Each line in the VCF file represents a single variant, with columns representing various features of that variant such as chromosomal number and position, reference allele, alternate allele, consequence, gene name. Following VCF generation, ANNOVAR tool is used for annotating the variants with allele frequency from different databases. It uses the Filter-based annotation to identify known rare dbSNP variants and their reported frequencies, the allele frequency in the 1000 Genome Project, Exome Aggregation Consortium (ExAC) or Genome Aggregation Database (gnomAD), calculate the SIFT/PolyPhen and other information (Wang et al., 2010) (Karczewski et al., 2017) (Karczewski et al., 2020) (Auton et al., 2015).

Then, we removed the "not pass" variations and ranked the variants in the annotated VCF files. The GATK hard filtering and the QC data given by the lab assay validation standard operating procedure determine the selection of these filtration criteria. Genomic Analysis Toolkit (GATK) is a tool for analyzing genomic data and finding variations in germline DNA and RNA.

Software versions

Bcl2fastq v2-20, which is a conversion software supported by Illumina to convert bcl files to FASTQ files. BWA-mem (aligner) v-0.7.12, Burrows Wheeler Aligner, is a program that uses algorithms to perform alignment and mapping of relatively similar sequences to a large reference genome.

Samtools v1.2, Sequence Alignment/Map, is a software tool with a collection of functions used to store nucleotide sequence alignment in a flexible format (SAM, BAM, or CRAM) that can be manipulated by sorting, combining, indexing, and swift reads retrieval.

GATK-HaplotypeCaller v-4.1.4, which utilizes local de-novo assembly of haplotypes to do simultaneous SNP and indel calling.

2.7 Sanger sequencing

Sanger sequencing was used to validate for identified TaqMan assay, NGS, and WES variants. Sanger sequencing was developed to read the sequence of individually amplified DNA samples. The chain termination method of sequencing is commonly called Sanger sequencing depends on using dideoxynucleotide triphosphate (ddNTPs) or any deoxynucleotide analogue that does not contain a 3'-OH. The process requires a template, a DNA polymerase, normal NTPs and labelled (irreversible) terminators. The base is chemically altered by adding a fluorescent adduct so it can be detected. The 2'-deoxyribose sugar is modified by substitution the 3'-OH with -H to produce a 2',3'-dideoxynucleotide. Incorporation of a 2',3'-dideoxynucleotide terminates DNA strand synthesis (Sanger et al., 1977). DNA can be accurately and sensitively separated by capillary electrophoresis using a viscous polyacrylamide matrix. In vertical slab gel electrophoresis, the gel remains static. The medium flows so that it is constantly refreshed to separately identify each ddNTP terminated DNA. In a reaction tube, Taq polymerase was added with primers, target DNA to be sequenced along with dNTPs and small amounts of ddNTPs chain terminators (each of which is fluorescently labeled with a different dye). The dNTPs: ddNTPs ratio is around 300:1. During the target amplification reaction, when a ddNTP is cognated to the strand, it will be fluorescently labeled at the 3' end and it will block any further dNTP extension and terminate the process. This allows the production of multiple terminated products from the same template (Sanger et al., 1977). Sanger sequencing uses an excess of template. Each copy should be terminated at some point and since we have so many copies, we can read the sequence. However, the efficiency of the termination reaction varies, so the peaks are uneven. These ddNTPs are each labelled with a different fluorescent label and therefore has a different emission spectrum, which is what is

read by the detector on excitation with a laser when they reach the end of the capillary. A detector will detect the dye emission color and computer software will translate it into digital readable data (Smith et al., 1986). The workflow of Sanger sequencing after an exponential PCR amplification of a genomic template, consists of three major steps that includes cycle sequencing, purification of cycle sequencing products, and drying and denaturing. (Figure 28)

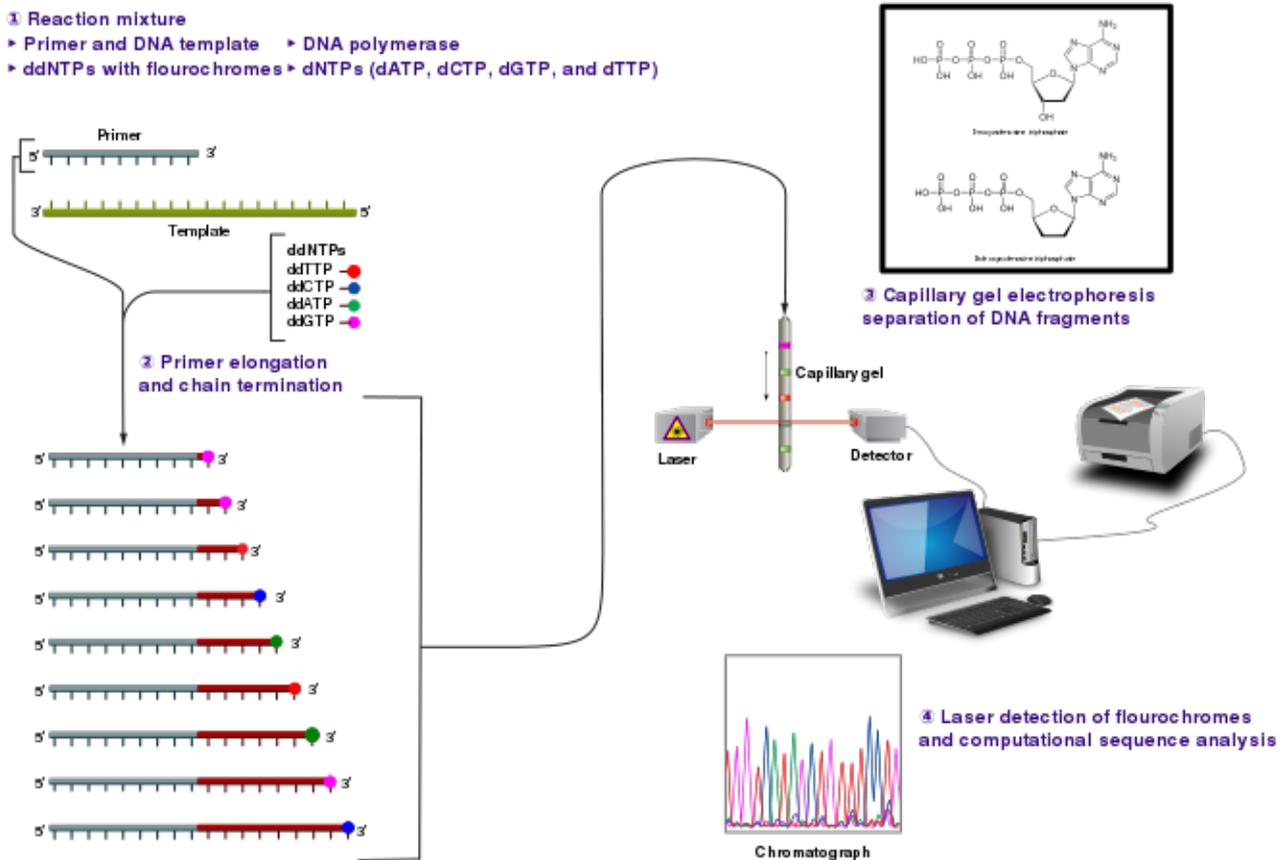


Figure 28: Principle of Sanger sequencing.

Target DNA is amplified and attached to a primer. With added dNTPs and ddNTPs, the complementary chain will grow until it encounters a termination dye. Then, the grown chain is denatured and electrophoresed. Afterwards, fluorochromes are detected by laser detection and a computational sequence analysis is performed.

Source: From Wikimedia Commons, the free media repository. This figure is licensed under the Creative Commons Attribution-Share Alike 3.0 license. creativecommons.org

2.7.1 Primers

I have designed the primers using a primer designing tool, Primer-Blast, and ordered them from Microgen.

HBB gene

Custom oligonucleotide primers were designed to cover the 5'UTR exon 1, intron 1, exon 2, intron 2, and exon 3 of the HBB gene. (Figure 29)(Table 7)

>NC_000011.10:c5227271-5225264 Homo sapiens chromosome 11, GRCh38.p14 Primary Assembly

ATATATCTTAGAGGGAGGGCTGAGGGTTGAAGTCCAACCTCCTAAGCCAGTGCAGAAAGAGCCAAGGACAGGTA
 CGGCTGTATCCTTAGACCTCACCTGTGGAGCCACACCTAGGGTTGGCCAACTACTCCCAGGAGCAGGGAG
 GGCAGGAGCCAGGGCTGGGCATAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAAGTGTG
 TTCCTAGCAACCTCAAACAGACACCATGTGCATCTGACTCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGG
 GCAAGGTGAACGTGGATGAAGTTGGTGGTAGGGCCCTGGGCAGTTGGTATCAAGGTTACAAGACAGGTTTAA
 GGAGACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTGGGTTTCTGATAGGCACTGACTCTCTGCC
 TATTGGTCTATTTCCACCCTTAGGCTGCTGGTGGTCTACCCTTGACCCAGAGGTTCTTTGAGTCTTTGGGGAT
 CTGTCCACTCTGATGCTGTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTCTGGTGCCTTTAGT
 GATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACA CTGAGTGAGCTGCACTGTGA CAAGCTGCA
 CGTGGATCCTGAGAACCTCAGGCTGAGTCTATGGGACGCTTGA TGTTCCTTTCCCTTCTTTCTATGGTTAAGT
 CATGTCATAGGAAGGGGATAAGTAACAGGGTACAGTTTGAATGGGAAACAGACGAATGATGTCATCAGTGTGG
 AAGTCTCAGGATCGTTTTAGTTCTTTATTTGCTGTTACATAACAATTGTTTTCTTTGTTAATTCTGTCTTTCTTTT
 TTTCTCTCCGCAATTTTACTATTACTTAATGCCTTAACATTGTGTATAACAAAAGGAAATATCTCTGAGATAC
 ATTAAGTAACCTAAAAAATTTTACACAGTCTGCCTAGTACA TFACTATTTGGAATATATGTGTGCTTATTTGCA
 ATTCATAATCTCCCTACTTTATTTCTTTTATTTTAAATTGATACATAATCATTATACATATTTATGGGTTAAAGTGA
 ATGTTTTAATATGTGTACACATATTGACCAATCAGGGTAATTTGCATTTGTAATTTAAAAAATGCTTTCTCTTT
 AATATACTTTTTGTTTATCTTATTCTAATCTTCCCTAATCTTTCTTTGAGGGCAATAATGATACAAATGATCAT
 GCCTCTTGCACCATTTAAAGAATAACAGTGATAATTTCTGGTTAAGGCAATAGCAATA TCTCTGCATATAAATA
 TTCTGCATATAAATTGTAAGTGTGTAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATTTGCT
 TTATTTATGTTGGGATAAAGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTGCTAATCATGTTACACCTCTTA
 TCTTCTCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGCCCCATCACTTTGGCAAGAATTCAACCCACC
 AGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCACAAGTATCAC TAA GCTCGCT
 TTCTGTGCTCAATTTCTATTAAGGTTCTTTGTTCCCTAAGTCCAACCTACTAACTGGGGGATATTATGAAGGG
 CCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTTTCATTGCAATGATGTATTTAAATTTTCTGAATATT
 TTAATAAAAGGGAATGTGGGAGTCAAGTGCATTTAAACATAAAGAAATGAAGAGCTAGTTCAAACCTGGGAA
 AATACACTATCTTAACTCCATGAAAGAAGGTGAGGCTGCAACA GCTAATGCACATTGGCAACAGCCCTGAT
 GCATATGCCTTATTCATCCCT

Figure 29: Primers' coverage of Sanger sequencing of HBB. Bases in red are untranslated regions (UTR). Highlighted bases in green are the starting codon of the gene. Bases written in bold green are exons 1,2, and 3 respectively. Bases highlighted in yellow are sequences for forward primers. Bases highlighted in blue are sequences for reverse primers. Bases highlighted in red is the stop codon of the gene. Bases written in black between the exons are the introns.

Table 7: HBB forward and reverse sequencing primers.

Primer	Sequence
Forward 1	CCAGAAGAGCCAAGGACAGG
Forward 2.1	CTGCTGGTGGTCTACCCTTG
Forward 2.2	CTGAGTGAGCTGCACTGTGA
Forward 3	CTTTACACAGTCTGCCTAGTACA
Reverse 1	TCAAGCGTCCCATAGACTCAC
Reverse 2	TATTGCTATTGCCTTAACCC
Reverse 3	TGTTTGCAGCCTCACCTTCT

PKLR gene (c.1015G>A) (Table 8)(Figure 30)

Table 8: Forward and reverse primers for PKLR gene.

Primer	Sequence
Forward	CTCAAGGCCTCACTCCAGAC
Reverse	CTGTCGCTATTCCCATCAC

>ref|NC_000001.10|:155263849-155264310 Homo sapiens chromosome 1, GRCh37.p13
Primary Assembly

AAGAGGCGGAGGTTGCAGTGAGCGGAGATTGCGCCACTGCACTCCAGCCTGGATGACAGAGTGAGACTCC
GTCTCAAAAAACAAAACAAAATAAACCCCTACAGTGTGGGTATTCACCCACAGGTGTCCCTAAAA
CCCACAGAGTGCCGAACCTCAAGGCCTCACTCCAGACCTGTGTGGCACAGACAACAGGCTTGCCCGCCAA
GTTGCAGCGCCCAATCATCATCTTCTGAGCCAGGAAAACCTTCTCTGCTGGGATCTCGATGCCTAGGTC
CCCCGTGCCACCATGATGCCGTGCTCACCTCCAGGATTTTCATCAAACCTGAGAGGTTGGGAGAATCAAG
GCAGAGGCAGGCAGGAGAAGAAAGTTCATCCCAATACCCCAAGGCCGAAGGGGAACAGAGCCCAAGCC
TCACCTCTTCACGCCTTCGTGGTTCTCAATTTTGCTGATGAT

Figure 30: Sanger sequencing primer design for confirmation of PKLR c.1015G>A homozygous mutation in sample 183.

FANCC gene (c.1015G>A) (Figure 31)(Table 9)

>NC_000009.11:97911800-97912800 Homo sapiens chromosome 9, GRCh37.p13
Primary Assembly

GCCCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAATCTCCGCCTCCAGGTGGAAGCGATCCTCCTGA
CTCAGCCCCCCTAGTAGCTGGGATTAGAGGCACACACCACCATGCTTGACTAATTTTTGCATTTTTAGTAGAGAC
AGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACCTTGACCTCAGGTGATCCACCCGCCTTGCCCTCCCAAAG
TGCTGGGATTACAGGTGTGAGCCACCACACCTGGCCGGGATACTCAGTAATTTTAAAGAGTATAAAGGGTACTGA
GACAAAAACGTTTGGACACTGCTGTGCTACAGTCTTTCCAACACACCACAGCCTTCTAAGAAAAAGGAAAAACGAC
GCAGGATGACAGGAAACATTTGCCACTTACAGCAAATGGCCTCGTTTACAGCCTCAAAGAACTCTGGCTGGAGG
ATTTCTGAGGTTTACGTCATGACAGATGAGGAGAGCCTCCACCAGGGGTCAACATCTGTCAGGGTAATAAGT
GGGACACAAACTCGTGACAGGGACGCCACTCGCTCGGGAGCCATTCTATGGAAGAAATAAGAAATAATCACTCAA
ATCTAAGAGCCATGCATAATTAAGGACATATAAAAACCTTCATGCTAAAAAGGTCAAAGACAGATCACCTGTTTT
GCCTCTAAATAAAGAGAATGACTCTAACACCATCGATGAACACTTCTCTTTTATCCAATCTTTAGGAAGAAATC
TTTTAGCTACCATGCTGTTTCAGTGTTCATTCATGGTCAGGACTTTGTTCTTTATATTACAGGCACTTAATATG
TCCAAAACACAACCTCCTGAATTCCTTCCAGTGTTCCTTTCCCCAGTGCTATCTTTGTTTGTCTCAGGCCCAAAT
CCTGATTCCCTCTCTCAGGCCCTCATCTGATCCATCAGTGAATACTAGCAACACATCTGACCTCTTCTG
AACACCTCCATCACTGTCCTCCTAAG

Figure 31: Sanger sequencing primer design for confirmation of FANCC: c.1015G>A heterozygous mutation.

Table 9: Forward and reverse primers for FANCC gene.

Primer	Sequence
Forward	TTCCAACACACCACAGCCTT
Reverse	GAGCAAACAAGATAGCACTG

SMC5 gene (c.676G>A) (Figure 32)(Table 10)

Fwd region: 1-400

Rev region 602-1001

```
>NC_000009.11:72893039-72894039 Homo sapiens chromosome 9, GRCh37.p13
Primary Assembly
GTAGATGTTTGA AATTGCCTTTTTTTTTTCTTTTTTAAACCCTCTAACATTTCAACTGATATGTAATAGACAGTGA
TCATTTTCATCTTGGGACTATGTGTTCTTTTATTCCCAAACCTTTATTATACATATGTAGACATTTCAACCTTGA
AATATTGGCACAGAAAATTTTAAAGATGATCAAAAAGGTAACAGTTTATATCCTAAGTTTATCTTCATTATTCT
AAAGAATAATGCACAAAAGGTTTATATCTCAAATGTTTTTTTTTAAATGAGTTTACCTTGACAAGTGTAAAGCATTG
AGTGAATAACATTTTGAAGATATATGTAATGAATGATAGTTGCTGATATTTAATTTTTTGTGCTCTAGGACAAAAG
TTGGAGAATTTGCTAAACTCAGCAAAATTGAACTCCTCGAAGCCACTGAAAAGTCAATTGGTCCCCCAGAAATGC
ACAAATATCACTGTGAACTCAAAAACCTTAAGGGAGAAAAGAAAACAGCTCGAGGTACTTTAAATAGACAACTCAT
TTGTATTGTTTCTTATTGATTTCTGTATCTCAGAACATGGGAGAGAGAGTAGTAGTATTGATTCATTGAGTGAAT
TCGTATTTGTGATTTTTTTTTCACTTAGGATTGTAAGTCTAATATCAGTTAAAGTGACAGAGTGAGAATGGATCTA
ATAATGTGAAGTTTAAAAAGGTTGACTATTTTCAGCAAAATACGTATTACTAATTGAATTTTTAAGGAACCAGTAAC
AGTATTATAAAAAAAATATATCTCAGGAAGTTATTCTTTCAATTCATTGAGTATCTATTAAACACCTACTATGTGC
CAGGCAGTGTGTAAGGCTTTGAGTAATAGAAGAGATACCATGTGCCATTTAATATAGCACTTCCGTTGAACCAAT
TCTCTCATGTATTTTTTTTCAGTCACATGAGGGTAGACTTTGGAAGAGAATTACACATTTAGTTGGAATGTTGGTT
ATTTTCACCATTAATAAACATAAAAA
```

Figure 32: Sanger sequencing primer design for confirmation of SMC5 gene (c.676G>A) heterozygous mutation.

Table 10: Forward and reverse primers for SMC5 gene.

Primer	Sequence
Forward	GGACAAAGTTGGAGAATTTGCT
Reverse	CCATTCTCACTCTGTCACTTTAACT

TALDO1 gene (c.604G>A) (Figure 33)(Table 11)

Fwd region: 1-400

Rev region 602-1001

```
>NC_000011.9:762986-763986 Homo sapiens chromosome 11, GRCh37.p13 Primary
Assembly
GGCCTCCCTTGCCTTTGTCTATTGTGTCATGTGGGTTGGAGAAAATCCTGGTCACCCATAACCCACTATCCACTT
AGAGCCCTCGAGTCTTACCGATTTGTTCTTAACAATGGCATGAGCCTGTTCCTGCCTTTTTAGTTTAAAGCCCCT
GAGCGTCTTTGGCAGCCTGTGGCTCTGGGTGGGCACCTGGCCTGCATTCACCTGCCCGCCCTCACCTGCCCGCC
CCTCACCTGCCCGCCCTCACCCGCCCGCCCTCACCCATCCCCGCCCTCACCGTCCCCGCCCTCACCCGCC
CCGCCCTCACCTGTCCCCGCCCTCACCTGCCCGCCCTCACCTGTCCCCGCCCGCAGGGAGCTCGAGGAGCAGC
ACGGCATCCACTGCAACATGACGTTACTCTTCTCCTTCGCCCAGGCTGTGGCCTGTGCCGAGGCGGGTGTGACCC
TCATCTCCCCATTTGTTGGGCGCATCCTTGATTGGCATGTGGCAAACACCGACAAGAAATCCTATGAGCCCCCTGG
AAGACCCTGGTGAGGGTCCCTCTGTGGTAATGGGGTAAGGGGAGCAGCCTCAGCAGCACCTCAGGCAGGAAGAGC
CCGAGACGGAGCTGCCGCATCAACAAGCAGTGAGGTGCACAGGTCGGCGCGGGGCAGGCAGGGGTGGCGGCTCAA
GGTAGTGCAGCCGGCAGGACTGGGAGGGCAGGTGGGGTGGTACCTCTGCCGAAGCCTTCTGGTACAGCTTGG
TCTCTTTCCAGGGGTAAGAGTGTCACTAAAATCTACAATACTACAAGAAGTTTAGCTACAAAACCATTTGTCAT
GGGCGCTCCTTCCGCAACACGGGCGAGATCAAAGCACTGGCCGGCTGTGACTTCTCACCATCTCACCCAAGCT
CCTGGGAGAGCTGCTGCAGGACAACGCCAAGCTGGTGCCTGTGCTCTCAGCCAAGGCGGGTGGAGCCCCACTGCC
CAGCTGTGGCTCCTGCTCGGGCAAGG
```

Figure 33: Sanger sequencing primer design for confirmation of TALDO1 gene (c.604G>A) heterozygous mutation.

Table 11: Forward and reverse primers for TALDO1 gene.

Primer	Sequence
Forward	GGCATCCACTGCAACATGAC
Reverse	CAGGAGCTTGGGTGAGATGG

I performed the whole procedure except loading the samples into the sequencer which was performed by the lab technician. As for the template preparation, DNA was first precipitated with ethanol by 2.5 volume of absolute ethanol, 1:10 volume of sodium acetate 3 M, pH 5.2 and 1:10 volume of EDTA 0.125 M, pH 8 for each sample. Samples were centrifuged at 14,000 rpm (using MiniSpin centrifuge by Eppendorf) for 30 min. The DNA pellet was washed with 100 µl of 70% v/v ethanol. Samples were then dried in open tubes at 65°C for 20 min. Water was then added to each sample (20 µl). Samples were gel electrophoresed for quality check.

2.7.2 Cycle sequencing

BigDye Terminator v3.1 Cycle Sequencing kit by Applied Biosystems was used. For each sample, 2 µl of 5x sequencing buffer, 1 µl of forward primers, 1 µl of BigDye terminator (cat #4336917), 5 µL water was added to PCR wells for each sample, then 1 µl of template was added. An amplification program was set and started as follows: (Table 12)

Table 12: Thermocycling used for amplifying PCR products.

This is a linear amplification done with a single primer. After 25 cycles, optimally, we might have 25 product strands off each template strand. Reaction volume was 10 µl with 2.63 °C/s ramp up temperature.

Stage	Cycle	Temperature	Time
1	1	96°C	5 min
2	25	96°C	45 s
		61°C	5 s
		60°C	4 min
3	1	4 °C	∞

Sequencing products were purified and dried by the same method as template preparation but two extra washes with 70% ethanol were used to remove soluble fluorescent material.

2.7.3 Product Denaturation

The final step before loading the samples in the sequencer (3500 Genetic Analyzer by Thermo Fisher Scientific/Applied Biosystems/Hitachi) was denaturation. After drying the samples on the thermal cycler, 15 µl of formamide was added to each sample (for resuspension after ethanol precipitation)

then transferred to a plate (loaded on plate vertically). The plate was sealed and incubated at 96°C for 5 min then immediately put on ice for 5 min then loaded into the sequencer (Gel POP7, capillary 50, and temperature ready at 60°C).

All steps were repeated for all samples for a second time but using the reverse PCR primers.

2.8 Multiplex ligation-dependent probe amplification (MLPA) of the alpha globin gene cluster

Alpha-thalassemia-causing rearrangements can be identified using MLPA, specifically large deletions or duplications encompassing the entire α cluster and/or deletions or duplications in the HS-40 region of the alpha globin genes. MLPA is a quantitative technique used to detect any copy number variations of at least 60 nucleotides. Here, I have employed MLPA on patients' samples II-17 and 11-14 with only one heterozygous beta thalassemia mutation to rule out co-inheritance of alpha thalassemia or alpha thalassemia triplication as a reason for their severe cases. The quantitative approach is necessary because there are four potentially functional *HBA* genes in a normal patient. MLPA reaction requires 50-250 ng of genomic DNA. It performs a simple PCR reaction amplification using multiple variable sized probes (up to 60 probes, 20-30 nt in length) each targeting a specific DNA sequence to evaluate and detect aberrant copy numbers like deletion or duplication of approximately 50-60 nucleotides in length. These probes target specific stretches of sequences on the target gene. MLPA probe pairs incorporate 5' terminal primer recognition sequences. Ligation of adjacent probes indicates that the template sequence is present, and it can be quantified after PCR amplification of ligated products (amplicons of 64-500 nucleic acid unique sequences).

2.8.1 Principle

Genomic DNA is denatured, then a mixture of single-strand MLPA probes is added to the denatured sample. Each of these probes contains two single-stranded DNA (a left and a right probe oligonucleotides) with PCR primer and DNA hybridization sequences, and a unique length of about 60-80 nt (to ensure identification of amplification products downstream the procedure) which will be ligated using a ligase that binds the left and right probe oligonucleotides (catalyzing the bond formation between the two) after their hybridization to directly adjacent sequences on the target gene.

A stuffer sequence of defined length between the primer's binding site and the sequence of the target, on the right probe oligo, helps with the determination of the probe's total length.

If a mismatch between the probe and its target at the ligation site of even a single nucleotide was found, the high specificity of the ligation enzyme will prevent it from binding and ligating the oligos. After ligation, and during the PCR reaction, exponential amplification of the ligated probes is achieved all at the same time using the same PCR primer pair, forward and reverse (of which the forward PCR primer is fluorescently labeled), complementary to ligated probes resulting in unique various-sized amplicons fragments.

Amplicons are then separated based on their length and visualized using capillary electrophoresis instrument that uses size standard fragments of known length labelled with different fluorescent dyes.

Determination of MLPA amplicons length is achieved by the measured fluorescence which yields a peak pattern of electrophoretogram. This electrophoretogram compares the migration of the MLPA amplicons to the size standard fragments. Each amplicon peak was then linked to the correct MLPA probe that can be quantified during data analysis by measuring each peak's fluorescent signal.

The raw data generated from the instrument was analyzed using a free MLPA analysis software (Coffalyser.Net) developed by MRC-Holland. It performs a comparison between each sample to reference samples calculating a ratio for each probe in each sample which can be visualized by a ratio chart that sorts the probe ratios on genomic location (Stuppia et al., 2012) (Kozlowski et al., 2008).

2.8.2 Procedure

For the MLPA reaction, I have used EK20-FAM SALSA MLPA reagent kit and SALSA MLPA Probemix P140 HBA by MRC Holland. Genomic DNA concentration was quantified using Qubit 2.0, then optimized to 250 ng in total of 5 μ l (using TE buffer). Samples were then placed in thermocycler for 5 min at 98°C for denaturation, then cooled down to 25°C.

I have prepared a hybridization master mix for each sample by mixing 1.5 μ l MLPA buffer + 0.5 μ l probe mix and added to each sample after denaturation then placed in the thermocycler for 1 min at 95°C, then 16-20 hours at 60°C to hybridize to the sample DNA.

After that, ligation of MLPA probes mix was prepared by adding 25 μ l water + 3 μ l Ligase Buffer A (coenzyme NAD) + 3 μ l Ligase Buffer B (Tris-HCl, MgCl₂, non-ionic detergent) + 1 μ l Ligase-65 enzyme

(which is a NAD-dependent ligase enzyme that denatures easily at high temperatures) then was added to each sample while the plate was still in thermocycler from last step, but after adjusting the temperature from 60°C to 54°C. The thermocycler was then programmed to 15 min at 54°C for ligation, then 5 min at 98°C for inactivation of Ligase-65 enzyme, then cooled to 20°C.

Next, polymerase primer master mix was prepared by mixing 7.5 µl water + 2 µl SALSA PCR primer mix (synthetic oligonucleotides with fluorescent dye, dNTPs, Tris-HCl, KCl, EDTA, non-ionic detergent) + 0.5 µl SALSA Polymerase (glycerol, non-ionic detergent, EDTA, DTT, KCl, Tris-HCl, polymerase enzyme) then was added to each MLPA reaction at room temperature. Immediately, the tubes were placed in the thermocycler and the program was set as follows, 35 cycles of 95°C for 30 s, 60°C for 30 s, 72°C for 60 s, then incubation period of 72°C for 20 min.

PCR products were stored at 4°C in a dark box (fluorescent dyes are light sensitive) to continue the next day. To avoid contamination, tubes were not opened in the room with the thermocycler after PCR, and different micropipettes were used. The injection mixture was prepared by mixing 0.7 µl PCR reaction, 0.3 µl ROX size standard, 9 µl formamide and was added to each MLPA reaction. The plate was then sealed and heated for 3 min at 85°C, then cooled down to 4°C for 2 min. Finally, the plate was loaded in the ABI-3500 Genetic Analyzer instrument for fragment separation by capillary electrophoresis, which uses FAM as the primer dye, 50 cm capillaries, run voltage of 15 kV, and an injection mixture of 0.7 µl PCR reaction, 0.3 µl ROX 500 size standard, and 9 µl formamide.

Coffalyser.Net was used for data analysis.

2.9 Affymetrix CytoScan HD array

The CytoScan HD kit (901835) for copy number analysis was used throughout the procedure with total number of about 6.5 M of probes (~25 bp), hg19 genome build, 25 kb minimum resolution for losses, 50 kb minimum resolution for gains, and 25 marker/100 kb (more than 2.4 M copy number markers) of genes covered. I have used and performed this technique on I-2 (the father) for CNV analysis. The CytoScan HD array is designed to measure deletion and duplication of genetic material (CNVs) as well as SNPs, loss or absence of heterozygosity, acquired UPD, and chromosomal aberrations detection using high-density pattern of oligonucleotide probes that are spotted onto a solid surface (DNA microarray) to hybridize complementary sequences of fragmented target genomic DNA. This platform

uses dual-probe design chosen for their minor-allele frequency for copy number and SNPs (25 bp in size) (Scionti et al., 2018).

2.9.1 Protocol

Extracted genomic DNA was quantified using Qubit for its concentration and nanodrop for its purity. Each sample was diluted to 250 ng/μl in a total volume of 5 μl using nuclease-free water.

2.9.1.1 Digestion of genomic DNA into fragments with NspI endonuclease

All work steps are done on ice. Each sample was prepared for digestion by adding 11.55 μl of Chilled Affymetrix Nuclease-Free Water, 2 μl of 10x NspI buffer (1000 mM NaCl, 100 mM Tris-HCl, pH 8.0, 50 mM MgCl₂, 10 mM 2-mercaptoethanol), 0.4 μl 50x BSA (Bovine Serum Albumin to prevent enzyme adhesion to tubes and pipettes), 1 μl NspI digestion endonuclease (a restriction enzyme that recognizes the sequence RCATG[^]Y and cuts best at 37°C generating 3' overhang) to 5 μl of sample and subsequently loaded onto the thermal cycler and run at 37°C for 2 hours, followed by 65°C for 20 min.

2.9.1.2 Ligation (digestion-ligation-amplification)

A master mix of 2 μl T4 DNA ligase (ligates two strands of DNA with either cohesive or blunt ends), 10x T4 DNA ligase buffer (0.2 μl), and 0.75 μl of 50 μM adaptor NspI (single-stranded oligo that recognize the cohesive four bp overhangs used as linker for ligation regardless of fragment size) was prepared and added to each sample, vortexed, and centrifuged then loaded onto the thermal cycler and run at 16°C for 3 hours, followed by 70°C for 20 min, then held at 4°C overnight (stopping point).

2.9.1.3 PCR

PCR: The next day, samples were diluted by adding 75 μl of NFW bringing up the total volume of each sample to 100 μl. Each ligated and diluted sample were transferred with the amount of 5 μl/ PCR tube to a total of 8 tubes for each sample. Next, PCR master mix was prepared with 19.75 μl of Chilled Affymetrix H₂O, 5 μl of 10x Titanium Taq PCR buffer, 10 μl of GC melt reagent, 7 μl of dNTP mixture (2.5 mM each), 2.25 μl of PCR primer, and 1 μl of 50x Titanium Taq PCR polymerase and added to each sample then loaded onto the pre-programmed thermal cycler as follows, one cycle of 94°C for 3 min, followed by 30 cycles of 94°C for 30 s, 60°C for 45 s, and 68°C for 15 s, then 1 cycle of 68°C for 7 min, and hold at 4°C.

Quality control check

PCR reaction was checked by running 2% TBE gel electrophoresis using 3 μ l of PCR product with 2 μ l of 5x RapidRun loading dye. The gel was run at 5 V/cm for 45 min. PCR products were between 150 bp-2000 bp.

2.9.1.4 Purification of PCR products from un-incorporated dNTPs

All 8 aliquots of each sample were pooled and collected in one tube. Purification Beads (720 μ l) was added to each pooled sample and incubated at room temperature (20°C -20°C) for 10 min then centrifuged for 3 min at maximum speed. Tubes were placed on magnetic stand for 3 min and the supernatant was discarded. The bead pellet was washed 2 times with purification wash buffer (1st with 1000 μ l, then with 500 μ l), vortexed for 2 min then centrifuged for 3 min then placed on magnetic stand for 3 min and supernatant of wash buffer was discarded. The tubes were removed from magnetic stand and incubated uncapped at room temperature for 10 min to allow the remaining wash buffer to evaporate (air-dry). Purified samples were then eluted by adding 52 μ l of elution buffer directly onto the beads pellet then placed on magnetic stand for 10 min. Supernatant of eluted sample was collected.

Quantitation of purified samples

Each sample was measured by diluting 1 μ l of sample in 9 μ l of NFW using Qubit (samples should yield > 3 μ g/ μ l).

2.9.1.5 Sample fragmentation

It is performed to reduce the purified PCR product size from 150-2,000 bp to 25-125 bp for optimal hybridization to the (25-mers) array probes. PCR tubes were chilled on ice for 10 min before use. A fragmentation master mix was prepared by mixing 12.38 μ l of Chilled Affymetrix NFW, 15.84 μ l of 10x fragmentation buffer, and 0.58 μ l of DNase I enzyme fragmentation reagent (2.5 U/ μ l).

After mixing, 10 μ l of fragmentation mix was transferred to cooled PCR tube and 45 μ l of purified PCR product (sample) was added to it and loaded onto the thermal cycle at 37°C for 10 min followed by 95°C for 15 min. Fragmentation reaction was checked by using 4% TBE gel electrophoresis (fragment distribution between 25 bp to 125 bp).

2.9.1.6 Labeling

Labeling master mix was prepared by mixing 14 µl of 5x TdT (Terminal Deoxynucleotidyl Transferase) buffer, 2 µl of 30 mM DNA labeling reagent (biotin based), and 3.5 µl of TdT (30U/ µl), then added to each sample, vortexed, and centrifuged for 1 min.

After that, samples were loaded onto the thermal cycler at 37°C for 4 hours followed by 95°C for 15 min.

2.9.1.7 Target hybridization

GeneChip Hybridization Oven 640 by Applied Biosystems was turned on and pre-heated to 50°C and set to 60 rpm an hour before hybridization. The CytoScan HD array which uses 49 format and a 5 µl feature size (array/ sample) was unwrapped (from 4°C) and placed on the bench for 10 min at room temperature. The hybridization master mix was prepared on ice by mixing 165 µl of Hyb buffer part 1, 15 µl of Hyb buffer part 2, 7 µl of Hyb buffer part 3, 1 µl of Hyb buffer part 4, and 2 µl of oligo control reagent 0100. The master mix was then added to each sample and incubated on thermal cycler at 95°C for 10 min for denaturing, then cooled down to 49°C for 1 min. The sample (200 µl) was then injected immediately into the array (with the septa covered by Tough-Spot to prevent evaporation and leakage) and placed into the hybridization oven for 18 hours (within 1 min of injection).

2.9.1.8 Washing, staining, and scanning arrays

GeneChip Fluidics Station 450 instrument by Applied Biosystems was used for washing, staining, and scanning the arrays. A protocol 'CytoScanHD_Array_450' was selected from the AGCC Fluidics Control Panel. For each array, 3 microfuge vials were loaded on the Fluidics instrument with 500 µl of Stain Buffer 1, 500 µl of Stain Buffer 2, and 800 µl of Array Holding Buffer. Washing buffers, A and B, were filled in their designated reservoirs. After array hybridization, the array was removed from the oven and the Tough-Spots were removed from the arrays. The whole amount of 200 µl of the hybridization cocktail was pipetted and discarded using any of the openings on the array. After that, the array was immediately inserted into the designated module of the Fluidics station and the Fluidics protocol was initiated. After completion of washing and staining, the array was prepared for scanning by cleaning off excess fluid from around the septa, then covering both septa with Tough-Spots. The Affymetrix GeneChip Scanner 3000 was turned on 10 min before use. The arrays were loaded onto the AutoLoader of the scanner and the scanning process was initiated, and results were analyzed using Chromosome Analysis Suite (ChAS) software to perform .CEL file analysis.

2.10 Whole genome sequencing

The procedure of "whole-genome sequencing" identifies the DNA sequence of a genome in its whole. DNA shearing, barcoding, sequencing, and data analysis are the basic stages in the whole genome sequencing.

For the whole genome sequencing from human blood samples, DNBSEQ-G400S from MGI was used. Whole genome kit was not available in our laboratory and the samples were sent out. High-fidelity PCR Enzymes and innovative adaptor ligation technology were utilized in the library prep set to greatly boost library yield and conversion rate from 5-400 ng of genomic DNA into a customized library. All of the reagents included in this package had completed rigorous quality control and functional verification processes, ensuring consistency and reliability of performance. A high quality of genomic DNA ($OD_{260}/OD_{280}=1.8 \sim 2.0$, $OD_{260}/OD_{230} > 2.0$) was used for fragmentation for each sample. The protocol of the whole genome sequencing included twelve steps: fragmentation, size selection and cleanup of fragmentation product, end repair, a-tailing, adapter ligation, adapter-ligated DNA cleanup, PCR amplification, clean-up of PCR product, quality control of PCR product, denaturation, single strand circularization, enzymatic digestion, and enzymatic digestion product clean-up.

2.10.1 Procedure

2.10.1.1 Fragmentation

At first, The Frag Buffer II was retrieved from storage and thoroughly mixed by vortexing before being added to fresh 0.2 mL PCR tubes with $\leq 45 \mu\text{L}$ genomic DNA (dilution buffer was added to DNA when volume was less than $45 \mu\text{L}$). After the samples were briefly centrifuged and set on ice, $15 \mu\text{L}$ of the fragmentation mixture ($10 \mu\text{L}$ Frag buffer II + $5 \mu\text{L}$ Frag enzyme II) for each sample was transferred, mixed by pipetting 10 times, and the solution at the bottom of the tubes was collected after centrifugation. Then, it was loaded onto the thermocycler at 30°C for 8 min, 65°C for 15 min, then cooled down to 4°C . After that, to gather the solution at the tube's bottom, the samples were briefly centrifuged.

Quality control

Quantification of the fragmented DNA samples to check that our DNA is at useful length. Samples were run on an Advanced Fragment Analyzer Automated CE System by Advanced Analytical Technologies with the DNF-930 dsDNA kit 75 bp – 20,000 bp manufactured by Agilent (expected DNA fragments range from 200-500 bp). All reagents were loaded in the instrument as instructed in the manual and

the results were looked at by ProSize 2 software. Our fragments size should be 100 bp- 1000 bp peaking at 300 bp -500 bp.

2.10.1.2 Size selection and fragmentation product cleaning

For the size selection, following the manual instructions, 36 μL + 12 μL beads were used to obtain the fragmentation products with a main peak of 330 bp, which is suitable for PE150. The DNA-clean beads were allowed to be warmed to room temperature, followed by mixing by the vortex. Next, the 1.5 mL centrifuge tubes holding 60 μL of fragmentation product received 36 μL of clean DNA beads, and mixed by pipetting 10 times, then incubated for 5 minutes at room temperature.

After centrifugation, the tubes were placed on the Magnetic Separation Rack for 2 to 5 minutes to allow the liquid to become transparent. The supernatant was then transferred to fresh 1.5 mL centrifuge tubes. After that, 12 mL of clean DNA beads and 96 mL of supernatant were added to centrifuge tubes, and the mixture was pipetted ten times and incubated at room temperature for 5 min. The tubes were then centrifuged and placed on a Magnetic Separation Rack for two to five minutes until the liquid became clear, followed by removing of the supernatant. The beads walls were then washed with 200 μL of freshly made 80% ethanol and supernatant was discarded. The washing step was repeated once and all the liquid from the tubes was removed. Following that, the tube was incubated for 5 minutes at room temperature with the lid open to air-dry the beads.

After that, the centrifuge tubes were removed from the Magnetic Separation Rack, 43 μL of TE Buffer was added to elute the DNA and incubated at room temperature for 5 min. Next, the tubes were centrifuged, and were placed onto a Magnetic Separation Rack for 2 to 5 minutes until the liquid became clear. Finally, new 0.2 mL PCR tubes were filled with 41 μL of supernatant. The purified fragmented DNA was quantified using Qubit 2.0 high sensitive dsDNA kit.

2.10.1.3 End repair and A-tailing

TE buffer (40 μL) and 100 ng of purified fragmented DNA were added to new 0.2 mL PCR tubes. Then, the end repair and A-tailing mixture were prepared on ice by adding 7.1 μL of ERAT buffer and 2.9 μL of ERAT enzyme mix. Then, 10 μL of this mixture was added to the fragmented DNA, vortexed, centrifuged, and the solution at the bottom was collected and placed into the thermocycler at 37°C for 30 min then 67 °C for 15 min, followed by centrifugation and fluid at the tubes' bottom were collected.

2.10.1.4 Adapter ligation

The MGIEasy DNA Adapters (diluent) were added to the PCR tubes in a volume of 5 μL , vortexed, centrifuged, and the bottom solution was collected. Following that, the adapter ligation mixture was prepared on ice by adding ligation buffer (23.4 μL) and DNA ligase (1.6 μL), then 25 μL of the mixture was added to the sample, vortexed, centrifuged, and solution at the bottom of the tube was collected and placed on the thermal cycler at 23°C for 30 min. The solution at the bottom was collected after centrifugation and 20 μL of TE buffer was added to make a total volume of 100 μL .

Adapter-ligated DNA cleanup

DNA Clean Beads (50 μL) were added to the adapter-ligated DNA and mixed by pipetting before being incubated at room temperature for five minutes. The samples were centrifuged and placed onto a Magnetic Separation Rack for 2–5 minutes (until solution was cleared). Next, the supernatant was discarded. After that, the tubes on the Magnetic Separation Rack were kept after the beads had been cleaned with 200 μL of freshly prepared 80% ethanol, and the supernatant was then removed after 30 seconds of incubation. The liquid was completely emptied from the tubes without disturbing the beads after repeating the washing step. The beads were air-dried on the Magnetic Rack, then 21 μL of TE buffer was used to elute the DNA. After mixing, the samples were incubated for 5 minutes at room temperature and 19 μL of the supernatant was collected after putting them back on the Magnetic Separation Rack for 2 to 5 minutes.

2.10.1.5 PCR amplification

The PCR amplification mixture was prepared on ice by adding PCR enzyme mix (25 μL) and PCR primer mix (6 μL). Following the transfer of 31 μL of the PCR amplification mixture to the PCR tubes, the mixture was vortexed and centrifuged and loaded onto the thermal cycler at 95°C for one cycle, then 98°C for 20 sec, 60°C for 15 sec, and 72°C for 30 sec for 3 to 12 cycles, then 72°C for 10 min. After a quick centrifugation of the samples, all of the solutions were then transferred to fresh 1.5 mL centrifuge tubes.

Clean-up of PCR product

Each sample was then given 50 μL of DNA Clean Beads, which were then transferred to the centrifuge tubes and thoroughly mixed by pipetting. This was then let to sit at room temperature for five minutes. After a brief centrifugation, samples were then put onto a Magnetic Separation Rack for 2–5 minutes until the liquid turned clear, after which the supernatant was discarded. Then the washing step using

freshly prepared 80% ethanol was repeated twice and the beads were air dried, eluted with TE buffer, mixed, incubated at room temperature, centrifuged, then put back on the rack and supernatant was collected.

Quality control of PCR product

Purified PCR products were quantified using Qubit 2.0 high sensitive dsDNA kit to achieve the required yield of 1 pmol. Then, using electrophoresis-based tools, the fragment size distribution of the purified PCR products was evaluated. (Bioanalyzer).

2.10.1.6 Denaturation

TE buffer (48 μL) was added to each sample's 1 pmol of purified PCR products, and loaded onto the thermal cycler at 95°C for 3 min. Then, the tubes were immediately put on ice for 2 minutes after the reaction was finished, followed by quick centrifugation.

2.10.1.7 Single Strand Circularization

The single-strand circularization mixture was prepared on the ice by adding Splint buffer (11.6 μL), and DNA rapid ligase (0.5 μL). After that, 12.1 μL of single-strand circularization mixture was transferred to the sample tube, mixed, centrifuged, and loaded to the thermocycler at 37°C for 30 min. Then, the tubes were put on ice.

2.10.1.8 Enzymatic Digestion

An enzymatic digestion mixture was prepared by adding Digestion buffer and Digestion enzyme (1.4 μL) and (2.6 μL) respectively. Each PCR tube contained 4 μL of the enzymatic digestion mixture, was vortexed, centrifuged, and then put into the thermocycler at 37°C for 30 min. Each sample's solution was then collected after centrifugation and transferred into new 1.5 mL centrifuge tubes after being vortexed with 7.5 μL of digestion stop buffer.

Enzymatic digestion product cleanup

The enzymatically digested product was mixed with 170 μL of DNA Clean Beads and incubated at room temperature for ten minutes. The samples were then quickly centrifuged, put on the Magnetic Separation Rack for 2–5 minutes until cleared, and the supernatant was removed. The washing step using freshly prepared 80% ethanol was done twice and the DNA was eluted with TE buffer (by repeating the same steps performed in cleaning up steps).

Quality control of enzymatic digestion product

The Enzymatically digested products were quantified using Qubit® 2.0 ssDNA Assay Kit. ssDNA, ng/ input products of PCR dsDNA, ng was quantitated (< 7%).

Samples were loaded on the sequencer machine after that.

2.11 In silico variants analysis

2.11.1 FANCC

Mutant protein model modelling

We used the ModWeb web server to develop a mutant model of FANCC, (Webb & Sali, 2021). As a template for the mutant model, we selected the experimental crystal structure of FANCC with PDB ID 7kzp and Chain A (Wang et al., 2021). This template structure was used to build the mutant model with a mutation at position 197 (E197). The mutation was introduced using the ModWeb server's built-in mutation tool. By adjusting the template structure's coordinates to reflect the mutant allele, the server constructed the mutant model using homology modelling techniques. We then used energy-minimization methods to further refine and optimize the mutant model, making it more accurate and stable structurally. The final output was a 3D structure of the FANCC mutant model with the E197 mutation, which can be further analyzed and studied for its potential effects on protein function and interactions.

DynaMut server prediction

Dynamut is a computational method used for predicting the effects of amino acid substitutions on protein stability of FANCC (Rodrigues et al., 2018). To determine whether a mutation would stabilize or destabilize a protein's structure, this method predicts the resulting changes in free energy ($\Delta\Delta G$). The server determines the free energy shift ($\Delta\Delta G$) due to a mutation by utilizing molecular dynamics simulations and statistical thermodynamics. mCSM, SDM, and DUET, which are structure-based predictions, are also provided by Dynamut. Using several distinct algorithms, each of these predictions make advantage of unique aspects of protein structure and sequence. The NMA-based predictions determine the mutation-induced free energy change ($\Delta\Delta G_{\text{ENCoM}}$) by employing normal mode analysis (NMA). To evaluate how much the flexibility of a molecule has changed as a result of a mutation, Dynamut also computes the difference in vibrational entropy energy between the wild-type and mutant structures ($\Delta\Delta S_{\text{Vib ENCoM}}$).

MD simulations

MD simulations were performed on both wildtype and mutant type of FANCC proteins using Desmond (<https://www.deshawresearch.com/>) (Roopa et al., 2023). At first, we utilized the system builder utility to get the system ready for simulation after having pre-processed the protein with the protein preparation wizard. Each protein was given its own spot inside an orthorhombic simulation box, 10Å from any of the box's edges. The system was solvated by a simple point charge (SPC) water model that had already been brought to equilibrium. To neutralize the system, Na⁺, Cl⁻, together with 0.15 M salt (NaCl) concentration, were supplied to the system. Desmond protocol was applied to relax the model system before the MD simulation. The first step in this six-step relaxation protocol is a 12-ps NVT ensemble simulation at 10K with restraints on the solute heavy atoms. This is followed by a 12-ps NPT ensemble simulation at 10K and pressure (P) = 1atm with the same restraints, a 12-ps NPT ensemble simulation with restraints on the nonhydrogen solute atoms, and a 24-ps NPT ensemble simulation without restraints. Using Desmond's factory settings for the barostat (Martyna-Tobias-Klein) and thermostat (Nosé-Hoover), we were able to get a stable environment in terms of both pressure and temperature. Finally, the protein's force field parameters were assigned using the OPLS3 force field (Harder et al., 2016), and a 100ns MD simulation in the NPT ensemble (T = 300K, P = 1atm) was run, with the trajectory recorded at 100ps intervals by producing about 1000 frames.

2.11.2 SMC5 and TALDO1

AI protein model

We obtained the protein sequences of SMC5 and TALDO1 from the UniProt database using their respective UniProt IDs Q8IY18 and P37837. To construct the three-dimensional (3D) structures of SMC5 and TALDO1, we used AlphaFold, a software that predicts the most likely 3D structure of a protein based on its amino acid sequence. AlphaFold utilizes two neural networks: one to predict a distance map between pairs of residues and another to refine the distance map into a 3D structure. To evaluate the stereochemical quality of the models, we utilized Ramachandran plots. These plots characterize the distribution of phi and psi angles of amino acid residues within a protein structure, allowing for the identification of areas with suboptimal stereochemistry. We also employed ProSA-web to test the structural compatibility of the models with the proteins' amino acid sequences. This tool generates a Z-score that indicates the degree by which a structure's energy deviates from the expected energy of a structure with a comparable amino acid constitution.

Mutated protein model

To generate SMC5 and TALDO1 mutant sequences, we manually corrected amino acids in the protein sequence. The residues chosen for mutagenesis were based on their known or projected significance in protein function. The Swiss Model server was used to model the mutant structures of SMC5 and TALDO1, wherein we provided the mutant sequences and our AlphaFold protein models as templates for the modeling process.

Protein stability

To evaluate the effect of amino acid changes on protein stability, we utilized Dynamut - a computational tool that predicts the change in free energy ($\Delta\Delta G$) produced by mutations. This assessment helps in determining whether the alterations are stabilizing or destabilizing in protein structures like SMC5 and TALDO1. Calculating $\Delta\Delta G$ for a given mutation was done just as FANCC.

MD simulations

Desmond was used to create MD simulations of both wild-type and mutant SMC5 and TALDO1 proteins (Roopa et al., 2023) the same was as FANCC.

2.11.3 Molecular docking

Molecular docking was performed for the normal PKLR protein and the mutated PKLR (Asp339Asn) domains. First, we used both NCBI PDB ID and UniProt ID databases for the biological data collection of normal enzymes. I-TASSER server was used for modeling the 3D protein structures of normal and mutated enzymes, and each compound structure was obtained from the PubChem database. For the energy-minimizing process, Swiss PDB Viewer (spdbv) was used to modify all compound structures. Format conversion from pdb to pdbqt was done by using Open Babel (Version 2.3.1) software. After that, Discovery Studio software (Version 2019) was used to handle the alteration of each protein and ligand by adding hydrogen atoms and metals, before the molecular docking process. The grid box with 1.00 Å spacing and a grid map of mutated with the projected substrate were defined using Auto Dock Vina (Version 2.0) (Mohamed et al., 2022).

CHAPTER THREE

**IDENTIFICATION AND CHARACTERIZATION OF GENOTYPIC
VARIATIONS IN PATIENTS WITH BETA THALASSEMIA MAJOR**

3.1 Brief introduction

Given the wide prevalence of haemoglobinopathies disorders, their severity, and their need for life-long health care, these disorders attract probably significant attention and impose a major burden on the health of the Saudi nation. Providing healthcare to patients suffering from such diseases and syndromes carries a high cost both in monetary and emotional terms.

The term *hemoglobinopathies* often refer to defects or mutations in one of the alpha or beta globin genes which will result in either the expression of an abnormal hemoglobin chain in the hemoglobin molecule that leads to a change in the translated amino acid a structural defect, or a change in the quantity of globin protein produced, caused by a mutation that lead to overproduction, underproduction, or no production of the normal globin protein subunit (Weatherall, 1980).

Sickle cell anemia was the first genetic disease to be attributed to a defect in a specific molecule, by Pauling and colleagues in 1949 and occurs as a result to a specific protein mutation (Pauling et al., 1949). The change of normal hemoglobin to sickle cell hemoglobin occurs by a single nucleotide change from GAG to GTG at codon 7 of the open reading frame of the beta globin gene (*HBB*) and results in changing the transcribed codon from hydrophilic glutamic acid to the hydrophobic valine (p.E7V). This mutation causes a dramatic change in the hemoglobin structure under conditions of deoxygenation, and a severe malfunction leading to anemia (Ingram, 1956).

Thalassemia, on the other hand, is characterized by reduced hemoglobin in the RBC causing inadequate oxygen transport to the body. There are many forms of thalassemia ranging from mild to major (Marengo-Rowe, 2007). In order to help aid in preventing the rising number of people affected with these hemoglobinopathies, researchers in Saudi Arabia are employing molecular screening techniques to identify the most common and rare pathogenic HBB variants within the population.

In this study, we aim to tackle the issue of hemoglobinopathy genetic testing by identifying all the mutations that are causing anemia in a set of 154 patients who are transfusion dependent and have been diagnosed, probably, as thalassemic, which will enable us to identify novel alleles and pathogenic combinations of alleles in the beta globin gene within Saudi population in Jeddah with the utilization of the most recent and advanced molecular techniques.

3.2 Use of TaqMan genotyping assays in identifying patients with the most frequently known β -thalassemia variants within Saudi population

TaqMan genotyping assays were custom designed using Life Technologies Design Panel to be performed on 154 samples. Reference sequence and variant-allele specific probes were designed for each of these previous mutations (reported common within Saudi population). Two different reporter dyes were used of which one probe labeled with VIC[®] dye to detect the reference sequence and one probe labeled with FAM[™] dye to detect variants. The QuantStudio 12K Flex of Applied Biosystems was used for qPCR genotyping. PCR cycles were used as indicated by the manufacturer and assay optimized conditions were followed as indicated in the Experimental Procedures.

As a result of this experiment, 93 samples were found to be homozygous for one of the variants used in SNV assays (homozygous sickle cell was found in 28 samples) which is as expected from consanguineous families (Figure 36). Also, compound heterozygous mutations were found in 7 samples (Figure 37), which would reflect the high prevalence of beta-thalassemia alleles in the population. This makes a total of 100/154 samples confirmed for beta thalassemia and SCD diagnosis using our custom designed variants assays, thus not investigated further. Figure 35 shows examples of the genotyping experiment results. Nevertheless, 20 samples of the remaining samples had single heterozygous mutation (Figure 38), which is conventionally seen as insufficient to cause thalassemia major, and 34 samples tested negative for all of these variants. These 54 samples were investigated further using a more comprehensive sequencing technique to identify possibly novel or unassigned pathogenic variants of beta-thalassemia (Figure 34).

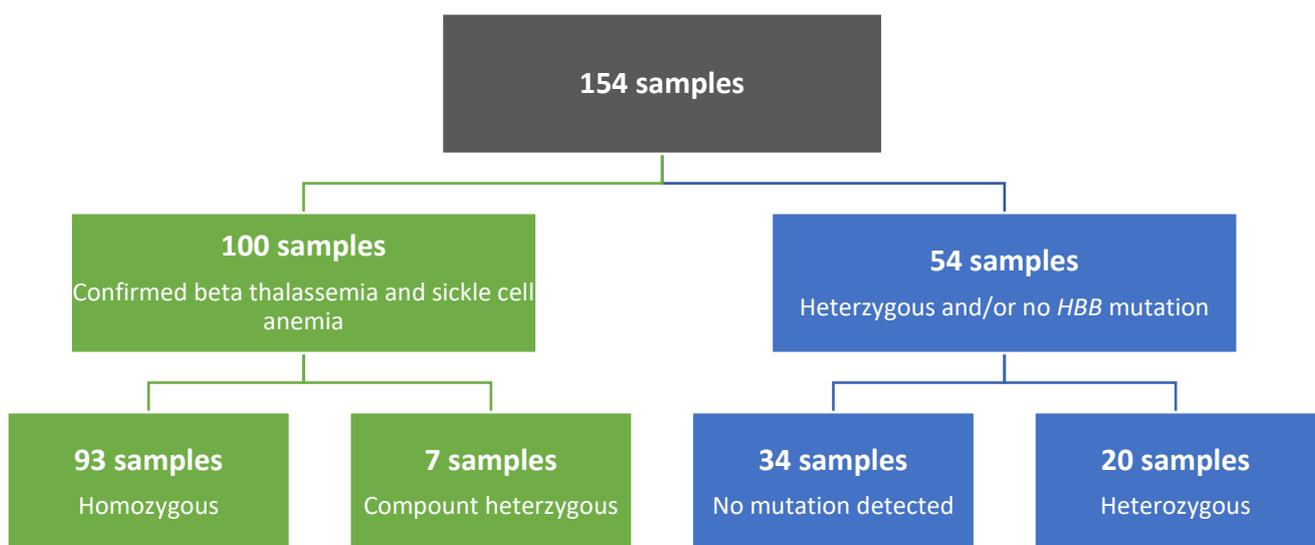


Figure 34: Genotyping result distribution

AMPLIFICATION PLOT

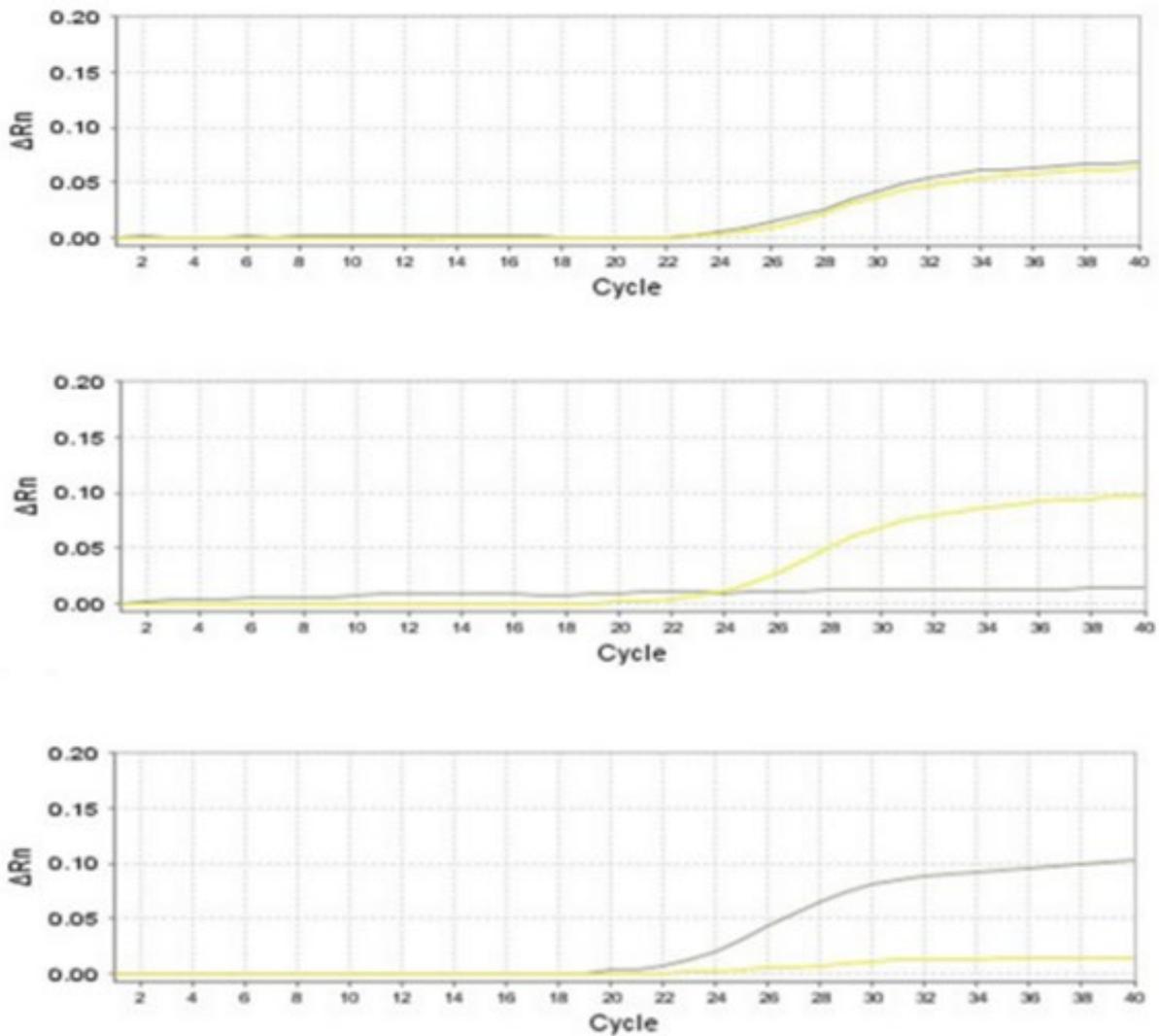


Figure 35: Example of genotyping of the rs33915217 SNP (c.92+5G>C).

Allele 1 appears in yellow and allele 2 appears in grey. The first plot is a sample with a heterozygous mutation. Both Allele 1 labeled with VIC dye and detects the reference sequence and Allele 2 labeled with FAM dye and detects the variation sequence were amplified in the qPCR. The second plot is a sample with a reference homozygous mutation. Allele 1 labeled with VIC dye and detects the reference sequence was the only probe amplified in the qPCR. The third plot is a sample with a homozygous mutation for the rs33915217 SNP. Allele 2 labeled with FAM dye and detects the variation sequence was the only probe amplified in the qPCR. ΔR_n is the value after normalizing the background (subtracting out the noise by baselining).

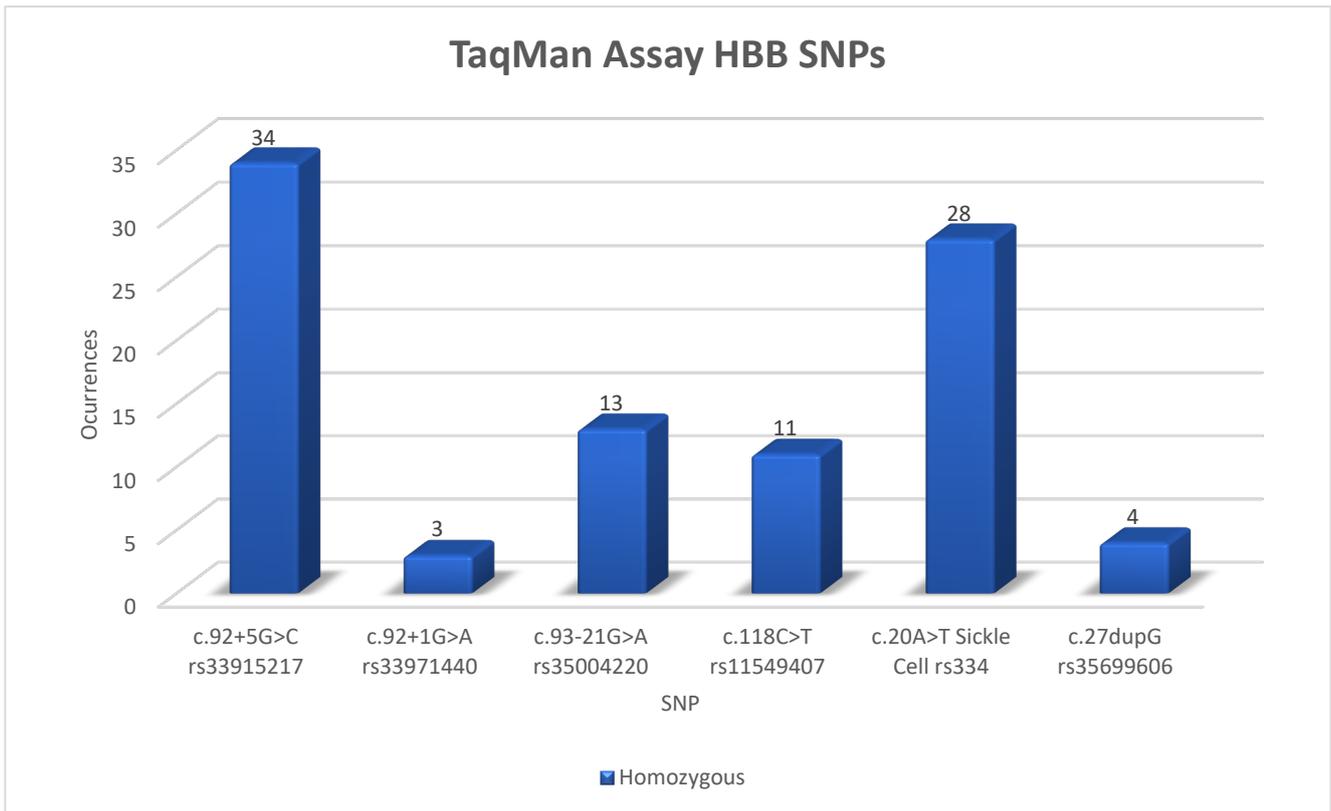


Figure 36: Observed occurrence of homozygous HBB mutations among 154 transfusion dependent beta thalassemia patients.

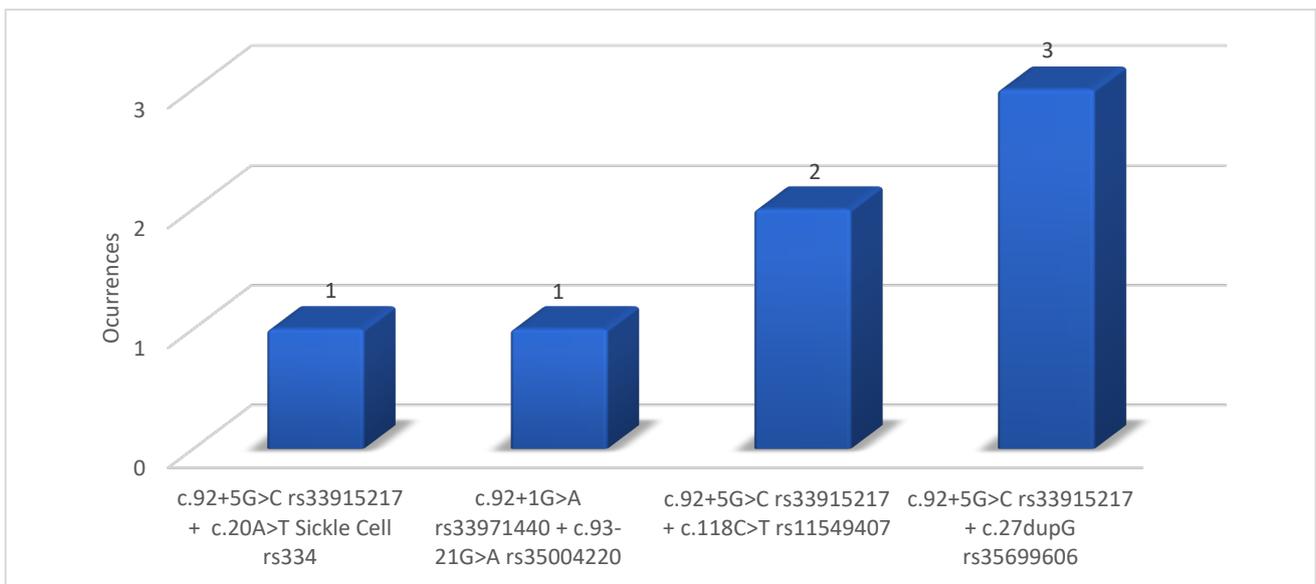


Figure 37: Observed occurrence of compound heterozygous HBB mutations found in seven samples among 154 transfusion dependent beta thalassemia patients.

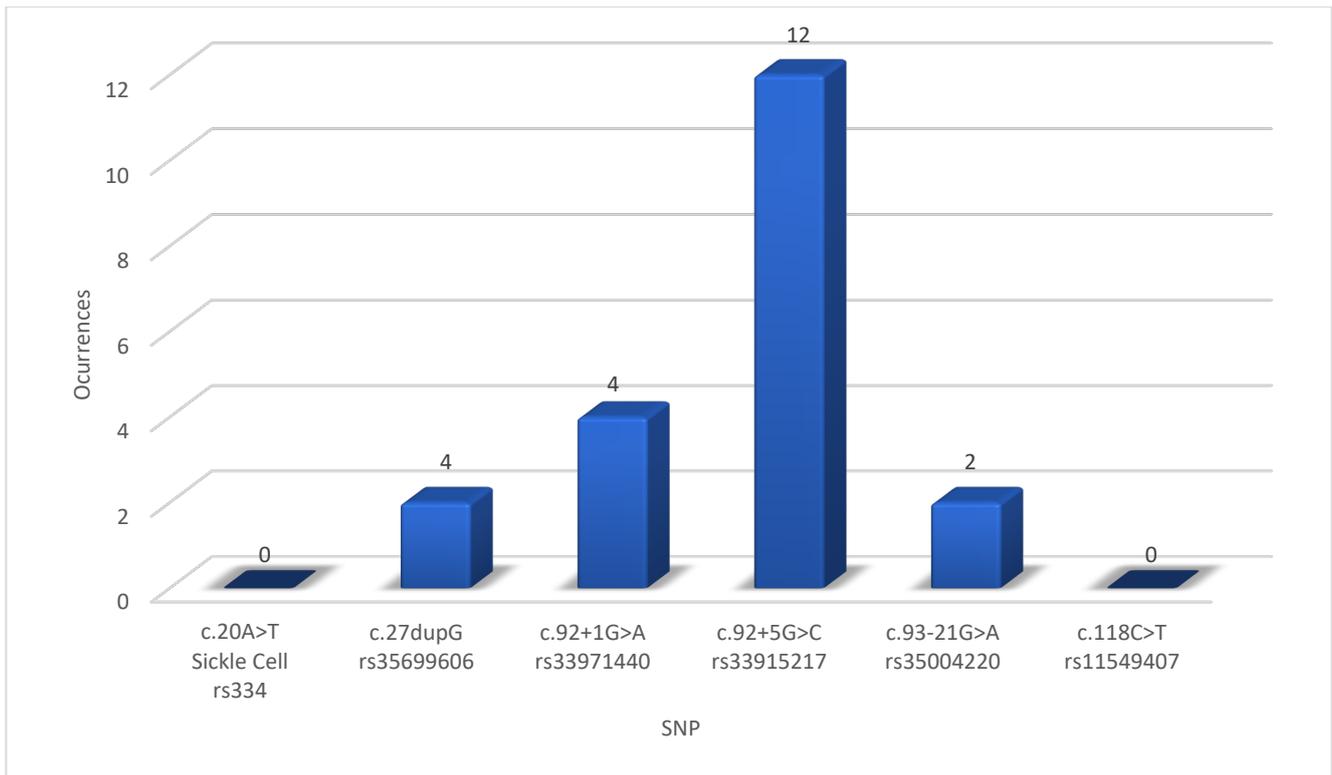


Figure 38: Observed occurrence of 20 single heterozygous *HBB* variants among 154 transfusion dependent patients.

The disease of these individuals is not fully explained.

3.3 Identification of beta globin gene variants in transfusion dependent thalassemia patients using targeted next generation sequencing by Ion Torrent PGM System

After TaqMan genotyping, 100 samples have been exempted from further screening of *HBB* gene by Next Generation Sequencing because they were either homozygous or had compound heterozygous mutants for the SNV assays used in genotyping experiment which were considered sufficient to explain their diagnosis. We have sequenced the *HBB* gene cluster of the 54 remaining samples using Ion Torrent PGM to identify the novel or unassigned mutations that complement beta thalassemia diagnosis.

PGM sequencing was performed on 54 samples where no mutation or a single heterozygous mutation had been detected by TaqMan genotyping for the six common variants from this population. Five additional positive control samples were added to the scan giving a total of 59 samples as a validation for the TaqMan genotyping results. Two of the controls were homozygous variants and three were heterozygotes.

Data from all 59 patients were analyzed. All expected incidences were recovered of the six variants previously detected by TaqMan. These included the 20 heterozygous variants and the variants expected from the controls. In addition, we detected 13 single incidences of other variants, all of which have been recorded previously. These *HBB* variants included: c.-138C>A, c.-137C>A, c.2T>C, c.17_18delCT, c.47G>A, c.51delC, c.79G>A, c.92+6T>C, c.93-1G>C, c.135delC, c.315+1G>A, c.316-14T>G, and c.316-3C>A (Table 13), (Figure 39). These variants involve single nucleotide changes (point mutations) that include missense, nonsense, upstream (5' end), base deletion, intronic mutations (which interfere with gene expression and likely lead to a change in the translated protein), and frameshift which may produce a premature termination codon, consequently altering function of resulting proteins. No discrepancies with the Taqman data were discovered during Ion Torrent sequencing.

Table 13: Identified mutations in 54 samples using Next Generation Sequencing by Ion Torrent. Mutations are mapped to HBB transcript RefSeq NM_000518.4.

HGVS name	Genomic Location (Hg19/GRCh37 chromosome 11)	Location on gene	Molecular consequence	Reference cluster ID (rs#)
c.-138C>A	5248389	Upstream	-	rs33944208
c.-137C>A	5248388	Upstream	-	rs33941377
c.2T>C p.Met1Thr	5248250	Exon 1	Missense	rs33941849
c.17_18delCT p.Pro6Argfs	5248233	Exon 1	Frameshift	rs34889882
c.47G>A p.Trp16Ter	5248205	Exon 1	Nonsense	rs63750783
c.51delC p.Lys18fs	5248200	Exon 1	Frameshift	rs35662066
c.79G>A p.Glu27Lys	5248173	Exon 1	Missense (HbE)	rs33950507
c.92+6T>C	5248154	Splice site 5'	Splice donor variant	rs35724775
c.93-1G>C	5248030	Splice site 3'	Splice acceptor variant	rs33943001
c.135delC p.Phe46Leuufs	5247986	Exon 2	Frameshift	rs80356820
c.315+1G>A	5247806	Splice site 5'	Splice donor variant	rs33945777
c.316-14T>G	5246970	Splice site 3'	Splice acceptor variant	rs35703285
c.316-3C>A	5246959	Splice site 3'	Splice acceptor variant	rs33913413

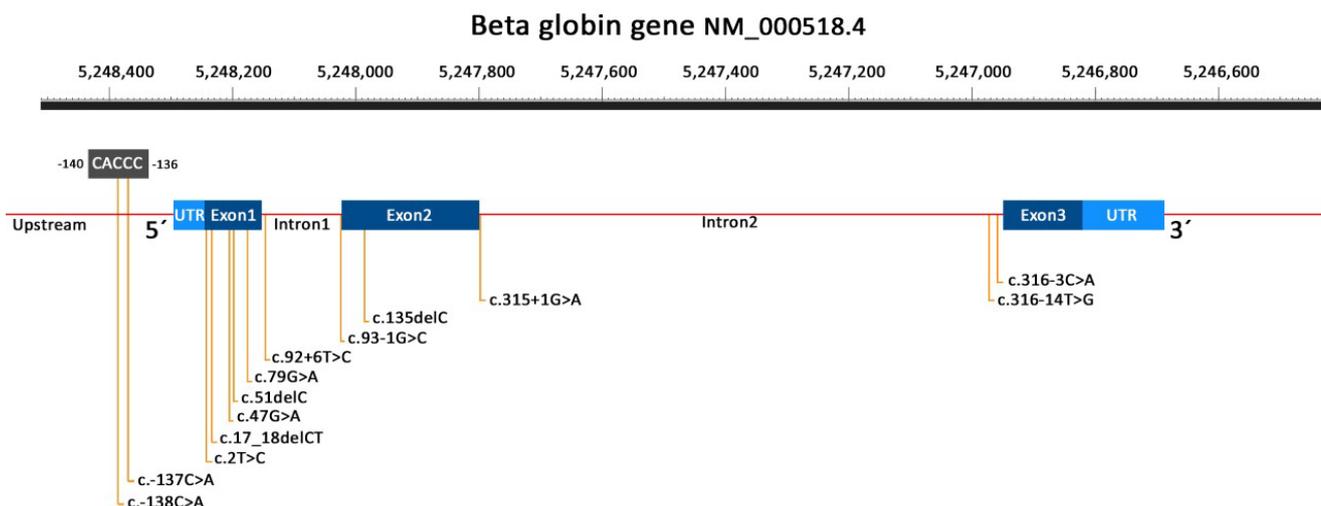


Figure 39: Location of mutations in *HBB* gene NP_000509.1. NM_000518.4 identified by Ion Torrent NGS.

Ion Torrent PGM has verified all of the mutations in the positive control samples that had been genotyped previously by TaqMan. NGS analysis provided biallelic *HBB* disease genotypes for 51 of the 54 patients. The disease occurrence was not fully explained in two individuals (II-14 and II-17), both of whom carried only the heterozygous variant c.27dupG, which had been found previously with TaqMan. Individuals who are heterozygotes (carriers) of beta thalassemia alleles are usually asymptomatic or have only mild clinical symptoms. So, the detected *HBB* genotype of these two patients does not fully explain their transfusion-dependency. These two cases will be discussed in chapter 4.

Additionally, the disease severity was unexplained as well in one other patient (sample # 14-0183), who had no detected mutations within the region of *HBB* covered by NGS which will be discussed in chapter 6.

3.4 Confirmation of discovered β -thalassemia variants using Sanger sequencing

Sanger sequencing was performed as a confirmation of the identified variants (Figure 40)(Figure 41). Samples were selected when *HBB* mutations had not been demonstrated or when only one heterozygous mutation had been identified. Random compound heterozygous alleles were also tested to confirm them. All variants were confirmed successfully, eliminating any false negative results from the Ion Torrent sequencing data or TaqMan Assay genotyping.

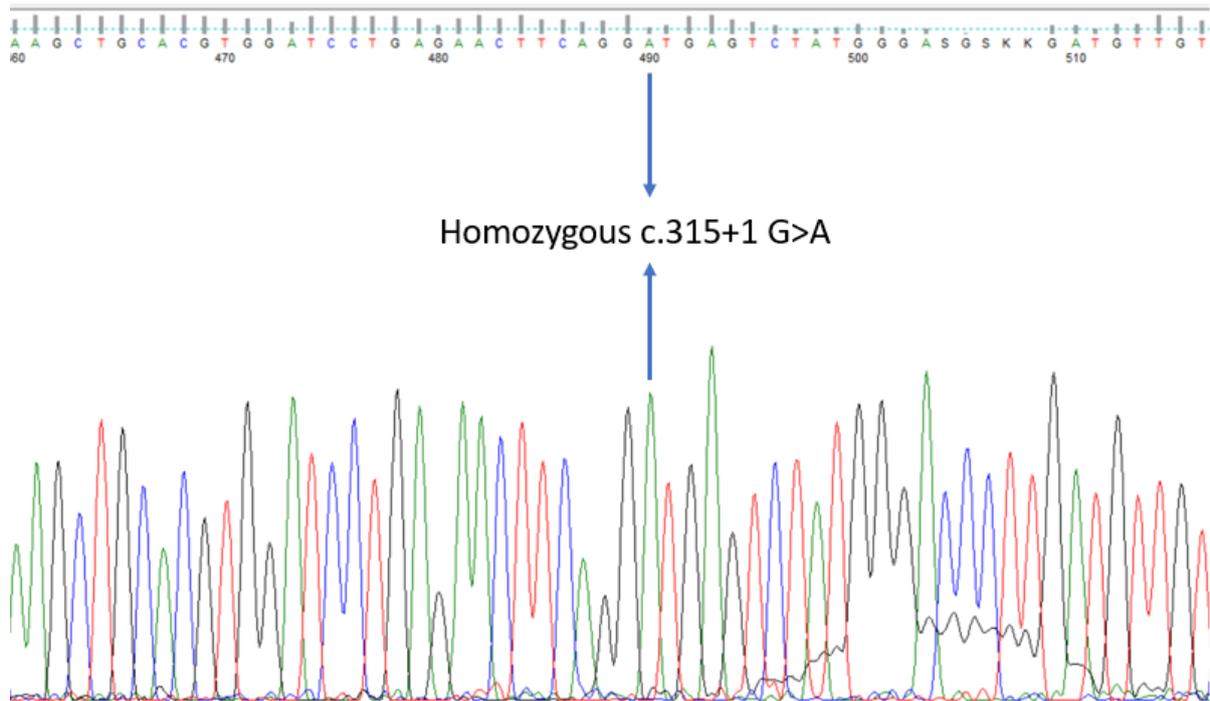


Figure 40: A chromatogram showing a beta thalassemia case with homozygous c.315+1 G>A variant.

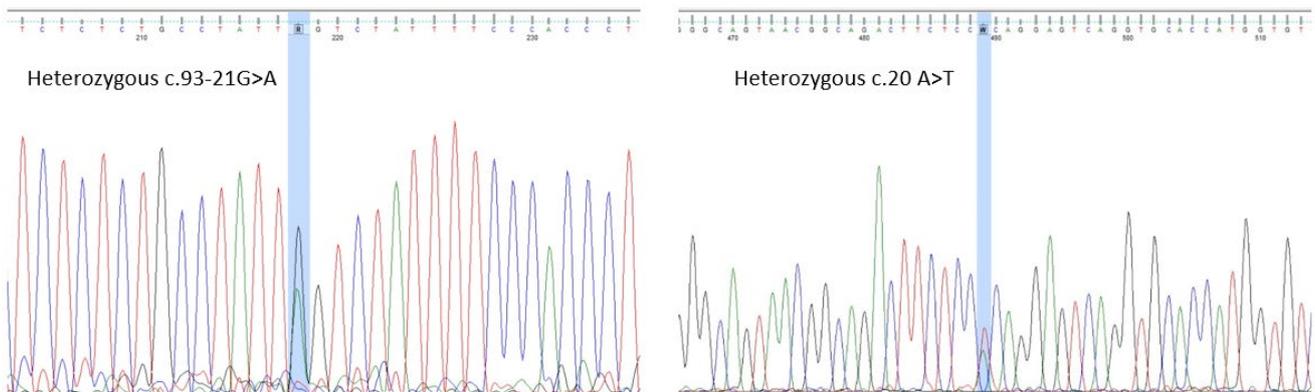


Figure 41: A chromatogram showing beta thalassemia case with compound heterozygous variations. c.93-21G>A on the left, and c.20 A>T on the right.

3.5 Discussion

3.5.1 Identification of genotype variations in transfusion-dependent beta-thalassemia patients

In this study, we aim to find the causative *HBB* mutations in 154 patients within Saudi population in Jeddah with the utilization of the most recent and advanced molecular techniques, in an attempt to develop a feasible approach that is able to facilitate and expedite the process of detecting and screening for these variations in a timely manner.

In Saudi Arabia, the first sickle cell case was reported in the 1960s (Jastaniah, 2011). Nowadays, a wide range of abnormal hemoglobin variants has been identified. An extensive screening program was conducted by many research groups and organizations all over Saudi regions to help identify the most common different variants present amongst Saudi population (El-Hazmi et al., 2011).

Various molecular techniques have been and are still widely used for the detection of beta thalassemia variants, including amplification refractory mutation system (ARMS) (Mirasena et al., 2008) (Varawalla et al., 1991) (Suwannakhon et al., 2019). ARMS was first described by Newton and colleagues (Newton et al., 1989). This method can detect single-base substitutions and minor deletions or insertions. It uses sequence-specific PCR primers that would allow generation of PCR product only if the sample contained the target allele. Although ARMS can reliably detect variants with high accuracy, it is technically-demanding and time-consuming when detecting several variants simultaneously because of the many assays that need to be prepared, a second set of ARMS-PCR for heterozygous/homozygous state discrimination, and the need for gel electrophoresis (Vo et al., 2018).

Another common method used in beta thalassemia detection of variations is reverse dot-blot hybridization (RDBH) (Zhou et al., 2019) (Maggio et al., 1993). In this technique, the target DNA sequence is amplified and labeled, then the resulting labelled amplicon is hybridized to a membrane with immobilized oligonucleotide probes (Kawasaki et al., 1993). RDBH is also technically demanding, requires intensively trained staff, and is time-consuming (Teh et al., 2015). Additionally, qPCR using high resolution melting analysis (HRMA) has been used as well (Hidayati et al., 2020), which analyzes the melt curves of PCR-amplified DNA products quantitatively. This method may lead to ambiguous results if the inspected mutations have similar melting temperatures.

Furthermore, denaturing high performance liquid chromatography (DHPLC) has been used in detecting beta thalassemia variations (Sahli et al., 2016), but it has narrow spectrum for mutation detection.

In the present study, I have used real-time PCR using TaqMan genotyping assays which is simple to setup and optimize. Generally, it involves designing and synthesizing specific probes and primers for the target genetic variation of interest (SNP), followed by PCR amplification and detection using fluorescence-based technologies in one step with the exclusion of gel-electrophoresis after sample processing. I have extracted good quality DNA from 154 transfusion dependent thalassemia and SCD patients. Then, I genotyped them using custom-made TaqMan fluorescent based real-time qPCR using thalassemia *HBB* variants found to be frequent in Saudi Arabia that included c.20A>T, c.27dupG c.92+1G>A, c.92+5G>C, c.93-21G>A, and c.118C>T (Abuzenadah et al., 2011) (Figure 42).

This technique is reliable and gives high throughput in a single run with maximum speed. With a simple workflow and decreased hands-on time, TaqMan provides real-time monitoring of the DNA target with high specificity and sensitivity. This technique has been used for the rapid diagnostic confirmation of beta thalassemia and SCD mutations in many laboratories worldwide using low DNA concentrations (Kho et al., 2013) (Breveglieri et al., 2017).

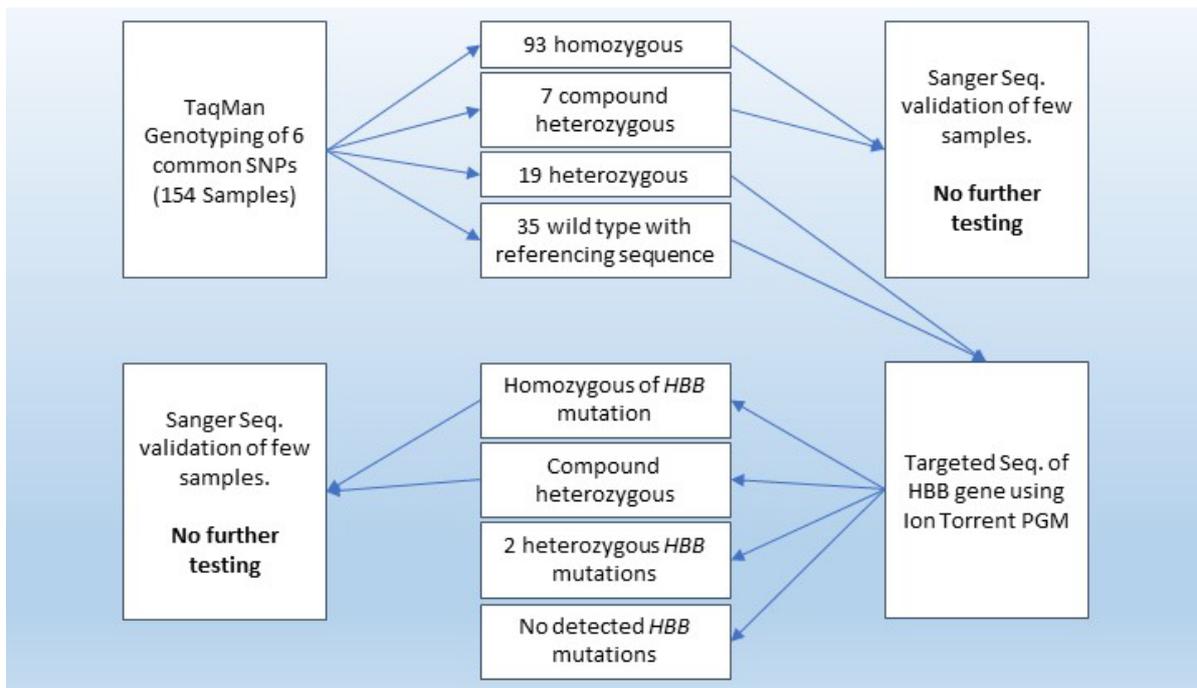


Figure 42: Workflow and molecular techniques used in the investigation of our cohort.

The diagram summarizes the steps and techniques used to nullify all previously known variants for the target of detecting pathogenic novel variants, reported variants, or known variants with unassigned influence on HBB gene.

TaqMan analysis using commonly known variants produced definitive results (which in this study were audited by Sanger sequencing) that allowed two thirds of the patients' samples (100 out of 154) to be excluded from further analysis. Following that, samples were only used in targeted NGS of the beta globin cluster region when TaqMan assay genotyping results were negative for the mutation or only found in heterozygous state (Figure 42). For NGS, customized targeted sequencing of the HBB gene was performed with Ion Torrent (PGM system) from Life Technologies. This technique uses a semiconductor chip that enables massively parallel sequencing of several samples simultaneously. It efficiently detects known common, rare, and novel variants with deep coverage and high specificity and sensitivity in the targeted DNA segments, which has proven to be successful and time saving. In

Oman, they have reliably identified *HBB* mutations using a cohort of 297 samples utilizing Ion Torrent technology that sequenced *HBB* gene in 100 samples in a single run (Hassan et al., 2014).

Table 14: 2d plot of all sample results.

	c.-138C>A	c.-137C>A	c.2T>C	c.17_18delCT	c.20A>T	c.27dupG	c.47G>A	c.51delC	c.79G>A	c.92+1G>A	c.92+5G>C	c.92+6T>C	c.93-21G>A	c.93-1G>C	c.118C>T	c.135delC	c.315+1G>A	c.316-14T>G	c.316-3C>A
c.-138C>A																			
c.-137C>A																			
c.2T>C															Homozygous				
c.17_18delCT															Compound Heterozygous				
c.20A>T					28														
c.27dupG	1	1				4													
c.47G>A							2												
c.51delC																			
c.79G>A								1											
c.92+1G>A								1		3									
c.92+5G>C	3	1	1	1	1	3	1		1		34								
c.92+6T>C						1						2							
c.93-21G>A										1			13						
c.93-1G>C														1					
c.118C>T											2				11				
c.135delC													2			1			
c.315+1G>A									1	1							15		
c.316-14T>G							2			2	3								
c.316-3C>A							5				1								

This study has successfully showed through NGS of 54 samples (no SCD variant amongst them) from transfusion dependent patients with beta thalassemia major, the presence of an additional 13 common and rare pathogenic beta *HBB* variants that included c.-138C>A, c.-137C>A, c.2T>C, c.17_18delCT, c.47G>A, c.51delC, c.79G>A, c.92+6T>C, c.93-1G>C, c.135delC, c.315+1G>A, c.316-14T>G, and c.316-3C>A (Table 14). These variants involve single nucleotide changes (point mutations), base deletion, missense, or nonsense mutations that interfere with gene expression and lead to a change in the translated protein or a frameshift or produce a premature termination codon, consequently altering the function of the resulting protein. All have previously been reported.

3.5.2 Genotype variation and phenotype correlation in transfusion-dependent patients

Accurate prediction of genotype-phenotype associations is a fundamental focus in medical genetics. Effective diagnosis of thalassemia patients is still challenging because of the complex genotype-phenotype correlation, its multiallelic origin, and the involvement of other genes and factors that influence the symptoms of thalassemia. On occasion, it might be difficult to make a correct diagnosis of hereditary anemia due to the fact that the clinical symptoms may overlap in cases of distinct etiologies, making it difficult to tell them apart using traditional diagnostic methods which may lead to misdiagnosis especially if there is a high prevalence of a specific type over another in a specific population as there is in this population (Kim et al., 2017) (Ma et al., 2018) (Hou et al., 2018) (Cil et al., 2020). Since many hemoglobin-related variants have been reported in a wide range of different genes, molecular identification of pathogenic variants is urgently needed for accurate diagnosis.

The *HBB* detected variants which included upstream 5', exonic, and intronic variants that lead to beta thalassemia major are well-studied and each sample has at least one variant (either in a homozygous state or combined with another variant) that is previously reported to be severe variant leading to severe molecular consequence and associated with beta thalassemia major. I started the variants investigation by starting with the variants c.-138C>A, and c.-137C>A, which are located on the proximal control element region upstream the 5' UTR of the *HBB* gene (Figure 43). The region consists of conserved sequences of beta globin core promoter region and many cis-acting regulatory sequences that include the TATA, CCAAT, and proximal and distal CACCC boxes which are required for the assembly and the initiation of the transcription factor complex (Martyn et al., 2017). The upstream variants found in this study are located within the CACCC box at positions -137 and -138 (Mondesert et al., 2022). These variants can disrupt the binding of the Krüppel-like Factor 1 (KLF1) trans-acting transcription factor to the CACCC motif, which leads to disrupted activation of the beta globin gene (Rund et al., 1991) (Efremov et al., 1991) (Martyn et al., 2017).

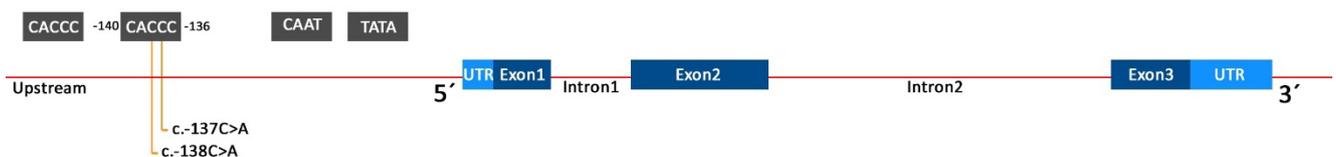


Figure 43: Beta thalassemia variants located upstream of 5' UTR region of *HBB* gene.

Variants are located within CACCC box motif of the promoter region found in patient's DNA in this study. The exons shown are those of the RefSeq beta globin transcript NM_000518.4.

I have identified several variants that are located on the coding region of the *HBB* gene (exons) that include c.2T>C, c.17_18delCT, c.20A>T, c.27dupG, c.47G>A, c.51delC, c.79G>A, c.118C>T, and c.135delC (Figure 44).

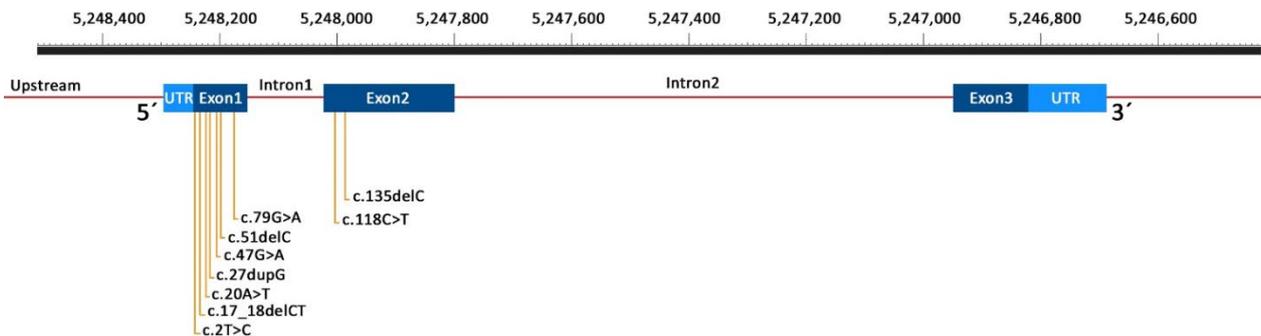


Figure 44: Beta thalassemia variants located in the coding regions of *HBB* gene found in patient's DNA in this study.

The exons shown are those of the RefSeq beta globin transcript NM_000518.4.

The variant causing red blood cells to sickle, c.20A>T missense mutation located on exon one of the *HBB* gene, is the most described variant in the literature. It causes a substitution of the nucleotides A>T in codon 7 that leads to the change in the amino acid from hydrophilic glutamine to hydrophobic valine. Individuals with this mutation have sickle cell disease where RBC sickle under deoxygenation stress, block vessels, and have a reduced lifespan which leads to anemia as described extensively in chapter one.

The variant c.2T>C is a missense variant located on exon 1 and causes a substitution in the initiation codon AUG which encodes methionine, to ACG which encodes threonine. The next AUG is located at codon 56. Southern blot analysis of this variation reveals that it abolishes the Nco I restriction enzyme site and results in the presence of two bands, 8.3 kb and 5.2 kb (Lam et al., 1990) (Jankovic et al., 1990). This variation manifests as β^0 -thalassemia phenotype and results in severe microcytic and hypochromic anemia and an in vitro β/α chain synthesis ratio that was close to 0.5 (Wildmann et al., 1993).

A single nonsense variation has been found amongst our samples as well. It includes the variant c.47G>A, which is located on exon 1 and causes an amino acid substitution from tryptophan to a termination codon (stop codon). Premature termination codons (PTC), depending on their position, can lead to activation of nonsense mediated decay pathway (NMD) that results in mRNA decay. The result of mRNA degradation is a combination of translational inhibition and an elevated mRNA

ribonuclease susceptibility which ultimately leads to the loss of protein expression (Kurosaki et al., 2019) (Nickless et al., 2017). Failure of the NMD pathway to remove these mRNAs with PTCs can lead to the production of aberrant proteins that, through dominant-negative or gain-of-function effects, could be harmful to cells (Udy & Bradley, 2022). HbVar has associated this variant with beta thalassemia major cases.

Several frameshift mutations that include c.17_18delCT, c.27dupG, and c.51delC on exon 1, in addition to c.135delC on exon 2 were also found within our cohort. A frameshift occurs when the number of bases in a gene DNA is altered by the addition or deletion of a base pair or more that is not a multiple of three, which leads to a shift in the reading frame and misreading of all subsequent codons and ultimately this usually causes termination at a premature stop codon in the new reading frame. This mutation also results in the production of short protein that is unstable and the mRNA is targeted for NMD. The variation c.27dupG is a well-described pathogenic variant (classified as beta0 variant) anticipated to result in premature protein termination in exon 1. The frame-shift occurs after codon 10 (serine), the shifted reading frame terminates after 11 more (out-of-frame) amino acids, CRYCPVGQGERG[stop at codon 22].

The missense variant c.79G>A, also called hemoglobin E, is located on exon 1 (codon 26) and causes the substitution of glutamic acid 27 with lysine which produces structurally defective hemoglobin and by activating a cryptic splice site that leads to improper processing of mRNA and reduced levels of the correctly spliced mRNA. The generation of a new stop codon renders the abnormally spliced mRNA inactive (Olivieri et al., 2011). Individuals with the HbE variant in a homozygous state have been reported to have mild beta thalassemia symptoms (Aldakeel et al., 2020), but inheriting HbE in a compound heterozygous state together with another β^0 thalassemia variant reduces beta chain production and increases the disease severity remarkably (Fucharoen & Weatherall, 2012). In our study, all thalassemic individuals with HbE have been found to coinherit β^0 thalassemia variant with it. Ten samples had a compound heterozygous of HbE with c.92+5G>C, one sample had HbE with the splice donor variation c.315+1G>A (destroying the conserved GU at the start of the intron), and one sample had HbE with the frameshift variant c.51delC. All these individuals come from South Asia (3 from Pakistan and 6 from Bangladesh) and Southeast Asia (3 from Myanmar) where HbE variant is highly prevalent.

On the other hand, c.92+1G>A, c.92+5G>C, c.92+6T>C, c.93-21G>A, c.93-1G>C, c.315+1G>A, c.316-14T>G, c.316-3C>A are all intronic variants (Figure 45).

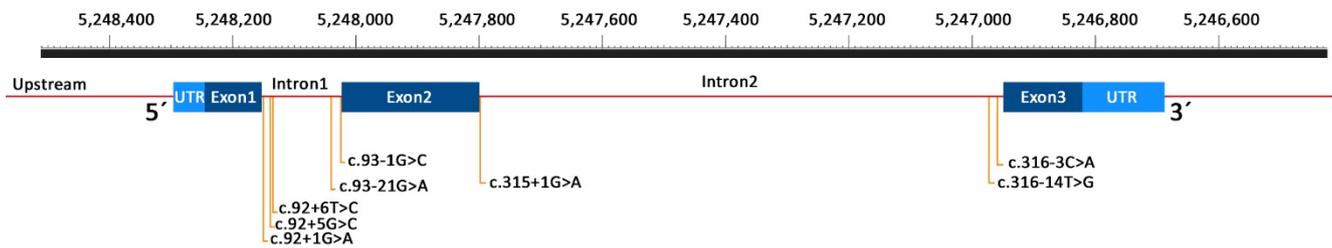


Figure 45: Beta thalassemia variants located in the intronic regions of the HBB gene.

The exons shown are those of the RefSeq beta globin transcript NM_000518.4.

The + sign indicates the variant location on the 5' side of the intron, whereas the - sign indicates the variant location on the 3' side of the intron. Therefore, the splice donor variants c.92+1G>A, c.92+5G>C, c.92+6T>C are located on intron 1 immediately after exon 1, while the splice acceptor variant c.93-1G>C is located on intron 1 immediately before exon 2. The splice site variants c.315+1G>A, c.316-14T>G, c.316-3C>A are on intron 2, where c.315+1G>A is after exon 2 and c.316-14T>G, c.316-3C>A are before exon 3. Mutations that occur in the beginning or the end of introns (donor and acceptor splicing sites) within the spliceosome recognition sequence can affect RNA splicing process and leads to impaired removal of introns and ineffective joining of exons during mRNA processing by exon skipping, intron inclusion, or cryptic splice site use (Lord & Baralle, 2021). Furthermore, variations found in the internal region of the intron (deep intronic) may also affect pre-mRNA processing and have pathogenic effect such as c.93-21G>A which is located on intron 1 before exon 2. This variant has been reported to be associated with beta thalassemia major because the nucleotide change from G to A at that position creates an identical sequence of the authentic splice site and renders it into a functional splice site that can be recognized by the splicing enzyme complex and the U-1 small nuclear RNA. Splicing of the intron at this position will produce an unstable mRNA (Spritz et al., 1981). As for the variant c.316-14T>G, computational investigations (Alamut v.2.11) suggest that this intronic variation may affect splicing by generating a unique cryptic acceptor splice site since it occurs in a moderately conserved nucleotide (Varawalla et al., 1991).

Targeted sequencing of specific gene regions by Ion Torrent PGM was a reasonable choice and certainly justified at the time we started this project to identify variants within *HBB* gene. Nevertheless, this technique, just like any other has its own limitations. The difficulty in differentiating homopolymer repeats, which leads to inadequate coverage of this specific region, is one of this system's limitations.

Intergenic and intragenic regions of low specificity and pooling loss were excluded from our coverage region which may harbor significant variants that would not be detected. Also, some clinically significant variations may go undetected in regions that include very high/low G-C content.

In a total of three out of the 54 samples, targeted NGS analysis failed to explain the disease and results showed inconsistency with the severity of their anemia. Sample # 14-0183 had no obvious or apparent mutation in the covered regions and two were only heterozygous for a pathogenic variant (II-14 and II-17 are both heterozygous for the same variant *HBB*: c.27dupG). Carriers of heterozygous beta-thalassemia (excluding dominant beta thalassemia), regardless of whether it is beta⁰ or beta⁺, is asymptomatic to mildly symptomatic and would not be expected to require transfusion from a clinical standpoint. Since samples were collected from transfusion-dependent patients with severe clinical complications, finding beta-thalassemia variants in only a heterozygous state makes them interesting and requires more investigation to uncover the genetic basis of their phenotype.

Several mechanisms have been identified that may explain why beta-thalassemia heterozygotes tend to have more severe disease at both the clinical and hematological levels. One is inheriting a triple or quadruple alpha globin gene arrangement in addition to inheriting heterozygous beta-thalassemia, which causes the increase of alpha chains production, and consequently the increase in alpha/beta chain imbalance that leads to the premature red blood cell destruction, because of the toxicity of free alpha (refer to page 43), thereby elevating the severity of their phenotype. Also, the inherited variants do not only cause the reduction in the beta globin synthesis but may also produce unstable beta globin chains. Before binding to alpha chains to form functional hemoglobin, unstable beta chains bind heme and precipitate in RBC precursors and lead to the production of inclusion bodies. Hyper-unstable beta chains result from dominantly inherited beta-thalassemia, occurring on exon 3 of the *HBB* gene. Mutations that cause premature codon termination which lie in exons 1 or 2 are responsible for most of the recessively inherited beta thalassemia as they activate nonsense mediated mRNA decay process which help in the prevention of abnormal beta globin mRNA accumulation in the cytoplasm of pre-RBC. On the other hand, mutations in exon 3 are linked to a high level of aberrant mRNA in the cytoplasm. This results in the production of truncated beta chains products, which are unstable and cause red blood cell death by acting in a dominant-negative mode (Lin et al., 2022) (Efremov, 2007). Diagnosing dominant beta-thalassemia is challenging because of the difficult detection of precipitated unstable beta globin chains in peripheral blood (Rizzuto et al., 2021). Moreover, genetic variables

located outside the beta globin gene cluster may also play a role in increasing the severity of beta thalassemia phenotype such as variants that affect transcription factors GATA1 (X-linked gene).

Since *HBB* variant c.27dupG is located on exon 1, we can safely exclude the possibility of dominance effect of this variant. The severity of the patients' cases might be the result of a number of factors, including the possibility that another *HBB* allele has not been found (e.g., due to insufficient sequencing coverage) or was detected but not recognized as causative (e.g., due to complicated structural rearrangements). Structural rearrangements will consist of variations having functional effects that are not detected by conventional variant filtering methods and variants in genes that have not yet been linked to the condition. Therefore, further candidate investigation(s) of variants mapping outside the beta globin cluster is necessary to correlate the manifested phenotype of the affected individuals with their genotype.

From this point, the first step forward was to investigate the family history of each remaining sample. Upon investigation of the history of these samples, the sample with no *HBB* variants had no history of beta thalassemia in her family, which raises the questions, might these samples be misdiagnosed? Should we be looking for other diseases that might manifest with the same phenotype and symptoms? Towards finding an answer to these questions, this sample will be discussed in detail in chapter 5.

Interestingly, on the other hand, the two samples with the same heterozygous *HBB* mutation, II-17 and II-14, were found to be a brother and sister who had five more affected siblings and half-siblings (total of 7). Since all previously mentioned assays have failed to explain the severity of the disease, we adopted another approach to search for novel or rare variants not previously known to be associated with thalassemia with regards to II-17 and II-14, on one hand, and to expand our search, depending on the availability of samples, to include their family of both affected and non-affected members, on the other hand which will be discussed in detail in chapter 6.

After years of exclusively relying on transfusions to treat anemia and inhibit ineffective erythropoiesis in β -thalassemia, numerous novel therapeutic approaches are presently under development. Numerous medication candidates are designed to stimulate hemoglobin F. The potential for permanent transfusion independence exists through the utilization of autologous hematopoietic stem cells that have been genetically modified. In order to induce HbF by silencing *BCL11A*, researchers are currently examining gene editing techniques such as CRISPR-Cas9 and zinc-finger nucleases (Langer & Esrick, 2021). Towards this end, in chapter 4, I have described my contribution in setting and starting

up a fully equipped laboratory in efforts to establish a protocol for reprogramming somatic cells to pluripotent stem cells. These stem cells can be used later on in an attempt to find a unified alternative universal therapy to treat β -thalassemia syndromes for all patients regardless of the underlying *HBB* gene mutation.

CHAPTER FOUR

**A THERAPEUTIC APPROACH FOR AUTOLOGOUS GENE THERAPY
OF SICKLE CELL AND BETA THALASSEMIA DISEASE**

4.1 Brief Introduction

Given the wide prevalence of hemoglobinopathies disorders, severity, and their need for long-life health care, these disorders attracted significant consideration. Allogeneic hematopoietic stem cell (HSC) transplantation is the only curative treatment available for patients with β -thalassemia or sickle cell disease. When bone marrow is unable to produce enough healthy blood cells, an allogeneic hematopoietic stem cell transplant can be used to replace it with healthy blood stem cells from a donor. A donor might be anyone from a close relative to a casual acquaintance to a complete stranger. However, allogeneic transplantation is hampered and limited and carries many challenges, potential hazards and risks including low rate availability of full-matched donors and possibility of immune rejection (Hoban et al., 2016). The need for alternative treatment and development of an effective and safe HSC autologous transplantation therapy is a revolutionary breakthrough. Autologous gene therapy is utilizing a patient's own cells to make the necessary genetic adjustments before reinfusing them into the patient. When these genetically manipulated stem cells differentiate into RBCs, they restore healthy levels of globin chains in the blood, relieving anemia. Scientists in many laboratories all over the world are working on moving toward this goal. With the substantial advances in the use of gene editing tools, it is imperative to establish robust means that have a pronounced potential to achieve clinical success, lower transfusion requirements, and lower the substantial cost of health care and disease management.

Site-specific gene correction and repair of causative mutations using gene editing offers the ultimate gene therapy solution. Gene editing is a set of techniques that researchers have developed to alter the genetic material of living organisms. These methods make it possible to insert, delete, or modify DNA at specific sites in the human genome. There are a number of different genome editing methods now in use. Gene correction in human HSCs has been reported and many labs have corrected the mutated *HBB* gene in induced pluripotent stem cells (iPSCs) isolated from sickle cell patients from a variety of cell types (Xie et al., 2014) (Sun & Zhao, 2014) (Song et al., 2015) (Niu et al., 2016) (Ali et al., 2021). Moreover, although sickle cell is defined by single base substitution mutation present in all SCD patients, β -thalassemia on the other hand is caused by hundreds of mutations with many different types such as SNVs, small deletions or frameshifts. Using site-specific gene correction, every single one of these mutations will have to be corrected individually (Finotti et al., 2015). This indicates the necessity for a unified alternative universal therapy to treat β -thalassemia syndromes for all patients regardless of the underlying *HBB* gene mutation.

Sickle cell and β -thalassemia patients have an improved life expectancy when they have high levels of fetal hemoglobin (as previously described chapter one). In the early 90s, Platt et al. studied the natural history of over 3000 patients with SCD and β -thalassemia and their pain episodes and mortality rate. They have concluded that even a little elevation in the level of fetal hemoglobin can decline the number of pain episodes and improve survival rate (Platt et al., 1991) (Platt et al., 1994). Furthermore, it has been observed that an increase of HbF level in sickle cell patients decreases intravascular sickling and reduces inflammation and disease severity (Fabry et al., 2001) (Thein et al., 2009) (Kaul et al., 2013). HPFH is a benign genetic condition in which production and expression of HbF hemoglobin maintain after birth and throughout life due to mutations that reduce switching of fetal γ -globin to adult β -globin (Forget, 1998). In this project, we will take advantage of these HbF features and try to mimic the naturally occurring HPFH-associated mutations to alleviate β -haemoglobinopathies as a therapeutic approach.

Two main goals have to be achieved for gene therapy to work properly (i) constant gene expression of manipulated long-term HSCs (ii) the safe and efficient transfer of the manipulated gene to the patient (Hoban et al., 2016).

4.1.1 CRISPR/Cas9

We have chosen to use class II clustered regularly interspaced palindromic repeats-associated protein 9 (CRISPR-Cas9) targeted nuclease model system in this PhD project. CRISPR/Cas is an adaptive immune system of prokaryotes against phages and foreign plasmids attack (Koonin & Makarova, 2009). There are three main steps in this immune process: (i) acquisition of spacer sequences (ii) CRISPR transcription (iii) interference (Hu, 2016). When a foreign phage invades a bacterial cell, the bacterium copies a fragment of the virus DNA (called spacer) and stores it within the short DNA repeats of the CRISPR locus of the bacterium genome, which enables it to defend itself against subsequent attack of the corresponding intruder by recognizing and cutting the virus DNA with the utilization of small RNAs and Cas9 endonuclease complex (Figure 46) (Hu, 2016) (Marraffini, 2015).

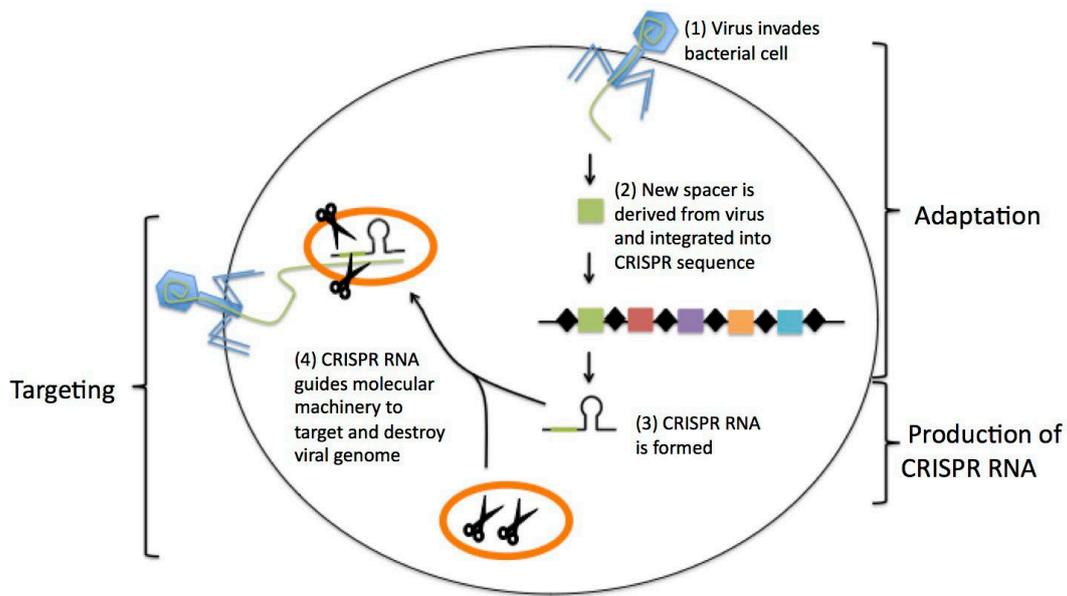


Figure 46: CRISPRs are viral defense mechanisms in the bacterial genome. Black diamonds within CRISPR region represent short DNA repeats, whereas colorful boxes represent spacers. When a new virus infects a bacterium, it contributes a new spacer to the mix of spacers already there. CRISPR RNA molecules are synthesized by processing the CRISPR sequence during transcription. The CRISPR RNA binds to the bacterial machinery and directs it to the viral target sequence that matches. The molecular machinery kills the foreign organism by breaking apart its genome. Source: <http://sitn.hms.harvard.edu/flash/2014/crispr-a-game-changing-genetic-engineering-technique/> Adapted with permission (Barrangou & Marraffini, 2014).

Scientists have figured out how to use CRISPR in the lab to produce exact alterations in the genes of organisms, and their use isn't limited to applications involving bacterial immune responses. Using CRISPR technology, researchers may alter individual genes while leaving the rest of the genome unaffected, shedding light on the causal relationship between a gene and its effect on the organism as a whole. CRISPR class II is the latest discovered tool for gene manipulation (at the time) as it was first realized and described by Doudna and her colleagues in 2012 (Jinek et al., 2012). Hundreds of papers have been published since then. Due to the simplicity, adaptability, ease of design and use, specificity and efficiency of CRISPR-Cas9 system in targeting and manipulating DNA, it is considered to be the tool of choice for precise genome editing for most researchers in all scientific departments (DiCarlo et al., 2013) (Cong et al., 2013) (Hwang, 2013) (Jiang et al., 2013) (Liang et al., 2015) (Port et al., 2014). This system has turned out to be a breakthrough discovery in molecular biology and biotechnology industries.

To utilize the system to edit a gene, two main components are required: (i) single-guide RNA (sgRNA) which is composed of crRNA and tracrRNA chimera. crRNA contains a user-defined 20 nucleotides

targeting sequence fused with tracrRNA which acts as a scaffold that holds crRNA in place (ii) Cas endonuclease. The nuclease Cas9 protein recognizes the target sequence (few bases upstream of a protospacer adjacent motif PAM sequence) using guide RNA. The endonuclease will then act as scissors and introduce a site-specific double-strand break to the target DNA sequence. Homologous end joining (HEJ) or nonhomologous end joining (NHEJ) will subsequently ligate and repair the cleavage site (Figure 47) (Sander & Joung, 2014) (Munshi, 2016) (Tasan et al., 2016).

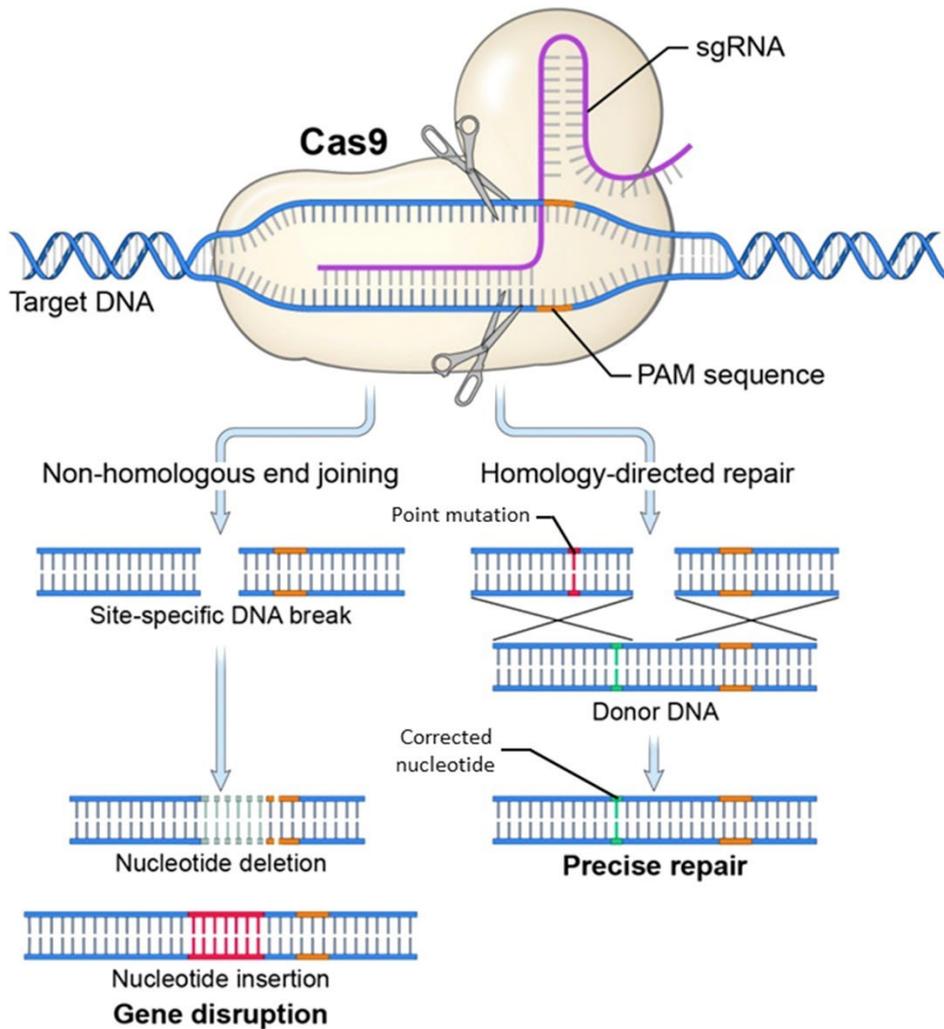


Figure 47: A chimeric single-guide RNA (sgRNA) facilitates Cas9 binding to the target DNA.

The structure is made up of a protospacer and a protospacer adjacent motif (PAM) that together recognize the target sequence. Both nonhomologous end joining (NHEJ) and homology-directed repair (HDR) employing a synthetic donor DNA template can introduce desirable sequence modifications after Cas9-induced double-strand break (DSBs). The gene is disrupted in NHEJ (error-prone) because the cell randomly inserts and deletes nucleotides when it ligates the DNA double-strand break (gene silencing). On the other hand, in HDR, the double-strand break is repaired by introducing a repair template into a cell that incorporates a desired sequence change into DNA during the repair process. The new DNA is modified to include the new sequence, resulting in a directed exact repair. Source: reproduced with permission (Guitart et al., 2016) (Savić & Schwank, 2016).

4.1.2 Generation of Induced Pluripotent Stem Cells

Embryonic stem cells are pluripotent. The inner cell mass of a blastocyst is where human embryonic stem cells (ESCs) originate. These cells may divide and replenish themselves indefinitely while also retaining the ability to develop into any specialized cell type. By initiating the "forced" expression of particular genes in differentiated mature adult somatic cells, researchers are able to return them to an embryonic-like state, creating what are known as induced pluripotent stem (iPS) cells. They have many properties with natural pluripotent stem cells like embryonic stem (ES) cells, including the ability to self-renew and differentiate into a wide variety of cell types. In order to replace the disease-causing gene with a healthy one, the cells must first be reprogrammed into iPSCs. The iPSCs that have been changed genetically are isolated and then differentiated into the impacted cell type. After that, the cells are reintroduced into the patient. Immunological resistance and graft failure is a typical consequence following allogeneic transplantation; however, this method of autologous transplantation may eliminate the need for such drastic measures. iPSCs may be created from the patient's own cells, thus are genetically identical to the recipient and may not be rejected by the immune system (Rattananon et al., 2021) (Zakaria et al., 2022) (Al Abbar et al., 2020).

Reprogramming factors, or a group of genes linked with pluripotency, are often introduced into an adult cell type to generate induced pluripotent stem cells. Comparable to human embryonic stem cells, these cells exhibit similar properties. octamer-binding transcription factor 3/4 (Oct3/4), sex determining region Y-box 2 (Sox2), cellular-Myelocytomatosis (c-Myc) and Krüppel-like factor 4 (Klf4) are the founding members of the original reprogramming factor (or Yamanaka factor) set. Unipotency can be converted to pluripotency by ectopic expression of these four identified transcription factors (Ghaedi & Niklason, 2019) (Al Abbar et al., 2020).

In 2009, Raya et al have shown that somatic cells isolated from individuals with Fanconi anemia may be reprogrammed to pluripotency and used to create iPS cells once the underlying genetic abnormality has been corrected. They demonstrated that these iPS cells may generate phenotypically normal myeloid and erythroid progenitors that are disease-free (Raya et al., 2009). Another group led by Ye et al used a patient with homozygous beta thalassemia (β^0) to generate iPS cells from skin fibroblasts and demonstrated that these iPS cells could be differentiated into hematopoietic cells that made normal hemoglobin (Ye et al., 2009).

4.2 Aim

1. Optimize protocols for the generation of iPSCs from whole blood collected from control samples then beta-thalassemia patients.
2. Designing and screening guide RNAs for generation of deletional HPFH as a therapeutic approach for sickle cell and thalassemia patients.
3. Electroporation of the CRISPR constructs into generated iPSCs and differentiation into hematopoietic stem cells.
4. Comparison between the expression of HbF (fetal hemoglobin) from clone cells with deletion and clone cells without deletion.

4.3 Methods

Equipment and consumables

Biosafety cabinet, centrifuge, mini-centrifuge, microscope, 37 °C–5 % CO₂ incubator, 37 °C water bath, 6-well plates, pipettes, 5 mL and 10 mL sterile serological pipettes, aspirating pipette, 15 mL and 50 mL conical tube (Corning, 430791), 2 mL centrifuge tubes, hemocytometer, and Neon electroporation system.

4.3.1 Reprogramming of whole blood derived cells to iPSCs using ReprRNA-OKSGM vector by StemCell

4.3.1.1 Preparation of reagents

A. Culturing fibroblast from whole blood medium

I have prepared the medium required for fibroblast culturing by adding Advanced DMEM (Dulbecco's Modified Eagle Medium) with 4500 mg/L D-Glucose + 10% FBS + 1% MEM Non-essential AA (100X) (nutrients for growth and viability stimulation) + 1% Penicillin/streptomycin (to control bacterial contamination) + 1% sodium pyruvate (source of energy and antioxidant) + 1% 200 mM L-glutamine (supports fast-replicating and high energy demanding cells- protein and nucleic acid synthesis).

After preparation, the medium was stored at 2-8°C (up to two weeks) and warmed to room temperature before each use.

B. Culture medium for erythroid progenitors' expansion

I used StemSpan™ Serum-Free Expansion Media SFEM II (Catalog #09605/09655) and StemSpan™ Erythroid Expansion Supplement (Catalog #02692) from StemCell Technologies.

Cultures from human PB can be used to selectively promote the expansion of erythroid progenitor cells when supplemented with the Erythroid Expansion Supplement, which contains an optimized combination of recombinant human cytokines (SCF, IL-3, and EPO). After 7-14 days in culture, populations of erythroblasts that are transferrin receptor (CD71) and Glycophorin A (CD235, GlyA) positive will remain.

The media was prepared by adding 10 mL of StemSpan SFEM II medium + 100 µl of Erythroid Expansion supplement (for 1 well of 6-well plate for 7 days expansion + Day 0 of transfection). For the 7-day expansion medium, 10% PS was added. The media was stored at 4 °C for one month.

After 14 days in culture, more than 80% of the cells produced with this procedure are CD71⁺GlyA⁺ erythroblasts. Pro-erythroblasts that are CD71⁺GlyA⁻ and normoblasts that are CD71^{-/lo}GlyA⁺ may also be present.

C. Growth medium

Preparation for 1 well of 6-well plate: I have prepared the growth medium by adding 10.38 mL Advanced DMEM + 1.16 mL Fetal Bovine Serum FBS (embryonic growth promoting factors) + 116.66 µl 200 mM L-Glutamine + 4 µl 0.5mg/mL B18.

After preparation, the medium was stored at 2-8°C (for 1 week) and warmed to room temperature before each use.

D. Growth medium with puromycin

When working with cell cultures, puromycin is an antibiotic that is used as a selective agent, which will kill all cells lacking resistance genes while leaving behind just its intended targets.

Puromycin was prepared by adding 1 mL of DMSO to 1 mg of puromycin. After that, the preparation of the growth medium with puromycin was done fresh daily by adding 1.2 µl of puromycin to 1.5 mL of growth medium for 1 well of 6-well plate.

E. ReprorNATM-OKSGM (catalog # 05930) cocktail

In addition to the puromycin-resistance gene, the single-stranded RNA replicon vector ReprorNATM-OKSGM contains the reprogramming factors OCT4, KLF-4, SOX2, GLIS1, and c-MYC. This RNA vector efficiently converts fibroblasts and other somatic cells into pluripotent stem cells.

While using RNase-DNase free tips and tubes, I have prepared the cocktail for 1 well of 6-well plate by adding the following reagents in order: 1 µl ReprorNA + 100 µl Opti-MEM 1 + 2 µl ReprorNA

transfection supplement + 2 µl ReproRNA transfection reagent. The cocktail was prepared immediately before transfection and incubated at room temperature for exactly 5 minutes before adding to cells.

F. Complete ReproTeSR medium (without B18)

Reprogramming medium for human iPS cell induction (Catalog #05926) from StemCell Technologies was used.

Complete ReproTeSR medium was prepared by adding 2.5 mL of 20X ReproTeSR supplement and 0.1 mL of 500X ReproTeSR supplement to 47.4 mL of basal medium and stored at 2-8°C for up to two weeks.

G. Complete ReproTeSR medium with recombinant B18R protein

During the first two weeks following transfection, recombinant B18R protein is added for interferon response inhibition and increasing cell survival.

Preparation was done by adding 4 µl of 0.5 mg/ mL B18R to 11.6 mL of complete ReproTeSR medium. After preparation, the medium was stored at 2-8°C (for 1 week) and warmed to room temperature before each use.

H. 6-well Plate coating

I coated the wells of the 6-well culture plate with vitronectin one hour prior to culture by diluting 40 µl of vitronectin with 1 mL of Cell Adhere dilution buffer for each well. The plate was then left for one hour at room temperature then the vitronectin was aspirated and the wells were washed using 1 mL of Cell Adhere solution. After washing, 1 mL of prepared warmed culture media was added to each well (so it won't dry off), and the plate was kept in the incubator (37 °C– 5 % CO₂) until used (adding cells).

4.3.1.2 Procedures

A. Isolation of mononuclear cells (MNC) from peripheral blood

Blood samples were collected in sodium heparin tubes by the amount of 15-20 mL (adult donor would yield 0.8-3.6x10⁶ MNC/ mL of whole blood, and I need at least 2x10⁶ MNC/ mL for nucleofection). I diluted the blood sample with equal amount of PBS at 1:1 ratio (diluent: PBS was prepared by adding 5 mL PBS + 5 mL DMEM medium + 10% FBS). After warming Histopaque reagent to room temperature, I added 15 mL of histopaque in a Falcon tube and added the diluted sample dropwise on the side of the tube at 2:1 ratio (blood:histopaque) carefully creating two distinct layers. The tube then was

centrifuged at 400 x g for 30 min at room temperature with the breaks off (DEC set at 0). The tube will have three layers, I pipetted out the first layer which is the plasma layer, then collected the second layer which is a thin white layer (buffy coat of WBCs and platelets) and transferred to a new tube. The third layer of RBCs was discarded. The collected WBCs were then washed twice using PBS by filling the tube and spinning at 200 x g for 10 min at RT. In the last wash, the supernatant was discarded, and the pellet was resuspended using 1 mL of PBS. Cells were counted using hemocytometer.

B. Cell counting using a hemocytometer

From the cell suspension, 10 μ l was transferred to 2 mL Eppendorf tube and mixed by pipetting with 10 μ l of trypan blue (stain). The glass hemocytometer slide is cleaned with alcohol before each use. The cover slip is moistened and affixed to the hemocytometer. To count the cells, 10 μ l of trypan blue treated cells is collected to applied to the hemocytometer filling the chamber underneath the coverslip. To count live cells (white unstained), I need to count all 4 corners (each corner has 16 sub squares) one by one using a microscope (with 10X objective) and a hand tally counter. To calculate the average cells, I summed up the number of live cells in all four corners and divided it by 4. To calculate the number of viable cells/mL, I used the formula= average live cells x 10^4 x 2 (dilution factor).

C. Culturing fibrocytes from whole blood

The 6-well plate was coated with vitronectin as described above. After buffy coat extraction from peripheral blood sample and after cells were counted, I used 1×10^6 cells/ well. The cells pellet was resuspended in culturing fibroblast medium. After that, I added 1 mL of resuspended cells to the coated plate (which already contains 1 mL/ well of medium) which brings the total to 2 mL /well. The cells were mixed in each well by shaking the plate gently in N-S-E-W motion then incubated at 37 °C– 5 % CO₂ incubator. On day 2, the media (non-adherent cells) was aspirated and discarded, and 2 mL of fresh media was added to the cells as fibrocytes are very adherent. The media was changed every other day (day 5, 7, 9, 12) for 8-14 days. The cells were checked by microscope for morphology (elongated and spindle shaped), viability, and attachment. To detach the fibrocytes from the plastic surface, they were incubated for 1.5 min at 37 °C with Trypsin-EDTA 0.05%, then trypsin was aspirated off and media with 10% FBS was added to cells (trypsin deactivation). After that, cells were scraped gently and harvested then washed with PBS. A microscope was used to confirm cells detachment.

D. Erythroid progenitors' expansion

After sample has been collected and PBMNCs have been isolated, the cells were counted using a hemocytometer and the cells pellet was resuspended to a density of 8×10^8 cells/2 mL in erythroid expansion medium and transferred to 6-well plate (Day0). Cells were incubated overnight at 37 °C–5 % CO₂. On Day1, the 2 mL of non-adherent cell suspension was transferred to a new well. On Days2, 4, 6, and 8, the cell suspension was collected, centrifuged at 300 x g for 5 minutes, and the pellet was resuspended in 2 mL/ well of fresh erythroid expansion medium and incubated at 37 °C–5 % CO₂. On Day9, the cells were collected from the well and washed twice using PBS and ready for transfection.

E. Flow cytometry for the confirmation of erythroid progenitors' expansion

I have performed flow cytometry twice on each sample, both before and after culturing for expansion of erythroid progenitors.

For staining cells and preparing for flow cytometry, cells were collected, washed twice with PBS, and counted (5×10^4 for unstained and $1-3 \times 10^5$ for mixed colors). The appropriate number of cells were transferred into FACS tubes and FACS buffer (PBS + 2% FBS + 0.5% EDTA) was added to reach a total of 1 mL. After that, cells were centrifuged at 300 xg for 5 min. The supernatant was decanted by inverting the tube quickly in one motion without shaking. Then, FACS buffer was added to the pellet by the amount of 100 µl and vortexed. Conjugated antibodies were added to each labelled tube (3-5 µl), vortexed, and incubated in the dark for 30 minutes at room temperature. After incubation, the sample was washed twice by using 1 mL of FACS buffer and centrifuged at 300 xg for 5 minutes, and the supernatant was decanted by tube inversion. The stained pellet was resuspended in 100 µl FACS buffer. The tubes were covered with foil and kept on ice and ready for flow cytometry. The flow cytometry machine was operated by the lab technician.

F. GFP Neon electroporation

Whole blood was collected in sodium-heparin tubes and MNCs were extracted using Histopaque. Cells' pellet was resuspended in 1 mL of media (DMEM + 10% FBS with no PS), then cultured in 10 mL of media in T25 flask and incubated at 37 °C–5 % CO₂ overnight. The next day, the cultured cells were washed twice using PBS and centrifugation at 1000 rpm for 10 minutes. Cells were counted using a hemocytometer. Depending on the cell count, the 24-well plate was selected for transfection to have 70-80% confluency.

For setting up the Neon electroporation device, depending on my cell type, the input was set at:

Pulse (voltage)	Pulse width (ms)	Pulse number	Neon tip type
2,150	20	1	10 μ l

Format	Cell type	DNA (μ g)	Volume plating medium	Cell number
24-well plate	Suspension/ human	0.5-3 (should not exceed 10% of total transfection volume)	500 μ l	1-2.5 x 10 ⁵

Resuspension Buffer T (for primary blood-derived suspension cells) was used with the amount of 10 μ l to resuspend the washed cultured cells pellet. GFP protein (our DNA) was added to the resuspended cells with the amount of 2 μ l of 0.7 μ g. Following that, electroporation was performed according to the manufacturer's manual. After transfection, the sample was immediately transferred into the prepared 24-well plate containing pre-warmed culture media (with no antibiotics) and the plate was gently rocked to assure even cells' distribution then incubated at 37 °C–5 % CO₂ overnight. The next day, fluorescence microscopy was used to determine the transfection efficiency.

G. CRISPR of erythroid progenitors expanded cells

The Blood sample was collected in sodium-heparin tube, MNCs were isolated using Histopaque, and the cells were cultured for 8 days for expansion of erythroid progenitors. On Day8 of expansion, cells were collected, washed twice with PBS, counted using hemocytometer, and prepared for transfection with ArciTect™ HPRT CRISPR (positive control) (catalog # 76013) using Neon electroporation.

For the preparation of HPRT 2 nmol crRNA (catalog # 76010) and 5 nmol tracrRNA (catalog # 76016), nuclease-free water was added to give a final concentration of 200 μ M with the amount of 10 μ l to crRNA and 100 μ l to tracrRNA. As for the preparation of 80 μ M gRNA, 4 μ l of 200 μ M of crRNA was combined with 4 μ l of 200 μ M of tracrRNA and 2 μ l of ArciTect™ annealing buffer (5X) (catalog # 76020) bringing the total to 10 μ l. The gRNA mixture was incubated after that at 95°C for 5 minutes followed by 60°C for 1 minute, then cooled to room temperature and placed on ice.

In the preparation of cells for transfection, after cell counting, 1×10^6 of cells in 500 μl was transferred to a new microcentrifuge tube and centrifuged at $300 \times g$ for 5 minutes. During centrifugation, 6-well culture plate (not coated) was prepared by adding 2 mL of erythroid expansion medium combined with 200 μl cloneR for each well and incubate it at 37°C until used for culturing after electroporation. Also, ArciTect™ CRISPR-Cas9 RNP complex mix was prepared for electroporation by combining 0.56 μl of 80 μg gRNA + 1.2 μl of ArciTect™ Cas9 Nuclease (catalog# 76002) (3 $\mu\text{g}/\mu\text{l}$) + 3.24 μl of resuspension buffer T (used for cells extracted from blood) which brings the total to 5 μl (1:2-1:4 molar ratio of Cas9 to gRNA is recommended). The RNA complex mix was incubated at room temperature for 20 minutes. The recommended electroporation conditions for knockout of peripheral blood cells using Neon device is to add 2.5 μl of resuspension buffet T to the 5 μl of the RNP complex mix for a volume total of 7.5 μl per transfection (complete RNP complex).

After centrifugation of 1×10^6 cells, supernatant was aspirated off and the pellet was resuspended in 7.5 μl of resuspension buffer T and mixed by pipetting. Then, the resuspended cell suspension was transferred to 7.5 μl complete RNP complex bringing the total to 15 μl . Using the 10 μl Neon pipette tip, 10 μl of the cell mixture was drawn (free of bubbles) and placed into the electroporation chamber which contained 3 mL of electrolytic buffer E. The cells were electroporated with electrical potential set at 2150 V, pulse width of 20 ms, and number of pulses of 1. Immediately after electroporation, the cells were transferred to the warmed prepared 6-well culture plate. The cells were mixed by gently rocking the plate back and forth, then incubated at 37°C –5 % CO_2 for 48-72 hours (or up to 7 days depending on confluency) for genome editing to occur. Cells were harvested for assessment of genome editing efficiency by PCR and stored at -80°C for downstream DNA extraction.

H. Reprogramming fibroblasts to iPSCs using ReprorNA™-OKSGM vector by StemCell

At Day0 of transfection, a vitronectin coated plate was prepared and 1.5 mL of growth medium (+ B18R) was added per well and incubated at 37°C –5 % CO_2 for at least 20 minutes. During media warming period, cultured fibrocytes were collected by centrifugation at 1500 rpm for 10 minutes and washed twice using PBS by spinning at 1500 rpm for 5 minutes and counted by hemocytometer (0.4-1 $\times 10^6$ / well of 6-well plate is needed for transfection). The cell pellet was then resuspended with 10 μl of resuspension buffer T, then 10 μl of ReprorNA™ was added to the sample mix. The sample was transfected with Neon with the same PBMNCs settings (Voltage 2150 v, Pulse= 20 ms, pulse#= 1), then the sample was immediately dispensed in the warm culture prepared plate. On Day1-5 after

transfection, growth medium (+B18R) with puromycin was prepared fresh on daily basis and media change was performed. The cells were checked under the microscope to see if they were attached or in suspension. If in suspension, cells were collected by media aspiration and centrifugation at 1500 rpm for 5-10 minutes and then pellet was resuspended in growth medium (+B18R) with puromycin. On the other hand, if the cells were already attached to the well, the medium was simply aspirated and replaced with freshly prepared one. On Day 6 and 7, media change was performed but with growth medium (+B18R) without puromycin. Day 8-14 after transfection, the medium was changed on daily basis by aspirating the old medium and adding 1.5 mL of complete ReproTeSR + B18R. Day 15-19 after transfection, medium was changed daily but with complete ReproTeSR without B18R. The media was changed on a daily basis after that (to pick colonies if formed) (Figure 48). Cell morphology was monitored using a microscope with each media change.

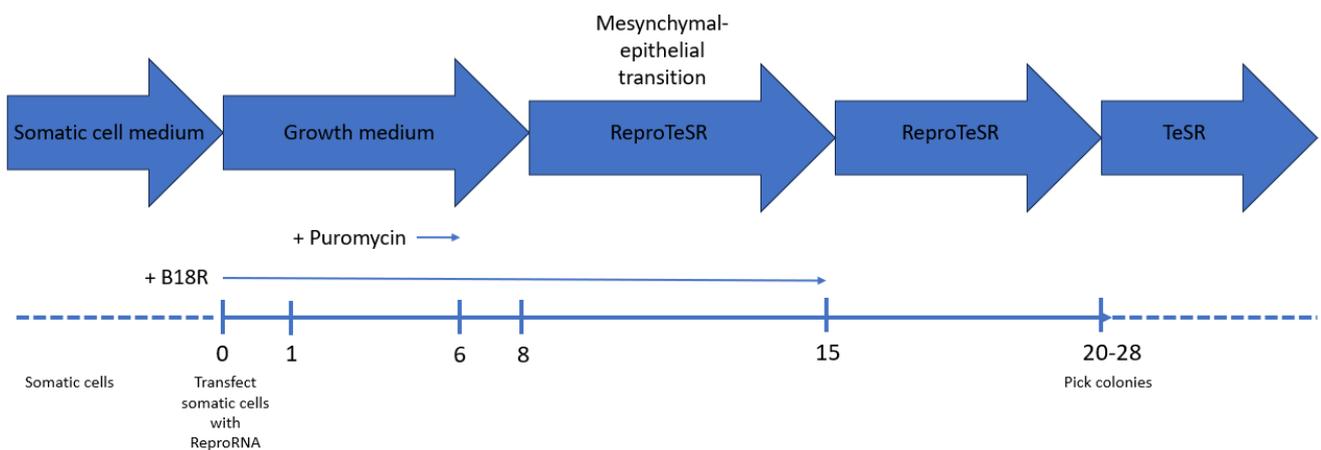


Figure 48: Reprogramming fibroblasts to iPSCs with ReproRNA-OKSGM timeline.

I. Reprogramming expanded erythroid progenitors to iPSCs using ReproRNA-OKSGM vector by StemCell

Vitronectin coated plate was prepared. After aspirating the vitronectin wash solution, in each well of 6-well coated plate, 2 mL of StemSpan SFEM II medium + 0.66 μ l was added and incubated at 37 °C–5 % CO₂ for at least 20 minutes. On Day0 of transfection, cells were collected after 9-days of culture in the expansion medium and washed twice using PBS. Then, cells were counted and resuspended in 10 μ l of T buffer with the amount of 1×10^5 cells/ well. 1 μ l of ReproRNA was added to the sample mix. After that, the cells were transfected by Neon electroporation on the same PB-MNCs settings (Voltage 2150 v, Pulse= 20 ms, pulse#= 1), then the sample was immediately dispensed into the pre-warmed 2

mL of expansion medium + B18R (with no antibiotics) and the cells were evenly distributed across the well of the coated plate.

A cocktail of B18R and Puromycin (BP) was prepared by mixing 8 μ l of B18R with 19 μ l of Puromycin to a total of 27 μ l to be used for the next 6 days of culture. On Day1 (the following day), 2.26 μ l of BP was added to the transfected well. On Day2, 1 mL of the same expansion medium was added to the well (without removing any medium from the well) + 3.4 μ l of BP. On Day3, 1 mL of ReproTeSR medium was added to the well (without removing any medium) + 4.53 μ l of BP. On Day4, 4.53 μ l of BP was added. On Day5, 1 mL of of ReproTeSR medium was added to the well (without removing any medium) + 5.66 μ l of BP. On Day6, 5.66 μ l of BP was added to the well. After that, the medium was aspirated, and 18 mL of ReproTeSR media with 6 μ l of B18R was prepared for the next 9 days of culture (Day7-15). Media was changed daily. After Day15, media was changed with ReproTeSR without B18R, and the cells were monitored for colonies appearance (Figure 49).

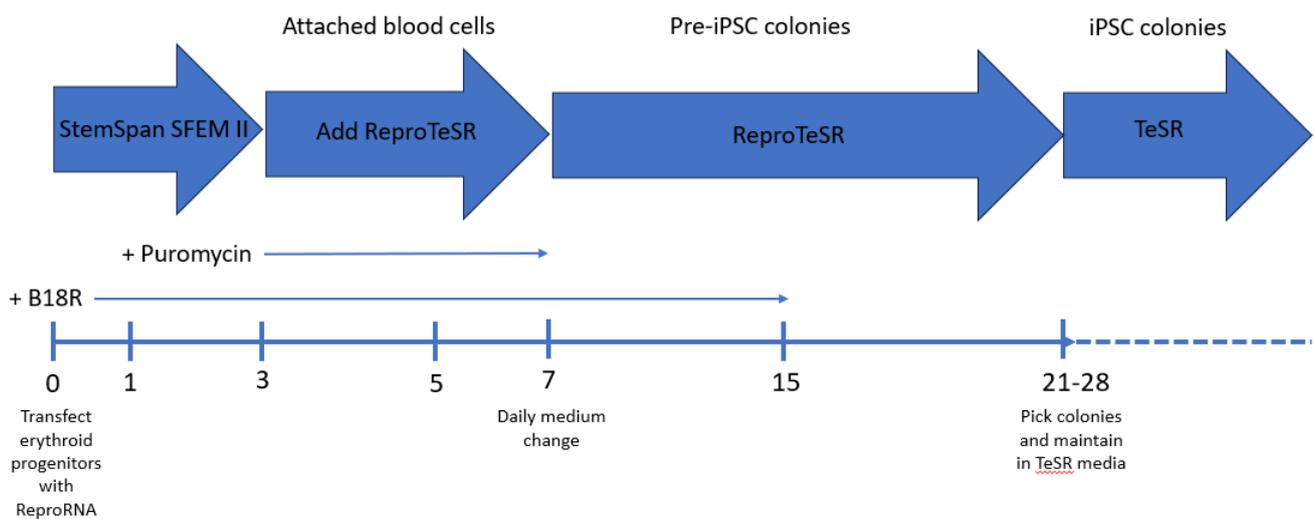


Figure 49: Reprogramming erythroid progenitors expanded cells to iPSCs with ReproRNA-OKSGM timeline.

4.3.2 Reprogramming of fibroblasts derived from whole blood to iPSCs using Epi5TM Episomal iPSC reprogramming kit from Invitrogen by Life Technologies

4.3.2.1 Preparation of reagents

A. Supplemented fibroblast medium

The following components were added to the fibroblast medium (DMEM + MEM non-essential AA + FBS), Basic Fibroblast Growth Factor (bFGF) to the final concentration of 4 ng/ mL + HA-100 (ROCK inhibitor) to the final concentration of 10 μ M.

B. N2B27 medium

For 250 mL of N2B27 medium, 238.75 mL of DMEM + 2.5 mL N-2 supplement (100X) + 5 mL of B-27 supplement (50X) + 2.5 mL MEM non-essential AA (10mM) + 1.25 mL GlutaMAX-I (100X) + 454.5 µL beta-mercaptoethanol (1000X) were mixed and stored at 2-8°C for up to one week.

C. N2B27 medium supplemented with CHALP molecules and bFGF

The following components were added to N2B27 medium freshly before each use:

Component	Final concentration
PD0325901 (MEK inhibitor)	0.5 µM
CHIR99021 (GSK3beta inhibitor)	3 µM
hLIF (Human Leukemia Inhibitory Factor)	10 ng/ mL
A-83-01 (TGF-beta/ Activin/ Nodal receptor inhibitor)	0.5 µM
HA-100 (ROCK inhibitor)	10 µM
bFGF	100 ng/ mL

D. Complete Essential 8 medium

Essential 8 supplement was thawed at 2-8°C and mixed by gently inverting the vial. After that, Essential 8 supplement was added to DMEM medium to the ratio of 1:1 and mixed by swirling the bottle. Complete Essential 8 medium was stored at 2-8°C up to two weeks. It was warmed at room temperature before each use.

4.3.2.2 Procedure

Vitronectin coated 6-well plate was prepared. After aspirating the vitronectin wash solution, 2 mL of supplemented fibroblast medium/ well was added and the plate was incubated at 37°C for at least 20 minutes. Four to two days before transfection, cultured fibroblasts were plated into T75 flask in fibroblast medium to achieve 75-90% confluency by the day of transfection (Day0). On transfection day (Day0), fibroblasts were collected and counted. 1×10^6 of cells which is needed for each transfection were resuspended in 10 µL of resuspension buffer T and 2.85 µL of Episomal iPSC Reprogramming Vector was added to the sample mix. After that, the cells were transfected by Neon electroporation with settings set up at (Voltage 1650 v, Pulse= 10 ms, pulse#= 3), then the sample was immediately dispensed into the pre-warmed 2 mL of supplemented fibroblast medium (cells were evenly distributed across the well of the coated plate) and incubated overnight at 37 °C–5 % CO₂.

On Day1 after transfection, the supplemented fibroblast medium was aspirated from the well and replaced with 2 mL of supplemented N2B27 medium. The media was changed every other day, up to

Day15. On Day15, the supplemented N2B27 medium was aspirated and replaced with supplemented Essential 8 medium. Following that, the medium was changed every other day, and the plate was observed for iPSC colonies formation (Figure 50).



Figure 50: Reprogramming fibroblasts derived from whole blood to iPSCs with Epi5TM Episomal iPSC vector timeline.

4.4 Limitations and delimitations

At the time of starting this project, researchers all around the world were racing to generate genetically edited iPSCs as a therapeutic approach for sickle cell and thalassemia patients and I was one of them. I spent a considerable amount of time in establishing a fully equipped tissue culture laboratory with limited access, for contamination control, to only personnel involved in the project. At that time, we were the first to start working on reprogramming somatic cells to iPSCs in our institution and research center with no previous attempts. Although this fact imposed a difficult challenge, I was eager to start the transfer of technology and optimize these protocols and get them working. I have used several different approaches using different reprogramming kits. Unfortunately, I have encountered many more challenges and shortcomings while working on this project than I have been anticipating rendering the progress of the project unable to proceed any further. Restrictions were imposed on my project as a result of many factors. The lack of experts in the field was an obstacle as I had to contact kits and reagents companies all the time for troubleshooting. Time constraints, resource intensiveness, and importation of reagents were all major factors that contributed to delaying the progress of the project as I had to wait for 8-12 weeks to receive the reagent every time I place a new order. Most of all, Covid and complete lock down had a severe impact and huge toll on the project's progress as well.

Sickle cell and beta thalassemia patients' fear of Covid infection hindered their participation even after restrictions are lifted. The obstacles and restrictions that I have encountered during the work on my project had a huge impact on the decisions that I made about the concentration of my research goals and forced me to make major adjustments to the aims and methodology to be able to present a well-executed work in a timely manner.

CHAPTER FIVE

CASE REPORT: IDENTIFICATION OF A RARE HOMOZYGOUS MISSENSE VARIANT IN THE PKLR GENE THAT CORRELATES WITH CHRONIC PYRUVATE KINASE DEFICIENCY ANEMIA

5.1 Brief Introduction

The most common erythrocytic glycolytic pathway defect connected with congenital non-spherocytic anemia is red cell pyruvate kinase deficiency. The condition which is inherited as an autosomal recessive Mendelian trait is caused by mutations in the PKLR gene located on chromosome 1q21. Our knowledge of the genetic diversity, pathophysiology, and hemolytic anemia complications brought on by red cell pyruvate kinase deficiency has grown over the past few decades. More than 300 distinct variants have been discovered since the discovery of the first PK deficiency pathogenic variants in 1991, and extensive research has been done on the investigation of its molecular mechanisms and the existence of genotype-phenotype connections (Bianchi & Fermo, 2020).

Pyruvate kinase deficiency (PKD) was initially identified in the early 1960s. De Gruchy and colleagues reported that adenosine triphosphate ATP, but not glucose, was able to reverse the symptoms of non-spherocytic hemolytic anemia in a subset of patients by performing autohemolytic test on patient's red blood cells alone at first, then compares it with autohemolytic rate of blood with added glucose, blood with added adenosine, and blood with added adenosine triphosphate using modified Dacie's method. This discovery came after Selwyn and Dacie's initial discovery of the connection between glycolysis and hemolytic anemia in the 1950s. PKD was soon identified as the molecular cause of this anemia by Valentine and Tanaka and their colleagues (Selwyn & Dacie, 1954) (Valentine, 1961) (Tanaka et al., 1962) (De Gruchy et al., 1960) (Grace & Barcellini, 2020).

Pyruvate kinase (PK) is crucial in the energy-producing glycolysis pathway (the Embden-Meyerhoff pathway) that provides the red blood cell with the primary source of energy (ATP) by breaking down the sugar glucose. PK catalyzes and synthesizes the conversion of phosphoenolpyruvate to pyruvate and one molecule of ATP (Figure 51) (Al-Samkari et al., 2020).

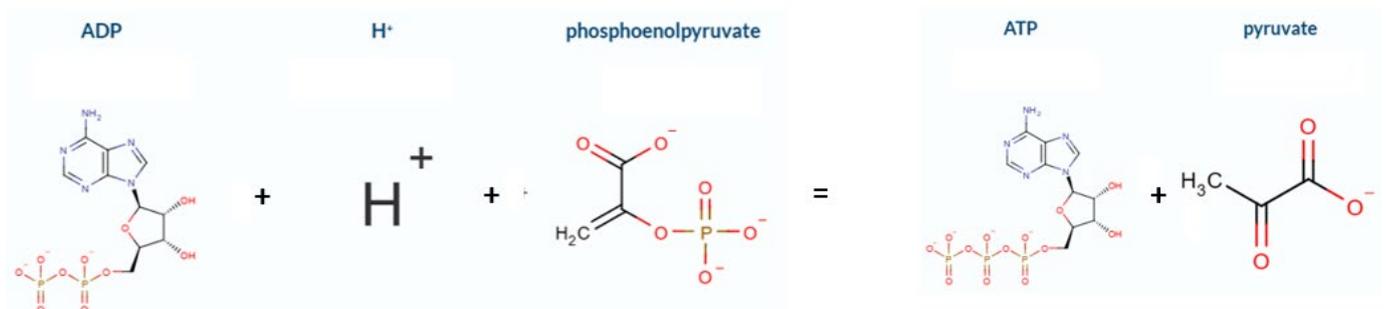


Figure 51: Equation of pyruvate kinase catalytic reaction.

The *PKLR* gene, which is 9.5 kb in size and has 12 exons, is found on chromosome 1q21. In accordance with tissue-specific promoters, the gene encodes for the enzyme's liver (L) and erythrocyte (R) isoforms. Both the erythrocyte and the hepatic isoenzyme share 10 of the 12 exons that make up the coding area, whereas exons 1 and 2 are unique to each isoform, respectively. The 2060 bp long cDNA for *PKR* codes for 574 amino acids (Bianchi & Fermo, 2020).

The *PKLR* gene, which is active in both the liver and red blood cells, codes for both pyruvate kinase isoenzymes PK-L and PK-R. Since mature RBC lack mitochondria and have a lifespan of 100 to 120 days, they rely on ATP for preserving their structural and functional integrity. Mutations in the *PKLR* gene result in abnormal or deficient PK enzyme which leads to insufficient ATP synthesis, loss of RBC membrane flexibility, cellular dehydration (As a result of the exhaustion of the energy supply for membrane ion transporters), and early destruction of RBCs in the spleen or liver causing hemolytic anemia and enlarged spleen. The hemoglobin-oxygen dissociation curve shifts rightward due to the accumulation of 2,3-bisphosphoglycerate when PK activity is inadequate. Shortness of breath, weariness, and pallor (pale skin) are symptoms of an insufficient amount of oxygen transporter RBCs (Roy et al., 2021). The severity of the resulting congenital non-spherocytic anemia varies from patient to patient. When RBCs are prematurely destroyed, iron and a chemical known as bilirubin are released, which causes an excess of these substances to circulate in the blood. Jaundice (yellowing of the eyes and skin) is a result of elevated bilirubin levels in the blood. If these levels went untreated, these levels could accumulate and have serious effect that can cause brain damage (acute bilirubin encephalopathy) in infants (very rare to develop in adults). Also, high levels of bilirubin make gallstones more likely to form (Porter & Dennis, 2002).

The diagnosis, treatment, and follow-up of children and adults with PK deficiency lack standards of care at present. Because it is unclear how people with PK deficiency differ in terms of risk of problems or clinical course, management techniques frequently mirror those used for patients with hereditary spherocytosis or thalassemia intermedia (Grace et al., 2018).

Worldwide reports of PKD patients have been made since the condition was first described. It is found in all ethnic groups. Although the frequency of PKD is not known with certainty, it is estimated to occur in 3-8 out of every 100,000 people (Secrest et al., 2020). Cases may be misdiagnosed due to its rarity, difficulties in making a diagnosis, and the wide range of clinical symptoms; as a result, this frequency

might be understated. Because heterozygous carriers frequently exhibit no symptoms, it is challenging to estimate the prevalence.

The number of *PKLR* gene pathogenic variants that are now known is constantly growing. Canu and colleagues have reported 260 mutations (Canu et al., 2016). On the other hand, HGMD has reported 290 pathogenic variants in their database. The coding regions are where the vast majority of pathogenic mutations are found, with most of them being missense mutations distributed throughout all *PKLR*'s exons. The majority of PK deficiency variations impact residues important to the enzyme's structure and/or function. The wide range of clinical manifestations associated with PKD reflect the significant molecular heterogeneity, and the pursuit of a link between genotype and phenotype has been the subject of research for many years.

Although both pyruvate kinase deficiency and beta thalassemia major are hereditary illnesses that impact red blood cells, the mechanism by which they manifest are distinct. Severe anemia results from a lack of beta-globin chains, a component of hemoglobin, which is a hallmark of beta thalassemia major. Conversely, pyruvate kinase deficiency is characterized by hemolytic anemia due to an insufficient amount of the enzyme pyruvate kinase, which is essential for the maintenance and stability of red blood cells. This, in turn, causes persistent hemolysis, which occurs at an earlier stage in the disease.

The pyruvate kinase enzyme has a new allosteric activator in the form of AG-348 (mitapivat), an innovative small chemical that is administered orally. It increases adenosine triphosphate (ATP) generation and decreases levels of 2,3-diphosphoglycerate by dramatically upregulating both wild-type and many mutant variants of erythrocyte pyruvate kinase (PKR). Clinical studies including mitapivat have been conducted in a variety of inherited hemolytic anemias, including PKD, sickle cell disease, and non-transfusion-dependent thalassemias. Mitapivat has become the first approved licensed treatment for PKD in adults (Al-Samkari & van Beers, 2021) (De, 2024).

5.2 Diagnosis

Hematologically, PK deficiency presents like any other hereditary hemolytic illness. Clinical presentation and test indicators of chronic hemolytic anemia, such as severe jaundice, splenomegaly, an increase in reticulocytes, and hyperferritinemia, might raise the possibility of PKD as a diagnosis. With the exception of miscarriages and affected siblings, the family history of PKD is largely unrevealing

because it is an autosomal recessive illness. The diverse range of inherited and acquired hemolytic diseases is part of the differential diagnosis. The diagnosis is made when more frequent causes of hemolysis have been ruled out, reduced PK enzyme activity has been shown (enzyme testing). Although low enzyme activity is correlated with PKD diagnosis, normal enzyme levels can be falsely diagnosed as normal in patients who have received blood transfusions because of contamination with normal donor cells. Additionally, transfusion dependency without obvious etiology since birth, with unexplained hyperbilirubinemia are strong indicatives of the disease. Molecular analysis and genetic diagnosis of PKD is considered the best way to confirm the disease if compound heterozygous or homozygous pathogenic mutations in the PKLR gene have been found (Bianchi et al., 2019).

5.3 Case Presentation

This case refers to sample 183 from chapter 3. Peripheral blood was collected from a female patient in her early 20s at the day care center of King Abdulaziz University Hospital (KAUH) where she receives her monthly blood transfusion. The patient was previously diagnosed as suspected beta-thalassemic with symptoms presented from 2-3 months and was administered regular continuous transfusions accordingly. Sample DNA was investigated for beta-thalassemia variations using several molecular techniques, including TaqMan genotyping, NGS targeted sequencing of *HBB* gene using Ion Torrent PGM, and Sanger sequencing. Variations on *HBB* gene were never identified to support the suspected diagnosis. Additionally, alpha thalassemia was investigated using MLPA technique but yielded negative results. Further investigation of the case and reviewing clinical data and family history became necessary to correlate her genotype with her clinical manifestation.

5.4 Clinical History

When the proband was born, she was discharged as a healthy new-born. Within a few days or weeks, the baby's skin and whites of her eyes started to appear yellowish with the family unaware of the severity of their daughter's condition, they did not seek medical attention until around 3-4 months old. Jaundice (yellow skin and eyes) is a condition in which the infant's liver may not be developed enough to effectively eliminate bilirubin which leads to accumulation of this substance in the infant's body. Kernicterus is a disorder that develops when severe jaundice is left untreated for an extended period of time and can cause brain defects, hearing impairment, and athetoid cerebral palsy. Unfortunately, when the family returned to seek medical care, brain damage and hearing impairment was already present in the proband and had severe hemolytic anemia of extremely low RBCs count

and low hemoglobin levels that required immediate blood transfusion. At that time, a diagnosis of suspected beta-thalassemia was given to her to fast-track the approval of her transfusion and regular monthly appointments were set on her behalf for continuous transfusions. The patient was not investigated further until she was transferred to be treated at KAUH (around the time of sample collection for our study at the age of 21). At that time, after sample collection and molecular characterization of the *HBB* gene, the hospital was informed that there was no genetic finding that correlates with beta-thalassemia. An attending hematologist reviewing her case file decided to investigate her symptoms and family history and performed Hb electrophoresis for the proband's mother (father deceased) and her healthy siblings and found no evidence of beta-thalassemia. The patient is the offspring of consanguineous parents and because of the rarity of pathogenic alleles in the population, it would be extremely rare to find cases in non-consanguineous families. Because it is a recessive disorder, there is unlikely to be any phenotype in the parents or even heterozygous siblings. There was no history of transfusion dependency in the family.

An interview with the proband's caring sister was conducted by phone to collect information on the family history of the disease, as well as the disease history of the proband. No history of the disease was detected in any of the proband's immediate or secondary family members. The parents were consanguineous but there was no evidence of the disease in other family members. Additionally, the attending hematologist was also contacted to further understand the proband's clinical manifestation. When an infant presents with hemolytic anemia of unknown etiology and there is no history of sickle cell disease, thalassemia, or any other genetic hemolytic anemia in their family, the diagnosis of PK deficiency is often made later on. Pyruvate kinase deficiency and beta thalassemia major share many of the general clinical manifestation, notably, thalassemia major and sickle cell disease do not cause hemolytic anemia in the neonatal period since they still depend on fetal hemoglobin until around the age of 6 months. On the other hand, severe pyruvate kinase deficiency can be manifested in neonates since it is an enzymatic disorder that breaks down red blood cells with no regard to hemoglobin type. Generally, kernicterus is not considered a high risk with thalassemia major. Bone abnormalities, enlarged spleens, and severe anemia are more common consequences of thalassemia major. On the other hand, the major symptom of PKD is persistent hemolysis, which increases the likelihood of developing jaundice. Elevated bilirubin levels, due to the continuous breakdown of red blood cells in PKD, increases the risk of kernicterus in newborns.

As mentioned in the case presentation above, several molecular techniques were used to investigate both beta thalassemia (*HBB*) and alpha thalassemia (*HBA*) genes. In order to identify a different genetic etiology, whole exome sequencing was performed rather than *HBB*-directed sequencing (Figure 52). NovaSeq 6000 DNA Exome by Illumina was used. Procedure was carried out as mentioned in the manufacturer's user guide (explained in detail in Experimental Procedure chapter). Software used for data analysis was Bcl2fastq v2-20, BWA-mem (aligner) v-0.7.12, Samtools v1.2, GATK-HaplotypeCaller v-4.1.4.

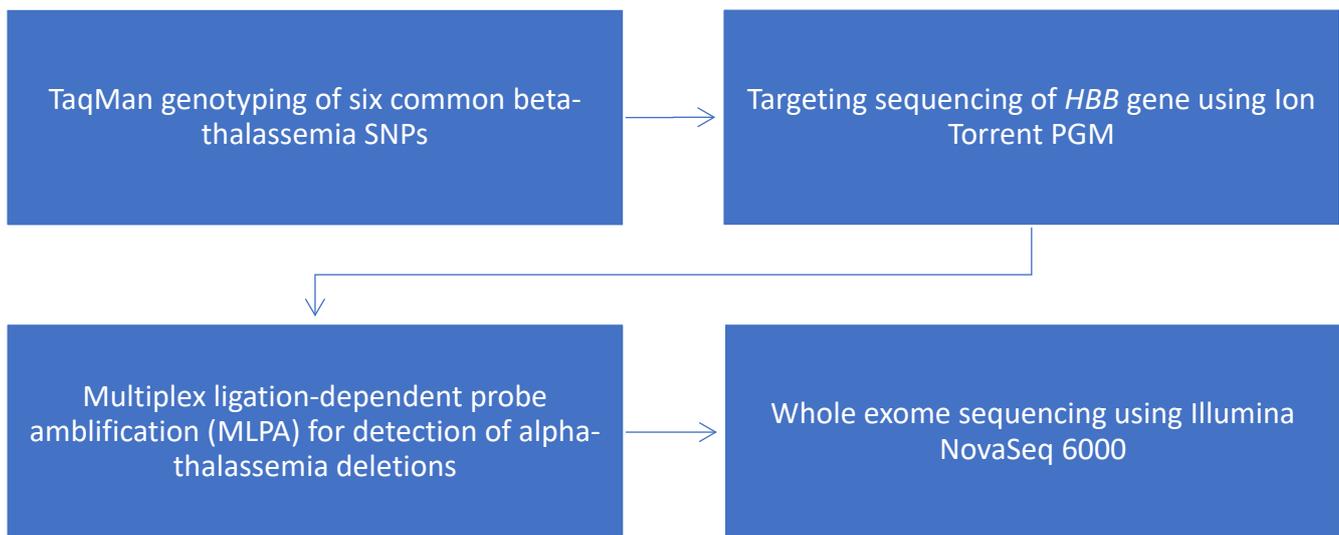


Figure 52: Visual diagram of molecular techniques utilized in attempt to identify disease causing variations of sample 183.

First, TaqMan genotyping was performed on DNA sample using six of the most frequent beta thalassemia variations in Saudi Arabia. With negative results from the genotyping, targeted sequencing of *HBB* gene using Ion Torrent PGM was performed but no variation was detected. After that, to rule out alpha thalassemia, we performed MLPA testing, and no deletions were detected. Finally, we detected the PKLR variant using whole exome sequencing.

5.5 Results

After molecular investigation using various sequencing techniques, neither TaqMan genotyping nor NGS *HBB* targeted sequencing revealed any mutation related to her initial suspected diagnosis of beta-thalassemia. Also, MLPA did not yield any deletions in her alpha globin genes related to alpha thalassemia (Figure 53).

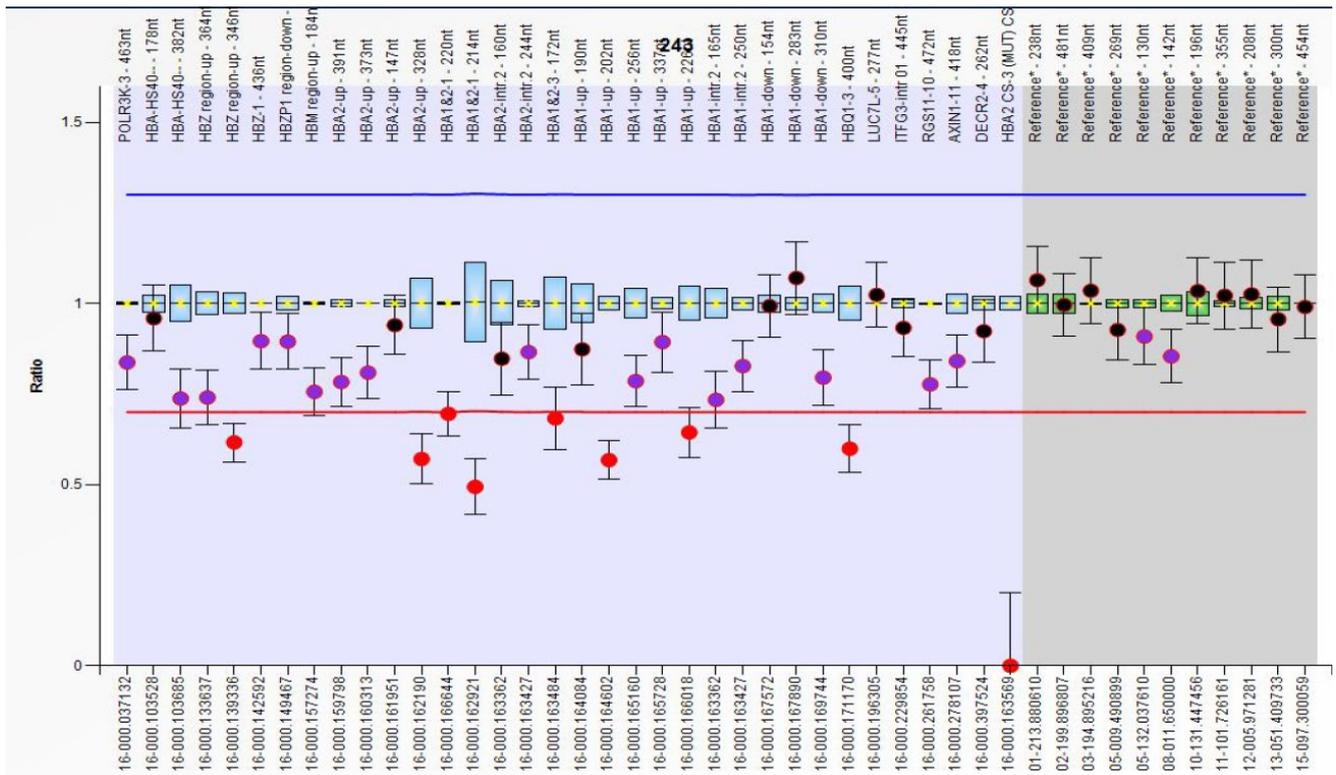


Figure 53: MLPA reaction result showing no significant deletions.

A score ratio of 1 indicates no change in copy number compared to reference sample. On the other hand, a score of 0.5 indicates a heterozygous deletion while a score of 1.5 indicates a heterozygous duplication. The blue boxes represent the 95% confidence interval of a probe over the reference samples. The colored dots are the calculated probe ratio. The error bars surrounding the dots represent a 95% confidence interval estimate for each probe in a sample. The red and blue lines display lower arbitrary border and upper arbitrary border respectively. These borders are +/- 0.3 from the average value of a probe over the reference sample. The black dot indicates no change in probe copy number because the 95% confidence interval estimates (the error bars) of the probe overlap with the 95% confidence intervals of the same probes over the reference sample. The purple dot indicates a decreased signals of more than two standard deviations compared to the reference samples, but the lower arbitrary border has not been crossed. The pink dot indicates an increased signals of more than two standard deviations compared to the reference samples, but the upper arbitrary border has not been crossed. The red dots indicate a decreased signals of more than two standard deviations compared to the reference samples and the lower arbitrary border has been crossed. The red dot far down in the ratio chart with one error bar indicates no signal found. Source: Coffalyser.Net reference manual v01.

WES analysis was then used to get a detailed view of the protein-coding regions of her whole genome. we have generated a gene list to include genes that have been reported to be involved in one way or another in hereditary anemia disorders. Generating gene panels to investigate inherited hemolytic anemia disorder have been previously employed (Agarwal et al., 2016) (Roy et al., 2016). Our list was

assembled by looking into literature and several gene panels that are used by clinical laboratories to test for anemia related disorders. This list included, ABCB7, ABCG5, ABCG8, ADA, ADA2, ADAMTS13, ADH5, AK1, ALAS2, ALDOA, AMN, ANK, ANK1, ATRX, BCL11A, BLM, BRCA2, BPGM, BRIP1, C15ORF41, CD59, CDAN1, CDIN1, CYB5A, CYB5R3, COL4A1, CUBN, DKC1, DHFR, DNAJC21, DNASE2, EFL1, EBP41, EBP42, ENO1, ERCC4, ERFE, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCM, FANCL, FTCD, G6PD, GATA1, GCLC, GIF, GLUT1, GLRX5, GPC, GPI, GSR, GPX1, GSS, GYPC, HEATR3, HK1, HSCB, HSPA9, HMOX1, KCNN4, KIF23, KLF1, LARS2, LPIN2, MTR, MTRR, NDUFB11, NHLRC2, NHP2, NT5C3A, NRF1, PALB2, PC, PDHA1, PDHX, PIEZO1, PFK, PFKM, PGK1, PKLR, PUS1, RAD51C, REN, RHAG, RPL11, RPL15, RPL18, RPL27, RPL31, RPL35A, RPL5, RPS10, RPS17, RPS19, RPS24, RPS26, RPS27, RPS28, RPS29, RPS7, SBDS, SEC23B, SLC19A1, SLC11A2, SLC19A2, SLC25A38, SLC4A1, SLC2A1, SF3B, SLX4, STEAP3, SPTA1, SPTB, SRP54, TCN2, TF, THBD, TMPRSS6, TRNT1, TSR2, TPI1, UMPS, UGT1A1, VPS4A, XK, YARS2. The advantages of using gene panels as a filtration search scenario include shorter turnaround times, easier data processing, more coverage in the regions of interest, and fewer incidental results (Wooderchak-Donahue et al., 2012) (Sun et al., 2015) (Kim et al., 2017).

A homozygous missense mutation was found on exon 7 of *PKLR* gene that has high potential to correlate and explain her symptoms and phenotype (Table 15) (Figure 54). This missense mutation leads to a change of amino acid number 339 and transforms its translated protein from aspartic acid (negatively charged) to uncharged asparagine (D339N). The overall allele frequency of this variation in the gnomAD v2.1.1 dataset is 0.002%. This variation has not been reported before in Saudi population. The variation has been verified using Sanger sequencing (Figure 55)(Figure 56).

Table 15: Whole exome sequencing result of patient 183.

Sample	HGVS name	Mutation Status	Genomic Location on Assembly GRCh37	Location on gene	Molecular consequence	Reference cluster ID (rs#)
14-0183	NM_000298.6 c.1015G>A	Homozygous	Chr1: 155264127	Exon 7	Missense SNV NP_000289.1 P.Asp339Asn	rs747097560 C-T

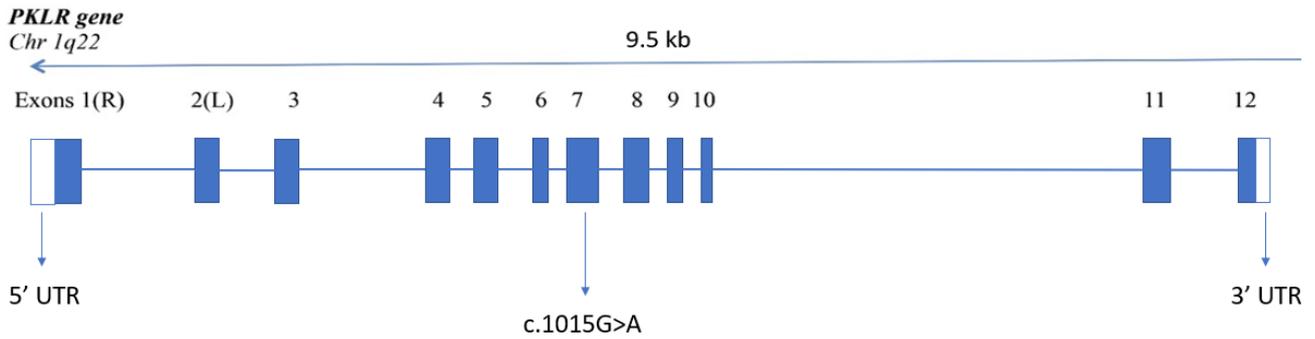


Figure 54: *c.1015G>A* variation mapped on *PKLR* gene.

R indicates erythrocytic isozyme, while *L* indicates hepatic isozyme which are both encoded by the same gene but with different promoters. With 574 amino acids, the *R*-type subunit is 31 residues longer than the *L*-type at its amino terminus.

The hematologist was presented with clinical manifestation supporting of PKD that includes hemolytic anemia with no known etiology, and severe neonatal jaundice that progressed to bilirubin encephalopathy with lifelong neurological damage, the hematologist investigated PKD as potential diagnosis which was clinically confirmed by testing pyruvate kinase enzyme activity levels (before transfusion) which appeared to be low.

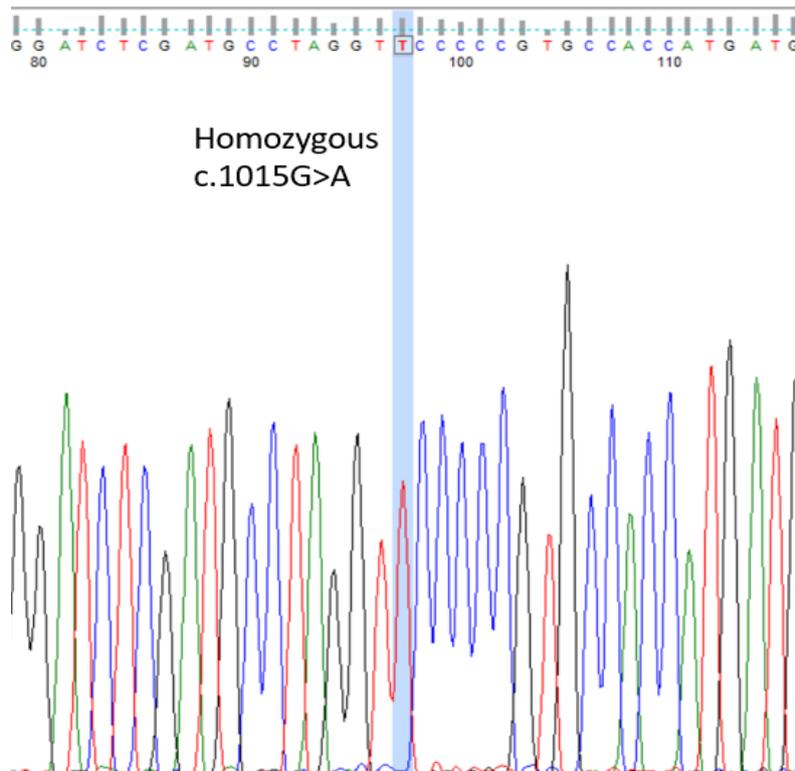


Figure 55: Sanger validation of *PKLR* mutation.

Score	Expect	Identities	Gaps	Strand
329 bits(178)	8e-86	180/181(99%)	0/181(0%)	Plus/Plus
Query 1	GTCGCTATTCCCCATCACCTTTCTTCTCCTGCCTGCCTCTGCCTTGATTCTCCCAACCTC	60		
Sbjct 7496	GTCGCTATTCCCCATCACCTTTCTTCTCCTGCCTGCCTCTGCCTTGATTCTCCCAACCTC	7555		
Query 61	TCAGGTTTGATGAAATCCTGGAGGTGAGCGACGGCATCATGGTGGCACGGGGGAACCTAG	120		
Sbjct 7556	TCAGGTTTGATGAAATCCTGGAGGTGAGCGACGGCATCATGGTGGCACGGGGGGACCTAG	7615		
Query 121	GCATCGAGATCCCAGCAGAGAAGGTTTTCTGGCTCAGAAGATGATGATTGGGCGCTGCA	180		
Sbjct 7616	GCATCGAGATCCCAGCAGAGAAGGTTTTCTGGCTCAGAAGATGATGATTGGGCGCTGCA	7675		
Query 181	A	181		
Sbjct 7676	A	7676		

Figure 56: Alignment of PKLR gene Sanger Sequence.

5.6 Data Analysis

During our study, the homozygous missense variant found in our proband had been also identified and associated with PKD phenotype in a case report in 2022. Rehman and his group have reported this bi-allelic variant in the PKLR gene and correlated them to PK deficiency found in a consanguineous Pakistani family (Rehman et al., 2022) in which there were four affected children.

The PKLR gene encodes for the pyruvate kinase enzyme, which catalyzes the transfer of a phosphate group from 2-phosphoenolpyruvate (PEP) to ADP, forming pyruvate and ATP. This is the final stage in the glycolytic process, and under normal conditions, the reaction is not reversed. Mammals, including humans, possess four distinct PK isozymes. Alternative splicing of *PKLR* gene transcripts results in mRNAs encoding either the M1 (muscle) or M2 (fetal) proteins. The sole difference between PKM1 and PKM2 mRNAs is whether they include one of two overlapping exons or not, PKM1 (8-9-11) while PKM2 (8-10-11) (M. Chen et al., 2010). Fructose 1,6-bisphosphate (FBP) and phosphoenolpyruvate (PEP) allosterically activate M2 PK, while simple saturation hyperbolic kinetics are exclusive to the M1 enzyme. Using tissue-specific alternative promoters, the same PKLR gene codes for the mammalian liver and erythrocyte PK isozymes. PEP and FBP stimulate isozymes in both erythrocytes and the liver (P. et al., 2019) (Schormann et al., 2019).

PK enzyme is a tetramer made up of four subunits (Figure 57). Each subunit has four domains: a domain with (β/α) 8 barrel topology called domain A, domain B which is located between alpha-helix 3 and beta-sheet 3 of domain A, domain C with $\alpha+\beta$ topology, and a small alpha-helical domain called the N-terminus domain (Figure 58) (Valentini et al., 2002).

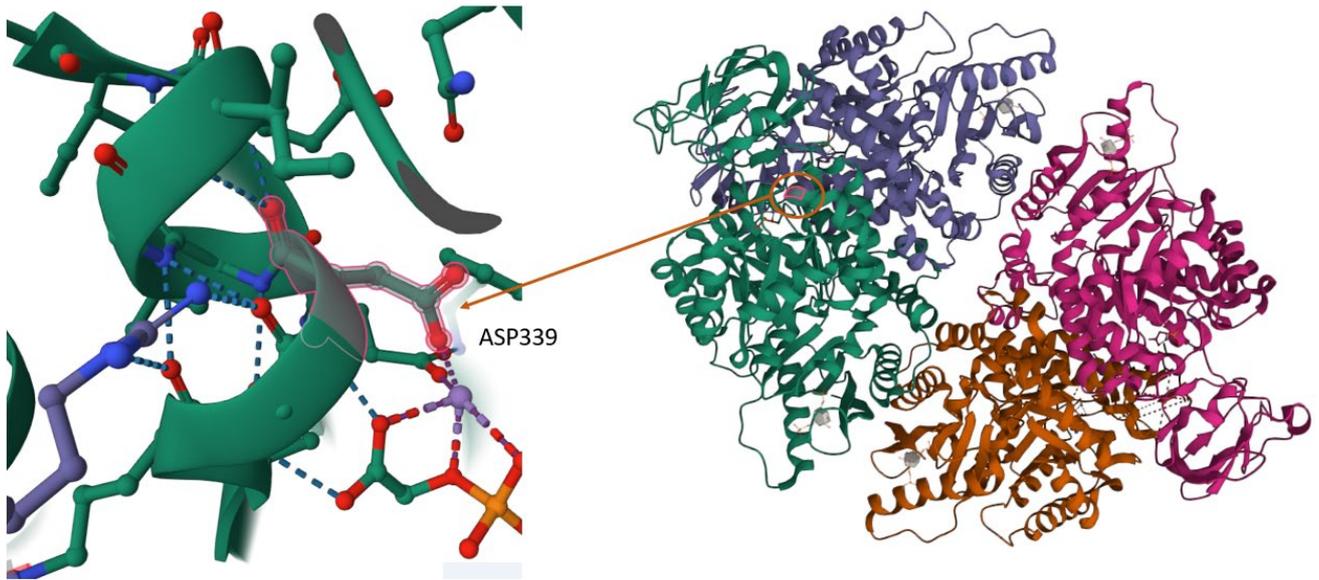


Figure 57: Pyruvate Kinase 3D structure.
Adapted from PDB database (2VGI) (Valentini et al., 2002).

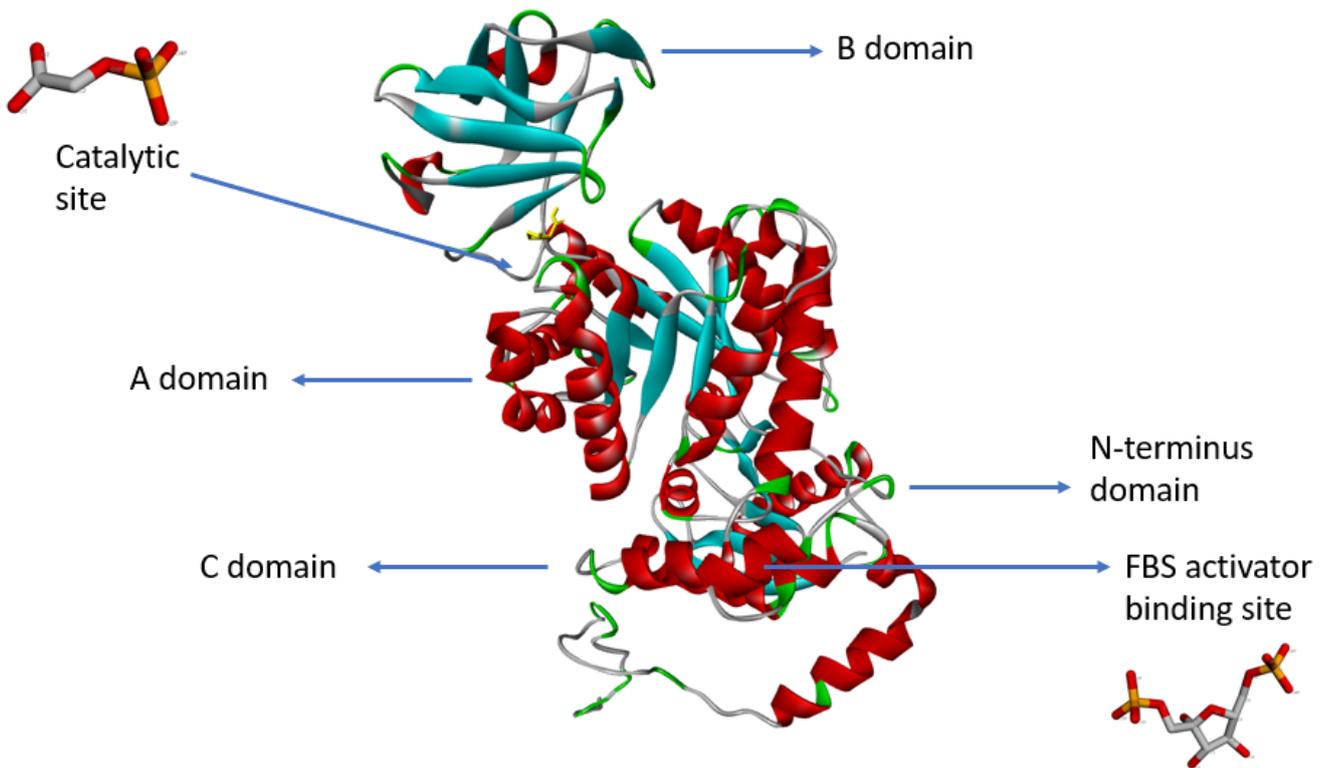


Figure 58: Pyruvate kinase 3D structure.

The control of PK activity relies heavily on this multidomain structure. The activation of the enzyme involves a series of rotations in the enzyme's domains and subunits, as well as changes to the geometry of the enzyme's active site. Residues at the domain and subunit interfaces play critical roles in this mechanism by mediating interactions between the activator-binding site (inside the C domain) and the

catalytic core (between the domains A and B) (Zanella et al., 2007) (Bianchi & Fermo, 2020). Fructose 1,6-bisphosphate acts as an allosteric activator of red cell pyruvate kinase, whose reaction kinetics follow a sigmoidal shape with respect to PEP.

The catalytic mechanism of pyruvate kinase involves metal binding, a Mg^{2+} and K^+ ions, that is coordinated by several residues in the active site. The ions play a critical role in stabilizing the negative charges that are found during the reaction, facilitating the transfer of the phosphate group. The metal binding properties of pyruvate kinase are important for understanding the enzyme's function and regulation (Murakami & Yoshino, 2017).

The carboxylate side chain of Asp339 (in active site) is involved in the coordination of the Mg^{2+} , which is critical for the enzyme's catalytic activity. In studies of X-ray crystal structure, the inhibitor 2-Phosphoglycolic acid is bound at the PEP-binding site of the A domain, where protein residues and the cations Mn^{2+} (or Mg^{2+}) and K^+ form a complex network of hydrogen bonds. K^+ ions and the side chain of Arg116 bind the phosphate group, whereas Mn^{2+} ions, the side chain of Thr371, and the main chain nitrogen atoms of Gly338 and Asp339 at the N terminus of a short helical segment belonging to loop 6 of the A domain, anchor the carboxylate moiety. These interactions help to stabilize the Mg^{2+} ion and orient it in the active site, allowing it to facilitate the transfer of the phosphate group from PEP to ADP (Figure 59)(Figure 60) (Schormann et al., 2019). Mutations in Asp339 which involves Mg^{2+} binding can impair the catalytic activity of pyruvate kinase and alter the substrate specificity or allosteric regulation of the enzyme, leading to changes in metabolic pathways and cellular functions.

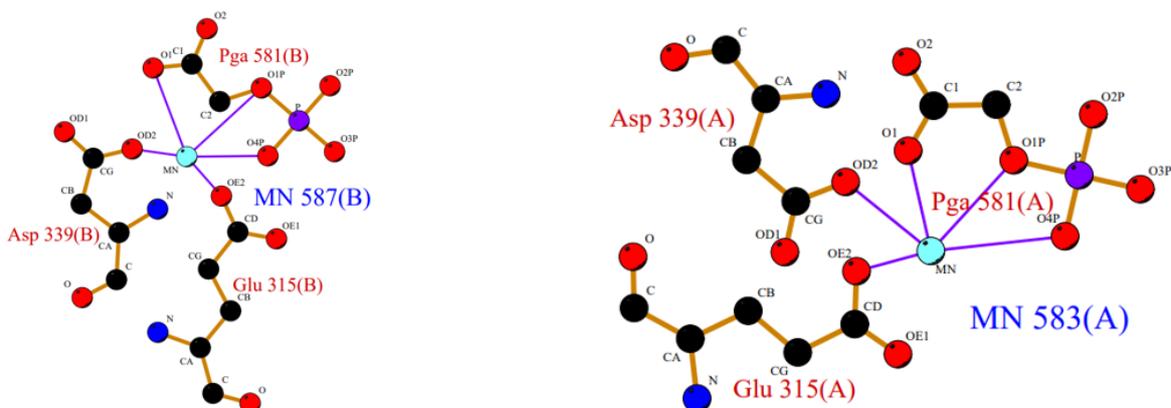


Figure 59: Crystallography Ligand with 2-phosphoglycolic acid/metal interaction involving metal ion Mn^{2+} in both chain A and chain B using LIGPLOT (Wallace et al., 1995).

The change from Asp to Asn would remove the negative charge from the ligand.

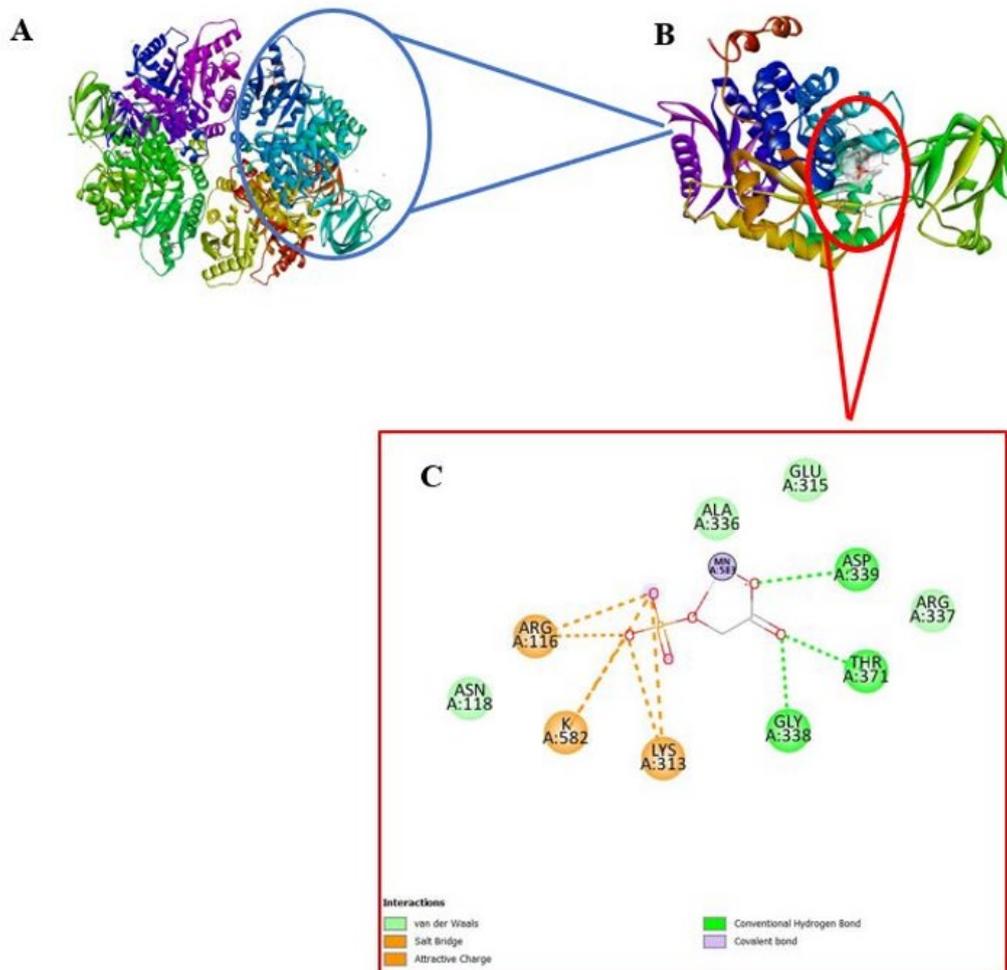


Figure 60: Pyruvate Kinase active site.

A represents the 3D structure of the enzyme tetramer. B is a single chain in the enzyme. C represents the 2D structure of the enzyme's active site bound to its substrate 2-phosphoglycolic acid. The homology modelled 3D structure of pyruvate kinase using I-TASSER server.

Asp339 is a highly conserved residue located in the active site of pyruvate kinase (Figure 61). Based on a search of the UniProt database, Asp339 is highly conserved across different organisms, suggesting its functional importance in the enzyme. Asp33 is conserved in all vertebrate pyruvate kinase isoforms (Figure 62), including human (P30613), mouse (P52480), zebrafish (Q919N4), Xenopus (Q90XJ9), and is also conserved in other organisms such as fruitfly (P07132) and yeast (P00887).

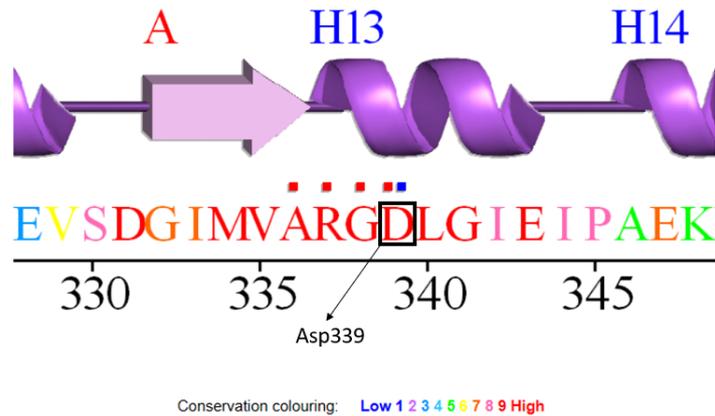


Figure 61: Amino Acid sequence of PKLR gene showing Asp339 residue conservation.
Residue conservation calculated by ConSurf-DB (Ben et al., 2020)

277	RRFD---	EILEASD	GIMVARGDLGIEIPAEKVF	LAQKMMI	GRGNR	AGKPVICATQ	MLESMIKKPRP	TRAEGSDVANAVLD	353	cat
248	NNFD---	EILKVDG	VMVARGDLGIEIPAEVLA	VQKLLI	AKSNL	AGKPVICATQ	MLESMTYNPRP	TRAEVSDVGNAILD	324	baker's yeast
228	NNFD---	EILEASD	GIMVARGDLGVEIPVEEVI	FAQKMMI	EKCIR	ARKVVITATM	MLDSMIKNPRP	TRAEAGDVANAILD	304	Escherichia coli
262	NNFD---	EILEETD	GVMVARGDLGIEIPAPKVF	IAQKMMI	AKCNI	KGKPVICATQ	MLESMTYNPRP	TRAEVSDVANAVLD	338	Emericella nidulans
271	NNFA---	EILEETD	GVMVARGDLGIEIPAAEVFA	AAQKMMI	AMCNI	AGKPVICATQ	MLESMIKNPRP	TRAEISDVGNVTD	347	Hypocrea jecorina
259	NNFD---	EILKETD	GVMVARGDLGIEIPAPQVF	IAQKQLI	AKCNL	AGKPVICATQ	MLDSMTYNPRP	TRAEVSDVGNVTD	335	Yarrowia lipolytica
248	INFD---	EILRETD	SFMVARGDLGMEIPVEKIF	LAQKMMI	YKCNL	LAGKAVVTATQ	MLESMIKSPAP	TRAEATDVANAVLD	324	potato
229	ANID---	EILEAAD	GLMVARGDLGVEIPAAEVL	LIQKLLI	KKCNML	GKPVITATQ	MLDSMQRNPRP	TRAEASDVANAIFD	305	Geobacillus stearo...
231	ANDEamd	DIILASD	VIMVARGDLGVEIGDPELV	GVQKLLI	RRSRQL	NRAVITATQ	MMESMISNPM	PTRAEVMDVANAVLD	310	Haemophilus influe...
229	DNID---	EILQVSD	GLMVARGDMGVEIPFINVP	FVQKTLI	KKCNAL	GKPVITATQ	MLDSMQENPRP	TRAEVTDVANAVLD	305	Lactobacillus delb...
248	NNFD---	EILKVDG	VMVARGDLGIEIPAEVLA	VQKLLI	AKSNL	AGKPVICATQ	MLESMTYNPRP	TRAEVSDVGNAILD	324	baker's yeast
228	NNFD---	EILEASD	GIMVARGDLGVEIPVEEVI	FAQKMMI	EKCIR	ARKVVITATM	MLDSMIKNPRP	TRAEAGDVANAILD	304	Escherichia coli
262	NNFD---	EILEETD	GVMVARGDLGIEIPAPKVF	IAQKMMI	AKCNI	KGKPVICATQ	MLESMTYNPRP	TRAEVSDVANAVLD	338	Emericella nidulans
271	NNFA---	EILEETD	GVMVARGDLGIEIPAAEVFA	AAQKMMI	AMCNI	AGKPVICATQ	MLESMIKNPRP	TRAEISDVGNVTD	347	Hypocrea jecorina
259	NNFD---	EILKETD	GVMVARGDLGIEIPAPQVF	IAQKQLI	AKCNL	AGKPVICATQ	MLDSMTYNPRP	TRAEVSDVGNVTD	335	Yarrowia lipolytica
248	INFD---	EILRETD	SFMVARGDLGMEIPVEKIF	LAQKMMI	YKCNL	LAGKAVVTATQ	MLESMIKSPAP	TRAEATDVANAVLD	324	potato
229	ANID---	EILEAAD	GLMVARGDLGVEIPAAEVL	LIQKLLI	KKCNML	GKPVITATQ	MLDSMQRNPRP	TRAEASDVANAIFD	305	Geobacillus stearo...
231	ANDEamd	DIILASD	VIMVARGDLGVEIGDPELV	GVQKLLI	RRSRQL	NRAVITATQ	MMESMISNPM	PTRAEVMDVANAVLD	310	Haemophilus influe...
229	DNID---	EILQVSD	GLMVARGDMGVEIPFINVP	FVQKTLI	KKCNAL	GKPVITATQ	MLDSMQENPRP	TRAEVTDVANAVLD	305	Lactobacillus delb...

Figure 62: The amino acid, D (Asp339), conservation across different organisms.
Source: CCD database (cd00288) (Lu et al., 2020).

Aspartic acid (Asp) and asparagine (Asn) are structurally similar amino acids, with Asn having an additional amide group in its side chain. However, Asp is negatively charged, while Asn is uncharged. Therefore, the substitution of Asp339 with Asn would eliminate the negatively charged carboxylate group that is crucial for the coordination of Mg^{2+} in the active site. This could lead to a reduction or even loss of Mg^{2+} modelled binding, impairing the catalytic activity of pyruvate kinase, potentially leading to significantly functional consequences, metabolic disorders, and physiological effects affecting its interaction with other molecules and regulatory factors.

On a similar note, in 1997, a variation that produces the amino acidic substitution Asp390Asn was found by Zanella and his research group while studying a group of Italian patients with pyruvate kinase deficiency (Zanella et al., 1997). At the A/A interface, Arg337 (positively charged) and Ser389 (uncharged) from two separate subunits form a hydrogen bond network that prevents Asp390

(negatively charged) from interacting with the solvent. Crystallographic analyses of *E. coli* PK reveal that Asp390, via connecting quaternary structural alterations with active site rearrangements, is essential for the enzyme's allosteric transition. Molecular investigation reveals that replacing Asp390 with Asn results in a nearly inactive protein that is. This suggests that the Asp390Asn mutation prevents the protein from transitioning to its active R state (Zanella et al., 2005) (Valentini et al., 2000) (Mattevi et al., 1995).

In our investigation of the variation, we have modelled of the mutated protein and compared it with the normal protein, which resulted in noticeable changes with regards to their protein-ligand 2D chemical interactions, hydrophobicity, H-bonds, and charges.

Molecular docking was performed for the normal PKLR protein and the mutated PKLR (Asp339Asn) domains. First, we used both NCBI (PDB ID: 2VGB) and UniProt (ID: P30613) databases for the biological data collection of normal PKLR enzyme. I-TASSER server was used for modeling of the 3D protein structures of normal and mutated PKLR enzyme, and each compound structure was obtained from PubChem database. For energy minimizing process, Swiss PDB Viewer (spdbv) was used to modify all compound structures. Format conversion from pdb to pdbqt was done by using Open Babel (Version 2.3.1) software. After that, Discovery Studio software (Version 2019) was used to handle the alteration of each protein and ligand by adding hydrogen atoms and metals, before the molecular docking process. , Auto Dock Vina (Version 2.0) was used to define the grid box with 1.00 Å spacing and a grid map of 72 X, 76 Y, 84 Z Å points for the mutated PKLR with the predicted new substrate (proline) and 14 X, 16 Y, 10 Z Å points for the mutated PKLR with its normal substrate. Auto Dock was used to rank Van der Waals interactions, binding energy and inhibition constant (Mohamed et al., 2022).

The analysis of the binding sites revealed the pockets for each protein and the nature of the pockets using the Discovery studio program. The protein pockets and the amino acids involved in binding can help us make an estimate of how H-bonds, chemical interactions, hydrophobic/hydrophilic, positively/negatively charged, and what type of substrate may bind to that specific protein.

The 2D chemical interactions comparison between the normal and mutated PKLR showed that the mutated form will not form regular interactions between the active site and its modelled substrate (2-phosphoglycollic acid). The normal substrate is 2-phosphoenolpyruvic acid. The phosphoglycollic acid is a substrate-model/inhibitor used in the crystallography because ADP cannot be added with PEP because they would react (Figure 63).

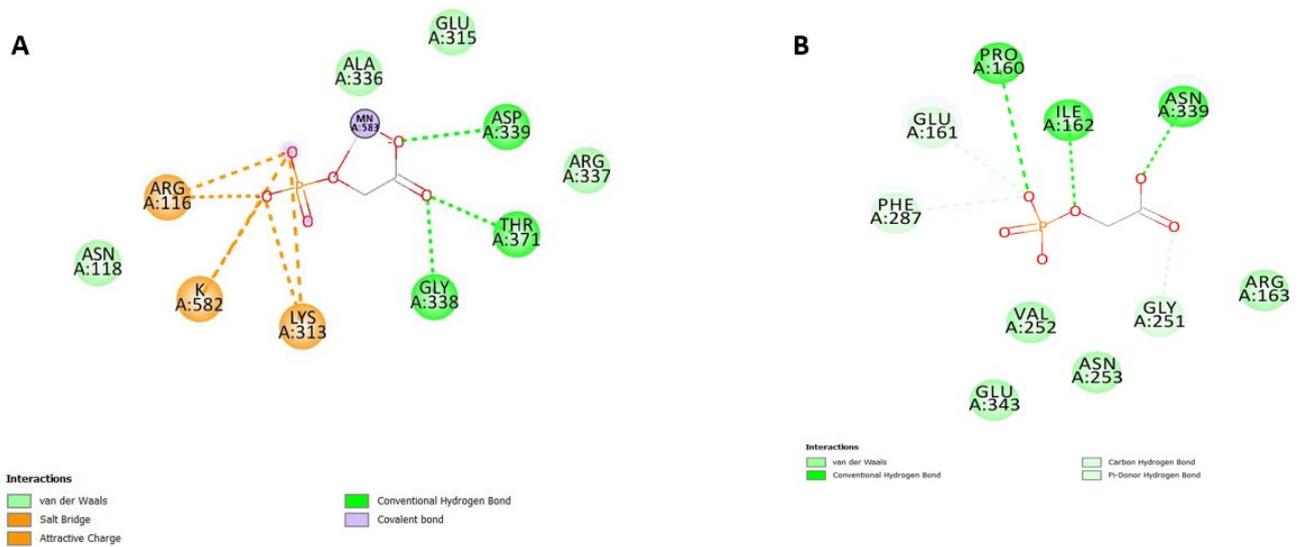


Figure 63: A comparison in 2D structure of the active site chemical interactions.

A) The normal domain of PKLR. B) The mutated domain with “2-phosphoglycolic acid”.

H-bonds are shown to control molecular interactions via a donor-acceptor pairing process that decreases water competition (Figure 64). Moreover, each charge within the protein will communicate with the surrounding solvent (Figure 65). When there are more charges present, those charges interact with one another, but the solvent dampens the intensity of those interactions. There is a substantial impact of charged particle interactions with solvents and solvent filtration of charge-charge interactions on the electrostatic energy of proteins. All these protein properties exhibited significant change between normal and mutated PKLR. Additionally, normal and mutated PKLR have also shown a significant difference in hydrophobicity (Figure 66).

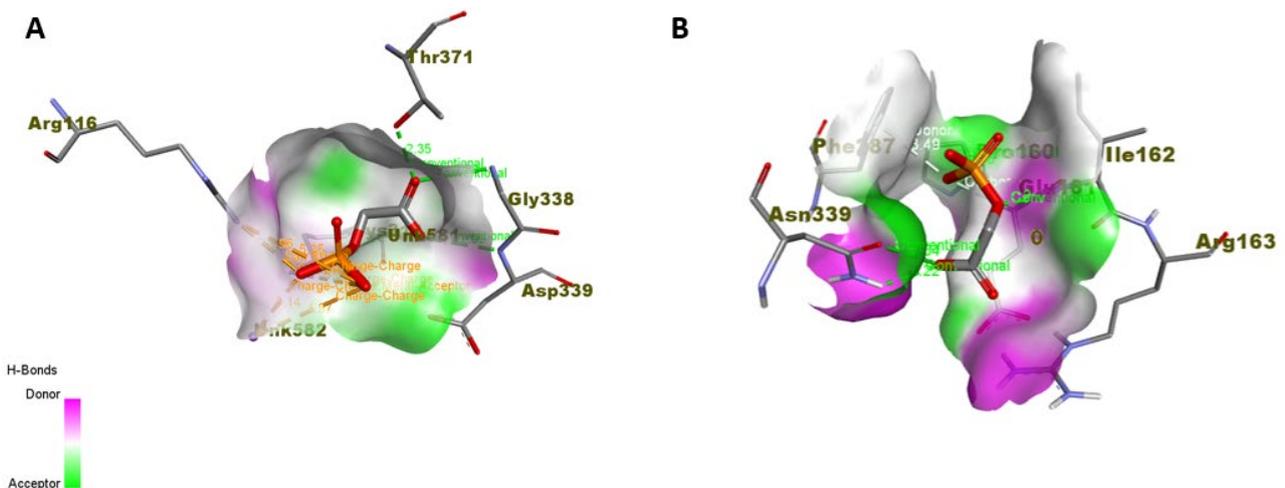


Figure 64: A comparison in H-bonds.

A) The normal domain of PKLR. B) The mutated domain with “2-phosphoglycolic acid”. H-bonds as donors appeared with pink color, while H-bonds as acceptors appeared with green color. H-bonds as donors through the docked of both mutated PKLR complex were more than the normal chain-substrate complex.

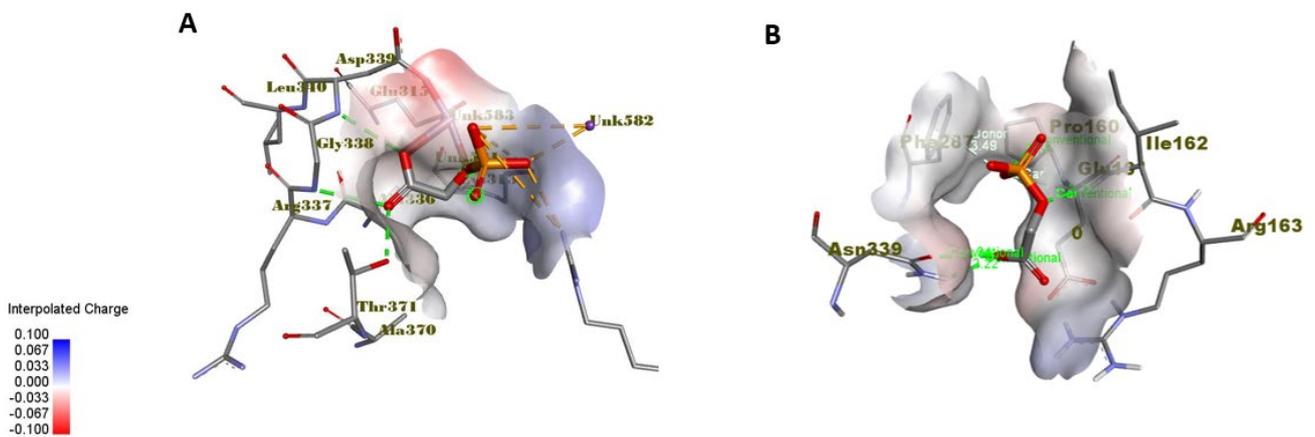


Figure 65: A comparison of protein's charges.

A) The normal domain of PKLR. B) The mutated domain with "2-phosphoglycolic acid". Positive charges appeared with blue color while negative charges appeared with red color. The docked chain A of normal PKLR both negative and positive charges, while the docked mutated PKLR charges were mostly neutral.

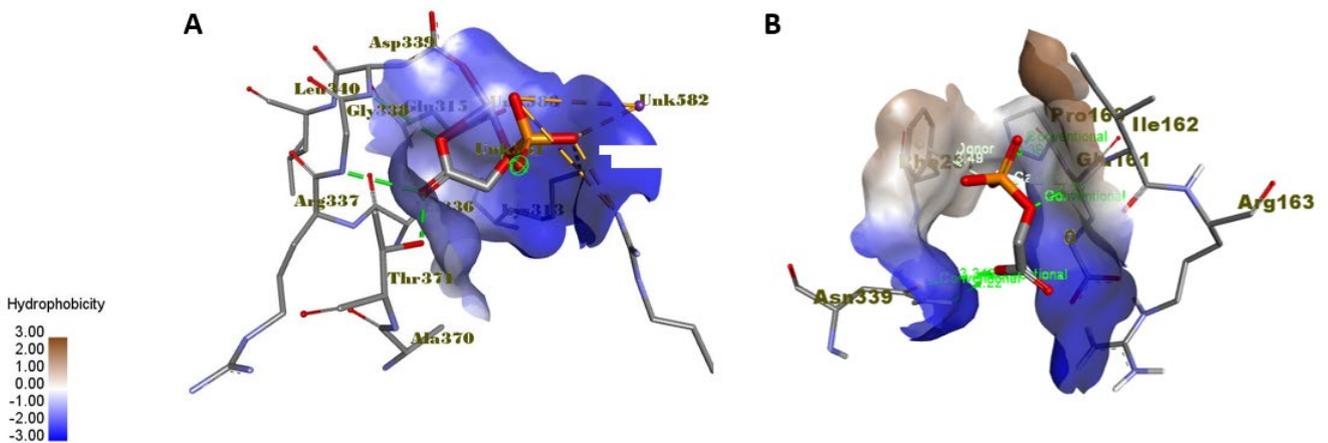


Figure 66 A comparison in hydrophobicity.

A) the normal domain of PKLR B) The mutated domain with "2-phosphoglycolic acid". Hydrophobicity appeared with brown color, while hydrophilicity appeared with blue color. The docked chain A of normal PKLR was mostly hydrophilic, while the docked mutated PKLR substrate complex had hydrophobic regions.

Additionally, according to the 3D modeling of the mutant protein by Rehman and colleagues, with the evaluating of protein-ligand interaction showed that the mutant protein established new connections through Arg216 and Glu347 and lost its regular interactions with phosphoenolpyruvate, which causes poor or nonexistent dephosphorylation of phosphoenolpyruvate that leads to deficiency in glycolysis's ability to produce energy (Rehman et al., 2022).

Furthermore, we have used several in-silico prediction tools available online to predict pathogenicity or damaging effect of the amino acid substitution on the protein. The substitution at position 339 from aspartic acid (D) to asparagine (N) is predicted to affect protein function with a score of 0.00 (predicted damaging if the score is equal or less than 0.05), and median sequence conservation of 2.99 (Prediction-sequence diversity is measured using this metric) by using SIFT (sorts intolerant from tolerant), which is a sequence homology-based tool used to estimate the potential phenotypic impact of a protein-level amino acid change. Protein function and evolutionary history are assumed to be linked in SIFT. Alignments of protein families should show conservation at functionally relevant locations while showing diversity at less critical places (Ng & Henikoff, 2003).

SNAP² (Screening for Non-Acceptable Polymorphisms) had predicted an effect on protein function as well with a score of 84 and expected accuracy of 91%. SNAP uses a neural network to anticipate how a SNP may influence protein's function. When available, SNAP makes use of evolutionary data to determine which residues within a given sequence family are conserved, as well as other features of the protein's structure and annotations (Bromberg & Rost, 2007).

Additionally, another in-silico tool was used to predict the effect of the amino acid substitution, PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms), which is a machine learning method that employs a supervised training algorithms to determine if a given SNP is connected with disease by analyzing the protein sequence, also resulted in predicting a disease-related polymorphism in corresponding to the change in the amino acid sequence D339N (Capriotti & Fariselli, 2017). Also, Mutation Assessor, which is based on the evolutionary conservation of the damaged amino acid in protein homologs, provides an approximate assessment of the chance that the mutation has a phenotypic effect at the organism level. This web-based tool has predicted a high functional impact of this missense mutation on the resulted protein (Reva et al., 2011).

PANTHER tool calculates the probability that a particular SNP will lead to a change in the protein's function and presents the outcome as a percentage. It does this by computing a score known as subPSEC, which stands for substitution position-specific evolutionary conservation, based on an alignment of evolutionarily related proteins using a hidden Markov model (Thomas et al., 2003). PANTHER predicted the variation as probably damaging with a probability of deleterious effect score of 0.95.

Moreover, MutPred2 was also used in pathogenicity prediction, which resulted in a score of 0.92 (equal or more than 0.5 is pathogenic) (Pejaver et al., 2020).

Furthermore, the mutation was also categorized as “Probably damaging” by using another in-silico software tool, PolyPhen-2, which is a Bayesian classifier that determines the likelihood that a given non-synonymous SNP is deleterious. The Location-Specific Independent Count (PSIC) score, which ranges from 0 to 1, is based on the evolutionary conservation of a protein sequence in the MSA (multiple sequence alignment) and the negative effect on the protein structure (Adzhubei et al., 2010). The use of different in-silico prediction tools has shown that all of them have predicted the identified variant to be deleterious, pathogenic, and probably damaging which makes it the best candidate to be associated with the disease.

5.7 Conclusion

With the dramatic increase of the number of molecular mutations, and clinical misdiagnosis, NGS technologies have led to increased knowledge of rare congenital conditions. While none of the proband’s family members were clinically affected or had the disease, it was present in the proband in a homozygous state. Our data collectively suggest that c.1015G>A missense variation of the PKLR gene may decrease or abolish PK enzymatic activity, resulting in PK deficiency in the affected individuals.

To further expand our knowledge of this variation and show its pathogenic character, functional consequences of the variant should be investigated experimentally such as PK enzymatic assays, biomarkers, and expression analysis should be performed in future work.

CHAPTER SIX

**EXPLORING THE MOLECULAR HETEROGENEITY OF A FAMILY
EXHIBITING THALASSEMIA MAJOR**

6.1 Use of whole exome sequencing to elucidate the genetic bases of transfusion-dependent patients

After failing to identify mutations on *HBB* gene that would explain the phenotype severity of three transfusion-dependent patients using multiple sequencing techniques, two heterozygote carriers and one with no identifiable *HBB* variants, we used whole exome sequencing (using the Illumina platform) to explore all mutants, not only in *HBB*, but also in all genes related to thalassemia and anemia.

The procedure was carried out as described in the Experimental Procedures of Chapter two. It starts with library preparation and enrichment. After that, the library was hybridized to an Illumina flow cell, sequenced and data is interpreted using Illumina BaseSpace (Table 16)(Table 17)

Table 16: QC metrics of thalassemia samples' results.

Sample	Percent Q30	Duplication rate	Total aligned reads	Total aligned targeted reads	% read enrichment	Mean region coverage depth
II-17	98.61	7.77	87548077	64417715	73.58	104.02
14_0183	98.75	9.78	114623265	88377592	77.10	139.66
II-14	98.62	9.67	130340744	98415991	75.51	155.12

Table 17: QC metrics of thalassemia samples' results continues.

Sample	Uniformity of coverage (Pct > 0.2*mean)	Depth of Sequencing Coverage at 30x	Depth of Sequencing Coverage at 100x
II-17	97.73	94.23	47.64
14_0183	97.59	96.53	68.85
II-14	97.76	97.29	74.80

When sequencing was completed, the run folder was copied from the NovaSeq 6000 machine to the University high performance computer (HPC). The in-house GenaTi NGS analysis pipeline reads the run folder. The main steps in the workflow of variant calling are described in chapter 2.

6.2 Results and data analysis

As mentioned before, we have two cases with only heterozygous beta-thalassemia pathogenic mutation, c.27dupG, which is located on exon 1 of *HBB* gene which is expected to cause very early premature termination. Upon clinical history investigation, we have found that these two cases are siblings from the same mother (samples II-14 and II-17), and that they have five more affected siblings and half-siblings. All seven were diagnosed with thalassemia major as they require frequent transfusions. Since the father is the common factor between the two families, our initial assumption

was that the father is the carrier of the frameshift *HBB* mutation, and we should be looking at the genetic contribution from the mothers that caused the children to exhibit thalassemia major. However, this did not prove to be the case.

Samples were collected from family members depending on their availability. A total of 14 samples were collected from the father, two mothers, and from their affected (five out of the actual seven) and non-affected children (six samples), and whole exome sequencing was performed on their DNA to analyze the sequences of both beta globin and alpha globin gene clusters in more detail. Interestingly, the father did not have a mutation within his *HBB* gene, instead, we found that each mother was a carrier of a different known pathogenic gene-disrupting alleles c.92+5G>C and c.27 dupG on *HBB* in a heterozygous state, while the father did not appear to have variations in any of the globin genes. However, he was hemizygous for the non-synonymous missense G6PD variant c.653 C>T (p.Ser218Phe) that has been reported to cause X-linked hemolytic anemia due to deficiency of G6PD activity when coupled with an increased reactive oxidative stress production caused by environmental triggers such as certain food or medication (Richardson & O'Malley, 2023). Mother I-1 is coincidentally homozygous for the same G6PD mutation (Figure 67).

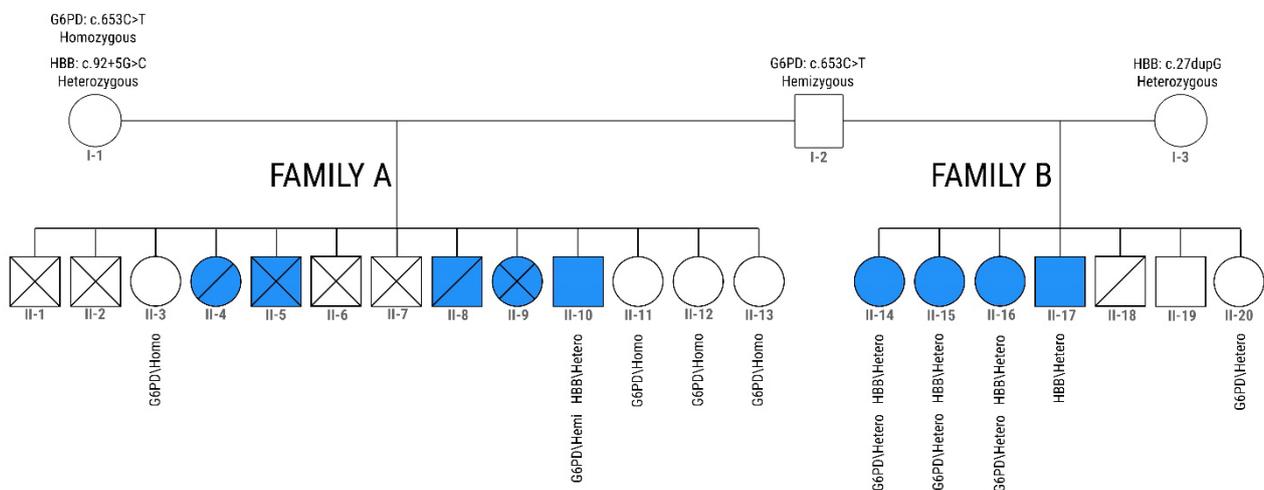


Figure 67: Pedigree of a family with beta-thalassemia major phenotype showing pathogenic variants in two genes, *HBB* and *G6PD*.

The Roman numbering, I and II, indicate the two generations. Shapes with diagonal lines are deceased. Blue color is affected (transfusion dependent) while white is unaffected. Crossed out shapes are not investigated.

Neither mother is closely related to the father (non-consanguineous marriages). From family A, five of the 13 children were transfusion dependent and so categorised as thalassemia major patients. Two of whom are deceased, two were unavailable for testing, while the remaining affected son (II-10) had both *G6PD* (hemizygous) and *HBB* pathogenic (heterozygous) variations with severe phenotype. On

the other hand, from family B, four of the seven children are transfusion-dependent patients, three of whom are female and one is male. The affected females (II-14, II-15, and II-16) had inherited both pathogenic variations, *G6PD* from their father (heterozygous) and *HBB* from their mother (heterozygous), whereas the male (II-17) had only inherited the heterozygous *HBB* mutation from his mother since *G6PD* is an X-linked gene and he could not have inherited a *G6PD* mutation from his father because he can only have inherited his mother's X chromosome. On the other hand, the unaffected children that were available for testing from family A were all females (II-3, II-11, II-12, II-13) that had inherited the missense *G6PD* variation (homozygous) and had no *HBB* mutation. On the other hand, II-19 (unaffected male) from family B had no apparent pathogenic variation and II-20 had inherited the *G6PD* from her father in a heterozygous state. (Table 18)(Table 19)

Table 18: Variations found in each affected member of families A and B by initial investigation of WES data.

ID	Gender	Phenotype	HBB variant	Evidence of HBB variant	G6PD variant	Evidence of G6PD variant
II-4	Female	Affected/ deceased	N/A	Not tested	Homozygous	Implied
II-5	Male	Affected	N/A	Not tested	Hemizygous	Implied
II-8	Male	Affected/ deceased	N/A	Not tested	Hemizygous	Implied
II-9	Female	Affected	N/A	Not tested	Homozygous	Implied
II-10	Male	Affected	Heterozygous	Tested	Hemizygous	Tested
II-14	Female	Affected	Heterozygous	Tested	Hemizygous	Tested
II-15	Female	Affected	Heterozygous	Tested	Heterozygous	Tested
II-16	Female	Affected	Heterozygous	Tested	Heterozygous	Tested
II-17	Male	Affected	Heterozygous	Tested	Wild type	Tested

Table 19: Variations found in each unaffected member of families A and B by initial investigation of WES data.

ID	Gender	Phenotype	HBB variant	Evidence of HBB variant	G6PD variant	Evidence of G6PD variant
I-1	Female (mother A)	Unaffected	Heterozygous	Tested	Homozygous	Tested
I-2	Male (father)	Unaffected	Wild type	Tested	Hemizygous	Tested
I-3	Female (mother B)	Unaffected	Heterozygous	Tested	Wild type	Tested
II-1	Male	Unaffected	N/A	Not tested	Hemizygous	Implied
II-2	Male	Unaffected	N/A	Not tested	Hemizygous	Implied
II-3	Female	Unaffected	Wild type	Tested	Homozygous	Tested
II-6	Male	Unaffected	N/A	Not tested	Hemizygous	Implied
II-7	Male	Unaffected	N/A	Not tested	Hemizygous	Implied
II-11	Female	Unaffected	Wild type	Tested	Homozygous	Tested
II-12	Female	Unaffected	Wild type	Tested	Homozygous	Tested
II-13	Female	Unaffected	Wild type	Tested	Homozygous	Tested
II-18	Male	Unaffected/ deceased	N/A	Not tested	Wild type	Implied
II-19	Male	Unaffected	Wild type	Tested	Wild type	Tested
II-20	Female	Unaffected	Wild type	Tested	Heterozygous	Tested

The *HBB* splice-site intronic variation c.92+5G>C which is carried by I-1 and her affected children, and the *HBB* exonic frameshift variation c.27dupG in I-3 and her affected children, have been both described in chapter 3 as severe variations. Carriers of either these variations in a heterozygous state can be asymptomatic or can have mild anemia (beta thalassemia minor or trait). A more severe spectrum of thalassemia (intermedia and major) is manifested if *HBB* variations are inherited in a homozygous state or in a compound heterozygous state (which was not suggested by the sequencing data). G6PD deficiency on the other hand is an X-linked recessive inherited disease that caused by a structural defect of the G6PD enzyme and reduces its activity. G6PD activity is necessary for RBC protection against hemolysis under oxidative stress or external triggers such as certain food, medicines, or infections. Most individuals with G6PD deficiency experience no symptoms unless exposed to triggers (Antwi-Baffour et al., 2019). Both I-1 and father I-2 have the same G6PD variation rs80356820, which is reported as pathogenic or likely pathogenic according to ACMG guidelines by numerous submitters with few conflicting reports of uncertain significance of pathogenicity in ClinVar (Richards et al., 2015). ClinVar is a publicly accessible database that stores findings on the associations between human germline or somatic variants of any kind, size, position and phenotypes, along with the supporting evidence (Landrum et al., 2018). Because of differences in polarity, charge, size, and maybe other features between the uncharged amino acid serine (polar) with its hydrophobic substitution phenylalanine (non-polar), this variation might affect secondary protein structure. As compared to the anticipated frequency of the associated disorder, the allele frequency of this variation is higher in the gnomAD v2.1.1 dataset. Yet, in the Mediterranean region, the Middle East, and the Indian subcontinent, this variation is the leading source of G6PD deficiency (Al-Jaouni et al., 2011) (Alfadhli et al., 2005) (Sukumar et al., 2004). Mediterranean G6PD is a well studied variant that has been linked to favism and other forms of severe hemolytic anemia (Pfeffer et al., 2022). The enzyme kinetics of G6PD Mediterranean have distinctively changed by reduced affinity to G6P due to the location of the amino acid replacement being nearest to lysine 205 which is believed to be involved in the binding of G6P. This variant also has decreased in vitro thermostability and decreased enzyme activity in circulating RBC (Vulliamy et al., 1988) (Moiz et al., 2012). However, the conflicting data on ClinVar that do not clearly show Ser218Phe to be a pathogenic mutation is based on inconsistent results from in silico prediction tools such as SIFT and Polyphen that predicts whether the mutation knocks out the protein function and not on its effects on human health. The clinical manifestation of G6PD deficiency is not consistent. Individuals with the same genotype may exhibit different levels of

severity. In the molecular characterization study of G6PD deficiency by Sathupak et al in 2021, G6PD mutations appeared to be responsible for the severe enzymatic activity observed in the Lue community, but these results did not correspond with the established WHO categories of Class II-severe or Class III-moderate-to-mild deficiency (Sathupak et al., 2021).

The inheritance of both G6PD deficiency with thalassemia is common in regions with high rates of both hematological diseases (Thiam et al., 2022). In a comparative study of hematological parameters between patients with only thalassemia and patients with co-inheritance of thalassemia and G6PD deficiency, Pornprasert and Phanthong have concluded that the severity of anemia in thalassemia patients was not worsened by G6PD deficiency (Pornprasert & Phanthong, 2013). In a contradicting study by Deng et al, showed that G6PD deficiency is commonly inherited together with HbAE (mild anemia) in the Kachin ethnic group, and that they may interact together to cause severe anemia in men (Deng et al., 2017).

Another recent study published in 2022 by Yang and colleagues about a Chinese male child with co-inheritance of *G6PD* gene pathogenic variation c.1225C>T (hemizygous) and *HBB* gene pathogenic variation c.316-197C>T (heterozygous). They believed that because the heterozygous state of the *HBB* mutation does not cause severe anemia, therefore the *G6PD* mutation is primarily responsible for the patient's clinical symptoms (Yang et al., 2022). Similarly, all of our affected cohort have a heterozygous pathogenic *HBB* mutation that is not enough on its own to raise their anemia manifestation from mild to major. Inheriting both *HBB* and *G6PD* mutations might actually be the explanation for their severity except for three concerning reasons. First, the mother from family A (I-1) has the same genotype as her not tested daughter but does not share their severe phenotype. Second, the affected females from family B are only heterozygous for the *G6PD* variation and since it is an X-linked gene, there is no evidence for illness in G6PD heterozygotes. Third, we have one affected male in family B that does not have the G6PD mutation which suggests that he needs to have inherited something else from the father that correlates with the severity of his case. The explanation for the phenotype in family B must be different, if we assume that G6PD with HBB explains the phenotype in family A. Either that or G6PD is irrelevant (or a contributor) in both families and the father has another mutation that we have not discovered yet.

Upon further investigation into the clinical details of the severity in affected members, several features of beta thalassemia major were discovered and summarized in the table below (Table 20).

Table 20: Clinical details of the severity in affected members.

ID	Gender	Sample availability	Age of transfusion dependency	Other clinical details	Reaction to favabeans
II-4	Female	Not tested	Around 2 years old	Deceased	No
II-5	Male	Not tested	Around 9 years old	Type II diabetes	No
II-8	Male	Not tested	Around 2 years old	Deasead	No
II-9	Female	Not tested	Around 2 years old	Delayed growth Splenectomy	No
II-10	Male	Tested	Around 2 years old	N/A	No
II-14	Female	Tested	Around 2 years old	Delayed growth Splenectomy Type I diabetes	No
II-15	Female	Tested	Around 2 years old	Delayed growth Splenectomy Type I diabetes	No
II-16	Female	Tested	Around 2 years old	Splenectomy	No
II-17	Male	Tested	Around 2 years old	Delayed growth Splenectomy	No

While some patients can be quickly diagnosed, others require a more extensive investigation into possible causes for their symptoms. Although exome sequencing is often the last step in the diagnostic process, it is not without its limitations. Therefore, deeper investigation into the genomic make-up of these individuals was needed to uncover the key players to explain their genotype/ phenotype correlation.

We have continued our family investigation by firstly validating WES data and confirming the presence of G6PD and HBB variations in the family members by performing Sanger sequencing. (Figure 68)(Figure 69)(Figure 70)

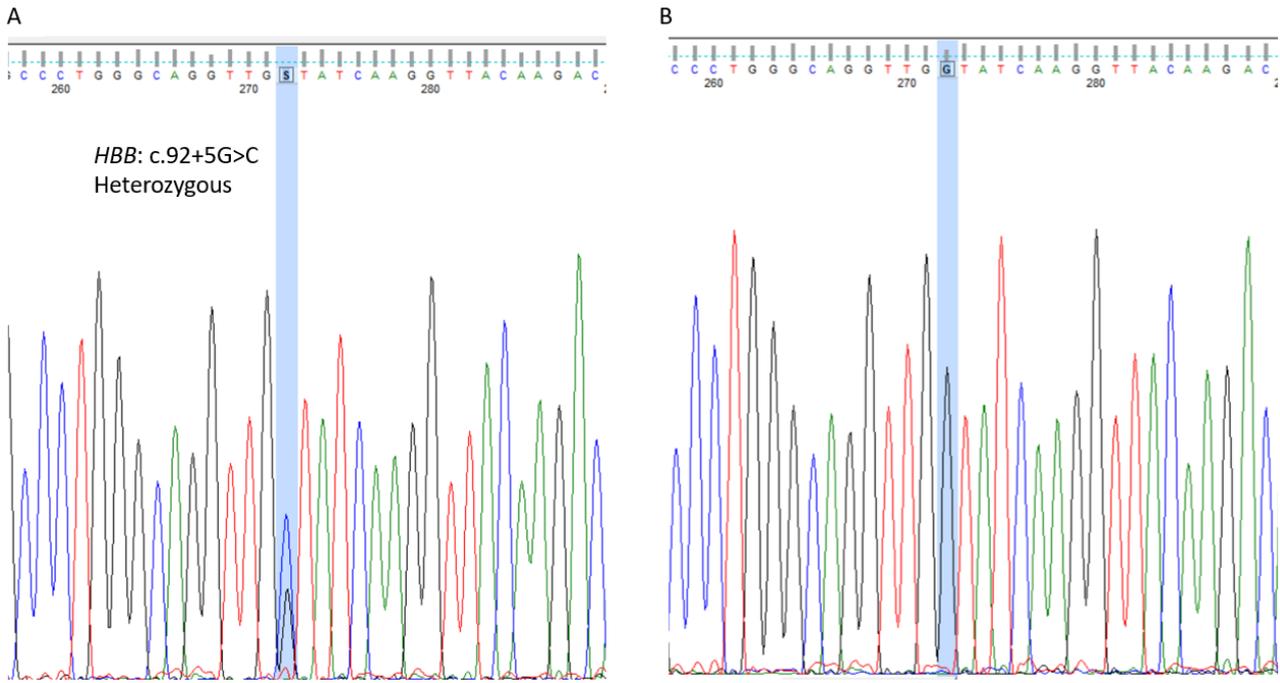


Figure 68: Comparison between A) the mother's sample (I-1) with the heterozygous splice region variation (c.92+5G>C) vs. B) the father's sample (I-2) with reference sequence.

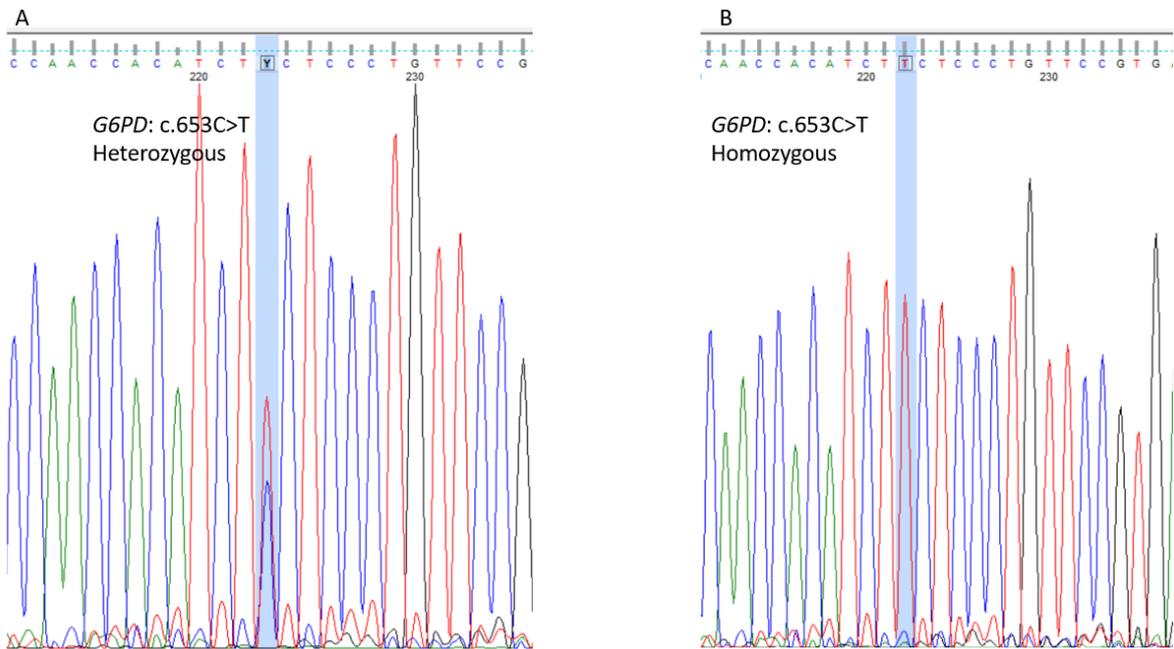


Figure 69: Chromatogram of G6PD c.653C>T variation with the chromosomal position of hg37: chrX:153762634.

A.shows a heterozygous mutation in sample II-14. B. shows a homozygous mutation in sample I-1.

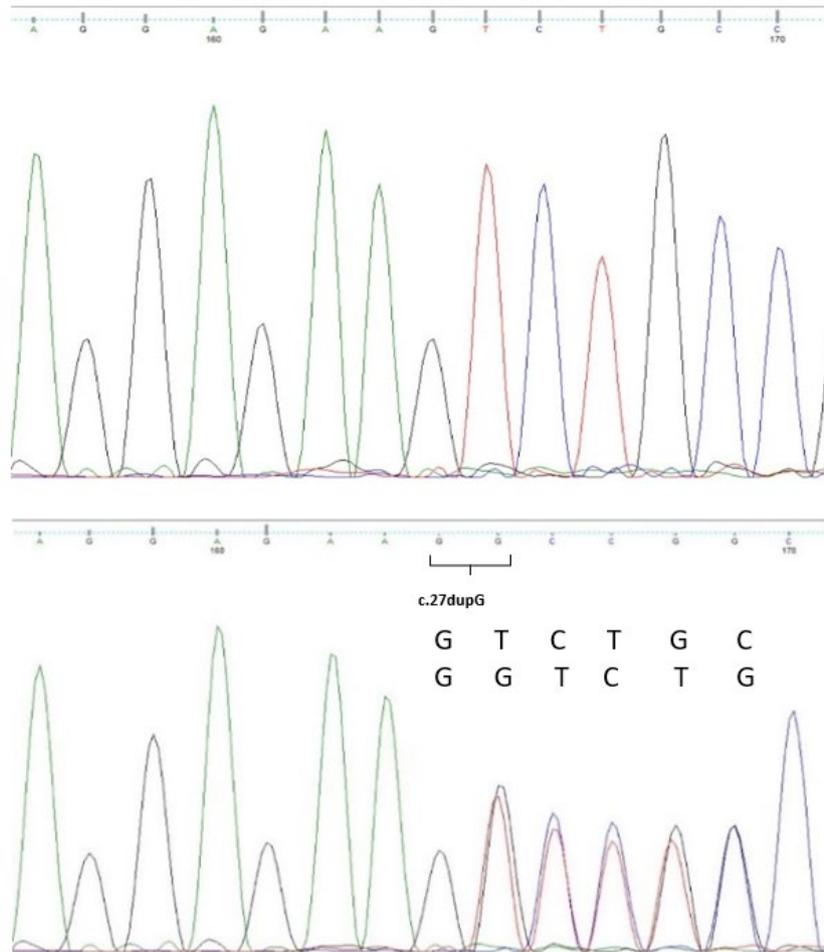


Figure 70: Comparison of sequencing chromatograms. Sequencing chromatogram comparison of wild type versus heterozygous variant of c.27dupG showing wild type case of reference sequence with no c.27dupG mutation (top figure) vs. heterozygous mutation of c.27dupG (bottom figure) which contains an insertion of a G nucleotide that shifts the open reading frame resulting in premature termination.

Between 6 and 24 months of age, when hemoglobin synthesis switches from fetal (HbF) to adult (HbA), patients with beta-thalassemia major often start presenting with clinical symptoms. On the other hand, the presence of beta-thalassemia minor (trait) is often found unexpectedly during a normal complete blood count. Mild anemic symptoms may exist in patients without any obvious abnormalities on physical examination. Individuals with intermedia phenotype may exhibit a wide range of symptoms and severity, possibly with the need for occasional but not routine transfusions. In order to diagnose asymptomatic carriers (sometimes minor symptoms), automated electronic cell counters are often used to perform complete blood count (CBC). Since we did not have access to clinical histories for the family members, we performed basic diagnostic clinical tests starting with complete blood count which

include hematological indices of RBC count, hemoglobin, average size of red blood cells MCV (mean corpuscular volume), amount of hemoglobin in red blood cells MCH (mean corpuscular hemoglobin), differences in the size and volume of erythrocytes RDW (red cell distribution width), and RBC morphology. Also, we measured the levels of different hemoglobin types (HbA, HbA₂, and HbF) by performing Hb capillary gel electrophoresis on haemoglobin released from a lysed blood sample. Finally, we assayed G6PD activity to evaluate if the detected variant does cause G6PD deficiency. Since our affected patients have blood transfusions on regular basis, their own blood will be contaminated with donor blood. To minimise this effect, blood samples were drawn in hospital immediately before transfusion when donor contamination will be at its lowest possible. Samples were collected from the parents, and from three of the unaffected offspring.

For hematological analysis, we have looked into the parameters of the red blood indices for each type of thalassemia (minor-intermedia-major) as a reference to be able to compare to our samples. For thalassemia minor (trait), an elevation in red blood cell count is accompanied by a lower mean corpuscular volume (MCV; 60–70 fl), lower mean corpuscular hemoglobin (19–23 pg), and slight elevation in RDW. Red blood cells may exhibit morphological abnormalities such as microcytosis, hypochromia, and anisopoikilocytosis (abnormal shapes and sizes). Beta-thalassemia minor have at least one beta-globin gene that is unaffected, which enables them to produce adequate amounts of hemoglobin to meet the routine demands of the body without significantly increasing the risk of erythroid hyperplasia. In addition to this, the decrease in hemoglobin is compensated for by a rise in other types of hemoglobin, most often HbA₂ ($\alpha_2\delta_2$). Quantitative determination of HbA₂ is the most helpful diagnostic tool with values between 3.6% to 7% with normal to slightly higher levels of HbF ($\alpha_2\gamma_2$). Thalassemia intermedia (non-transfusion dependent) should be considered in people presented at a later age with milder clinical symptoms than thalassemia major. They have mild to moderate anemia, with hemoglobin levels between 7 and 10 g/dL (even in the most severe cases), with MCV levels that ranges between 50 to 80 fl, and MCH between 16 to 24 pg. Thalassemia intermedia is characterized by HbA₂ level of more than or equal to 3.5 percent and HbF level between 10 and 50 percent. The levels of HbA₂ and HbF are affected by both the severity of the genetic abnormality and the level of erythropoiesis. Thalassemia major (transfusion dependent) patients have severe normocytic/microcytic anemia at a young age. There is a consistent decrease in hemoglobin below 7 g/dL with low MCH and MCV values, 12-20 pg and 50-70 fl respectively. In beta thalassemia major with homozygous variants, HbA ($\alpha_2\beta_2$) is completely absent, and HbF represents 92 to 95% of the total hemoglobin. On the other hand, in beta thalassemia major with compound heterozygous

variants, the HbA levels are between 10 and 30%, while HbF represents 70 and 90% of total hemoglobin (Khan & Rehman, 2022) (Needs et al., 2022). Listed below are the results of our tested samples (Table 21).

Table 21: Complete blood count test results.

Sample	RBC 10 ⁶ /uL	HGB g/ dL	MCV 80-100 fL	MCH 23-31 pg	RDW %	RBC morphology
II-19	5.88	10.7	59	18.2	16.8	Microcytosis
I-2	5.99	11.7	65.1	19.5	13.8	Microcytosis
I-3	4.66	8.3	63.3	17.8	21.7	Microcytosis, anisocytosis, hypochromia
I-1	4.69	6.6	51.8	14.1	24.6	Microcytosis, anisocytosis, hypochromia
II-3	3.9	8.4	75.4	21.5	14.3	Hypochromia
II-11	4.55	10.9	80.9	24	13.4	Normal

We have also measured the quantitative determination of G6PD in RBC using Randox G6PDH assay. The enzymatic activity is determined by measuring the change in absorbance rate at 340 nm due to NADP reduction using RX Daytona analyzer. The diagnostic lab performing the assay has a standard of testing subjects and control samples in parallel. In case of positive deficiency results, the test is repeated twice for the same samples for result confirmation.

Table 22: Hb electrophoresis and G6PD deficiency blood test results.

None of these individuals require transfusion.

Sample	G6PD value 116-408 mU/10 ⁹ RBC	Hemoglobin A (α2β2) 95.8-98 %	Hemoglobin A ₂ (α2δ2) 2-3.3 %	Hemoglobin F (α2γ2) 0-0.9 %
II-19	163	91.7	6.7	1.6
I-2	1	93.6	6.4	0
I-3	240	95.3	4.7	0
I-1	4	96.3	3.7	0
II-3	1	97.6	2.4	0
II-11	1	97.6	2.4	0
II-12	NA	98	2	0
II-13	NA	97.8	2.2	0

Hemoglobin electrophoresis for II-19, I-2, I-3, and I-1 showed a mild increase in HbA2 with associated microcytosis, suggesting the presence of beta thalassemia trait. Though neither I-2 nor II-19 have any detected HBB mutation. As for II-3, II-11, II-12, and II-13 the levels of all hemoglobin types were normal. The Hb electrophoresis result for the father (I-2) corresponds with having children affected with beta thalassemia major by compound heterozygosity, but both I-2 (the father) and II-19 had no beta thalassemia pathogenic variants in their WES results (Table 22). It is clear that there must be another causative mutation in family B. As for the G6PD activity test, II-19 and II-3 had normal levels while the results for samples I-2, II-3, II-11, and I-1 was less than 2.5 % of normal enzyme activity which corresponds with having a pathogenic variant that caused a severe deficiency in the enzyme activity (Table 22).

Our investigation has sought single nucleotide polymorphisms or small deletions/insertions. The identification of structural changes is constrained by WES due to its low sensitivity. The technological limitations of WES suggest that even if it can detect some copy number variations, such as indels and duplications, it is likely missing others. In 2020, Burdick and colleagues were studying exome sequencing and its limitations in diagnosing disease by using participants from the Undiagnosed Disease Network. In their study, they found that whole exome sequencing they were able to diagnose 67% of the cases, while the variations missed by WES (33%) required further approaches to identify and/or validate their functional consequences. According to these findings, The functional impact of the variations overlooked by exome sequencing required several approaches to discover and validate. nondiagnostic preliminaries should be taken into consideration following testing for elusive variations (Burdick et al., 2020).

As a result of their sequence-based construction, some genomic areas cannot be analyzed effectively using WES. Large insertion/deletions, copy number variants, and repeat expansions are examples of the sorts of genomic abnormalities that WES cannot identify (Shakiba & Keramatipour, 2018).

There are other genetic modulators that have a direct role in the beta globin chains imbalance (mentioned in chapter 1). Many published studies have discussed and correlated the increase severity in beta thalassemia heterozygotes with the increase of alpha globin chain production from alpha thalassemia gene triplication which develops during meiosis when homologous regions in the -globin gene cluster are misaligned and undergo an inefficient crossover. The alpha triplication is thought to have a significant role in the phenotypic severity of heterozygous/ homozygous beta thalassemia, making the disease even more severe by further destabilizing the globin chain balance (Farashi et al.,

2015). In 2022, Ropero and colleagues have conducted a retrospective study in Spain that involved 73 patients with alpha thalassemia triplication (or quadruplication) and beta thalassemia. They analyzed the hematological and phenotypic characteristics of the Spanish population with thalassemia intermedia to see how the increase in alpha genes among individuals with a single mutation in a beta globin locus affects the disease. They concluded that several variables contribute to the thalassemia intermedia phenotype, one of which is the severity degree of the beta thalassemia mutation (type and location) that ranges from mild reduction in beta globin chain production to diminished beta globin chain production (absent), and the elevation in alpha globin chains production that results in the increased balance of alpha and beta chain (Ropero et al., 2022). Triplication of the alpha globin gene has been reported in numerous studies to aggravate beta thalassemia symptoms in certain individuals, however this is not universally accepted because the projected worsening of anemia has not happened in all patients (Luo et al., 2021). In a contradicting study that compared hematological parameters and phenotypical changes in a cohort of 4005 with beta thalassemia trait, 455 sickle cell carriers, and 2000 healthy individuals and discovered that triplication of alpha thalassemia frequencies were 1.67%, 0.88%, and 0.9% respectively. They found that none of the blood indices changed noticeably in healthy people or those who carried the sickle cell or beta thalassemia trait due to this mutation and concluded that alpha thalassemia triplication does not increase the severity of anemia (Hamid et al., 2021).

With conflicting conclusions and debatable findings, nevertheless, our next step was to perform MLPA alpha thalassemia testing to rule out deletions or duplications in alpha thalassemia that might have an impact on the severity of our cases. For the MLPA reaction, EK20-FAM by MRC Holland was used. Coffalyser.Net (free analysis software developed by MRC-Holland) was used for data analysis as it ensures quality requirements are met by performing control fragment checks automatically through MLPA probe mixes (that contain quality control fragments). Coffalyser.Net only interprets data that meets the quality control requirements. For copy number calculation, relative probe signals from each undigested reaction were compared between reference samples and a test sample to allow for the calculation of the relative changes in probe signals in order to determine the relative copy number of the target sequences (Figure 71)(Figure 72)(Figure 73). Negative controls were used, while positive controls of alpha duplications (triplicated and quadruplicated alpha globin gene) were not available.

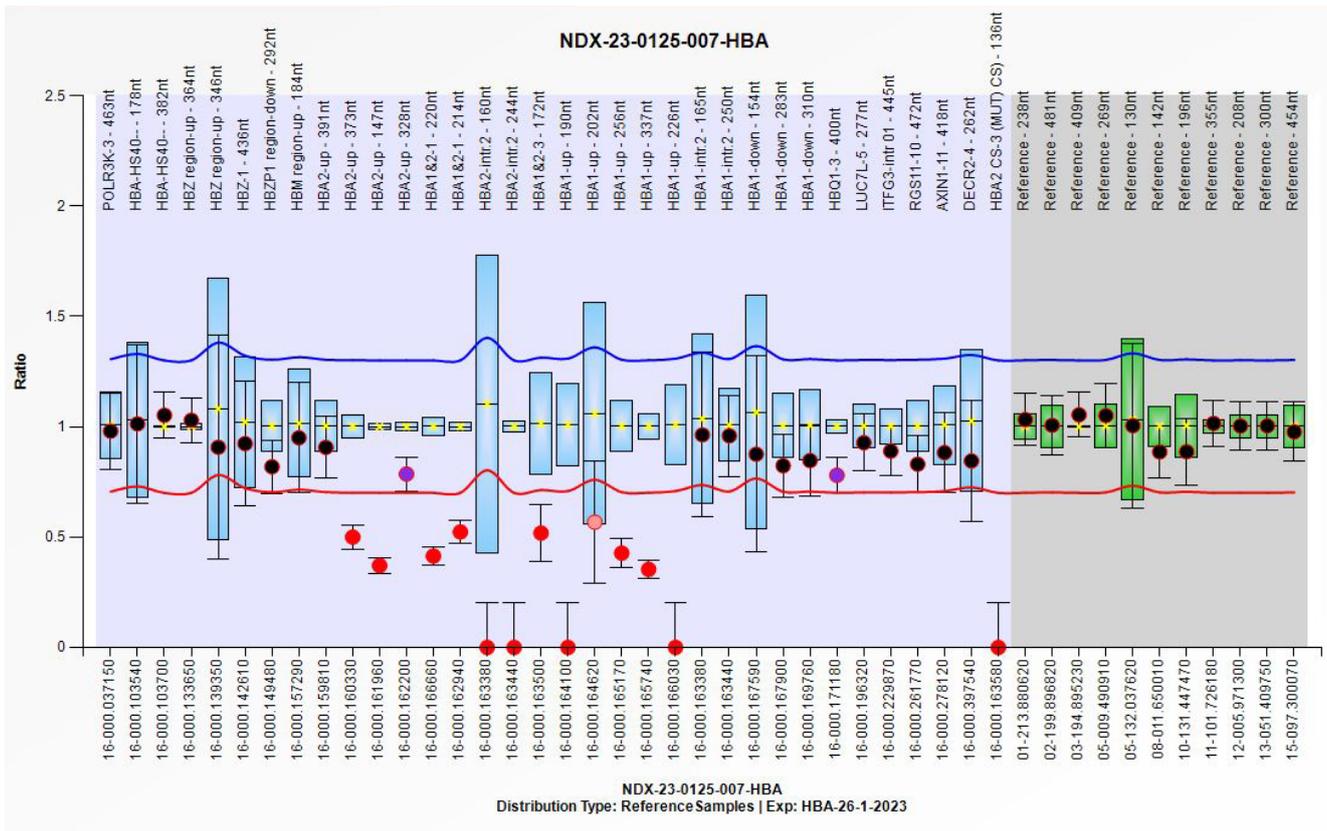


Figure 71: MLPA reaction result of sample II-9 (transfusion dependent female) indicating heterozygous deletion.

A score ratio of 1 indicates no change in copy number compared to reference sample. On the other hand, a score of 0.5 indicates a heterozygous deletion while a score of 1.5 indicates a heterozygous duplication. The blue boxes represent the 95% confidence interval of a probe over the reference samples. The colored dots are the calculated probe ratio. The error bars surrounding the dots represent a 95% confidence interval estimate for each probe in a sample. The red and blue lines display lower arbitrary border and upper arbitrary border respectively. These borders are +/- 0.3 from the average value of a probe over the reference sample. The black dot indicates no change in probe copy number because the 95% confidence interval estimates (the error bars) of the probe overlap with the 95% confidence intervals of the same probes over the reference sample. The purple dot indicates a decreased signals of more than two standard deviations compared to the reference samples, but the lower arbitrary border has not been crossed. The pink dot indicates an increased signals of more than two standard deviations compared to the reference samples, but the upper arbitrary border has not been crossed. The red dots indicate a decreased signals of more than two standard deviations compared to the reference samples and the lower arbitrary border has been crossed. The red dot far down in the ratio chart with one error bar indicates no signal found. Source: Coffalyser.Net reference manual v01.

D [nt]	Gene-Exon	Chr.band	hg18 loc.	Height	Area	Ratio ^H	Stdev	[REF]	Width	d[nt]	[Mut details]
463	POLR3K-3	16p13.3	16-000.037150	3213	25960	0.98	0.09	=	70	0.11	-
178	HBA-HS40--	16p13.3	16-000.103540	11542	69171	1.01	0.18	=	57	0.01	-
382	HBA-HS40--	16p13.3	16-000.103700	1663	12494	1.05	0.05	=	43	0.02	-
364	HBZ region-up	16p13.3	16-000.133650	3334	24329	1.03	0.05	=	59	0.01	-
346	HBZ region-up	16p13.3	16-000.139350	3780	27113	0.91	0.25	=	69	0.05	-
436	HBZ-1	16p13.3	16-000.142610	1839	14793	0.92	0.14	=	62	0.11	-
292	HBZP1 region-	16p13.3	16-000.149480	7035	48339	0.82	0.06	=	66	0.04	-
184	HBM region-up	16p13.3	16-000.157290	7621	44482	0.95	0.12	=	36	0.02	-
391	HBA2-up	16p13.3	16-000.159810	2568	18880	0.91	0.07	=	48	0.06	-
373	HBA2-up	16p13.3	16-000.160330	1965	14217	0.5	0.03	<<*	46	0.05	-
147	HBA2-up	16p13.3	16-000.161960	6743	40450	0.37	0.02	<<*	43	0.04	-
328	HBA2-up	16p13.3	16-000.162200	2493	17865	0.79	0.04	<<	51	0.11	-
220	HBA1&2-1	16p13.3	16-000.166660	4833	31361	0.41	0.02	<<*	69	0.05	-
214	HBA1&2-1	16p13.3	16-000.162940	3893	24512	0.52	0.03	<<*	45	0.01	-
160	HBA2-intr.2	16p13.3	16-000.163380	0	0	0	0	<<*	0	0.00	-
244	HBA2-intr.2	16p13.3	16-000.163440	0	0	0	0	<<*	0	0.00	-
172	HBA1&2-3	16p13.3	16-000.163500	7556	46807	0.52	0.06	<<*	46	0.01	-
190	HBA1-up	16p13.3	16-000.164100	0	0	0	0	<<*	0	0.00	-
202	HBA1-up	16p13.3	16-000.164620	5225	31136	0.57	0.14	<*	38	0.03	-
256	HBA1-up	16p13.3	16-000.165170	1942	13185	0.43	0.03	<<*	42	0.06	-
337	HBA1-up	16p13.3	16-000.165740	1760	12940	0.35	0.02	<<*	58	0.05	-
226	HBA1-up	16p13.3	16-000.166030	0	0	0	0	<<*	0	0.00	-
165	HBA1-intr.2	16p13.3	16-000.163380	10510	66188	0.96	0.19	=	53	0.06	-
250	HBA1-intr.2	16p13.3	16-000.163440	3418	24112	0.96	0.09	=	48	0.08	-
154	HBA1-down	16p13.3	16-000.167590	11014	66849	0.88	0.22	=	73	0.02	-
283	HBA1-down	16p13.3	16-000.167900	4417	28977	0.82	0.07	=	57	0.01	-
310	HBA1-down	16p13.3	16-000.169760	3284	22553	0.85	0.08	=	57	0.02	-
400	HBQ1-3	16p13.3	16-000.171180	2423	18953	0.78	0.04	<<	53	0.05	-
277	LUC7L-5	16p13.3	16-000.196320	7399	49421	0.93	0.06	=	63	0.05	-
445	ITFG3-intr.01	16p13.3	16-000.229870	3126	25562	0.89	0.06	=	74	0.07	-
472	RGS11-10	16p13.3	16-000.261770	1657	13811	0.83	0.06	=	48	0.00	-
418	AXIN1-11	16p13.3	16-000.278120	2864	22422	0.88	0.09	=	51	0.04	-
262	DEC2R-2	16p13.3	16-000.397540	9580	62226	0.84	0.14	=	55	3.52	-
136	HBA2 CS-3 (MU)	16p13.3	16-000.163580	0	0	0	0	<<*	0	0.00	CS
238	Reference	01q41	01-213.880620	12094	77829	1.03	0.06	=	70	0.03	-
481	Reference	02q33.1	02-199.896820	4439	37160	1.01	0.07	=	67	0.00	-
409	Reference	03q29	03-194.895230	6168	48582	1.05	0.05	=	101	0.01	-
269	Reference	05p15.2	05-009.490910	9994	65747	1.05	0.07	=	57	0.03	-
130	Reference	05q31.1	05-132.037620	13836	81227	1	0.19	=	44	0.03	-
142	Reference	08p23.1	08-011.650010	18895	115489	0.89	0.06	=	69	0.07	-
196	Reference	10q26.3	10-131.447470	11238	69922	0.89	0.08	=	65	0.02	-
355	Reference	11q22.2	11-101.726180	8109	56191	1.01	0.05	=	72	0.04	-
208	Reference	12p13.31	12-005.971300	13241	81852	1	0.06	=	57	0.02	-
300	Reference	13q14.3	13-051.409750	4142	27603	1	0.06	=	65	0.01	-
454	Reference	15q26.3	15-097.300070	2765	22070	0.98	0.07	=	52	0.11	-
Median value all probe values:				4428	30056	0.9	0.06		57	0.04	

Figure 72: A table that shows genomic locations of probe values for sample II-10 from MLPA reaction. = indicates no change in copy number. << indicates a decrease of more than two standard deviations compared to the reference samples has been calculated. <<* indicates a decrease of more than two standard deviations compared to the reference samples has been calculated with the lower arbitrary border has been crossed (heterozygous deletion). <<** indicate that no signal has been found. The Constant Spring probe does not give a signal unless the sample contains the point mutation. Source: Coffalyser.Net reference manual v02.

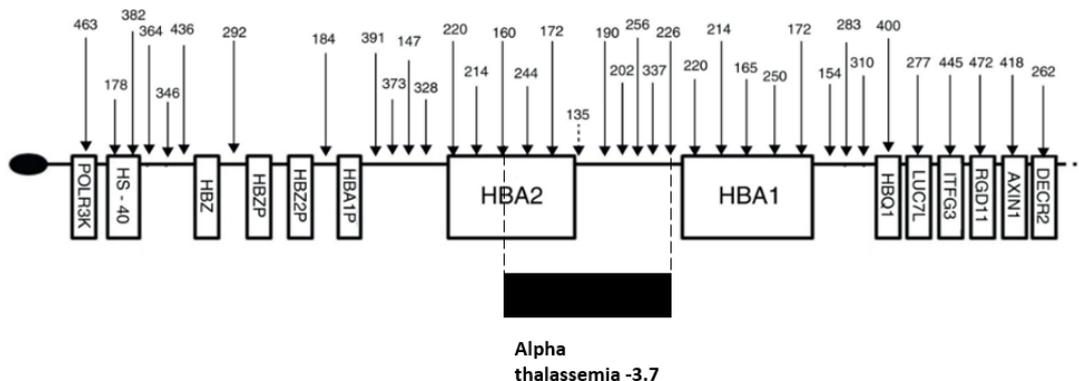


Figure 73: MLPA probes mapped on alpha globin cluster. The arrows indicate the probe positions while the boxes are the genes. The black oval shape represents the telomeric region. The black box represents the position of alpha thalassemia deletion (-3.7). The mapped probes adapted from the MLPA kit provider.

By analyzing the MLPA results as shown below (Figure 74), it is evident that the family are carriers of the alpha globin 3.7 kb deletion (-3.7α) either in a heterozygous $-3.7\alpha2/\alpha1\alpha2$ (I-2, II-9, II-16, I-1, II-19, II-11, II-10 and II-17) or homozygous $-3.7\alpha2/-3.7\alpha2$ (II-14, I-3, II-15, and II-3) state. This alpha globin deletion has the highest prevalence in Saudi population (Al Asoom et al., 2020) (Borgio et al., 2018) which makes it common to be also present with beta globin variation.

				Status	U	U	U	A	A	A	A	U	A	U	U	A
				I-2 Father	I-3 Mother	B I-1 Mother	A II-17	II-14	II-15	II-16	II-19	II-9	II-11	II-3	II-10	
chr16	37132	37152	19236-125316	POLR3K	0.65<<<	0.62<<<		0.59<<<	0.53<<<	0.55<<<	0.44<<<	0.66<<	0.52<<<			
chr16	103528	103548	04799-104797	HBA-HS40	0.58<<<	0.48<<<										1.01
chr16	103685	103705	04800-104175	HBA-HS40			1.6>>>	1.82>>>		1.57>>		1.71>>>				1.05
chr16	133637	133657	04926-123886	HBZ region			<<<		0.66<<<	0.64<<<				1.31>>		1.03
chr16	139356	139356	04622-104001	HBZ region	0.52<<<	0.66?		0.61?	0.49<<<	0.29<<<	0.48<<<	0.5<<<	0.43<<<			0.91
chr16	142592	142612	17214-SP0457-L20489	HBZ	0.53<<<	0.63<<<	0.59<<<		0.53<<<	0.53<<<	0.48<<<	0.5<<<	0.43<<<	0.69<<<	0.66<<<	0.92
chr16	149467	149487	04624-104004	HBZP1 region	5.3	20.5	BRIT	MED	FIL	SEA	THAI					0.81
chr16	157274	157294	04637-104018	HBM region	5.3	20.5	BRIT	MED	FIL	SEA	THAI					0.95
chr16	159798	159818	18097-122521	HBA2	5.3	20.5	4.2	BRIT	MED	FIL	SEA	THAI				0.91
chr16	160313	160333	18099-108415	HBA2	5.3	20.5	4.2	BRIT	MED	FIL	SEA	THAI				0.91
chr16	161951	161971	18098-122522	HBA2	5.3	20.5	4.2	BRIT	MED	FIL	SEA	THAI				0.91
chr16	162190	162210	18092-122516	HBA2	5.3	20.5	4.2	BRIT	MED	FIL	SEA	THAI				0.91
chr16	162921	162941	18881-106288	HBA1&2	20.5	4.2	HBA;	BRIT	MED	FIL	SEA	THAI	3.7			0.5<<<
chr16	163362	163382	08498-108422	HBA2	20.5	4.2	HBA;	BRIT	MED	FIL	SEA	THAI	3.7			0.5<<<
chr16	163427	163447	04633-123748	HBA2	20.5	4.2	HBA;	BRIT	MED	FIL	SEA	THAI	3.7			0.5<<<
chr16	163484	163504	15857-121812	HBA1&2	20.5	4.2	HBA;	BRIT	MED	FIL	SEA	THAI	3.7			0.5<<<
chr16	164084	164104	18096-122520	HBA1	20.5	4.2	BRIT	MED	FIL	SEA	THAI	3.7				0.5<<<
chr16	164602	164622	18880-124428	HBA1	20.5	4.2	BRIT	MED	FIL	SEA	THAI	3.7				0.5<<<
chr16	165160	165180	08494-108417	HBA1	20.5	4.2	BRIT	MED	FIL	SEA	THAI	3.7				0.5<<<
chr16	165728	165748	14855-123604	HBA1	20.5	4.2	BRIT	MED	FIL	SEA	THAI	3.7				0.5<<<
chr16	166018	166038	18093-122517	HBA1	20.5	4.2	BRIT	MED	FIL	SEA	THAI	3.7				0.5<<<
chr16	166644	166664	18099-122524	HBA1&2	20.5	4.2	HBA;	BRIT	MED	FIL	SEA	THAI	3.7			0.5<<<
chr16	167166	167186	08498-121607	HBA1			HBA;	BRIT	MED	FIL	SEA	THAI				0.96
chr16	167237	167257	04633-123600	HBA1			HBA;	BRIT	MED	FIL	SEA	THAI				0.96
chr16	167572	167592	08499-123594	HBA1			HBA;	BRIT	MED	FIL	SEA	THAI				0.88
chr16	167890	167910	04638-123602	HBA1			HBA;	BRIT	MED	FIL	SEA	THAI				0.82
chr16	168744	168764	04639-104020	HBA1			BRIT	MED	FIL	SEA	THAI					0.85
chr16	171170	171190	19233-125313	HBO1			HBA;	BRIT	FIL	SEA	THAI					0.78<<
chr16	196305	196325	15859-121960	LUC7L												0.93
chr16	229854	229874	17227-120554	ITFG3												0.89
chr16	261758	261778	18102-120488	RGS11												0.83
chr16	278107	278127	17212-133993	AXIN1												0.88
chr16	397524	397544	17613-123601	DEC2R												0.84
chr16	163569	163589	50585-SP0043-109493	HBA2 CS												0.84

Figure 74: MLPA mapping sheet for comparison with common deletions that cause alpha thalassemia. Samples found with heterozygous deletion $-3.7\alpha2/\alpha1\alpha2$ are highlighted in bright yellow, while samples found with homozygous deletion $-3.7\alpha2/-3.7\alpha2$ are highlighted in bright blue. U indicates unaffected individual while A indicates affected (transfusion-dependent) individual.

We have updated our pedigree to include the alpha globin gene deletion (Figure 75).

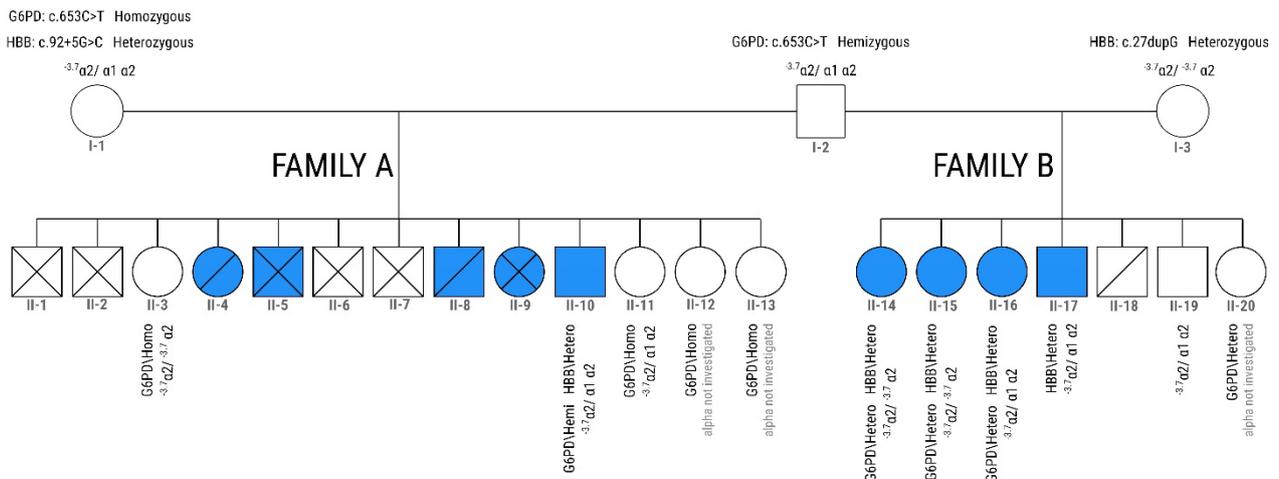


Figure 75: Pedigree of a family with beta-thalassemia major phenotype showing pathogenic variants in three genes, HBB, G6PD, and alpha globin gene.

The deletion of the alpha-gene has also been linked to a reduction in the severity of beta-thalassemia and sickle cell anemia and improvement of their hematological conditions (Hassan; et al., 2014). In Saudi Arabia, research into the inheritance patterns of HBA1 and HBA2 gene variants, as well as HBB gene variants, in the Saudi population has been conducted. Patients with transfusion-dependent beta-thalassemia major in the Eastern Province of Saudi Arabia were found to co-inherit 12 distinct alpha-globin gene variants, as reported by Akthar et al. Among these 12 co-inherited alpha-globin gene variants, the -3.7 deletion was shown to be the most common and patients were reported to have decreased incidence of osteoporosis and splenectomy (Akhtar et al., 2013). We must keep in mind that the degree of alpha/non alpha globin chain imbalance is the primary pathophysiological predictor of the severity of the beta-thalassemia syndromes in order to comprehend the clinical-molecular linkages. As a result, the clinical picture may improve if the alpha/beta chain imbalance is reduced by any available factors. Therefore, the coinheritance of homozygous beta-thalassemia with an alpha-thalassemia determinant that lowers alpha chain production leads to moderate beta-thalassemia by lowering the alpha/beta chain imbalance. The clinical phenotype of homozygous beta thalassemia can be improved by the deletion of a single alpha globin gene, on the other hand, in β^0 thalassemia, two alpha globin gene deletions may be required (Cao & Galanello, 2010).

Moreover, to further investigate any large deletions or insertions (copy number variations) that may possibly have affect on the phenotype, we performed an Affymetrix HD Cytoscan array. To reduce the cost and save time, we began with sample I-2 (the father) since we were investigating what variation did he pass on to his affected children (Figure 76)(Table 23).

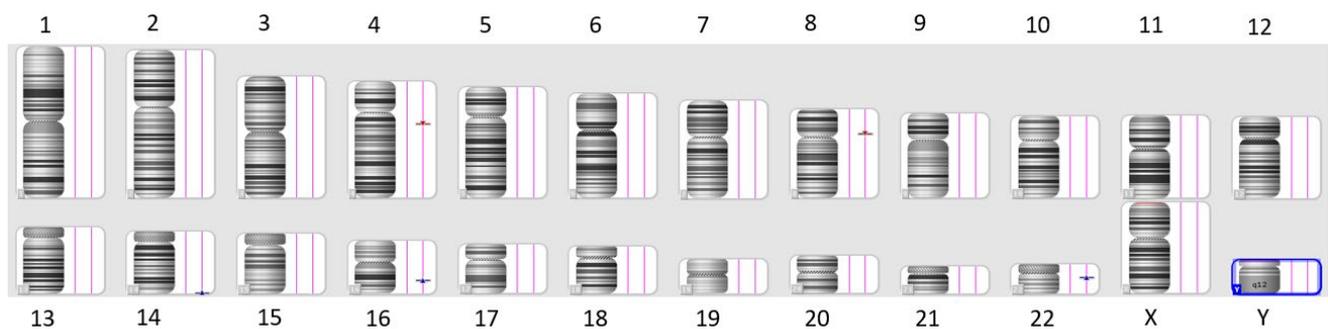


Figure 76: Chromosomal image of sample I-2 from microarray. The blue boundary indicates the Y chromosome. Red and blue marks indicate genomic loss and gain respectively.

Table 23: Copy number variation (CNV) results of sample I-2 from microarray.

CNVs									
File	CN State	Type	Chromosome	Cytoband Start	Size (kbp)	Size (kbp)	Marker Count	Gene Count	Genes
Father	3	Gain	14	q32.33	707.531	707.531	244	8	MIR4539, MIR4507, MIR4538, MIR4537, KIAA0125, ADAM6, LINC00226, LINC00221
Father	3	Gain	22	q11.22	452.643	452.643	244	7	ZNF280B, ZNF280A, PRAME, LL22NC03-63E9.3, POM121L1P, GGTL2, MIR650
Father	0	Loss	8	p11.22	158.454	158.454	79	2	ADAM5, ADAM3A
Father	3	Gain	16	q22.1	108.411	108.411	76	5	FAM96B, CES2, CES3, CES4A, CFBF
Father	1	Loss	4	q13.2	106.075	106.075	52	1	UGT2B15

Copy number variations are duplications, deletions, and other types of genomic rearrangements that result in the acquisition or loss of genetic material (Shaikh, 2017). By checking all the genes found in each gain or loss and their involvement with anemia and red blood cells, we have found a recent published paper that describes overexpression of *KIAA0125* (also known as *FAM30A*) gene, which is located on chromosome 14, and its involvement with acute myeloid leukemia and myelodysplastic syndromes (Hung et al., 2021; Wang; et al., 2021). Other than that, no significant large deletions were found in father (I-2).

Our next step was to explore every possibility and include the analysis of all kinds of variations found within the beta globin cluster.

It could be proposed that synonymous mutations that do not affect promoters, enhancers or splicing might alter the expression of genes in humans, as has been reported in some prokaryotic systems. In our cohort, four patients (II-14, II-15, II-16, and II-17) from family B with the heterozygous pathogenic frameshift variant, have also inherited the synonymus *HBB* variant (c.9T>C, rs713040) in a heterozygous state. The patient II-10 from family A did not inherit the synonymus variant. This variant is present in other members of the family including the mother A I-1 (heterozygous) which was the only one with mild symptoms, while the father I-2 (Heterozygous), mother B I-3 (homozygous), II-19 (heterozygous), II-20 (heterozygous), II-3 (homozygous), II-11 (homozygous), II-12 (homozygous), II-13 (homozygous) have no related symptoms. This synonymus variant is reported in clinvar as a benign variant. In 2020, Sabiha and colleagues have studied a family with a similar genotype to ours where several members had compound heterozygous mutation of rs713040 (either in a homozygous or heterozygous state) with the frameshift mutation rs35699606. All members were healthy except for the son that had inherited the rs35699606 variant in a homozygous state. They have concluded that manifestation of beta-thalassemia phenotype occurs only with the inheritance of rs35699606 variant in a homozygous state and that the synonymus variant did not contribute in the clinical manifestation

of the disease (Sabiha et al., 2020). However, as far as we are aware, such examples of disease causing synonymous variants do not exist.

The beginning of translation is a crucial step in protein synthesis that is typically controlled at the molecular level. Both cis-regulatory elements in the 5' and 3' untranslated regions and trans-acting factors are involved in this regulation. Failure of this control system can cause cellular metabolic disruption and related pathologies. The translation rate of mRNAs is greatly affected by their UTRs' structure. Mutations in the cis-regulatory regions of the UTRs might affect the translation and hence the expression of certain genes. Variations of this kind have the potential to tip the physiological scales toward pathological conditions (Chatterjee & Pal, 2009). Many 5' and 3' UTRs variations in the *HBB* gene have been reported that are linked to beta-thalassemia minor, intermedia, and major (Thein, 2018). Since we are investigating the father's contribution in the disease severity of his affected children, going back to the father's WES data, investigation of the beta globin cluster did not yield any 5' nor 3' UTR variations.

Alternatively, the more we learn and better understand the structure and function of promoters, the more whole-genome sequencing is becoming valuable and useful because it can reveal non-coding mutations that may be responsible for the interference of transcriptional factor occupancy sites and cis-regulatory regions. For genetic illnesses where pseudogenes or heavily duplicated sequences are present, long-read WGS will be useful for mapping large scale variations (third generation sequencing). Interpreting variants from non-coding regions from research to potential future diagnosis and treatment requires a deeper understanding of how transcription factors and cis-regulatory elements impact and drive the highly regulated processes of erythroid development. Several hematological illnesses have been linked to mutations that interfere with transcription factor binding regions. Severe pyruvate deficiency has been correlated to mutations in GATA1 motif binding site of the *PKLR* gene promoter region upstream from the initiation codon (Manco et al., 2000). Also, reduced *BCL11A* expression and increased fetal hemoglobin have both been linked to motif-disrupting variants in the enhancer of *BCL11A*, a crucial regulator of fetal Hb levels revealed by genome-wide association studies (Bauer et al., 2013).

Pseudogenes, a term coined by Jacq et al, initially identified using single-gene cloning, are sequences of genes in the genome that are closely similar to the coding gene sequence but are not ordinarily expressed (Jacq et al., 1977). Pseudogenes originate from real genes but have had their open reading frames (ORF) modified in some way during the course of evolution, maybe by introducing frameshift or nonsense mutations, rendering them incapable of encoding proteins with the intended function

because there is no selective pressure to maintain their open reading frames (Magrinelli & Lohmann, 2022). They can be found in nearly every genome. Approximately 15,000 pseudogenes have been identified in the human genome so far, according to the reference annotation project GENCODE (v39). Short-read NGS-based evaluations might be complicated by the significant sequence similarity of parental genes, in which mutations can contribute to human illness and the non-coding pseudogene(s). Pseudogenes were considered "junk DNA" for a long time. However, many pseudogenes are translated into RNA, and several have been found to serve crucial roles in the gene control of their parent genes. Pseudogene transcripts may lower the amount of miRNA in a cell or generate small interfering RNA, which means that oncogene activity may be regulated by pseudogenes (Tutar, 2012) (Tutar et al., 2018) (Carron et al., 2022). In a recent study, Maurer and colleagues have identified copy number variations in *ADGRL3* and 2 pseudogenes in a juvenile patient diagnosed with autism and ADHD (Maurer et al., 2022).

As for beta globin cluster, it consists of five functional genes and a single pseudogene *HBBP1* designated psi-beta1 ($\psi\beta1$) between gamma and delta globin genes. This pseudogene predates recent gene duplication demonstrated by comparative sequencing analyses. It has shown to be old and conserved in the same position as humans throughout all great ape species (Harris et al., 1984). Multiple patches of epigenetic evidence indicating functionally active chromatin overlap two key regulatory regions including active transcription factor binding sites in the *HBBP1* locus (Tomkins, 2013). In 2012, Giannopoulou and her group correlated a single nucleotide variation in the *HBBP1*, rs2071348 (g.5264146A>C), with a milder disease phenotype of beta thalassemia. They were studying the association of the *HBBP1* variation with the beta thalassemia disease severity by comparing groups of patients that share the same ethnic background (Asian) with beta thalassemia major, intermedia, mild, and normal samples and found that this variation is associated with milder phenotype and high levels of fetal hemoglobin.

This finding suggests that there is a possibility that this region has a regulatory function in the expression of the γ -globin gene and that it might, as a result, be connected with the severity of β -thalassemia disease (Giannopoulou et al., 2012). In a recent study, using genome wide analysis of human pseudogenes coupled with expression network analysis and functional assays to study the beta globin pseudogene *HBBP1* in depth, Yanni Ma and colleagues have demonstrated that the pseudogene has tissue-specific expression in the bone marrow. In addition, they have found that an important regulator of erythropoiesis, TAL1 (T cell acute lymphocytic leukemia 1), is upregulated in part by *HBBP1* competitively binding to the RNA-binding protein (RBP) HNRNPA1 (RNA-heterogeneous nuclear

ribonucleoprotein A1) which demonstrates HBBP1 essentiality in the process of erythropoiesis. Also, when HBBP1 and TAL1 interact, individuals with beta-thalassemia experience less severe symptoms (Ma et al., 2021). Since HBBP1 has been associated with HbF elevation, Shuang-Ping Ma et al wanted to study the mechanism underlying HBBP1 association with the increase in fetal hemoglobin production by using bioinformatic analysis tools. First, they found that HBBP1 is important in gamma-globin synthesis, and then they provisionally verified this discovery in K562 cells. Overexpression of HBBP1 in HUDEP-2 cells led to elevated gamma-globin expression at both the mRNA and protein levels. Their second discovery was that ELK1, a transcription factor, binds to the proximal promoter of HBBP1. This binding greatly enhances the activity of the HBBP1 promoter (Ma; et al., 2021). These were interesting findings that made us look back at our WES results. Again, we looked at the father's sequence result, which turned out that it lacks any variants in the *HBBP1* pseudogene.

Furthermore, the beta globin locus control region, which is a very important region that plays a crucial role in beta globin expression and development was not covered in the whole exome sequencing because it is a non coding region (deep intergenic) that spans for about 30 kb upstream of the beta globin genes. For that reason, we have decided to perform whole genome sequencing to explore this vast location for any possible variants that may cause the increase of disease severity. Since WGS is not available in our laboratory, two samples, I-2 and II-14, have been sent out for WGS. Again, disappointingly, no reported pathogenic variants were found in the LCR of beta gene.

Deletions, insertions, splice-site modifications, frameshifts, and nonsense alterations may be easily predicted and tested. However, it is far more challenging to identify predisposing or protective genes since their influence on the phenotype is generally small and can be found in a variety of genomic regions including the coding sequence, the 5' and 3' untranslated regions, introns, and even binding regions of transcription factors or control elements that are many kilobases distant from the actual gene. Rare benign variations cannot be easily distinguished from subtle mutations.

6.3 Analysis of whole exome sequencing data

6.3.1 First scenario

Our first scenario was to filter out the variants by following the mode of inheritance as autosomal recessive considering the allele state of homozygous variant alleles (Alfares, 2018). Starting with affected members of family B, using chromosomal locations, we have used Venn diagrams to show the overlapped shared variants between affected individuals (II-14, II-15, II-16, and II-17) that are in a

homozygous state, then exclude the variants that are shared with unaffected controls (II-19 and II-20), and finally to find overlapped variants between the shared affected and the parents (I-2 and I-3), where the variants of the parents are in a heterozygous state (Figure 77). In case we found a disease-causing autosomal recessive variant, we can go back to II-19 and check if he has it in a heterozygous state to correlate with his Hb electrophoresis result (as a carrier).

The same filtration criteria was applied to family A.

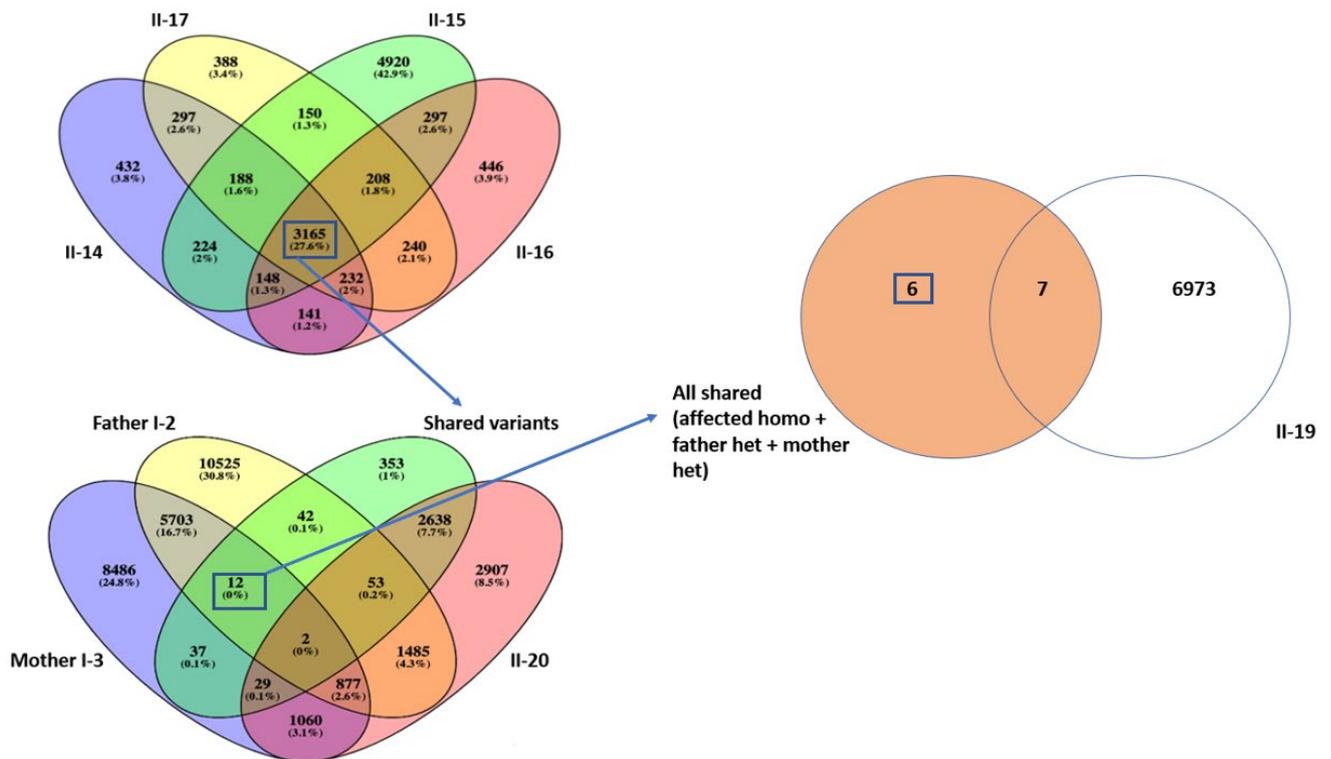


Figure 77: Filtered nucleotide variants from exome data, based on chromosomal positions, which showed autosomal recessive inheritance model of homozygous state in non-consanguineous thalassemia disease family B.

As for family B, after variants filtration in the autosomal recessive disease scenario, we were looking for rare variants, therefore, we checked the minor allele frequency for each variation and excluded all variants with $MAF > 0.02$. This scenario yielded no candidate genes for family B.

As for family A, the filtration yielded two candidate genes RGS12 and CEP290 (Table 24).

Table 24: List of two rare nucleotide variants from exome data which showed autosomal recessive inheritance model in family (A).

Gene	Chr#	Position (GRCh37)	Rs#	HGVS	Protein change	Molecular consequence	gnomad	1000 genomes	ClinVar
RGS12	4	3429844	rs138418915	c.3359C>T	p.Pro1120Leu	Missense	0.001	0.0032	Not reported
CEP290	12	88472996	rs61941020	c.5237G>A	p.Arg1746Gln	Missense	0.01	0.01	Not reported

RGS12 (Regulators of G Protein Signaling) gene comprises five functional domain that has potential roles as both a transcriptional repressor and a guanosine triphosphatase (GTPase) activator. Signal transduction inhibition of G-coupled receptor proteins is achieved by converting G α subunits into their inactive GDP-bound form as a result of increasing the hydrolysis of GTPase activity (Schroer et al., 2019). This gene is reported to be involved in osteoclastogenesis (Ng et al., 2019), pathogenesis of inflammatory arthritis (Yuan et al., 2020), osteoarthritis (Yuan et al., 2022), and regulation of myoblast proliferation and differentiation (Schroer et al., 2019).

CEP290 is a gene that encodes a ciliogenesis and centrosomal protein that localizes in actively dividing cells. This gene has been associated with Leber congenital amaurosis, a disease characterized with severe retinal dystrophy that may lead to blindness in infants (den Hollander et al., 2006), Joubert syndrome, a neurological disease (Sayer et al., 2006) (Abdelgadir et al., 2019), defects in Sonic Hedgehog machinery due to disruption of primary cilium (Kilander et al., 2018), and the embryonically fatal Meckel syndrome which is characterized with multisystem fibrosis (Peng et al., 2022).

Our investigation of these genes led to their exclusion as candidate genes for being functionally unrelated to our phenotype.

6.3.2 Second scenario

The second scenario was to filter out the variants by following the mode of inheritance as autosomal recessive considering the allele state of compound heterozygosity. Starting with family B, we have put together all the heterozygous variants shared between the affected siblings (II-14, II-15, II-16, and II-17) and excluded the ones shared with their unaffected siblings (II-19 and II-20). Then crossed the results with heterozygous variants from the father (I-2) and the mother (I-3), then check the compound heterozygous variants of the affected that are found on the same gene (different position on each allele) Figure 78. The same filtration criteria were applied to family A and resulted in the identification of different genes.

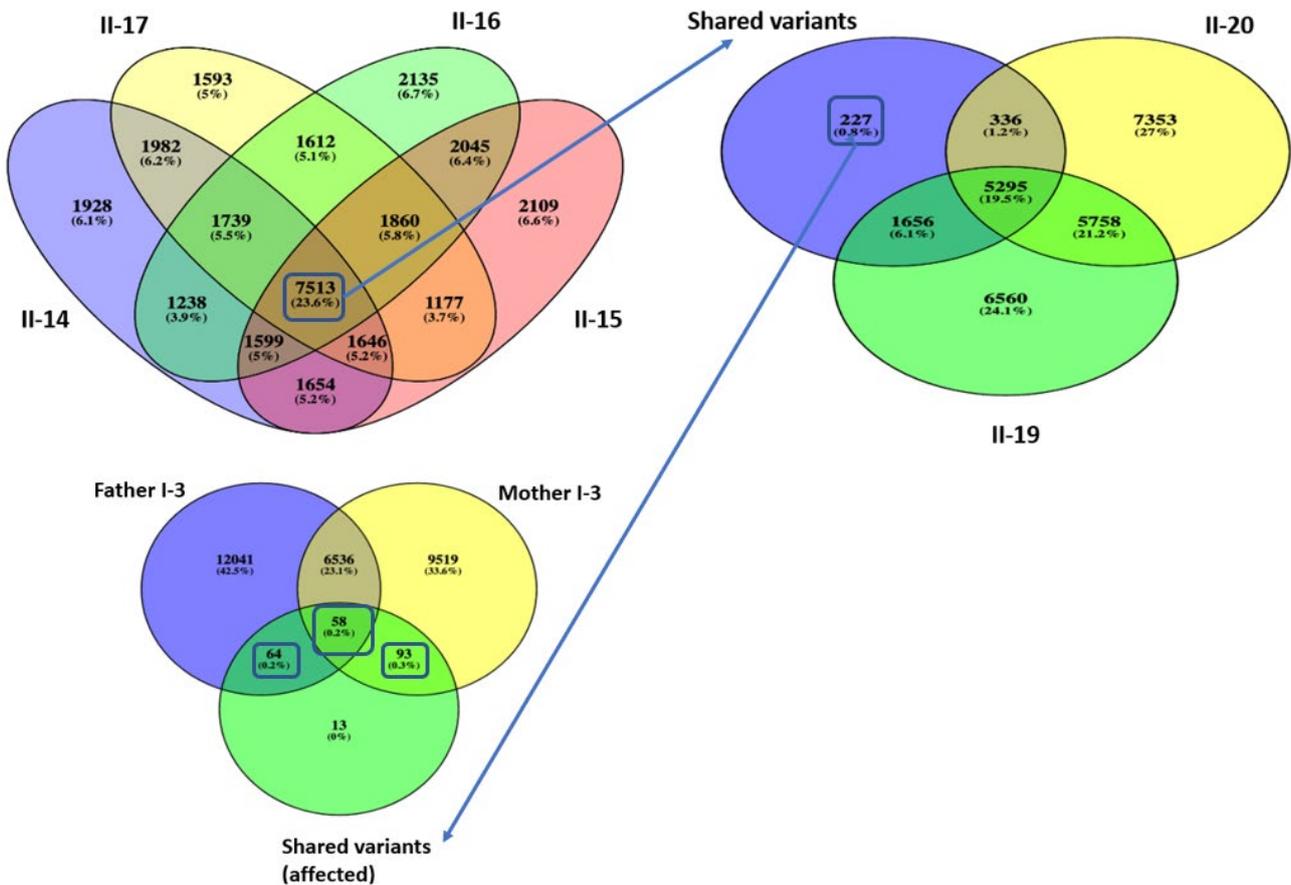


Figure 78: Filtered nucleotide variants from exome data, based on chromosomal positions, which showed autosomal recessive inheritance model of compound heterozygosity in non-consanguineous thalassemia disease family B.

As for the gene list of compound heterozygous variants in family B, three genes have been found which are *ZNF215*, *C11orf16*, and *DUX4L4* (Table 25).

Table 25: Filtered nucleotide variants from exome data, based on chromosomal positions, which showed autosomal recessive inheritance model of compound heterozygosity in non-consanguineous thalassemia disease family B.

Gene	Chromosome	Position (GRCh37)	HGVS	Info	Rs#
ZNF215	11	6976988 (Father)	c.780A>T	expressed in the testis	rs2239731
		6977175 (mother)	c.967G>C		rs2239730
C11orf16	11	8947586 (Father)	c.628G>C	chromosome 11 open reading frame 16	rs11042127
		8942942 8947283 (Mother)	c.1325G>A c.931C>G		rs2653601 rs3751066
DUX4L4	4	191003045 191003298 (Father)	g.873G>A c.1209G>A	Pseudogene	rs797032630 rs797038082
		191003081 191003118 191003123 (mother)	c.992G>C c.1029G>A c.1034C>G		rs782249755 rs376224108 rs782484580

ZNF215 is a zinc finger transcription factor imprinted in a tissue-specific manner predominantly expressed in the testis. Genomic imprinting refers to the phenomenon in which an individual expresses just one copy of a gene (inherited from either parent) while the other copy is silenced. Genomic imprinting does not modify the underlying DNA sequence. Instead, epigenetic insertion of chemical tags to the DNA during egg or sperm development silences its gene expression. *ZNF215* has been found to be associated to impaired spermatogenesis that is caused idiopathically or due to isolated (nonsyndromic) cryptorchidism as a biological cause of male factor infertility (Gianotten et al., 2003) (Massart & Saggese, 2010). It has been also associated with Beckwith-Wiedemann syndrome which is characterized mainly with overgrowth as children and possible hemihyperplasia (Alders et al., 2000). Additionally, upregulation of *ZNF215* was observed in patients with cytogenetically abnormal-acute

myeloid leukemia and patients with CA-AML who lose imprinting at the ZNF215 locus tend to have a poor five-year survival rate (Yang et al., 2021).

C11orf16 (Chromosome 11 Open Reading Frame 16), is a gene with low expression in most tissues. *C11orf16* may have a role in tumor suppression since its predicted protein shares 19.61% sequence similarity with p53-binding protein, another tumor suppressor (extracted from swiss-model data).

DUX4L4 (Double Homeobox 4 Like 4) is a pseudogene that is located in the subtelomeric region of chromosome 4q. While RT-PCR and in vitro expression tests suggest that a telomeric paralog of this gene is transcribed in some haplotypes where an algorithm generated description predicted that it has a role in transcription regulation, there is no evidence for transcription of the gene at this location (extracted from alliance of genome resources).

As for the gene list of compound heterozygous variants in family A, only one gene have been found which is *MUC4* (Table 26).

Table 26: Filtered nucleotide variants from exome data, based on chromosomal positions, which showed autosomal recessive inheritance model of compound heterozygosity in non-consanguineous thalassemia disease family A.

Gene	Chromosome	Position (GRCh37)	HGVS	Info	Rs#
MUC4	3	195510022	c.8429C>T	Glycoproteins that protect epithelial surfaces	rs201092610
		195512510	c.5941C>A		rs552200694
		195506258	c.12193C>A		rs11929196
		195515435	c.3016G>A		rs199540819

MUC4 is a member of different types of glycoproteins called mucins that are produced by different types of epithelial cells. Secreted or membrane-associated glycoproteins (viscous secretion), these molecules are thought to be crucial to the protection, differentiation, renewal, and function of epithelial surfaces (Debailleul et al., 1998) (Moniaux et al., 2000). In Chinese population, variants in *MUC4* have been associated with lung cancer risk (Zhang et al., 2013). Inhibition of apoptosis and stimulation of proliferation in epithelial and cancer cells are the results of CDKN1B down-regulation caused by the development of a *MUC4* ligand-receptor complex with ErbB2. *MUC4* has a critical role in promoting cancer progression and tumor development through its anti-adhesive properties and

inhibition of apoptosis rather than proliferation (Carraway et al., 2002) (Karg et al., 2006) (Gao et al., 2021).

Our investigation of these genes led to their exclusion as candidate genes for being functionally unrelated to our phenotype.

6.3.3 Third scenario

The clinical phenotype of the patients being investigated is used to filter WES data using a virtual gene panel consisting of clinically relevant genes (Neveling et al., 2013). This raises the likelihood of detecting disease-causing variations in the WES data by narrowing the focus of the analysis to just relevant candidate genes.

First, based on chromosomal positions, an average of ~47000-48000 variants were detected in each exome. Filtration criteria of these variants included sequence variants that are exonic, non-coding UTRs, known promoter regions, and splice sites with the exclusion of intergenic and intronic variants. Then, filtering based on minor allele frequency (MAF < 0.05) has resulted in approximately 2100 variants per exome. By filtering out nonsynonymous variants, we ended up with total filtered variants of about 1100-1200 (Figure 79)(Table 27).

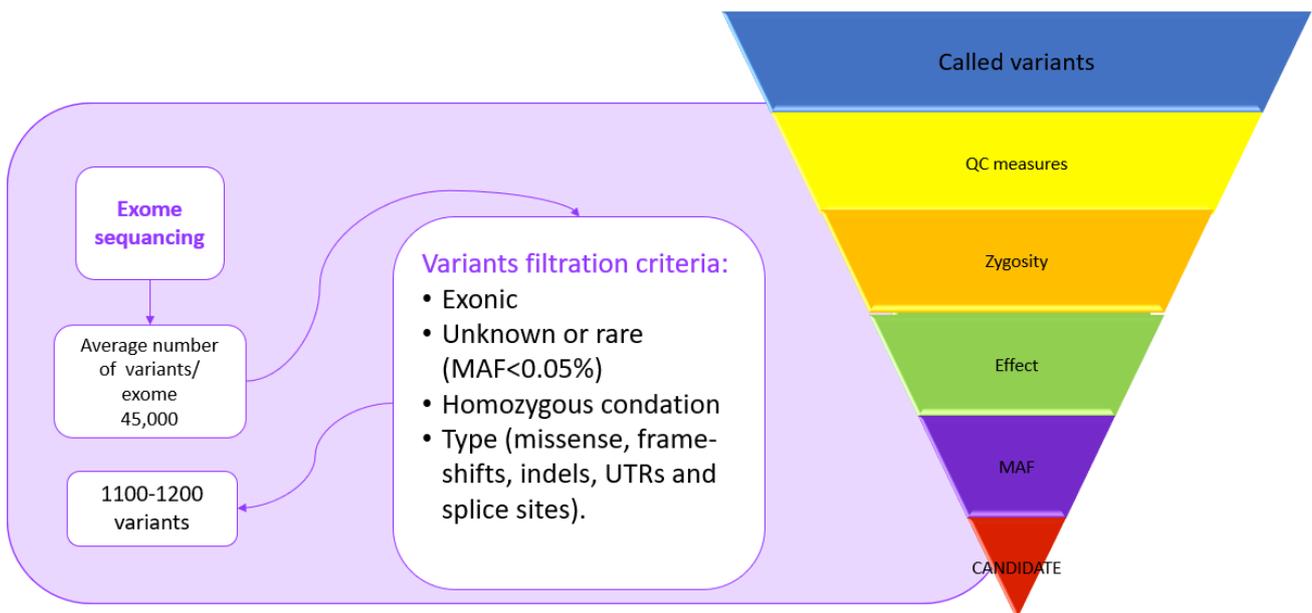


Figure 79: Variants filtration criteria.

Table 27: Filtering criteria for variants.

'Coding' indicates variants in coding regions such as missense, frameshifts, indels, and also UTRs and splice sites were included. 'Novel' indicates variants that are unknown or extremely rare (MAF < 0.05%). 'Filtered variants' are the total filtered variants with exclusion of synonymous variants. 'Homozygous' are the variants found in homozygous state in total filtered variants. 'Heterozygous' are the variants found in heterozygous state in total filtered variants.

Family member	Family	Status	Total	Coding	Novel (MAF)	Filtered variants	Homozygous state	Heterozygote
I-2	A/B	Unaffected	47445	7260	2147	1210	130	1080
I-3	B	Unaffected	43124	6479	2066	1040	100	940
II-20	B	Unaffected	46641	7198	2000	1107	97	1010
II-14	B	Affected	47665	7189	2200	1255	133	1122
II-17	B	Affected	47394	7236	1987	1163	139	1024
II-15	B	Affected	48255	7398	2187	1170	110	1060
II-16	B	Affected	48564	7245	2180	1159	101	1058
II-19	B	Unaffected	47741	6980	1983	1150	102	948
I-1	A	Unaffected	45457	5986	1845	1087	89	999
II-10	A	Affected	47268	7190	2140	1184	128	1056
II-11	A	Unaffected	45404	7100	2090	1205	112	1093
II-12	A	Unaffected	45677	5890	1960	1139	100	1039
II-13	A	Unaffected	45301	5890	1897	1155	106	1049
II-3	A	Unaffected	46663	6700	1899	1164	111	1053

After that, we have generated a gene list to include genes that have been reported to be involved in one way or another in hereditary anemia disorders. Generating gene panels to investigate inherited hemolytic anemia disorder have been previously employed (Agarwal et al., 2016) (Roy et al., 2016). Our list was assembled by looking into literature and several gene panels that are used by clinical laboratories to test for anemia related disorders. This list included, *ABCB7, ABCG5, ABCG8, ADA, ADA2, ADAMTS13, ADH5, AK1, ALAS2, ALDOA, AMN, ANK, ANK1, ATRX, BCL11A, BLM, BRCA2, BPGM, BRIP1, C15ORF41, CD59, CDAN1, CDIN1, CYB5A, CYB5R3, COL4A1, CUBN, DKC1, DHFR, DNAJC21, DNASE2,*

EFL1, EBP41, EBP42, ENO1, ERCC4, ERFE, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCM, FANCL, FTCD, G6PD, GATA1, GCLC, GIF, GLUT1, GLRX5, GPC, GPI, GSR, GPX1, GSS, GYPC, HEATR3, HK1, HSCB, HSPA9, HMOX1, KCNN4, KIF23, KLF1, LARS2, LPIN2, MTR, MTRR, NDUFB11, NHLRC2, NHP2, NT5C3A, NRF1, PALB2, PC, PDHA1, PDHX, PIEZO1, PFK, PFKM, PGK1, PKLR, PUS1, RAD51C, REN, RHAG, RPL11, RPL15, RPL18, RPL27, RPL31, RPL35A, RPL5, RPS10, RPS17, RPS19, RPS24, RPS26, RPS27, RPS28, RPS29, RPS7, SBDS, SEC23B, SLC19A1, SLC11A2, SLC19A2, SLC25A38, SLC4A1, SLC2A1, SF3B, SLX4, STEAP3, SPTA1, SPTB, SRP54, TCN2, TF, THBD, TMPRSS6, TRNT1, TSR2, TPI1, UMPS, UGT1A1, VPS4A, XK, YARS2. The advantages of using gene panels as a filtration search scenario include shorter turnaround times, easier data processing, more coverage in the regions of interest, and fewer incidental results (Wooderchak-Donahue et al., 2012) (Sun et al., 2015) (Kim et al., 2017).

The search through this gene list did not yield any direct known reported pathogenic variants (in a heterozygous state) that is shared between the father and all his affected children. Nevertheless, a reported pathogenic variant on *SBDS* (Shwachman-Bodian-Diamond syndrome) gene was found to be shared between the father and three of his affected children, II-15, II-16, and II-17 and one unaffected II-20 from family B, and absent from all other family members (affected and unaffected). *SBDS* encodes a conserved protein that is involved in the facilitating of ribosome assembly by removing eIF6, which inhibits the interaction between the ribosome's large and small subunits. Mutations in *SBDS* has been associated with aplastic anemia and Shwachman-Diamond syndrome (and has more than 80 identified variants), an autosomal recessive disorder that is predominantly characterized by hematological abnormality, increased risk in developing hematological malignancies, and pancreatic exocrine insufficiency (Kuijpers et al., 2005) (Boocock et al., 2003) (Wong et al., 2010). Ninety percent of individuals with Shwachman-Bodian-Diamond syndrome have a chromosomal 7 *SBDS* gene mutation. Though typically seen in younger age groups, acquired bone marrow failure or aplastic anemia can manifest at any age. The hematologic abnormalities in people with this syndrome are presented with thrombocytopenia which may present with life-threatening bleeding. Also, up to 80% of patients have macrocytic or normocytic anemia (Farooqui et al., 2023) (Taha et al., 2022). During times of rapid cell division, telomeres lengthen to safeguard the ends of chromosomes and prevent cellular senescence. Granulocyte telomeres were found to shorten abnormally in *SBDS* heterozygous individuals. (Thornley et al., 2002). Shortening of leukocyte telomeres via a telomerase-independent process appears to be a mechanism by which *SBDS* deficiency contributes to bone marrow failure (Calado et al., 2007).

The identified stop gained variation in our samples is a well studied variant that produces a truncated protein product due to translation interruption by the insertion of two nucleotides which created a premature stop codon (Table 28)(Nelson & Myers, 2018) (Wu et al., 2022) (Spangenberg et al., 2022).

Table 28: Pathogenic variant found in a heterozygous state shared between the father and two of the affected children.

Gene	Chromosome	Position (GRCh37)	HGVS	Rs#	Protein change	Molecular consequence
SBDS	7	66459273..66459274	c.183_184delinsCT	rs 120074160	p.Lys62Ter	Stop gained

In addition, we have identified a novel variant (unknown significance) in the *FANCC* gene (Table 29) in a heterozygous state that is shared between the father and II-14, II-15, II-16, II-17, II-19, II-20, II-11, II-12, and II-13 and absent from I-1, I-3, II-3, and II-10. This variant was confirmed using Sanger sequencing (Figure 80).

Table 29: Pathogenic variant found in a heterozygous state shared between the father and all affected individuals of family B.

Gene	Chromosome	Position (GRCh37)	HGVS	Rs#	Protein change	Molecular consequence
FANCC	9	97912300	c.591C>A	rs1457631500 C>A Novel	p.Asp197Glu	Missense

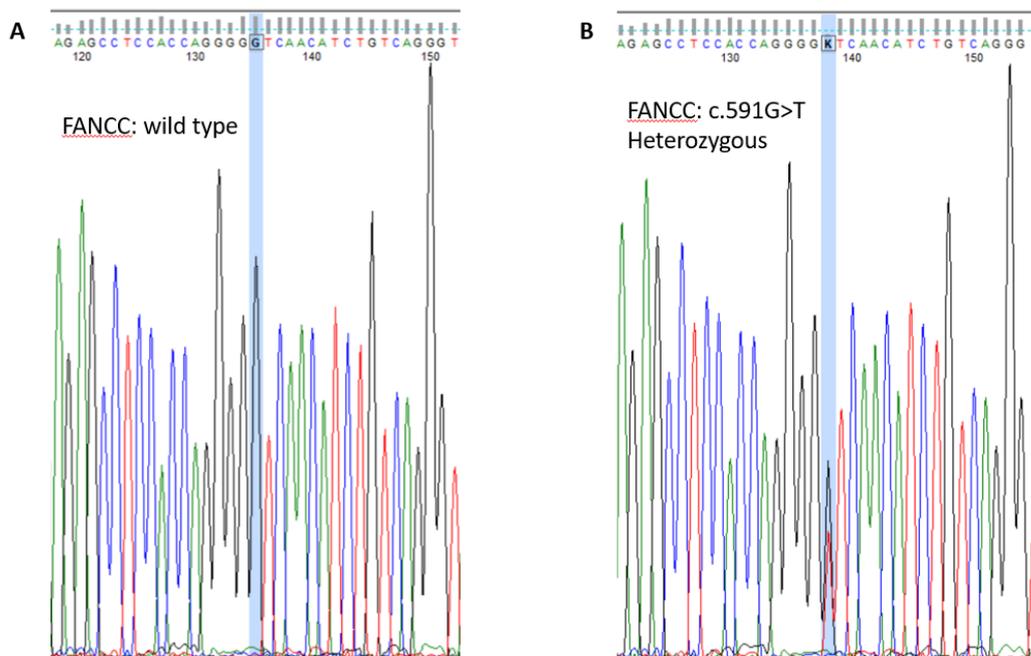


Figure 80: Sanger validation of FANCC variant.

A. Variant is not present in I-1 (mother of family A) B. Heterozygous variant in sample II-14.

FANCC (Fanconi anemia complementation group C) gene encodes a protein that potentially serves as a cell cycle checkpoint, delays DNA damage-induced apoptosis and stimulates homologous recombination repair preserving chromosomal integrity (the Fanconi anemia pathway) (Pang et al., 2001). Genomic instability is a hallmark of Fanconi anemia, which is caused by mutations in up to 22 Fanconi anemia genes and characterized by predisposition to tumor formation, bone marrow failure, and hematological abnormalities of all three blood cell lines (red blood cells, white blood cells, and platelets) (Bhandari et al., 2023). When DNA replication is blocked as a result of DNA damage, the FA (Fanconi anemia) pathway is activated. Interstrand cross-links (ICLs) are a specific form of DNA damage that elicits a response from the FA pathway. ICLs are caused by the aberrant attachment or linking of two nucleotides on opposing strands of DNA, which prevents DNA replication from proceeding normally. Mutations in this gene may lead to reduced or absent protein production, which lead to chromosomal breakage and impaired DNA repair and disruption of Fanconi anemia pathway which result in Fanconi anemia. When DNA replication is halted by the ICLs, it leads to either aberrant cell death from a failure to produce new DNA molecules or unchecked cell expansion and cancer predisposition from a failure of DNA repair mechanisms. Rapidly dividing cells, such those in bone marrow or a fetus, are especially vulnerable. Impairment of these cells causes the abnormally low number of blood cells of Fanconi anemia. Leukemia and other malignancies can arise when DNA damage accumulates and causes uncontrolled cell growth (Yousoufian, 1996) (Joenje & Patel, 2001) (Wang & Gautier, 2010) (Ceccaldi et al., 2016) (Liu et al., 2020). *FANCC* is a member of the Fanconi anemia core complex. Following its association with damaged chromatin, the FA core complex initiates the Fanconi anemia pathway by monoubiquitylating the heterodimer FANCI-FANCD2 (ID2) through UBE2T (*FANCT*). Upon activation, monoubiquitylated ID2 binds with the ICL to stimulate crosslink repair via subsequent effector proteins (Schubert et al., 2022). Mutations in *FANCC* may also enhance proapoptotic kinase (PKR) binding to *FANCC* and PKR activation (represses translation upon activation), ultimately resulting in growth suppression of hematopoietic progenitors and bone marrow failure (Zhang et al., 2004) (Liu et al., 2020).

6.3.3.1 In-silico investigation of FANCC variant effect

3D model generation

The mutant model (E197) of FANCC was successfully developed using the ModWeb web server. The experimental crystal structure of FANCC (PDB ID: 7kzp, Chain A) was selected as a template to build the mutant model. ModWeb generated the mutant model of FANCC based on the highest MPQS and sequence identity. The best model was selected based on the DOPE score of -250 Kcal/mol, which is a measure of the model's structural quality. The negative DOPE score indicates a good-quality model with low energy. Overall, the results suggest that the mutant model of FANCC developed using ModWeb is reliable and can be used for further analysis.

Dynamut Analysis

Based on the results obtained from the Dynamut server, it appears that the FANCC mutation of Asp197 to Glu197 is predicted to be stabilizing, with a $\Delta\Delta G$ of 0.932 kcal/mol (Figure 81). This suggests that the mutant protein may be more stable than the wild-type protein. However, it's important to note that other structure-based predictions suggest that the mutation may be destabilizing, with a $\Delta\Delta G$ mCSM of -0.848 kcal/mol and a $\Delta\Delta G$ SDM of 1.130 kcal/mol. Additionally, the NMA-based prediction using ENCoM suggests a small destabilization with a $\Delta\Delta G$ of 0.282 kcal/mol. Furthermore, the vibrational entropy analysis shows that there is a decrease in molecule flexibility with a $\Delta\Delta S_{\text{Vib}}$ ENCoM of -0.352 kcal.mol⁻¹.K⁻¹. The fluctuation map also shows that flexible nature of mutant protein model compared to wildtype structure.

The wild-type FANCC (Asp197) is found to form ionic interactions with several amino acids, including Leu199, Leu224, Thr194, Leu234, and Glu201. However, with the FANCC mutation (Glu197), the number of amino acids involved in ionic interactions reduces to only three, namely Thr194, Leu234, and Glu201. These results suggest that the mutation may have a stabilizing effect on the protein model, possibly due to the changes in the ionic interactions formed. Taken together, these results suggest that the impact of the mutation on protein stability is complex and may depend on other factors such as the specific conformation and context of the protein. Further MD simulation validation may be necessary to fully understand the effects of the Asp197 to Glu197 mutation on the protein's stability and function.

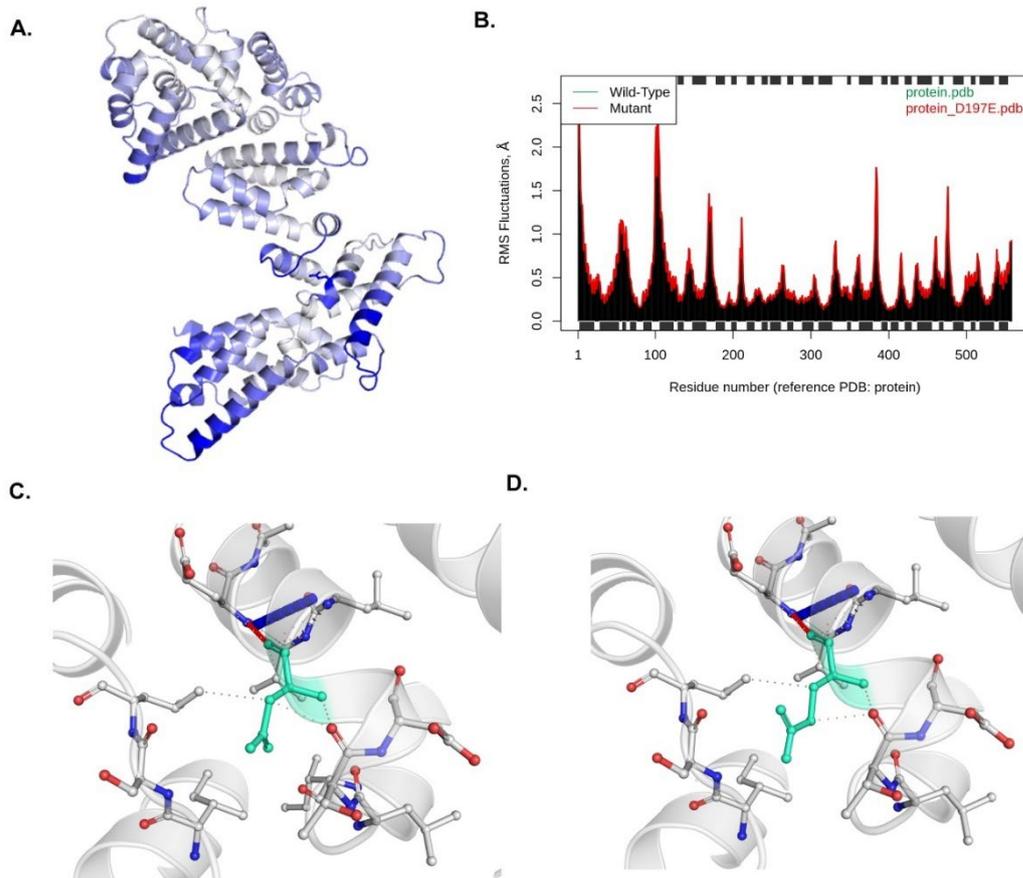


Figure 81: A. The mutant FANCC Δ vibrational entropy energy can be visually represented by coloring the amino acids based on the vibrational entropy change caused by the mutation.

Amino acids colored in blue indicate a rigidification of the structure, while those in red suggest a gain in flexibility. B. The wild-type and mutant sequences were extracted from their respective 3D structures and aligned to create an ensemble of normal mode data. The graph displays the results of the normal mode analysis for each sequence. C-D wild and mutant residues ionic interactions in FANCC.

MD Analysis

Molecular Dynamics (MD) simulations are a powerful tool used to study the behavior and dynamics of proteins over time. They provide information about the movement of atoms, the stability and flexibility of the protein structure, and the fluctuations in its conformation. In this study, MD simulations were carried out for both the wildtype and mutant forms of FANCC, a protein associated with Fanconi anemia. The simulations were conducted for a duration of 100 nanoseconds, during which the protein structures were allowed to evolve dynamically. RMSD (Root Mean Square Deviation) and RMSF (Root Mean Square Fluctuation) were calculated for both the wildtype and mutant forms of FANCC to compare their structural stability and fluctuations. The RMSD plot shows that the average RMSD of the wildtype FANCC protein is 3.88Å, while that of the mutant form is 10.96Å. This indicates that the mutant protein is more structurally unstable and undergoes larger structural changes compared to the

wildtype protein. Additionally, the RMSF plot shows that the wildtype protein exhibits less fluctuation in the region from AA 158 to 196 and from AA 479 to 497 when compared to the mutant form. This suggests that these regions may be important for maintaining the stability of the protein structure and may be affected by the E197 mutation (Figure 82). Overall, these results provide valuable insights into the structural dynamics of *FANCC* and the impact of the E197 mutation on its stability and fluctuations.

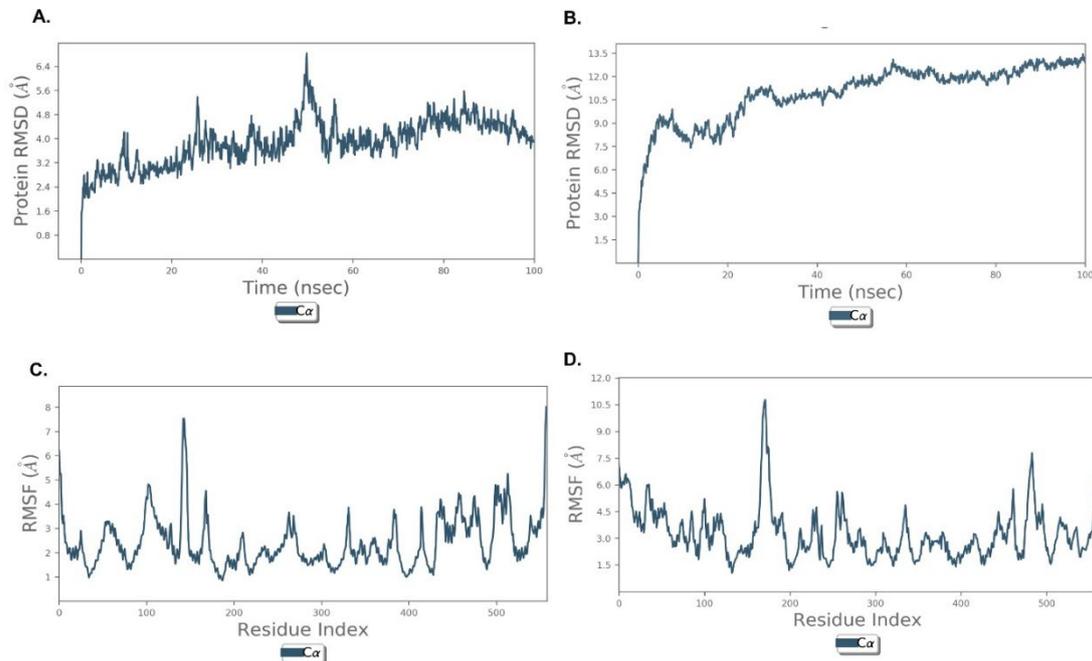


Figure 82: A-D RMSD and RMSF of *FANCC* wild and mutant models MD simulation for 100 ns.

The Desmond MD analysis was conducted to evaluate the secondary structure elements (SSE) contributing to the overall protein stability. The results showed that most of the proteins maintained an average of around 51.63% SSE, primarily composed of helices rather than strands and loops, except for the Asp197Glu mutation. Upon further investigation, it was revealed that the mutant exhibited a lower SSE percentage (30.46%) due to the conversion of residues present at positions 120 from strands to loops in comparison to the wild type (Figure 83).

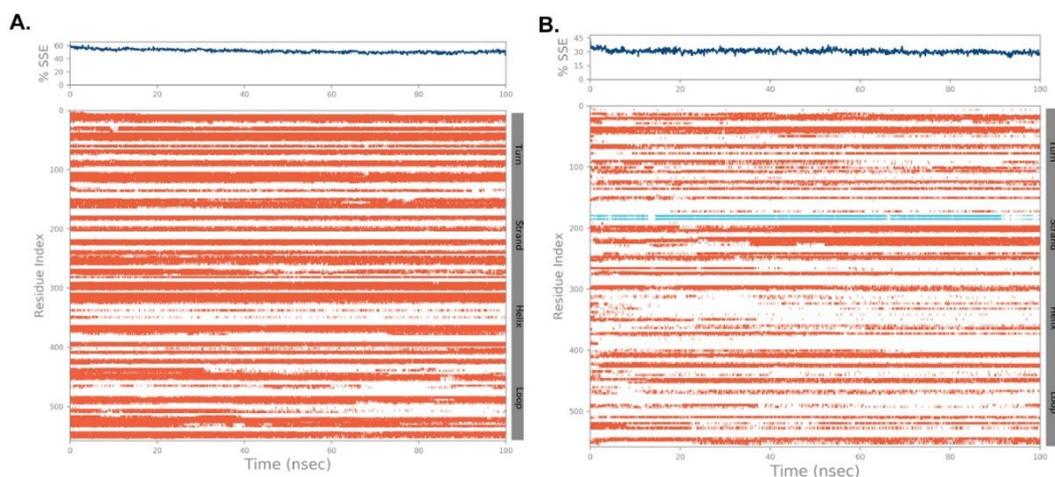


Figure 83: SSE components of wild and mutant models for 100ns MD simulation period.

This conversion resulted in the loss of SSE elements, which potentially led to damage to the protein's overall stability and conformational status (Figure 84).

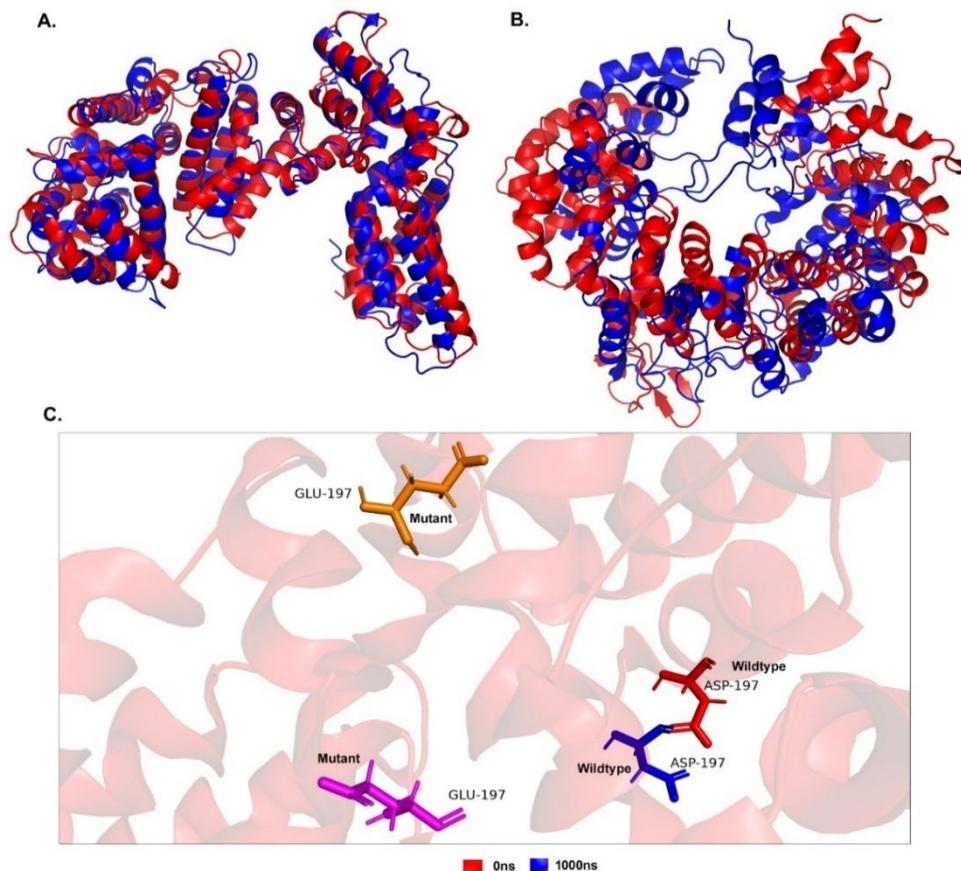


Figure 84: Structural conformation of wild and mutant FANCC protein structure in between 1ns to 100ns C. Mutant residues localization during simulation period.

This analysis provides valuable insights into the structural dynamics of FANCC and the impact of the E197 mutation on its stability and fluctuations. The results indicate that the mutant model is reliable and that the E197 mutation may have a stabilizing effect on the protein's structure, possibly due to changes in the ionic interactions formed. However, the impact of the mutation on protein stability is complex and may depend on other factors such as the specific conformation and context of the protein. MD simulations revealed that the mutant protein is more structurally unstable and undergoes larger structural changes compared to the wildtype protein. The loss of SSE elements in the mutant protein may also have led to damage to the protein's overall stability and conformational status. Therefore, the results suggest that further validation using more advanced MD simulation techniques, as well as experimental studies, may be necessary to fully understand the effects of the Asp197 to Glu197 mutation on the protein's stability and function. Moreover, the results obtained from different computational tools showed some discrepancies, which indicate that multiple computational methods

should be employed to obtain a more comprehensive understanding of the impact of mutations on protein stability. Overall, this analysis highlights the importance of computational approaches in predicting the effects of mutations on protein structure and stability, but it also emphasizes the need for further experimental validation.

6.3.4 Fourth scenario

Investigation of complex diseases is currently prioritized as a means of discovering highly penetrant variants. In our case, we decided to adapt an option that was proposed by Peterson and colleagues that suggested sequencing large families with several afflicted members in order to reduce the number of candidate variations in the dataset to only those that are present in all of the affected people (Petersen et al., 2017).

Failure of our three previous scenarios to explain the severity observed in our affected patients has led to a fourth scenario. Our fourth scenario was based on the fact that the father has to contribute a major thalassemia contributory allele to roughly half of his children of both sexes (and that cannot be G6PD) in the form of an allele of one specific gene. We sought to identify candidate genes with a heterozygous variant in the father, because he is just a carrier, shared by all the affected that is combined with either of the *HBB* severe variants results in an increased severity of hemolytic anemia (compound heterozygosity state coming from different genes).

First, we needed to make a list of heterozygous variants that are shared by the father (I-2), and all the affected children (II-10, II-14, II-15, II-16, II-17). Then, we had to exclude these variants from both mothers (I-1 and I-3). As for the rest of the unaffected siblings, if they do carry the father's allele, they would have the same genotype as the father (carrier state). Therefore, looking back at the hb electrophoresis results, which indicated that II-19 is a carrier, while II-3, II-11, II-12, and II-13 are normal, and since II-19 does not have the *HBB* variant, then we can assume that he has inherited the heterozygous pathogenic allele from his father. We decided to further narrow our list of candidate genes by including II-19 in the filtration to be added with the shared variants of the affected and father. Additionally, we will exclude II-3, II-11, II-12, and II-13 with I-1 and I-3. II-20 will be disregarded from the filtration since we do not know whether she is normal or a carrier (we do not know her hematological status) (Figure 85).

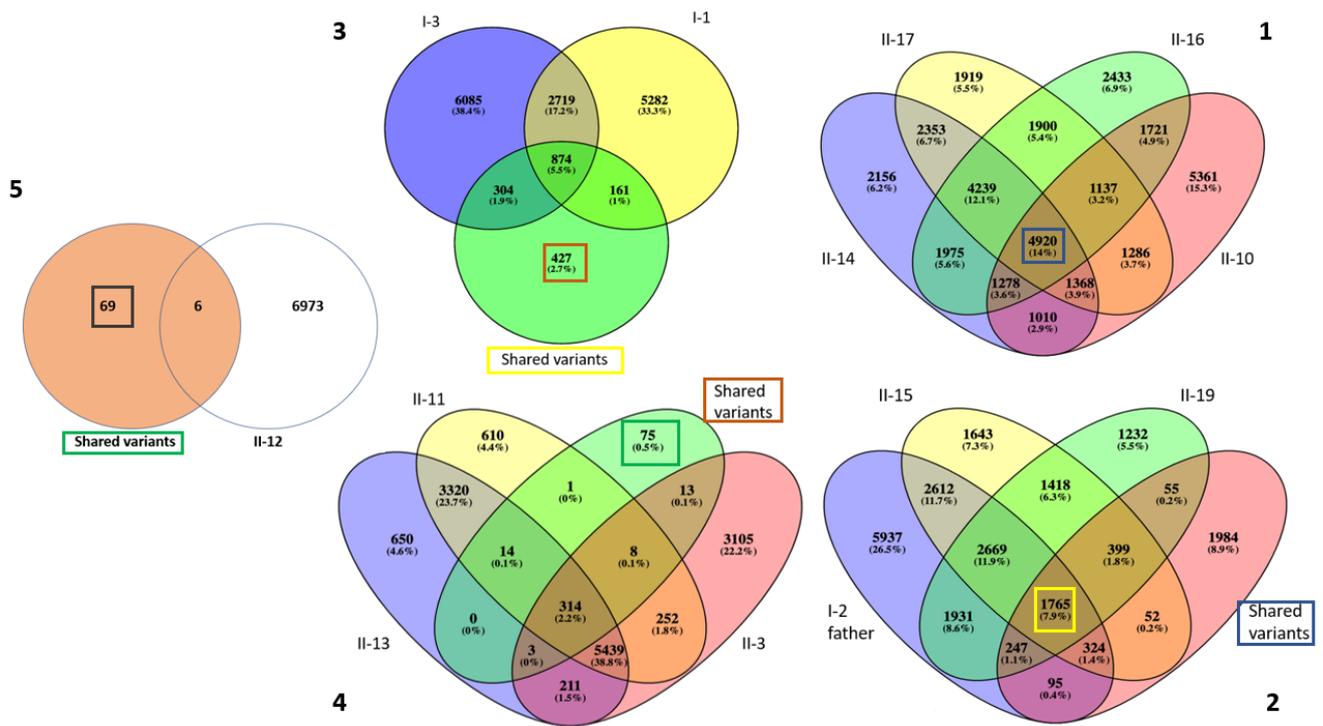


Figure 85: Filtered nucleotide variants from exome data, based on chromosomal positions, which showed shared heterozygous variants exclusively between the father, all the affected individuals, and I-19. Both mothers, II-11, II-12, II-13, and II-3 were excluded.

After filtration, we ended up with 69 variant positions that belong to 44 different genes. To further narrow our list of candidate genes, we excluded all common variants (MAF>0.02) and we ended up with 10 candidate genes (Table 30).

Table 30: Gene list of filtered nucleotide variants from exome data, based on chromosomal positions, representing scenario 4.

Gene	Chr#	Position (GRCh37)	Rs#	HGVS	Protein change	Molecular consequence
ASCL1	12	103352047	rs2136784885	c.25A>G	p.Ser9Gly	Missense
SBF2	11	9861208	rs117957652	c.3292C>G	p.Leu1098Val	Missense
TALDO1	11	763486	rs140985565	c.604G>A	p.Asp202Asn	Missense
OR5P3	11	7846682	rs200018758	c.838G>A	p.Val280Met	Missense
OR52B2	11	6191100	rs199830120	c.457C>T	p.Arg153*	Stop gained
OR52I2	11	4608908	rs138302486	c.866T>A	p.Ile289Asn	Missense
IFITM10	11	1769085	rs190833491	c.349C>T	p.Arg117Trp	Missense
UNC13B	9	35403767	rs142697917	c.4513C>T	p.Pro1505Ser	Missense
SMC5	9	72893539	rs141003657	c.676G>A	p.Glu226Lys	Missense
HLA-C	6	31239114	rs1071649	c.355C>A	p.Leu119Ile	Missense

6.3.4.1 Investigation of candidate genes

As for the gene list of heterozygous variants shared exclusively between the father, his affected children, and II-19, a list of 10 candidate variants was created. We started by looking into functionally related genes in an attempt to narrow the list and find variants that may have contributed in elevating the severity level of the cases.

ASCL1 (Achaete-scute homolog 1 gene) encodes basic-helix-loop-helix transcription factors that are associated with the differentiation, commitment, and activation of neuronal pathways (Soares et al., 2022). Glioblastoma (malignant brain tumor) stem cells' transcriptional subgroup affiliation is controlled by the proneural gene *ASCL1*, which does so by directly suppressing the mesenchymal gene *NDRG1* (Narayanan et al., 2019) (Vue et al., 2020). It has been also demonstrated to have a dedifferentiation role in mammalian Muller glia, a retinal neurodegenerative disease (Stanchfield et al., 2020). Also, *ASCL1* was found to be overexpressed in cases with chronic lymphocytic anemia (Malli et al., 2018).

SBF2, also known as *MTMR13*, encodes a member of the myotubularin-related protein family, is expressed in cytoplasm and nucleus of most tissues and regulates phosphatase activity. This gene has been associated with neuropathy and neurodegenerative diseases. Mutations in *SBF2* genes have been related to Charcot-Marie-Tooth Disease, a neuromuscular disorder with early onset of glaucoma (Laššuthová et al., 2018). Also, Abuzenadah and colleagues have associated this gene with severe congenital thrombocytopenia (decreased levels of platelets) by uncovering a homozygous mutation using WES data of an affected child and his father (Abuzenadah et al., 2013).

TALDO1, which is found on chromosome 11, encodes an enzyme that plays an important role in the pentose phosphate pathway (PPP) to produce ribose-5-phosphate, which is used in producing purine and pyrimidine precursors required in nucleic acid synthesis, and NADPH, which is used in lipid biosynthesis. Pentose phosphate pathway is located in the cytosol, and is a side-step from glycolysis with no ATP being created or destroyed along this pathway. However, if the organism only needs ribose 5-phosphate and not NADPH, then the non-oxidative phase of the pentose phosphate pathway may be used to produce 3 molecules of ribose 5-phosphate from 2 molecules of fructose 6-phosphate and 1 molecule of glyceraldehyde 3-phosphate. Conversely, if the cell needs NADPH but not ribose 5-phosphate (like a red blood cell or the liver after a high-carbohydrate meal), the oxidative and non-

oxidative parts can be combined to convert a molecule of G-6-P into 12 molecules of NADPH, 6 molecules of carbon dioxide, and a phosphate (Bender & Mayes, 2018).

Producing erythrose-4-phosphate, which is a prerequisite for the synthesis of aromatic amino acids such as tryptophan, phenylalanine, and tyrosine, is one of the functions of this pathway as well. In addition, it helps in the protection of cellular integrity from reactive oxygen radicals (Banki et al., 1997). Red blood cells rely heavily on the pentose phosphate pathway to withstand oxidative stress and other types of damage. Hemolysis can be prevented due to NADPH (produced from NADP⁻), which inhibits redox processes that generate significant amounts of reactive oxygen species (ROS) by scavenging the -SH groups in hemoglobin.

TALDO1 mutations have been associated with risk of hepatocellular carcinoma (Grammatikopoulos et al., 2022). Mutations in *TALDO1* causes transaldolase deficiency which is an inborn error of pentose metabolism presented with hepatosplenomegaly, anemia, thrombocytopenia, and liver dysfunction (Banne et al., 2016) (Tylki-Szymańska et al., 2009) (Tylki-Szymańska & Stradomska, 2011) (Valayannopoulos et al., 2006). Transaldolase deficiency is a rare autosomal recessive disease that was first described in 2001 by Verhoeven et al in a patient with liver disease (Verhoeven et al., 2001). In a retrospective study that involved 34 patients with transaldolase deficiency, anemia was seen in about 75% of the patients of both early and late onset of the disease (Williams et al., 2019). Moreover, a case report of a fetus with a prenatal diagnosis of transaldolase deficiency was manifested with hepatosplenomegaly and hemolytic anemia among other clinical features (Xue et al., 2022).

OR5P3, *OR52B2*, and *OR52I1* are all genes that encode olfactory (odorant) receptors located on chromosome 11. Olfactory receptors in the nose engage in a chemical reaction with odorant molecules, which in turn activates a neural response that results in the experience of smelling. The olfactory receptor gene superfamily is the largest in the human genome with about 800 OR identified genes localized in different chromosomal loci but only roughly 400 are intact, and between any two individuals, there will be a 30% difference in how their OR alleles function (Malnic et al., 2004) (Olender et al., 2008) (Li et al., 2022). Mutations in olfactory genes would alter odor perception (Keller et al., 2007). In 2010, a genome wide association studies of 848 blacks with sickle cell anemia, conducted by Solovieff and colleagues have suggested that olfactory gene cluster may have a role in the regulation of fetal hemoglobin (2 α -2 γ) levels by regulating the expression of gamma globin gene. They suggest that the chromatin structure of the beta-globin cluster might be modified by polymorphisms in the olfactory receptor region (*OR51B5* and *OR51B6*) located upstream of the beta globin cluster, hence

affecting HbF levels which may play a role in the severity of the disease. They validated their most significant SNPs by testing them on subjects with heterozygous beta thalassemia or HbE (Solovieff et al., 2010). In 2011, Ziffle, Yang, and Chehab studied patients with the homozygous 118 kb beta globin deletion that causes beta thalassemia (due to the deletion). This deletion extends to include four functional olfactory receptor genes (*OR52Z1*, *OR52A1*, *OR52A5*, and *OR52A4*). Since *OR52A1* has been discovered to include a gamma globin enhancer, which allows for the constant expression of the fetal gamma globin genes, they speculated that the loss of the gamma globin enhancer element contributes to the anemia severity in beta thalassemia patients who are homozygous for the 118 kb deletion (Van Ziffle et al., 2011).

IFITM10 (Interferon Induced Transmembrane Protein 10) gene has been associated with the modulating of innate immunological and inflammatory responses to viruses (Aizaz et al., 2023).

UNC13B (unc 13 homologue B) gene encodes munc13-2 protein, which is a presynaptic protein that is highly expressed in the cerebral cortex of the brain and is thought to influence neuronal excitability by priming and fusing synaptic vesicles. In a recent study, mutations in *UNC13B* have been found in twelve people with partial epilepsy and/or febrile seizures from eight different families with focused seizures and focused discharges in EEG recordings, however the patients ultimately had no signs of intellectual or developmental delay (Wang et al., 2021). Additionally, this gene is increased by hyperglycemia and is expressed in the cortical epithelial cells of the kidney. A correlation between polymorphisms in the *UNC13B* gene and the risk of developing diabetic kidney disease in a Chinese Han population with type 2 diabetes have been demonstrated by Wang et al in 2019. They found a strong correlation between having the haplotype GGCCG and an increased chance of developing the disease (Wang et al., 2019). *UNC13B*, was reported to have elevated mRNA expression in patients with T-cell acute lymphoblastic leukemia (Wang et al., 2022). In 2020, Edemir demonstrated that there is a predictive relevance in knowing whether or not a given tumor expresses organic anion or cation transporters, and he identified the expression of *UNC13B* (anion transporter) with favourable prognostic outcome (Edemir, 2020).

SMC5 (Structural maintenance of chromosomes protein 5) gene is a family of ubiquitous ring-shaped proteins that include cohesins, condensins, and smc5/6 complex. They are predicted to be involved in repair of homologous recombination of double strand DNA breaks, ATP binding activity, chromatin folding in interphase, genome stability, structural maintenance of chromosomes, and involved in immunity by preventing episomal DNA transcription (Venegas et al., 2020) (Diaz & Pecinka, 2018)

(Decorsière et al., 2016). Mutations in components of smc5/6 complex causes neurodevelopmental abnormalities and anemia (Atkins et al., 2020) (Grange et al., 2022). Furthermore, Rossi and colleagues have suggested in 2020 that the smc5/6 complex supports DNA repair and genome stability by working in tandem with the Fanconi anemia pathway, and variations in the smc5/6 complex may exhibit as Fanconi anemia related diseases (Rossi et al., 2020). The DNA replication machinery frequently encounters RNA-DNA hybrids as a natural obstacle. Accumulation of RNA-DNA hybrids causes replication stress, DNA damage, and disruption of genomic integrity when RNase H enzymes are absent. In the absence of RNase H enzymes, Lafuente-Barquero et al. showed that Mph1, the yeast homolog of Fanconi anemia protein M (FANCM), is essential for cell survival. The buildup of RNA-DNA hybrids and RNA-DNA hybrid-dependent DNA damage, as measured by Rad52 foci, requires the Mph1 helicase domain to remain intact. When RNA-DNA hybrids build up, Mph1 condenses into foci. The Smc5/6 complex is required to control the activity of Mph1 at hybrids. This is supported by the finding that the buildup of hazardous recombination intermediates causes Mph1-dependent synthetic lethality upon simultaneous silencing of RNase H2 and Smc5/6 (Lafuente-Barquero et al., 2017).

HLA-C (Human leukocyte antigen c) gene is expressed in almost all cell types and belongs to the class I major histocompatibility complex. It plays an important role in adaptive and innate immunity, antiviral immunity and reproduction, and protection against cancer and tumorigenesis (Castro et al., 2019). Bone marrow and kidney transplant recipients are regularly typed for polymorphisms that may be found in exons 2 and 3 of *HLA-C* gene for its implication in allograft rejection. The major histocompatibility complex can bind and recognize unlimited types of endogenous and exogenous antigens. *HLA-C* have been shown to correlate with many types of autoimmune disease such as rheumatoid arthritis and Crohn's disease (Siegel et al., 2019). Many research groups have investigated the involvement of HLA class I mutations in acquired aplastic anemia and demonstrated that certain HLA class I alleles (specifically *HLA-A* and *HLA-B*) enhance the chance for clonal complications and the severity of the disease was presented in an age-dependent manner (Babushok et al., 2017) (Olson et al., 2022).

6.4 Functional- related Candidate genes selection

In the scenario of compound heterozygosity between different genes, many genes in the generated list have been ruled out for being functionally unrelated. The remaining genes are *SMC5* and *TALDO1* (Table 31)(Figure 86)(Figure 87).

Table 31: Two functionally related gene candidates from the scenario of compound heterozygosity between different genes.

Gene	Chr#	Position (GRCh37)	Rs#	Aggregate allele frequency	1000 genome x30	GnomAD	ClinVar clinical significance
SMC5	9	72893539	rs141003657	0.0001	0.0006	0.0001	Not reported
TALDO1	11	763486	rs140985565	0.0001	0.0002	0.0001	Uncertain significance

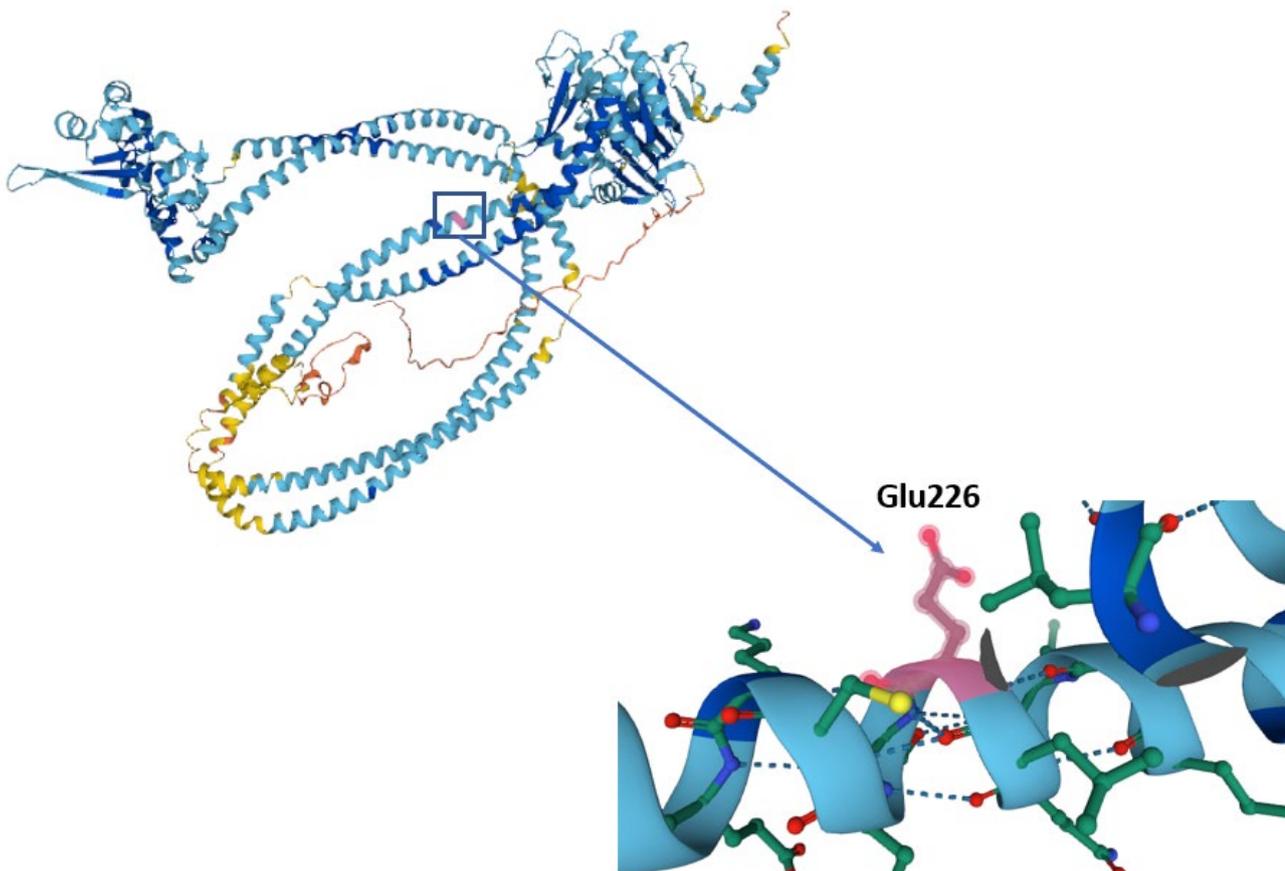


Figure 86: *SMC5* protein 3D structure with zoomed in residue position 226 (Glu).

Source: UniProtKB reviewed (Swiss-Prot) (Q8IY18).

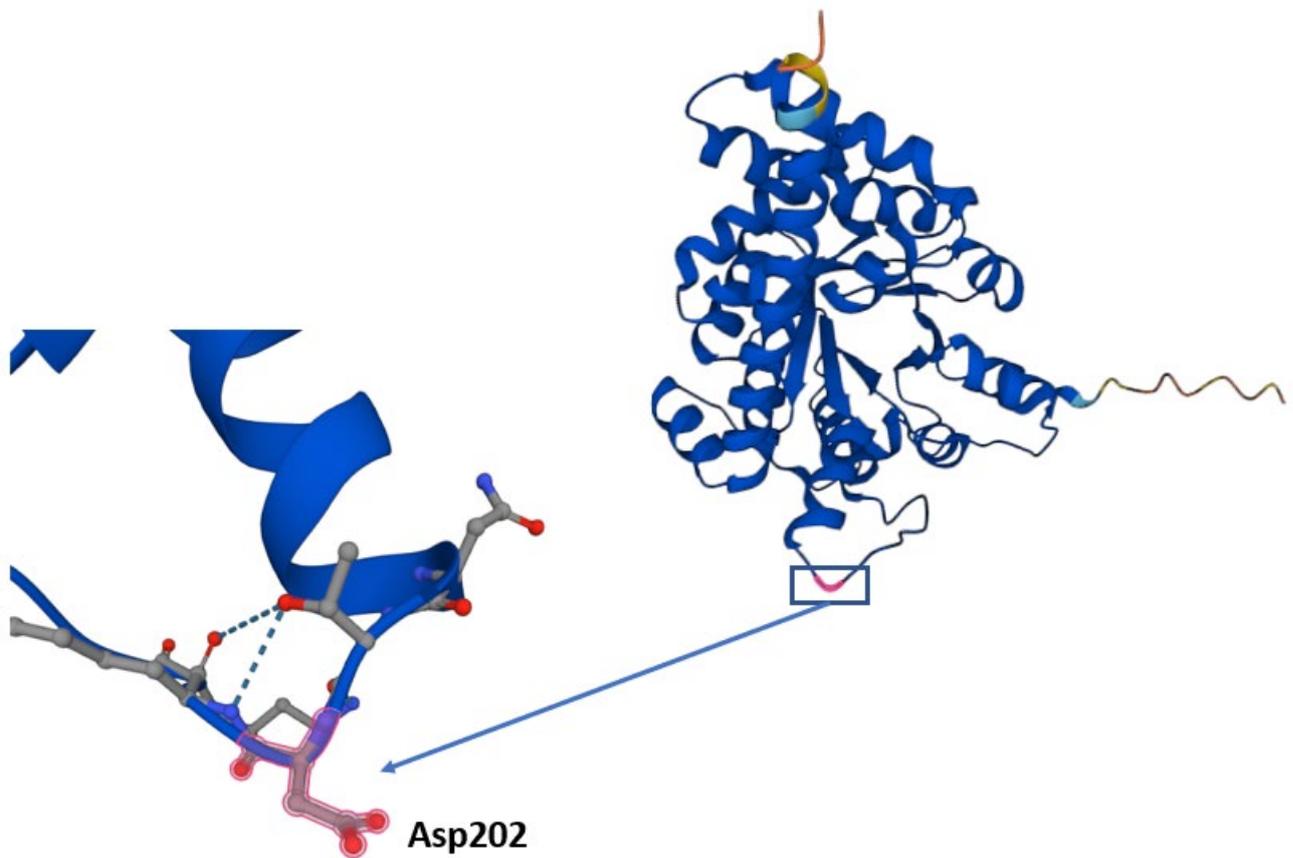


Figure 87: TALDO1 protein 3D structure with zoomed in Asp residue in position 202 (PDB ID: 1f05).

Sanger validation (Figure 88)(Figure 89)

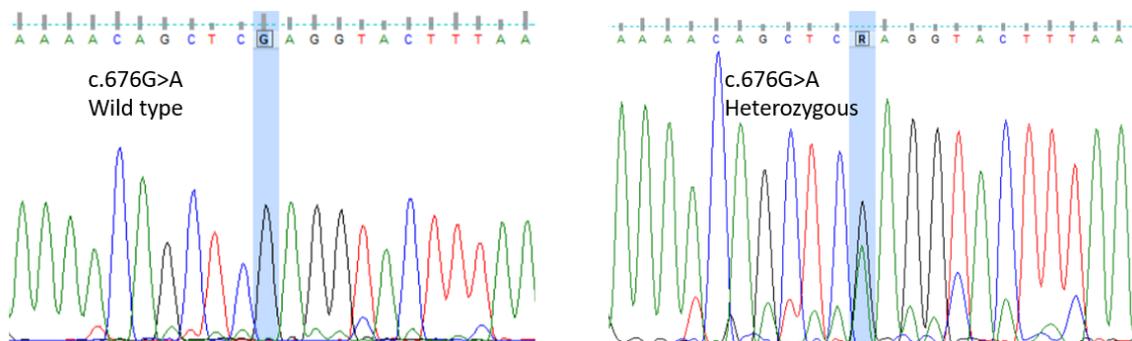


Figure 88: Sanger validation of TALDO1 variant.

Variant is not present in I-1 (mother of family A) on the left. Heterozygous variant in father's sample I-2 on the right.

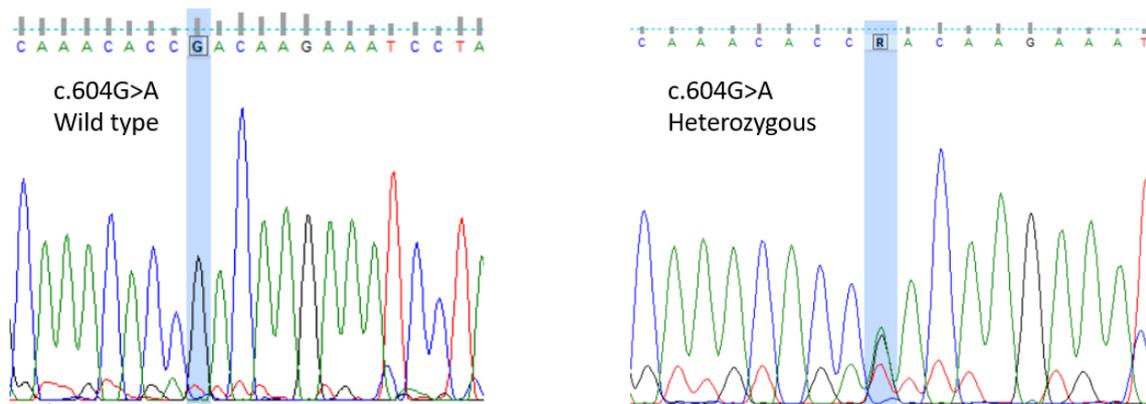


Figure 89: Variant is not present in I-1 (mother of family A) on the left. Heterozygous variant in father's sample I-2 on the right.

To further investigate these variants, they were screened by diverse systems biology approaches, such as gene expression, genotype-phenotype correlations in knock-out mouse models, and pathway enrichment analysis. Sequence and structure based approaches were both adopted to predict the impact of deleterious missense variations in the human *SMC5* and *TALDO1* genes on the protein's structure and function using various computational modeling and in-silico analysis (Hossain et al., 2020).

6.4.1 Pathway analysis

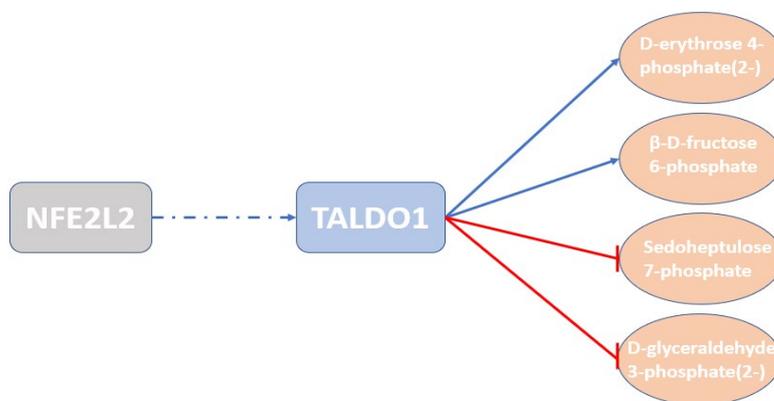


Figure 90: Transaldolase and metabolic stability along the pentose-phosphate pathway.

The rectangular shapes represent proteins, while oval shapes represent molecule/chemical. Blue lines indicate upregulation, red lines indicate downregulation, and dashes indicate binding.

In the pentose phosphate pathway, TALDO1 which encodes the enzyme Transaldolase upregulates and downregulates the quantity of a number of molecules involved in the pathway by chemical reactions (Figure 90), while NFE2L2 (nuclear factor (erythroid-derived 2)-like 2), also called NRF2, is a transcription factor that binds to TALDO1 to upregulate its expression (Cano et al., 2021) (Wu et al., 2011). NFE2p45 is the founding member of NRF2. The transcription factor NFE2p45 was initially

discovered to bind to the NFE2 element, which is part of the erythroid gene regulatory network. Researchers have discovered that human Nrf2 is a transcription factor that interacts with the beta globin gene's regulatory region (Kasai et al., 2018).

On the other hand, the SMC5/6 complex of structural maintenance of chromosomes is an important component in DPC (DNA-protein cross-links) repair; this complex has been preserved throughout evolution. DPCs are proteins that are covalently bonded to chromosomal DNA, and they are extremely harmful DNA lesions. DPCs, if left unrepaired, create a physical barrier that prevents DNA replication and transcription from occurring (Dvořák Tomašíková et al., 2023). Smc5/6 is a crucial complex that protects the integrity of homologous recombination and is activated in response to replication stress (Etheridge et al., 2021). With the use of single-molecule imaging and force manipulation, Tanasie et al. studied how the Smc5/6 interacted with structures resembling fork junctions, which are involved in DNA replication and repair. Smc5/6 is shown to recognize replication stress or DNA lesions by targeting and stabilizing these DNA substrates (Tanasie et al., 2022). Crabben and colleagues report a chromosomal breakage condition linked to severe pulmonary illness in infancy has been discovered. There was evidence of mixed T and B cell immunodeficiency in the lungs of four infants from two unrelated families who died of viral pneumonia. These infants died at early age after developing bonemarrow failure and increased susceptibility to infection. NSMCE3 (also known as NDNL2), encoding a member of the SMC5/6 complex was shown to have homozygous missense mutations by whole exome sequencing (van der Crabben et al., 2016).

6.4.2 Gene expression

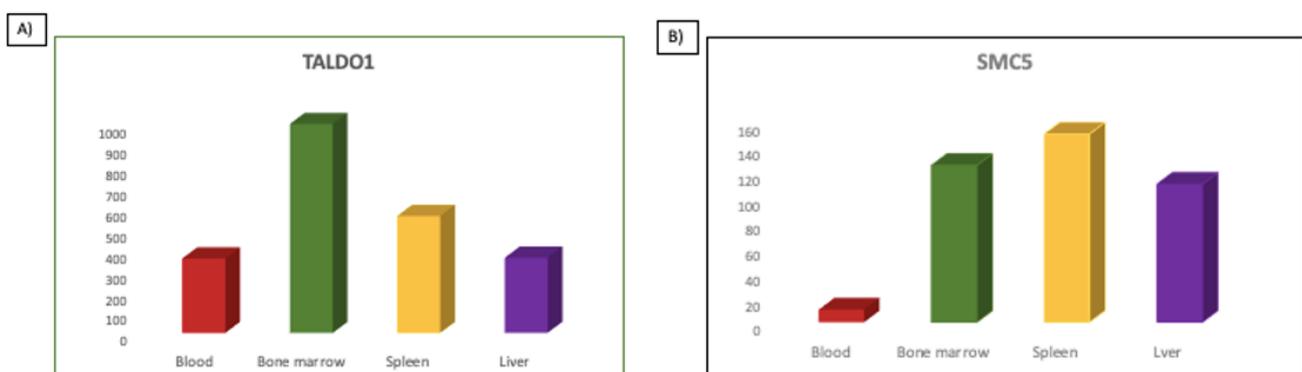


Figure 91: (A&B) Ensembl based transcriptomics expression data of different genes across different tissues and organs.

Tissues include spleen, liver, bone marrow, and blood. Transcription levels are in FPKM (fragments per kilobase of exon model per million mapped reads) and TPM (transcripts per million). Transcription scale: low (0-10), medium (11-1000), and high (>1000).

The Ensembl (<https://www.ensembl.org/index.html>) database was used to determine the expression status of potential candidate genes across different tissues (blood, bonemarrow, spleen, and liver). From the output, we selected only the expression data of query genes that have shown the significant expression (TPM of >40). Both genes show medium level of expression in all selected tissues, except for TALDO1 which showed high expression in bonemarrow (Figure 91).

6.4.3 Expression conservation in primate tissues

Although data on *SMC5* and *TALDO1* expression in bone marrow and spleen were not available for comparison, expression of these two genes was found conserved in both whole blood and liver tissues (Table 32) (Thierry-Mieg & Thierry-Mieg, 2006).

Table 32: Expression/ conservation of Taldo1 and SMC5 in primate tissues.

Gene expression in 15 primates, 16 tissues, from the NHPRTN project in sFPKM															
TALDO1	HUM	CHP	PTM	JMI	RMI	RMC	CMM	CMC	BAB	SMY	MST	SOM	OWL	MLM	RTL
Cerebellum		65.5	53.2	106	65.5							57.1		12.4	
Brain	65.5	30.6	65.5	57.1	57.1		40.3	61.1	61.1	30.6		70.2		17.6	
Pituitary		26.6	53.2	61.1	57.1		70.2	75.3	49.7	70.2	70.2				
Ovary	61.1				53.2										
Testis	65.5				37.6										
Kidney	122	57.1	70.2	99.3	75.3		80.7	80.7	92.7	92.7	114	122		30.6	
Liver	99.3	46.3	65.5	92.7	65.5	86.5	92.7	80.7	80.7	75.3	140	122		53.2	
Heart	46.3	26.6	61.1	80.7	65.5		70.2	75.3	61.1	70.2	65.5	75.3			
SkeletalMuscle	80.7	26.6	43.2	53.2	40.3		40.3	46.3	40.3	213	65.5	35.1		46.3	
Lung	75.3	26.6	106	92.7	80.7		92.7	99.3	92.7	173	106	131		80.7	
Colon	65.5	23.2	75.3	92.7			122	99.3	86.5	75.3	161	199		32.8	
BoneMarrow		161	371	371	426			456	323	281	213	397			
Spleen		28.5	99.3	131	86.5		106	114	131	185	199	161		28.5	
LymphNode	75.3	26.6	80.7	99.3			92.7	92.7	86.5	70.2	106	151			
Thymus		49.7	92.7	92.7	99.3		106	114	106	86.5					
WholeBlood	281	122	524	1589	978	978	795	524	562	1291		426	371		456

Gene expression in 15 primates, 16 tissues, from the NHPRTN project in sFPKM															
SMC5	HUM	CHP	PTM	JMI	RMI	RMC	CMM	CMC	BAB	SMY	MST	SOM	OWL	MLM	RTL
Cerebellum		10.8	37.6	21.6	26.6							30.6		37.6	
Brain	13.3	8.78	21.6	16.4	17.6		20.2	16.4	9.41	28.5	14.3	15.3		23.2	
Pituitary		37.6	53.2	37.6	43.2		40.3	35.1	28.5	46.3	40.3				
Ovary	28.5				70.2										
Testis	26.6				43.2										
Kidney	21.6	17.6	26.6	18.8	18.8		18.8	20.2	10.1	37.6	26.6	23.2		32.8	
Liver	14.3	23.2	12.4	9.41	10.1	6.65	8.78	8.78	5.40	17.6	10.8	14.3		16.4	
Heart	18.8	21.6	20.2	14.3	17.6		16.4	14.3	8.19	23.2	24.8	12.4			
SkeletalMuscle	23.2	9.41	14.3	13.3	12.4		13.3	11.6	6.21	28.5	13.3	6.65		24.8	
Lung	21.6	26.6	46.3	40.3	37.6		37.6	43.2	32.8	32.8	28.5	30.6		46.3	
Colon	13.3	17.6	28.5	28.5			32.8	26.6	16.4	28.5	16.4	16.4		21.6	
BoneMarrow		23.2	40.3	32.8	24.8			35.1	11.6	53.2	40.3	30.6			
Spleen		30.6	53.2	43.2	46.3		46.3	49.7	24.8	37.6	28.5	35.1		65.5	
LymphNode	17.6	46.3	53.2	65.5			53.2	65.5	28.5	53.2	32.8	35.1			
Thymus		35.1	53.2	49.7	49.7		49.7	49.7	28.5	49.7					
WholeBlood	26.6	30.6	49.7	24.8	35.1	28.5	32.8	26.6	32.8	37.6		13.3	13.3		37.6

6.4.4 Knockout mouse model

To understand the phenotypic and functional characteristics of the query genes, we used the Mouse Genome Informatics database (MGI) (<http://www.informatics.jax.org/>), which is a comprehensive resource of different mouse strains- inbred and KO to investigate homozygous knockout of both SMC5 (Table 33)(Figure 92) and TALDO1 (Figure 93).

SMC5 homozygous knockout in embryonic stem cells causes abnormal mitosis, increased apoptosis and a shift from pluripotency to differentiation (Blake et al., 2021).

Table 33: Phenotypes associated with SMC5 gene using mouse model (*Smc5^{tm1b(KOMP)Wtsi}*).

Phenotype	System	Zygoty
Preweaning lethality and complete penetrance	Mortality and aging	Homozygous
Decreased erythrocytes	Hematopoietic	Heterozygous
Decreased hematocrit	Hematopoietic	Heterozygous
decreased circulating glucose level	Homeostasis/ metabolism	Heterozygous
Decreased circulating bilirubin	Homeostasis/ metabolism	Heterozygous
Abnormal mitosis	Cellular	Homozygous

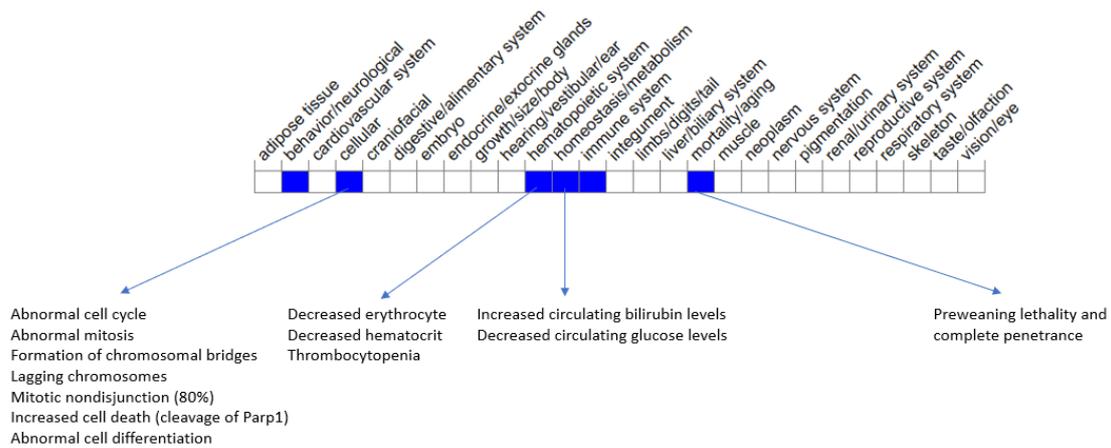


Figure 92: Phenotype overview of SMC5 homozygous knockout mouse (Blake et al., 2021).

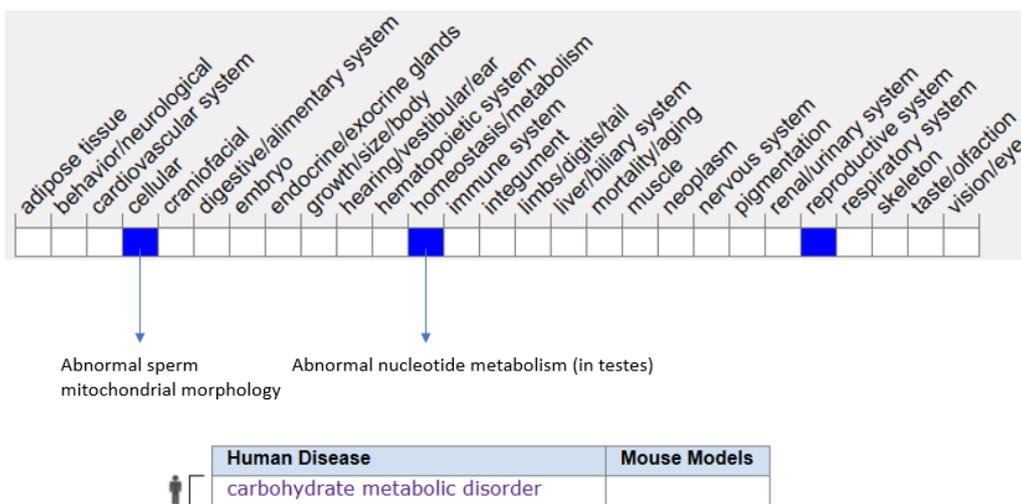


Figure 93: Phenotype overview of TALDO1 homozygous knockout mouse. Phenotypes associated with *Taldo1^{tm1Perl}/Taldo1^{tm1Perl}*.

6.4.5 Damage effect of the variant on the translated protein

6.4.5.1 Protein Modeling

The AlphaFold-generated model of TALDO1 had >90% of residues with pLDDT scores greater than 90, whereas the model of SMC5 had >90% of residues with pLDDT scores greater than 70. Upon analyzing the Ramachandran plot for both models, the majority of residues were situated in authorized and preferred regions, with only a small number in prohibited areas. The ProSA-web analysis showed that both models had high-quality crystal structures, with Z-scores of -8.39 and -10.91 for TALDO1 and SMC5, respectively. These results indicate that the AlphaFold algorithm is useful in predicting 3D protein structures and provides significant insights into their structure-function relationships. The GMQE score for the mutant models of both TALDO1 and SMC5, created using the Swiss Model server, was 0.61 with a QMEANDisCo Global score of 0.62 (Figure 94).

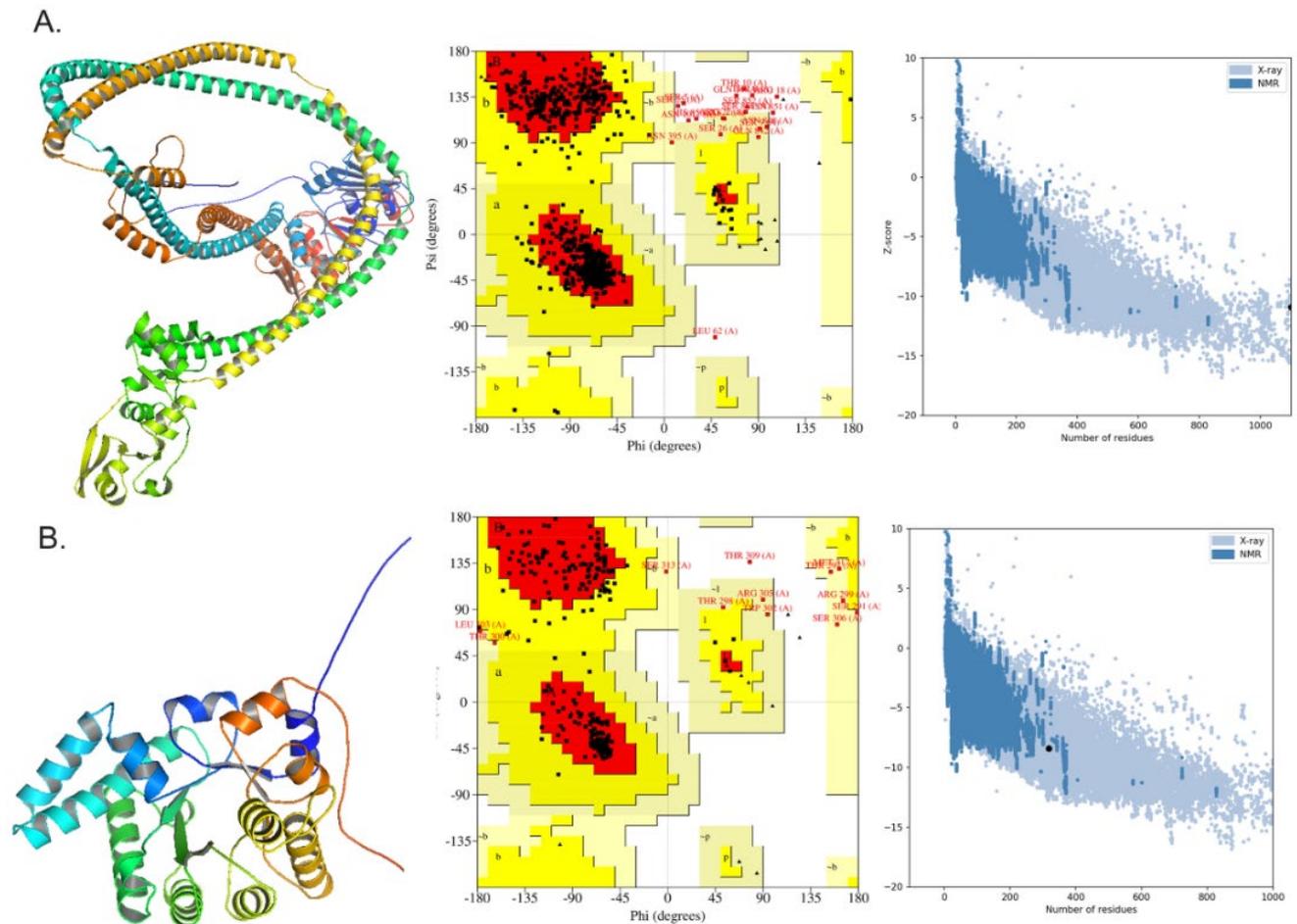


Figure 94: SMC5 and TALDO1 protein model, Ramachandranplot and Prosa-Web plots.

6.4.5.2 Mutation Stability

DynaMut mutation analysis was performed on two proteins, TALDO1 and SMC5. The mutation Asp202Asn was analyzed, and the results indicated that the mutation was stabilizing with a $\Delta\Delta G$ value of 0.618 kcal/mol. However, the NMA based prediction using $\Delta\Delta G$ ENCoM showed that the mutation was destabilizing, with a value of 0.062 kcal/mol. The other structure-based predictions using $\Delta\Delta G$ mCSM and $\Delta\Delta G$ DUET were also stabilizing, with values of 0.219 kcal/mol and 0.439 kcal/mol, respectively. Only $\Delta\Delta G$ SDM was destabilizing, with a value of -0.070 kcal/mol. The $\Delta\Delta S_{Vib}$ ENCoM indicated a decrease in molecule flexibility, with a value of -0.077 kcal.mol⁻¹.K⁻¹.

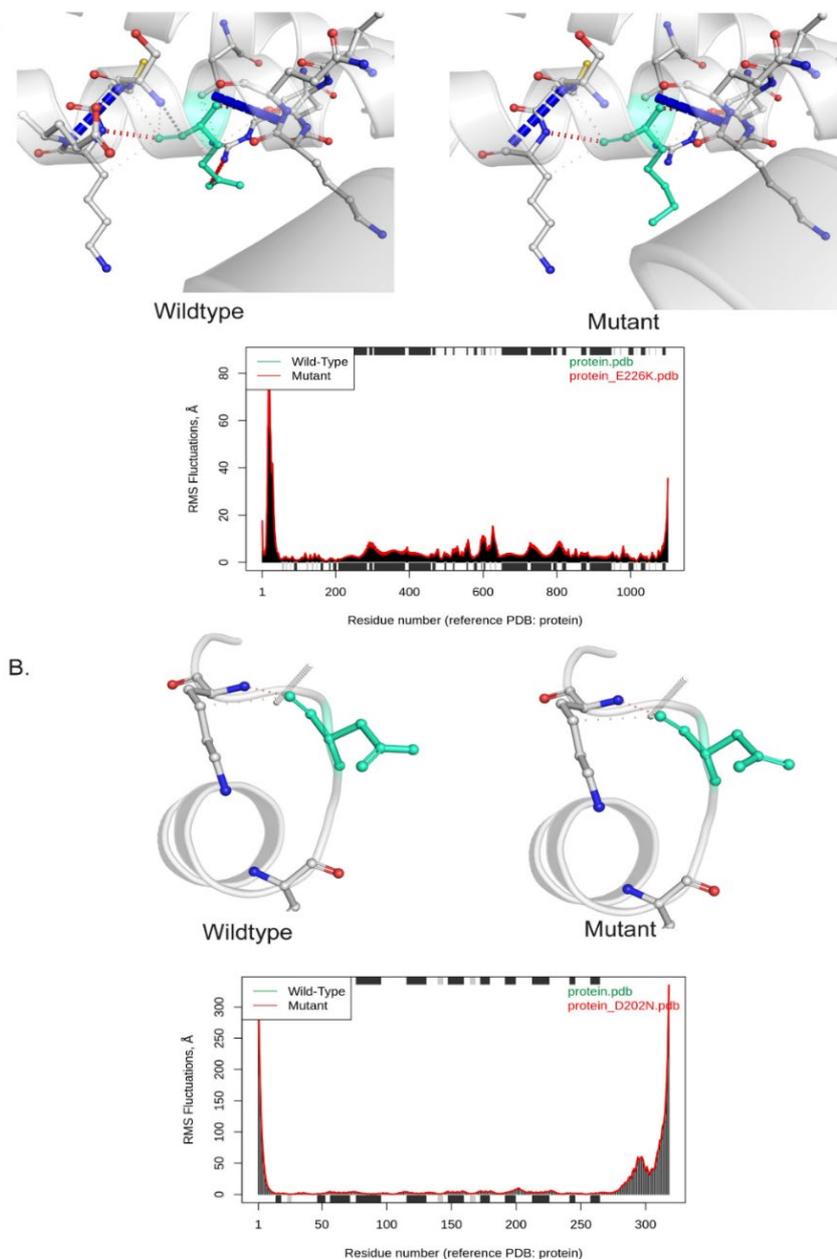


Figure 95: Shows the vibrational entropy change caused by mutations in SMC5 and TALDO1. The amino acids are color-coded to indicate their contribution to the change in entropy. Normal mode analysis was performed on wild-type and mutant sequences. The graph shows the results for each sequence, including the ionic interactions in both proteins.

The mutation Glu226Lys was analyzed, and the results indicated that the mutation was destabilizing with a $\Delta\Delta G$ value of -0.260 kcal/mol. The NMA based prediction using $\Delta\Delta G$ ENCoM showed that the mutation was also destabilizing, with a value of -0.164 kcal/mol. The other structure-based predictions using $\Delta\Delta G$ mCSM, $\Delta\Delta G$ SDM, and $\Delta\Delta G$ DUET were also destabilizing, with values of -0.263 kcal/mol, -0.770 kcal/mol, and -0.024 kcal/mol, respectively. The $\Delta\Delta S_{Vib}$ ENCoM indicated an increase in molecule flexibility, with a value of 0.205 kcal.mol⁻¹.K⁻¹. Overall, these results suggest that the mutations in TALDO1 and SMC5 have varying effects on protein stability and molecule flexibility, as determined by the different methods used in the DynaMut analysis (Figure 95). Further studies are needed to validate these predictions and investigate the functional implications of these mutations.

6.4.5.3 Molecular dynamics simulations of SMC5 and TALDO1 proteins

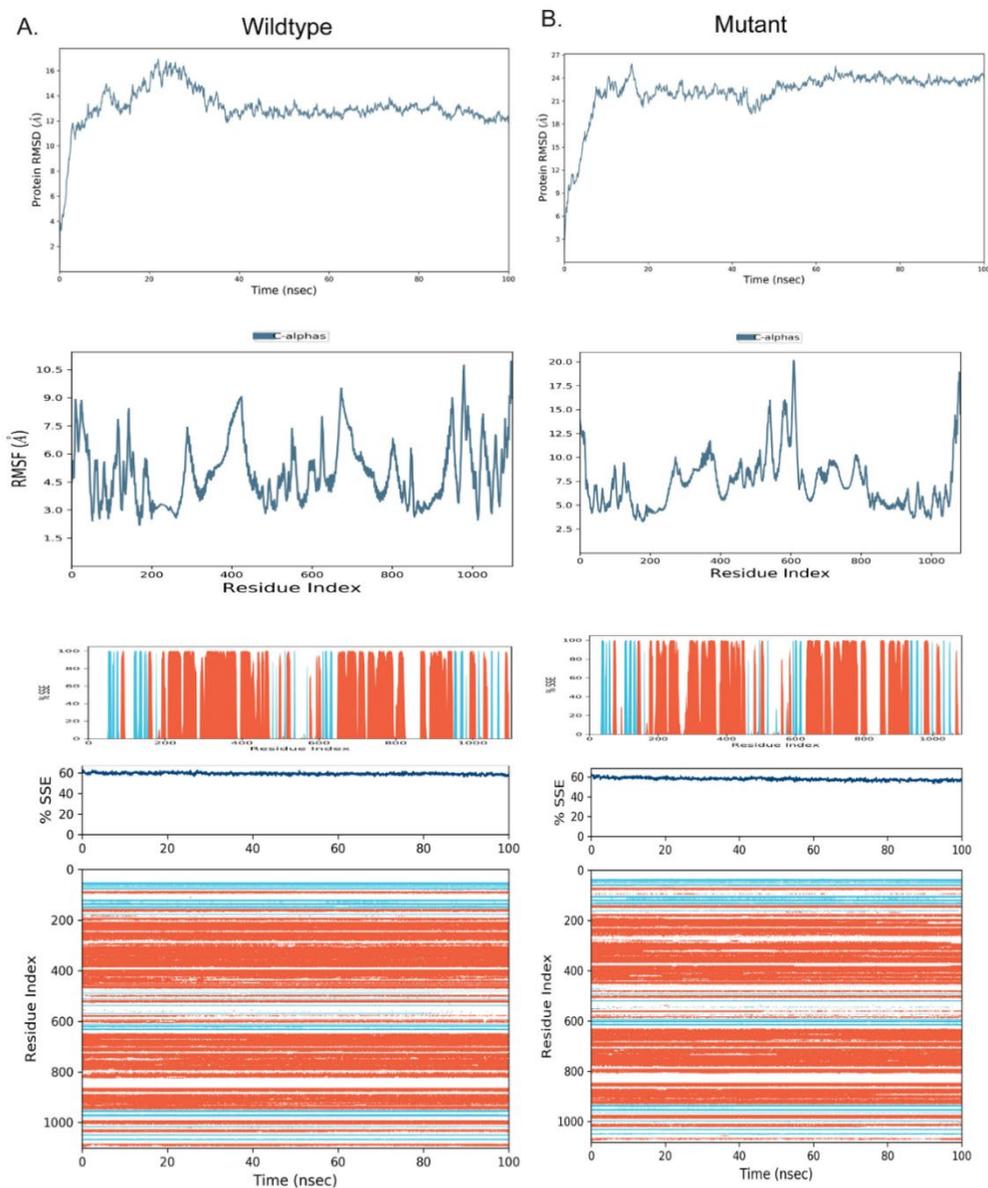


Figure 96: MD analysis plots RMSD, RMSF, SSE of wildtype and mutant model of SMC5

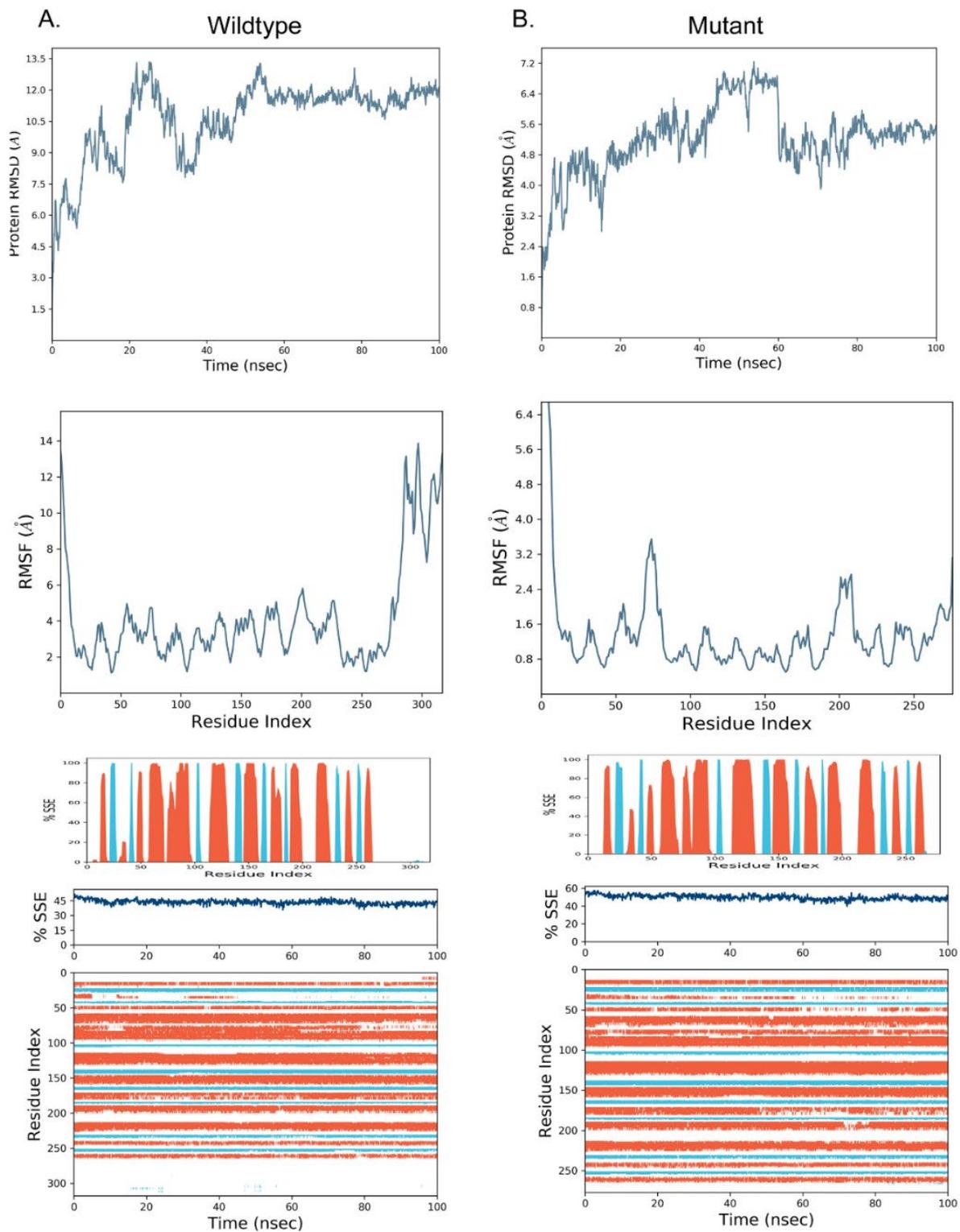


Figure 97: MD analysis plots RMSD, RMSF, SSE of Wildtype and mutant model of TALDO1.

Molecular dynamics simulations were performed on both wild-type and mutant versions of the SMC5 (Figure 96) and TALDO1 (Figure 97) proteins to examine their stability and flexibility. The wild-type SMC5 model reached equilibrium after 60 ns and remained stable, while the mutant model showed an increased RMSD and equilibrated at 60 ns. The wild-type model was more flexible than the mutant model, as shown by the RMSF analysis. The secondary structure elements analysis revealed that the wild-type model had a higher percentage of helices and strands than the mutant model. The TALDO1 wild-type model fluctuated until 50 ns but stabilized after 60 ns, while the mutant model showed a decrease in RMSD and equilibrated after 80 ns. The wild-type model was rigid throughout the simulation, with only the C-terminal showing flexibility, while the mutant model remained rigid with some regions showing increased flexibility. The secondary structure elements analysis showed that the wild-type model had a lower percentage of helices and strands than the mutant model.

In summary, the molecular dynamics simulations showed that mutations in SMC5 and TALDO1 proteins can affect their stability and flexibility. The wild-type models were more rigid, however the mutant variants were flexible. These findings imply that additional investigation is required to understand the functional consequences of these mutations.

6.4.6 Further in-silico analysis

To further investigate our identified variants, we have decided to use additional different computational methods such as I-TASSER server and Discovery studio program. I-TASSER is used to predict 3-dimensional structure of proteins (after predicting their 2-dimensional structure by machine learning) and their biological functions by using the amino acid sequence of the protein. Iterative structural assembly simulations and threading alignments are used to produce 3-dimensional models from a query protein sequence (using similar structures found in the PDB database). Based on these similarities to known proteins, conclusions about their function can be made (Yang et al., 2015) (Roy et al., 2010) (Zhang, 2008). The analysis of the binding sites revealed the pockets for each protein and the nature of the pockets by molecular visualization using the Discovery studio software. The protein pockets and the amino acids involved in binding can help us make an estimate of how H-bonds, chemical interactions, hydrophobic/hydrophilic, positively/negatively charged, and what type of substrate may bind to that specific protein.

6.4.6.1 SMC5

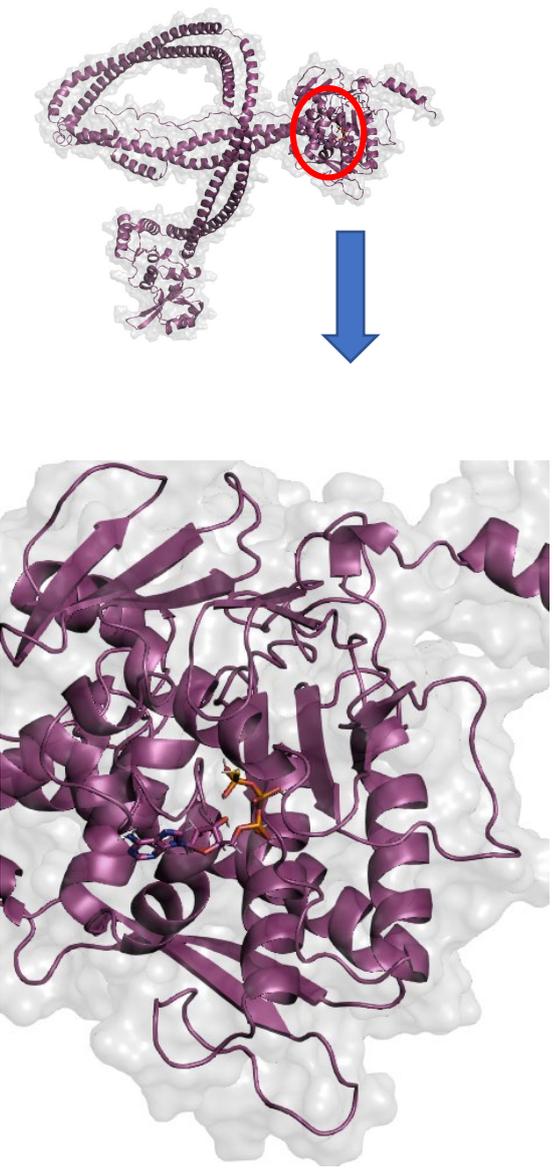
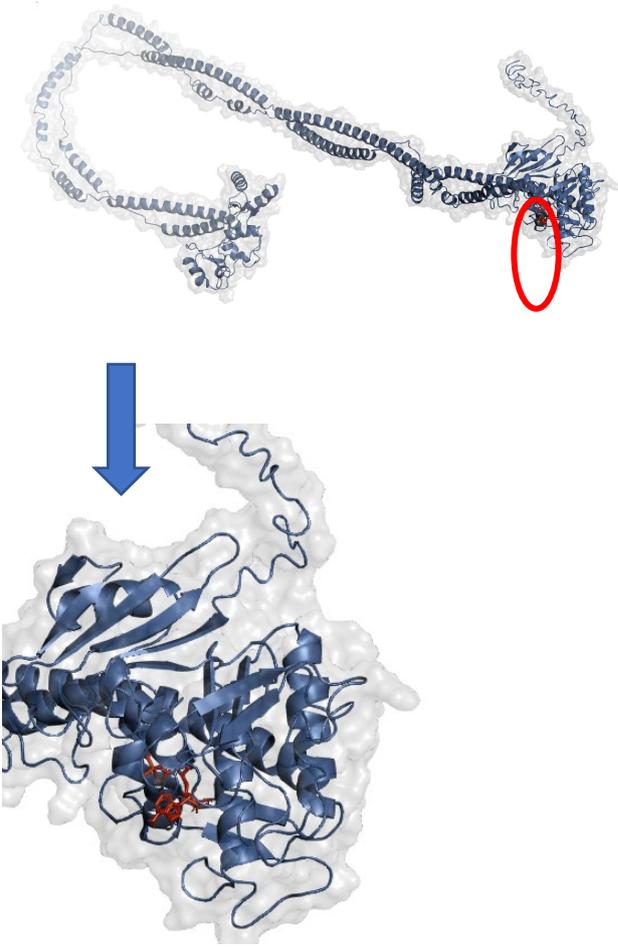
Gene	Mutation
<p data-bbox="156 324 263 369">SMC5</p> <p data-bbox="686 336 742 392">A</p>  <p>The wildtype SMC5 protein structure is shown in purple. A red circle highlights the active site region. A blue arrow points down to a close-up view of the protein with the substrate (phosphothiothosphoric acid-adenylate ester) bound in the active site, shown in orange and red.</p>	<p data-bbox="810 324 1045 369">rs141003657</p> <p data-bbox="1332 336 1388 392">B</p>  <p>The mutated SMC5 protein structure is shown in blue. A red circle highlights the active site region. A blue arrow points down to a close-up view of the protein with the substrate bound in the active site, shown in orange and red.</p>

Figure 98: SMC5 3D structure (UniProt ID: Q8IY18) with its predicted substrate (phosphothiothosphoric acid-adenylate ester) in A) Wildtype protein, and B) Mutated protein.

The 3D structure of SMC5 domain (Uniprot ID: Q8IY18) in a complex with the predicted substrate, phosphothiophosphoric acid-adenylate ester (an analog of ATP), which was predicted by I-TASSER to be the normal substrate for the wildtype and the mutated showing the difference in the folding between the normal and the mutated gene, also the position of the active site (Figure 98). SMC5 is a big protein with 1101 amino acids with several domains and active sites binding to several substrates (SMC5/6 protein complex). The domain chosen by I-TASSER is the closest to the input variant.

The predicted ligand binding site residues for the wild and the mutated enzymes were at the amino acid residues: 62,63,82,83,84,85,86,87,88,112,114,115,116,181,1020. The 2D chemical interaction between SMC5 domain (Uniprot ID: Q8IY18) in a complex with the predicted substrate phosphothiophosphoric acid-adenylate ester, showing 10 conventional hydrogen bonds at GLU A: 1020, LYS A: 86, SER A: 87, GLN A: 181, ARG A: 106, GLY A: 85, and one carbon-hydrogen bond at GLY A: 105 in the wildtype protein, while the mutated SMC5 (rs141003657) showed four conventional hydrogen bonds at LYS A: 86, GLU A: 1020, GLN A: 1023, GLN A: 181, and one carbon-hydrogen bond at GLU A: 1020. Notably, the number of hydrogen bonds in the wildtype was more than in the mutated enzyme (Figure 99).

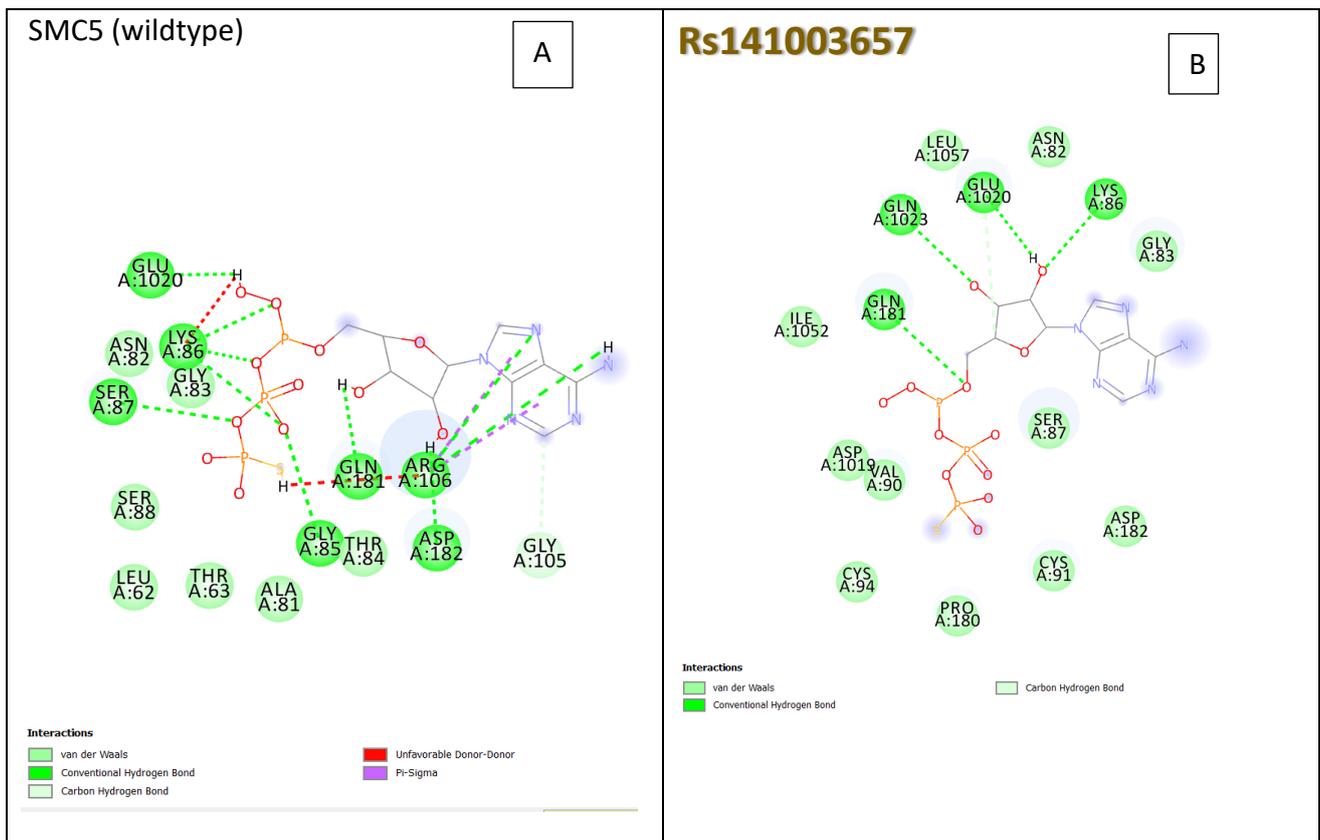


Figure 99: 2D chemical interaction in A) wildtype, and B) mutated proteins.

The H-bonds interaction of the docked SMC5 domain (Uniprot ID: Q8IY18) in a complex with the predicted substrate was compared between the wildtype and mutated proteins. Notably, the hydrogen acceptor area (in green color) was in the SMC5 domain more than the mutated SMC5 (rs141003657) (Figure 100).

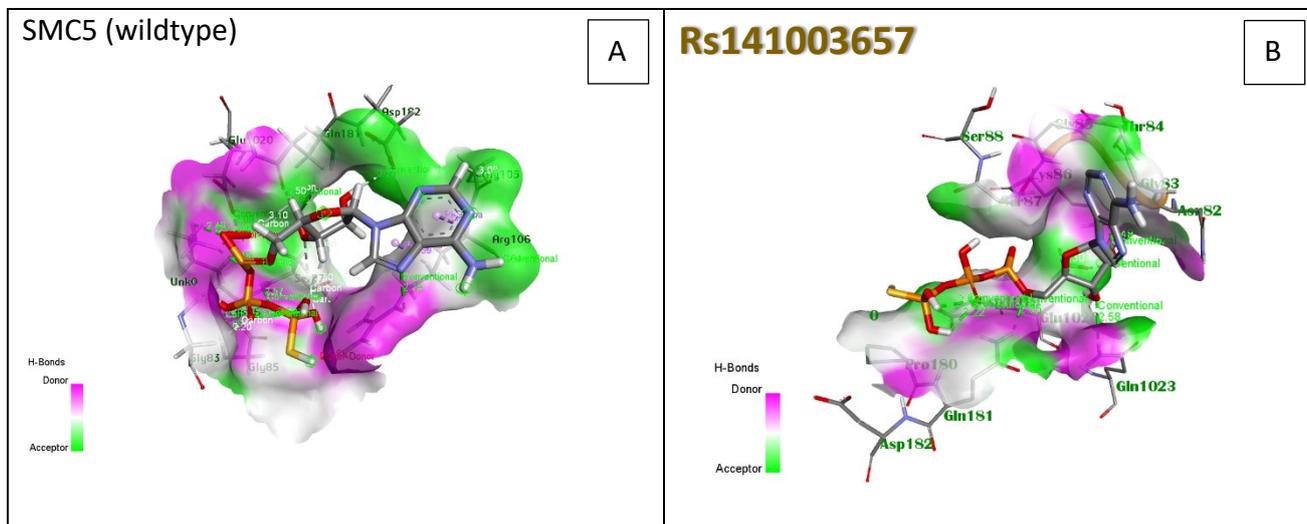


Figure 100: Comparison of hydrogen bonds between the wildtype (A) and the mutated (B) proteins. Pink color represents a donor while green color represents an acceptor.

In addition, the charges of the docked complexes, in the wildtype and mutated domains with their predicted substrate were compared. Notably, the charges area in the SMC5 wildtype domain was positive with slightly negative charges in the edge, while the charges area in the docked mutated SMC5 (rs141003657) was slightly negative and tended to be neutral in most of it (Figure 101).

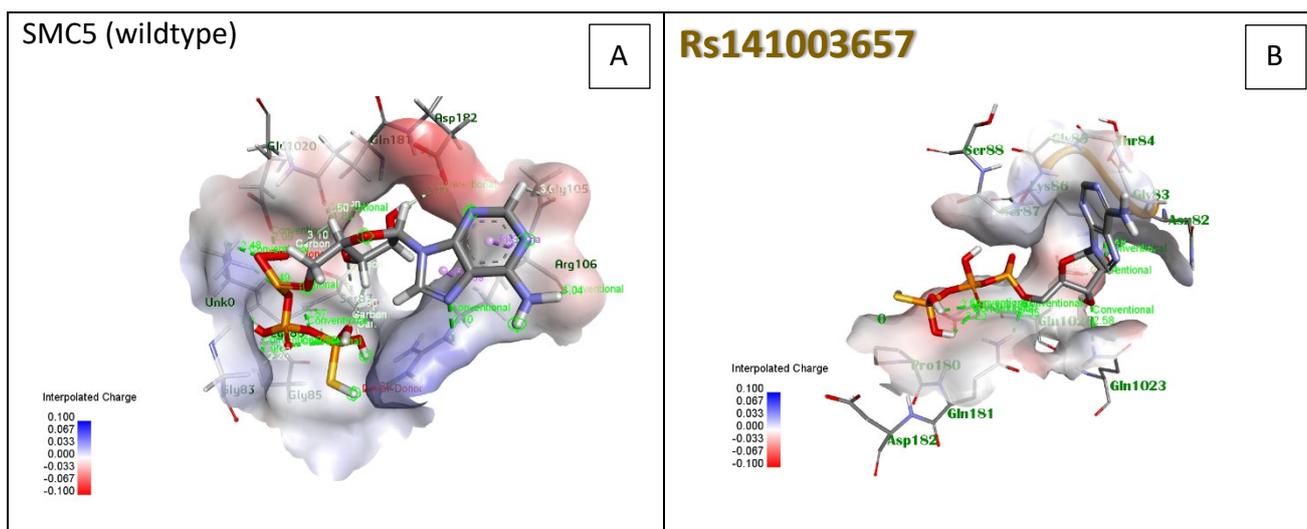


Figure 101: Charges of the domain area of the docked complexes in both the wildtype (A) and mutated (B) with their predicted substrate. Red color indicates negative charges, while blue color indicates positive charges.

Moreover, the hydrophobicity of the docked complexes, in the wildtype and mutated domains with their predicted substrate were compared. Notably, most of the docked SMC5 and the docked mutated SMC5 (rs141003657) domain areas were hydrophilic with blue color (Figure 102).

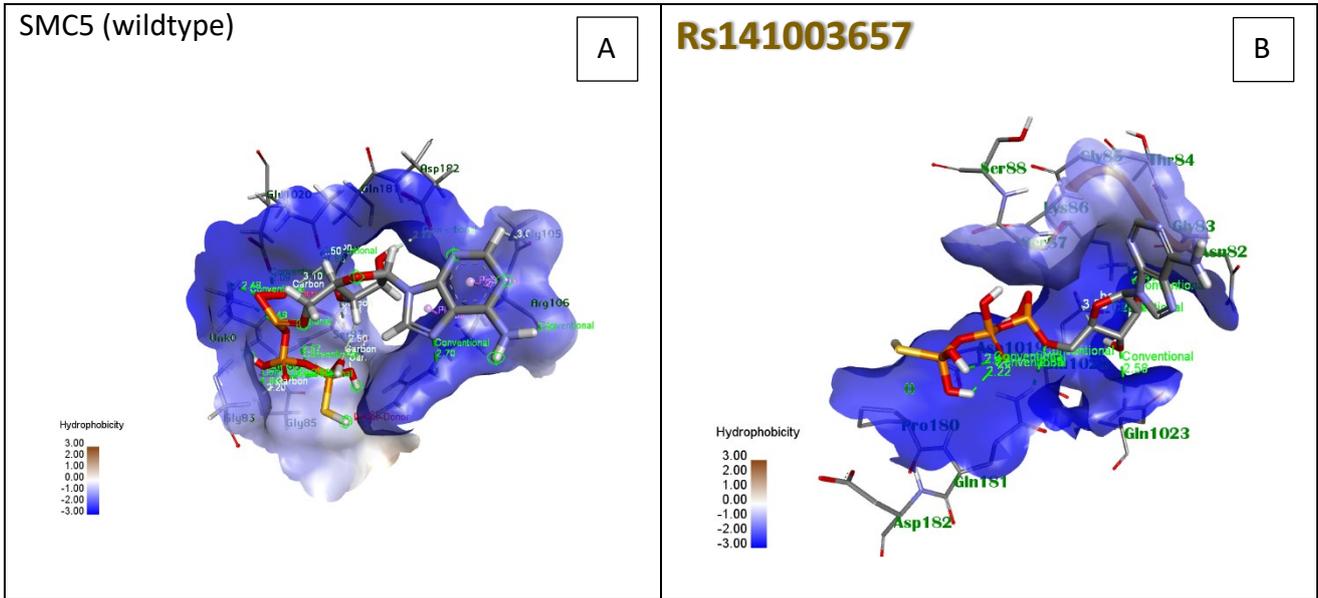


Figure 102: The hydrophobicity of the docked complexes, in the wildtype and mutated domains with their predicted substrate. Blue color indicates hydrophilic regions, while brown color indicates hydrophobic regions.

Furthermore, ionizability of the docked complexes, in the wildtype and mutated domains with their predicted substrate were compared. Notably, the normal SMC5 docked complex was more basic in most of the interacted area except a little area in the edge (surface) that was acidic. The mutated ionizability area was acidic more than the normal enzyme (Figure 103).

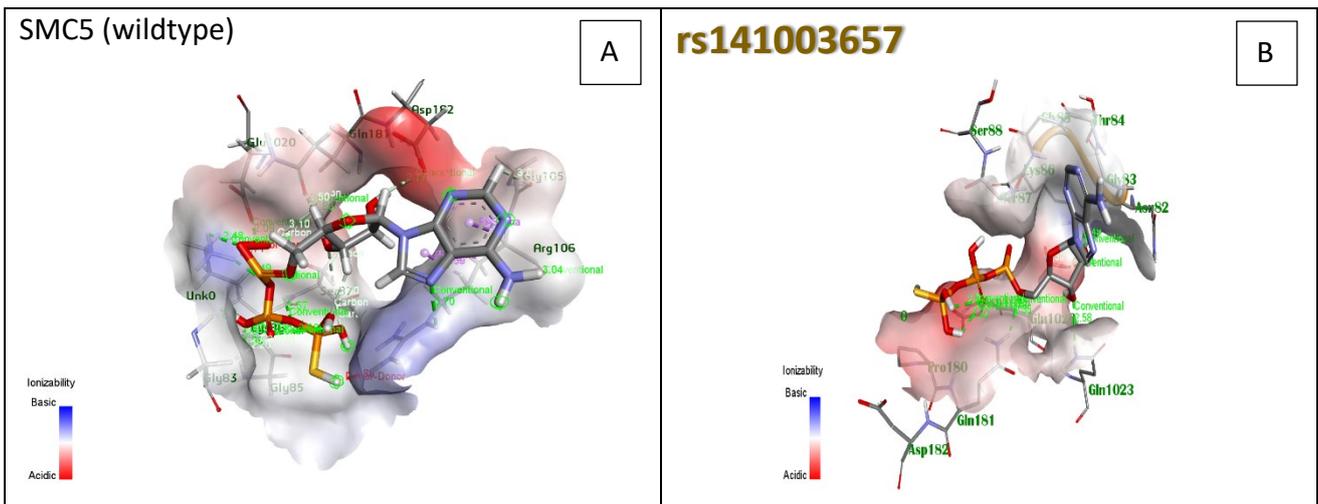


Figure 103: Ionizability of the docked complexes, in the wildtype and mutated domains with their predicted substrate. Blue color indicates basic region, while red color indicates acidic region.

SAS (solvent accessible surface) interactions of the docked complexes, in the wildtype and mutated domains with their predicted substrate were compared. Notably, the SAS interactions among the normal SMC5 docked complex were less than the mutated docked complexes (Figure 104).

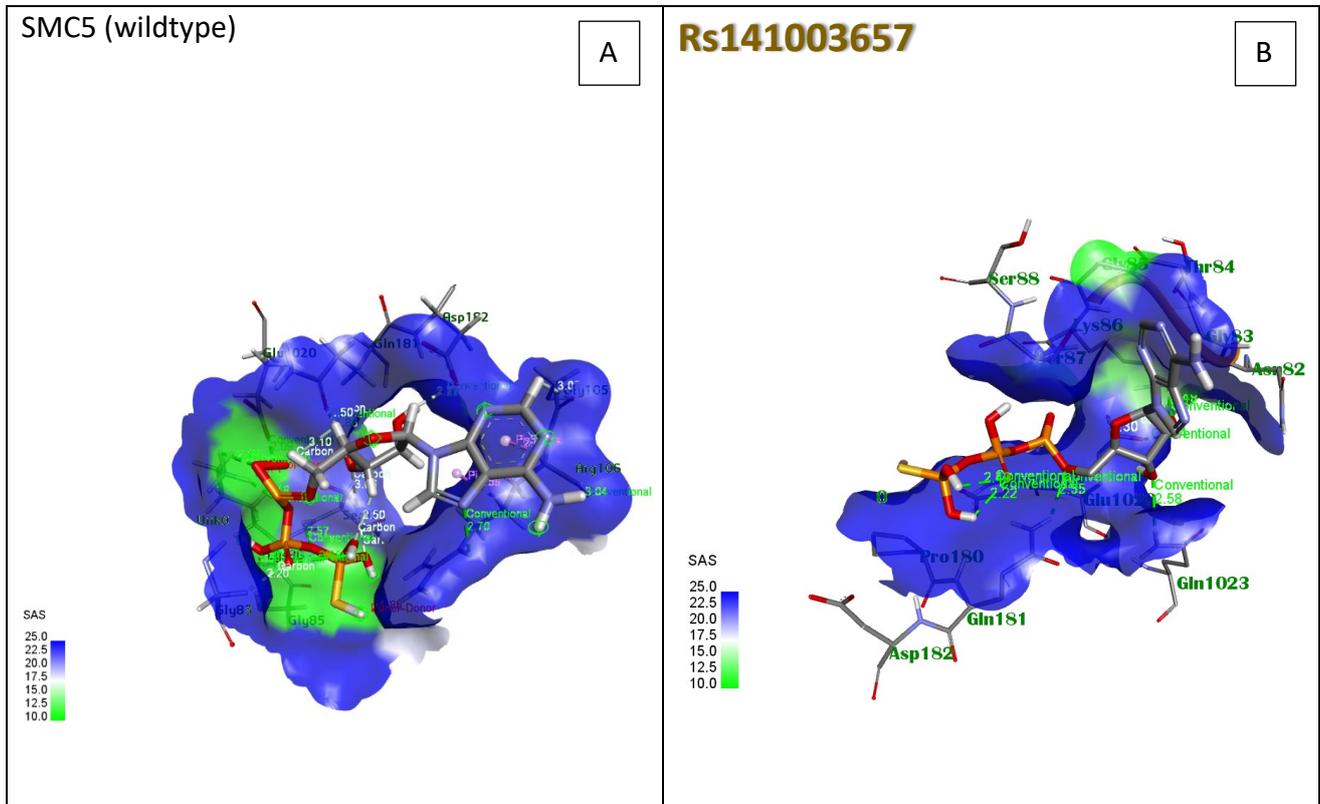


Figure 104: SAS interactions of the docked complexes, in the wildtype and mutated domains with their predicted substrate.

Green color indicates more SAS interactions, while blue color indicates less interactions.

Hydrogen bonds strengthen molecular connections, which in turn improves a wide range of biological activities. H-bonds are shown to control molecular interactions via a donor-acceptor pairing process that decreases water competition.

Each charge within the protein will communicate with the surrounding solvent. When there are more charges present, those charges interact with one another, but the solvent dampens the intensity of those interactions. There is a substantial impact of charged particle interactions with solvents and solvent filtration of charge-charge interactions on the electrostatic energy of proteins. Most of these protein properties exhibited noticeable change between normal and mutated SMC5.

6.4.6.2 TALDO1

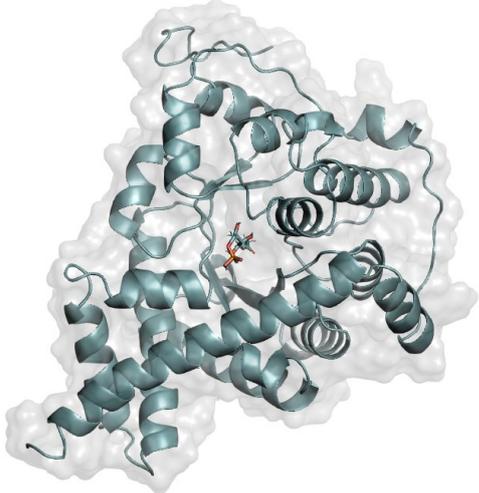
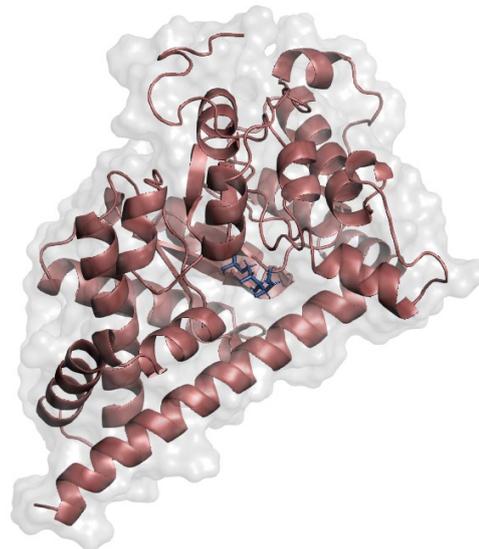
Gene	Mutation
<p>TALDO1</p> <div style="text-align: right; border: 1px solid black; width: 20px; height: 20px; display: flex; align-items: center; justify-content: center; margin: 0 auto;">A</div> 	<p>rs140985565</p> <div style="text-align: right; border: 1px solid black; width: 20px; height: 20px; display: flex; align-items: center; justify-content: center; margin: 0 auto;">B</div> 

Figure 105: The 3-dimensional structure of normal TALDO1 protein vs mutated protein with their substrate fructose-6-phosphate predicted by I-TASSER.

The 3D structure of TALDO1 chain A domain (PDB ID: 1f05) complex with fructose-6-phosphate (the 1f05 domain did not have any substrate, the fructose-6-phosphate was predicted by I-TASSER to be the normal substrate). The mutated TALDO1 (rs140985565) 3D structure in complex with its predicted substrate fructose-6-phosphate, showing a slightly different interactions area from the normal TALDO1 domain (Figure 105).

The 2D chemical interaction between TALDO1 domain (PDB ID: 1f05) in a complex with fructose-6-phosphate shows 4 conventional hydrogen bonds at ASN A: 165, LYS A: 142, ASP A: 27, ASN A: 45 residues, and two carbon-hydrogen bonds at SER A: 187, and LYS A: 142. The mutated TALDO1 (rs140985565) 3D structure in complex with its predicted substrate shows 5 conventional hydrogen bonds at ASP A: 27, THR A: 44, ARG A: 239, ARG A: 192, SER A: 237 residues. Notably, in the mutated interactions there was one unfavorable doner-doner in the LYS A: 142 residue (Figure 106).

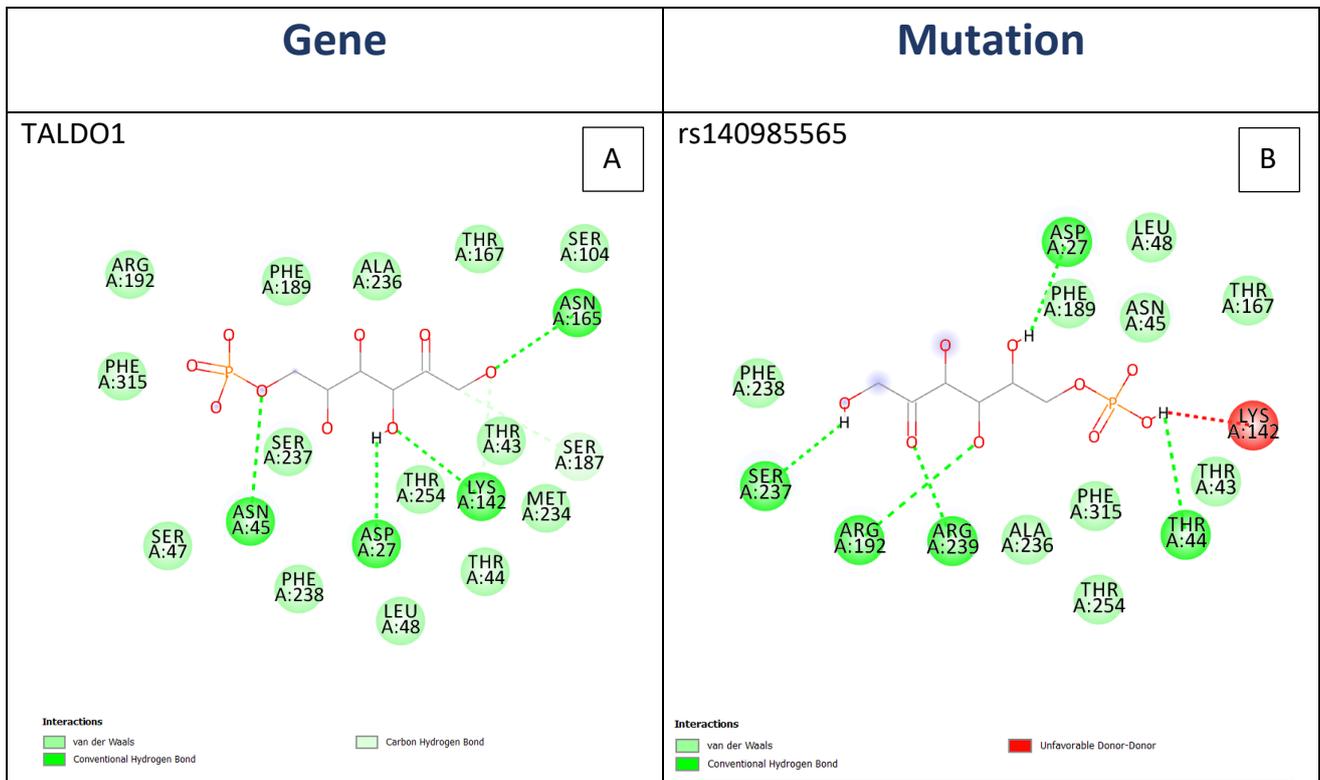


Figure 106: Comparison of 2D chemical interactions of TALDO1 active site between wildtype and mutated proteins.

The H-bonds interaction shows the region of the interaction of the docked TALDO1 domain (PDB ID: 1f05) in complex with the fructose-6-phosphate predicted substrate and the docked mutated TALDO1 (rs5030868) 3D structure in complex with its natural substrate fructose-6-phosphate. Notably, the hydrogen donor area was slightly more in the mutated domain than the normal TALDO1 (Figure 107).

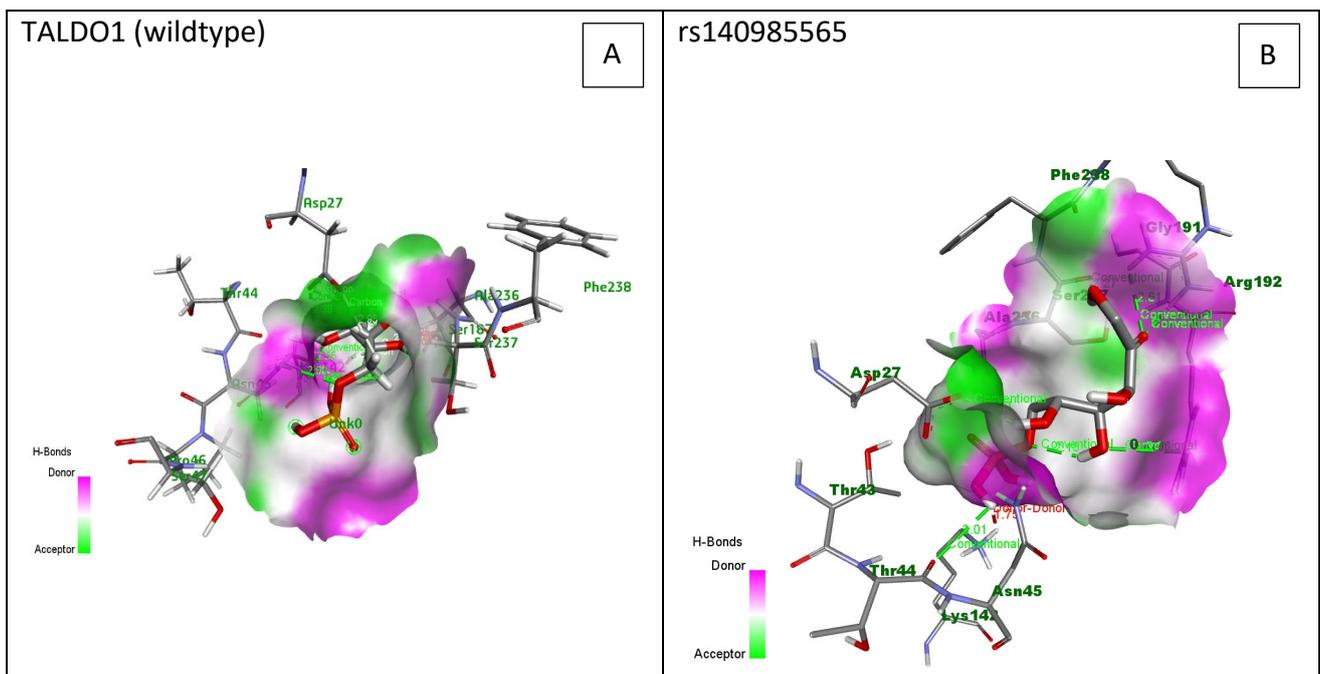


Figure 107: Hydrogen bonds of the docked complexes, in the wildtype and mutated domains with their predicted substrate.

The charges of the docked complexes in the wildtype and mutated domains with their predicted substrate were compared. Notably, almost all the charges area in the TALDO1 domain was neutral with slightly negative charges on the edge, while the charges area in the docked mutated TALDO1 were positive in charges with a little area with negative charges at the end of their edge surface (Figure 108).

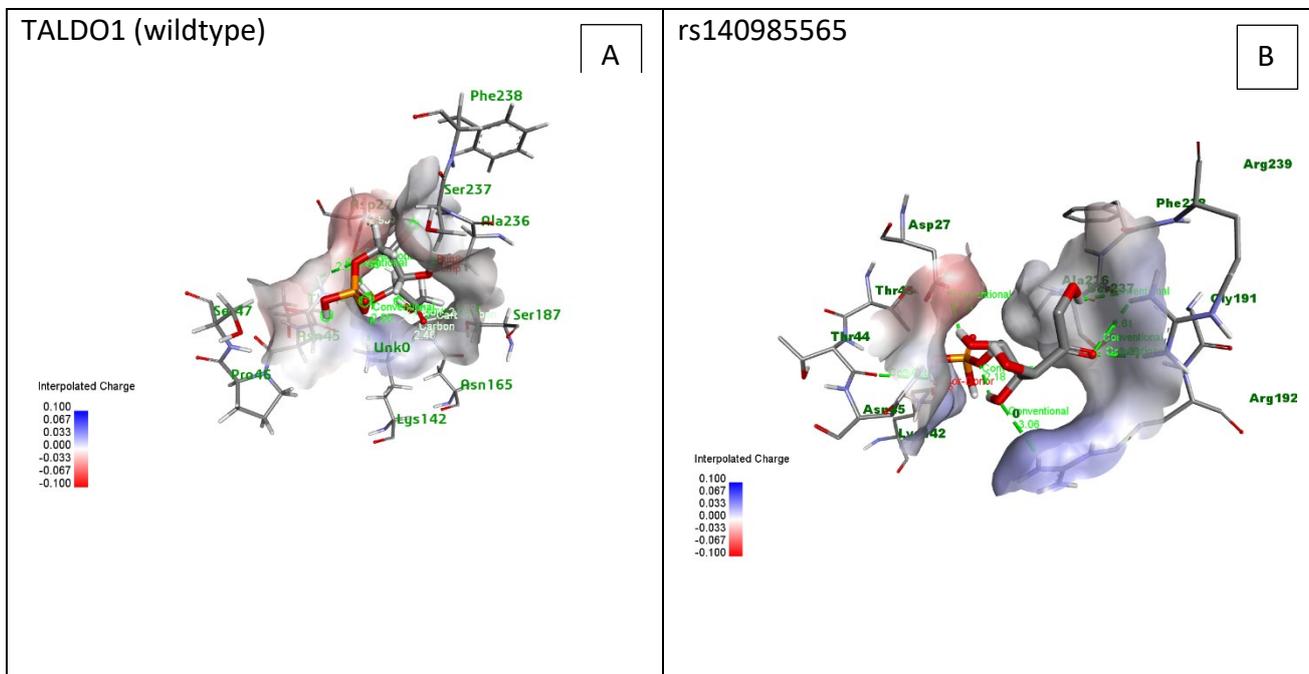


Figure 108: The charges of the docked complexes in the wildtype and mutated domains with their predicted substrate.

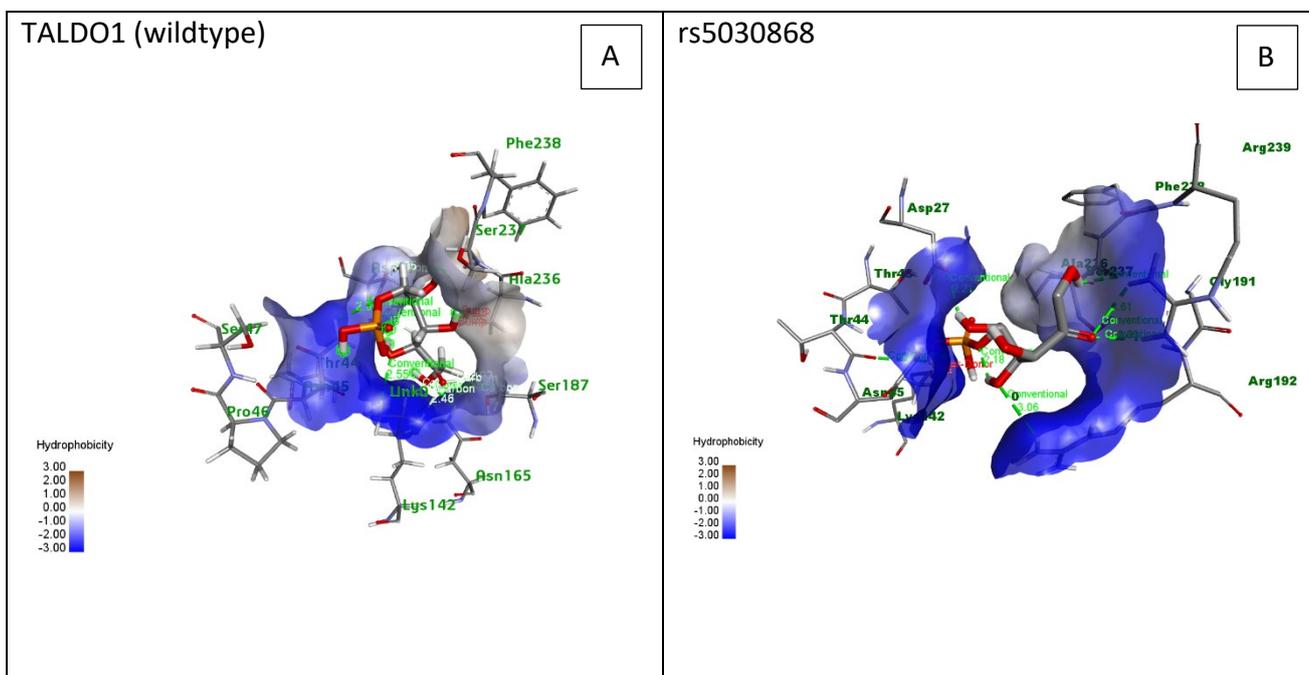


Figure 109: The hydrophobicity of the docked complexes in the wildtype and mutated domains with their predicted substrate were compared.

The hydrophobicity of the docked complexes in the wildtype and mutated domains with their predicted substrate were compared. Notably, most of the docked TALDO1 domain interactions areas were hydrophilic, while the docked mutated TALDO1 (rs5030868) interacted area was totally hydrophilic (Figure 109).

The ionizability of the docked complexes in the wildtype and mutated domains with their predicted substrate were compared. Notably, the normal TALDO1 docked complex was more neutral in its area's surface (with slightly negative and positive charges) rather than the mutated which was basic in most of its interactions area except a little area with acidic interactions (Figure 110).

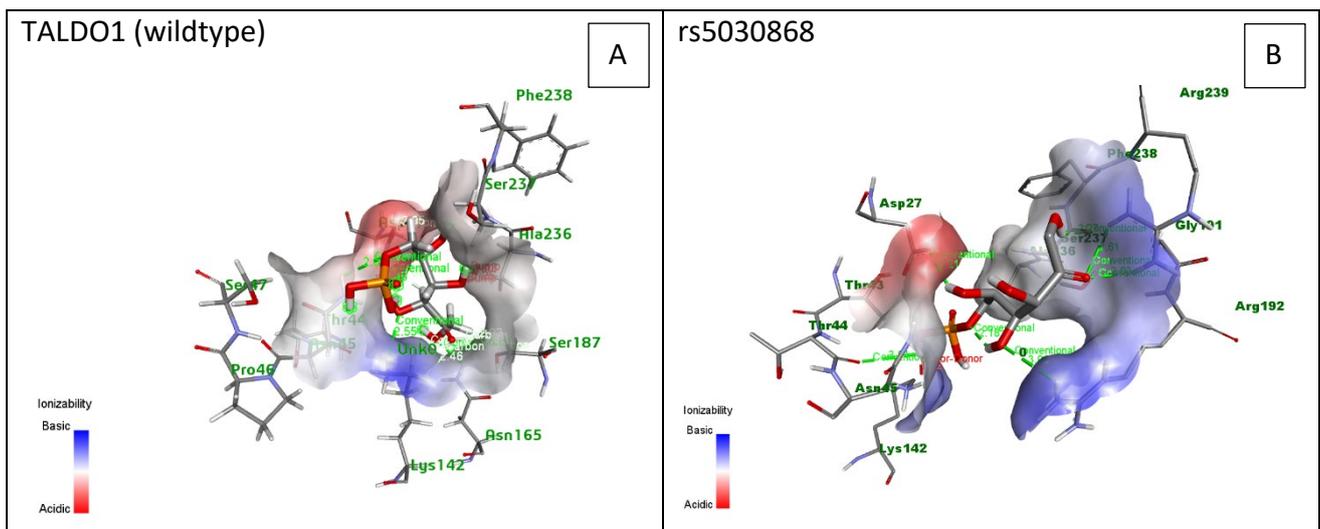


Figure 110: The ionizability of the docked complexes in the wildtype and mutated domains with their predicted substrate.

SAS interactions of the docked complexes in the wildtype and mutated domains with their predicted substrate were compared. Notably, the SAS interactions among the normal TALDO1 docked complex were less than the mutated docked complexes (Figure 111).

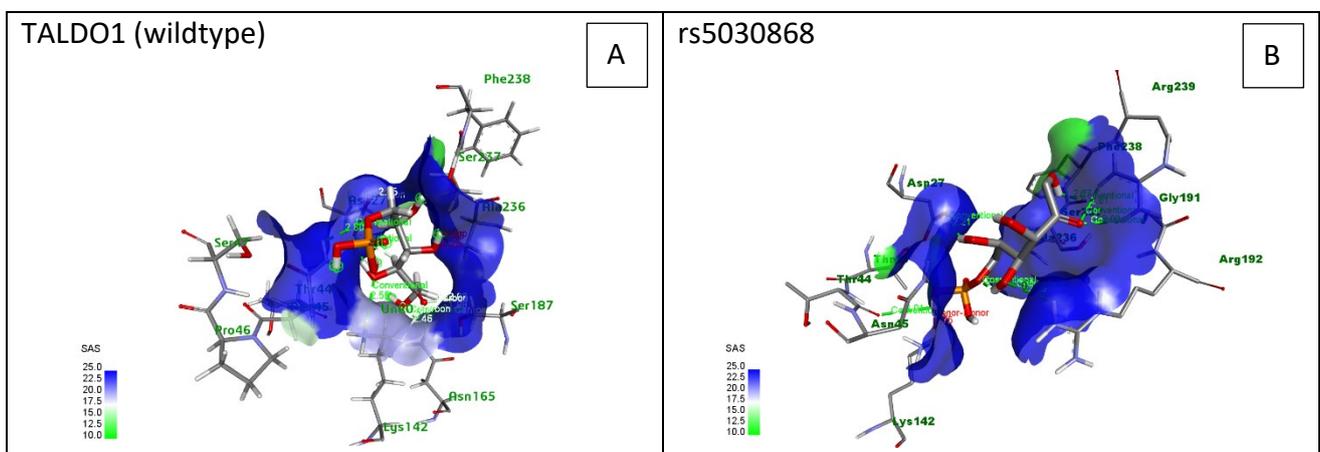


Figure 111: SAS interactions of the docked complexes in the wildtype and mutated domains with their predicted substrate.

6.4.7 Identification of deleterious SNPs using online in-silico prediction tools (Table 34).

Table 34: in-silico prediction tools

Gene	I-Mutant 2.0 (structural stability)	PolyPhen-2	Panther	SNPs&GO	MutPred2 (0.5<damaging)
SMC5	Decreased	Probably damaging (0.999)	Probably damaging (0.85)	Neutral	0.213
TALDO1	Decreased	Benign (0.096)	Probably damaging (0.57)	Neutral	0.411

Proteins are organized by the PANTHER software according to their evolutionary history, molecular activities, and interactions with one another. Position-specific evolutionary conservation scores are determined by aligning proteins with similar evolutionary histories and are then used to conduct a substitution analysis (Thomas et al., 2003). PolyPhen-2 uses sequence-based characterization⁵¹ to predict the functional effect of amino acid changes on the structure and functions of proteins (Adzhubei et al., 2010). The SNPs&GO server uses support vector machines (SVM) to provide estimates of human disease-related mutations (Calabrese et al., 2009). I-Mutant2.0 is a web service that uses a Support Vector Machine to forecast how SNP mutations in a protein will affect its stability. ProTherm, the largest repository of experimental data on protein mutations, served as the training set for the program (Bava et al., 2004). The MutPred2 technique and software package use machine learning to combine genomic and molecular data and make probabilistic inferences regarding the pathogenicity of amino acid changes (Pejaver et al., 2020). Different tools resulted in conflicting results. Further in-vitro and in-vivo investigations are required to evaluate the functional effect of these variants.

6.5 General discussion

Saudi Arabia has passed a long way regarding medical advances. The Kingdom has decreased the prevalence of many infectious diseases and overcame many obstacles to elevate the health status in its population. Nevertheless, Saudi Arabia is no different from its adjacent Mediterranean and Asian neighbors in having a high prevalence of sickle cell trait and thalassemia. Many environmental, cultural, and religious factors contribute to the reason why hemoglobinopathies in these countries are present at high rates. The “malaria hypothesis” suggests that heterozygosity for various red cell defect alleles has provided significant protection against lethal malaria particularly in children and pregnant women and consequently there has been strong selection for alleles which when homozygous may be severely damaging (Piel et al., 2010). The most evident factor of the high prevalence rate is the tradition of marrying from the same family (particularly first cousins), which accounts for up to 60% of Saudi marriages and because hemoglobinopathies are recessively inherited diseases, close family member marriages will increase the possibility of passing on the disease to the offspring (El-Hazmi et al., 2011).

Hemoglobinopathies impose a huge health and economic burden on the government to give each affected individual the optimum care needed to improve their health status and quality of life and to try and prolong their life expectancy as much as possible. A cross-sectional retrospective study was conducted in Al-Ahsa, Saudi Arabia at the Hereditary Blood Disease Center on patients with beta thalassemia major to assess their need and access of lifelong transfusion of packed red cells and found a huge gap between the demand for transfusions and supply of blood donations (Albagshi et al., 2021). Additionally, a retrospective observational study conducted by Bin Zuair et al at King Saud University Medical City to calculate the average of annual visits of 160 patients with SCD to the emergency department resulted in an average of four visits with a mean stay of six days (Bin Zuair et al., 2023).

Individuals have the right to know if they carry the risk of passing on such conditions to their children. Religious and ethical restrictions within Muslim societies prohibit abortion of the affected fetus. On the other hand, in the last few years, pre-implantation genetic diagnosis (PGD) has been used in Saudi Arabia by couples at risk of passing on pathogenic variants to their offspring to select mutation-free embryos. Although individuals carrying hemoglobin mutations can use PGD, this technique is offered in a limited number of medical and research centers due to its delicate procedure and its high cost (Abotalib, 2013) (Suhaimi et al., 2022). The most essential function in preventing the problems of hemoglobinopathies like thalassemia is primary prevention through population screening, premarital

and prenatal diagnosis, and the avoidance of marriages between carriers. Primary care physicians also play an essential role in detecting people at risk for or presenting with a clinically severe type of thalassemia, as well as in providing continuous education and support for patients and their families. In the effort to help control and reduce the spread of abnormal hemoglobin variants, many countries have initiated a pre-marital screening program (PMSP) in which some are even mandatory. In 2002, Saudi Arabia passed a Royal decree to make the public aware of the benefits of genetic premarital screening. It includes the necessity of explaining genetic disorders and their serious effects on public health. Also, it gives laboratories all over Saudi Arabia the time to be prepared to carry out premarital genetic tests and to make them available for any Saudi citizen that wishes to benefit from this program (El-Hazmi, 2004).

In 2003, a second decree was issued that makes pre-marital screening for sickle cell disease and thalassemia mandatory to all men and women before marriage free of charge to reveal at-risk partnerships and offer genetic counseling and was implemented in 2004. Regional new born screening is also offered which provides a great benefit of identifying affected babies and would give a good chance to manage and control the disorder and give proper health care at early stages of the disease (Al-Shahrani, 2009). The pre-marital screening program does not actually reflect the real prevalence of SCD because it only takes into account adults at the age of marriage and does not take into account the severely affected individuals, affected individuals that do not wish to marry, and children (Jastaniah, 2011). The compulsory pre-marital screening program in Saudi Arabia is offered by the Ministry of Health which provides health care, counselling and treatment free of charge for all of the Saudi population (Alswaidi & O'Brien, 2009).

The need has increased to consider these conditions as a public health problem that needs to be prioritized in the context of prevention. The introduction of premarital screening and genetic counseling in Saudi Arabia, education and awareness of the complications and the chronic illness associated with the disease have essential roles and have proved a degree of effectiveness in reducing the incidence of the disease through encouraging the voluntary cancellation of marriages (Jastaniah, 2011). In addition to the proper early diagnosis of such diseases, this comprehensive effort will diminish the chance of increasing frequency of deleterious alleles in the population. Nevertheless, even with the compulsory premarital screening program implementation, many couples do not cancel their wedding plans due to different reasons that include family and social pressure, cultural and religious norms, and upholding marriage commitments; a retrospective analysis conducted between

2017 and 2020 on > 32000 high performance liquid chromatography tests at the Armed Forces Hospital in the Southern region of Saudi Arabia has revealed that hemoglobin abnormalities remain prevalent (Makkawi et al., 2021).

Inherited hemolytic anemia is a spectrum of illnesses characterized by an elevated rate of red blood cell destruction. Hemoglobin defects, enzyme, and membrane defects in red blood cells are the most prevalent underlying causes of inherited hemolytic anemia. The degree of this defect determines the severity of the anemia. Traditionally, the diagnostic workflow of hereditary anemia conditions begins by gathering a thorough history from the patient, then doing biochemical and morphological assessments that includes RBC morphology, blood count, hemoglobin electrophoresis, RBC enzyme levels, and bonemarrow biopsy. At times, it might be difficult to make a correct diagnosis due to the fact that the clinical symptoms may overlap in cases of diverse etiologies, making it impossible to tell them apart using traditional diagnostic methods. However, recent changes to the diagnostic paradigm have made genetic testing an acceptable method for making a differential diagnosis in such cases and help us in predicting the effect of frequently encountered variant alleles. Single-gene testing is still advised for individuals with fully phenotypically characterized conditions and if the disorder has been associated with a specific gene. The polymerase chain reaction (PCR) may be used in a variety of different ways to identify mutations that are quite common within a population. The procedures of reverse dot blot analysis, primer-specific amplification using a set of primers, and real-time PCR (genotyping) are the ones that are utilized the majority of the time. If the mutation was not detected using targeted variations analysis, sequence analysis can be employed (Cao & Galanello, 2010). Genetic testing by first generation sequencing which employs Sanger chain-termination technology is used as a confirmation test but due to the genetic diversity of many illnesses, this method often fails to identify the relevant mutations in the target gene, leading to incorrect or absent molecular diagnosis. Nowadays, specially with transfusion dependent cases, genetic testing is used early in the diagnostic workflow, relieving the patients from the need for bonemarrow aspiration (Roy & Babbs, 2019). Since 90% of beta thalassemia alleles are point mutations that may be easily discovered by Sanger sequencing or other specialized approaches, comparative genomic hybridization arrays and MLPA are molecular approached can be used in the discovery of the remaining 10% of alleles that are deletions. However, it is essential to keep in mind that the implementation of all of these methods should not be seen as exclusive or sequential, but rather as synergistic.

Our main focus of this current study is beta thalassemia, an autosomal recessive hereditary anemia and one of the most common disorders worldwide is characterized by quantitative defect of hemoglobin synthesis resulting from mutations in the beta globin gene that leads to hemolysis, probably as the result of the accumulation of alpha globin dimers and the resulting oxidative damage to the cell membrane of erythrocytes. The clinical course of beta-thalassemia is very variable, from almost asymptomatic to life-threatening, even among people with the identical variants. Inter-patient heterogeneity in its clinical presentation has spurred intense interest in the search for genetic modifiers of disease severity. The search for these genetic modifiers is motivated by the possibility that they may lead to more accurate (or tailored) prognosis tests and direct the creation of more targeted and efficient treatments (Steinberg et al., 2009). We investigated five affected individuals whom are all diagnosed with beta thalassemia major, but upon thorough genetical investigations of *HBB* gene, beta globin cluster and their promoters, beta globin locus control region, and *HBA1/HBA2* genes, we found that they were only carriers of a well-studied severe pathogenic beta thalassemia variant but in a heterozygous state, which makes them only carriers (beta thalassemia trait) with no/mild symptoms and explanation of their severe symptoms had to be made. Also, they were carriers of the -3.7 alpha globin deletion, in either heterozygous or homozygous state, which may contribute in lowering the severity risk instead of increasing it.

The clinical and hematological severity of beta-thalassemia in heterozygotes may be increased by two pathways that have been discovered. The first is an excess of unassembled alpha chains, which can lead to the premature destruction of red blood cell precursors, and is associated with the inheritance of both heterozygous beta-thalassemia and a triple (alpha alpha alpha/alpha alpha) or quadruple (alpha alpha alpha alpha/alpha alpha) alpha globin gene arrangement (Kulozik et al., 1987). The second mechanism that may contribute in the increasing severity in heterozygotes is that the variation in the beta globin gene results in high instability of the beta globin chains due to production of inclusion bodies, the result of precipitation of unstable beta chains bound to heme before assembling with alpha chains such as in dominantly inherited beta thalassemia and de novo variants (Thein, 1992) (Murru et al., 1992). In addition, fetal hemoglobin is a major modulator in the severity of beta thalassemia as discussed in earlier chapters. In the case of our affected family, neither mechanisms were present in the affected samples, but instead, we have found heterozygous and homozygous alpha thalassemia deletions that should have helped in reducing the disease severity. After that, we had to continue our investigation and explore other possible causes that contributed in raising the severity of their symptoms.

Traditionally, it has been assumed that the transmission of a trait within a family is equivalent to the transmission of a single molecular mutation; this has been the basis for the identification of numerous single-gene illnesses. Illnesses that are produced by variations in a single gene are known as monogenic disorders. These illnesses are often identifiable due to the distinctive family inheritance patterns that they exhibit. Although it is still the case that some features are transmitted in a monogenic form, with individual alleles segregating into families according to Mendel's rules, the number of diseases for which a mutation at a single locus may sufficiently explain the entire phenotype is growing less. While research into multilocus illnesses has advanced, detecting and recognizing polygenic disorders has proven more difficult and challenging. Even while Mendelian models are helpful for pinpointing the major genetic etiology of familial illnesses, they may not accurately represent the underlying cellular and physiological basis of the problem. In recent years, it has become clear that a very small number of loci are responsible for the onset or progression of a wide variety of illnesses that were once thought to be monogenic. The term "polygenic disease" has been used to characterize these conditions, which span a wide range of phenotypes (Badano & Katsanis, 2002).

More than 60 years ago, Defrise–Gussenhoven hypothesized, that the inheritance of the investigated disease within a family pedigree would be explained more accurately with less reduced penetrance if investigated as digenic inheritance (two locus model) rather than only monogenic. 'Reduced penetrance' is when a proportion of individuals, who have a genotype that is known to cause a disease, do not develop the disease. The proportion who do develop the disease is the numerical value of the penetrance, which allows for imperfect genotype-phenotype correlation (Defrise-Gussenhoven, 1962) (Schaffer, 2013). When it comes to genetically complicated disorders, digenic inheritance is the simplest type of oligogenic inheritance, and it was defined by Schaffer as 'Inheritance is digenic when the variant genotypes at two loci explain the phenotypes of some patients and their unaffected (or more mildly affected) relatives more clearly than the genotypes at one locus alone' examples where one locus is predominant but has varying expressivity on its own are included by this definition, as are cases when the two loci are nearly equivalent in weight (Schaffer, 2013) (Gazzo et al., 2016). In 1994, Kajiwara and his colleagues have documented the first digenic inheritance report which included three family pedigrees where affected individuals developed retinitis pigmentosa only with a double heterozygous mutation in two different alleles that involve photoreceptor-specific genes ROM1 and peripherin/RDS (Kajiwara et al., 1994). Human digenic inheritance has been steadily discovered since then, the rate of reporting of digenic disorders is growing exponentially and includes deafness (Lerer et al., 2001), Pendred syndrome (Li et al., 2020), Charcot Marie Tooth disease (Fierro et al., 2020). The

reporting of oligogenic disorders in the literature has also been increasing recently. In 2007, Hoefele, et al, found evidence of oligogenic inheritance that correlates to a recessive cystic renal disease, nephronophthisis, which involves NPHP1-4 genes (Hoefele et al., 2007). Another example is the findings of Lenglet and colleagues in 2018, which suggested that the dominantly inherited acute intermittent porphyria with incomplete penetrance may not adhere to the classical autosomal dominant model, but may instead be modulated by strong environmental and genetic factors independent of the hydroxymethylbilane synthase gene (*HMBS*), as evidenced by the striking difference they found in the penetrance of *HMBS* mutations between the general population and their cohort of French families. It is possible that the environmental modifiers of the oligogenic inheritance model can better explain the penetrance and heredity of AIP (Lenglet et al., 2018). Another group has studied a cohort of 47 patients with hypogonadotropic hypogonadism phenotype using targeted next generation sequencing and found that while it is well established that "*SPRY4/SEMA3A, SRA1/SEMA7A, CHD7/SEMA7A, CCDC141/POLR3B/POLR3B, and PROKR2/SPRY4/NSMF*" genes have a role in the hypogonadotropic hypogonadism phenotype, they suggest the existence of novel "partners" involved in digenic and trigenic patterns of inheritance (Gach et al., 2022).

Discoveries in human digenic and oligogenic inheritance and identification of modifier genes have been sped up with the use of next generation sequencing which can find disease-causing mutations in two genes or even more with a single experiment. In 2019, patients with congenital dyserythropoietic anemia type II were analyzed for modifier genes using a target panel of 81 genes, and mutations in some of these genes were shown to account for observed clinical variation amongst patients. One such variant, A260S, was found in 12.5% of patients with CDAll who had severe symptoms and is a recurrent low-frequency mutation in the *ERFE* gene. The *ERFE*-A260S mutation is linked to the increased levels of *ERFE*, which severely impairs iron regulatory mechanisms in the liver. Characterization of its action in a hepatic cell culture revealed a modifier role for it in iron overload via the bone morphogenetic protein (*BMP*)/*SMAD* pathway (Andolfo et al., 2019). Another recent study led by Aggarwal, et al, on 73 Asian families using next generation sequencing as a diagnostic approach, found that in 15% of affected patients, co-inheritance of *G6PD* deficiency with hereditary spherocytosis increased their transfusion rate which led to more severe phenotype (Aggarwal et al., 2020). Mutations in the *HBB* gene have been shown to be inherited along with mutations in other genes in transfusion dependent Saudi patients. These additional genes include *ATRX*, *HBA1*, and *AHSP* (Alaithan et al., 2018) (Borgio et al., 2016) (Akhtar et al., 2013) (Al-Nafie et al., 2015) but there was no further investigations on the degree of these genes' contribution to the severity of the disease. Many beta thalassemia and sickle

cell disease studies have investigated the coinheritance of other gene mutations in terms of their contribution in elevation the severity of the disease by increasing the risk of associated complications such as stroke, renal failure, leg ulcer, pain crisis, osteonecrosis, priapism, and acute chest syndrome (Lette, 2012).

Genetic technologies have served as valuable resources for helping us better comprehend the genetic elements that influence the severity of many diseases. On the other hand, digenic and oligogenic inheritance makes it more challenging to determine which mutations are responsible for the phenotype, a challenge that is not resolved by NGS and require complex analysis.

Indeed, until now, the underlying genetic causes of hereditary anemia have been linked to over 70 genes involved in RBC physiology and function and can be roughly divided into four main categories according to their clinical presentations and morphological RBC alterations, diseases that affect red blood cell and hemoglobin production such as thalassemia, defects in the structural organization or transport function of RBC membrane, hyporegenerative anemia such as congenital dyserythropoietic anemia, and enzymopathies of RBC such as PKD and G6PD (Russo et al., 2020).

Up until recently, a clinical diagnosis might be confirmed by a single gene test used in molecular diagnostics. This conventional sequencing method yields a very sensitive and specific test for single gene diseases with obvious phenotypes, such as beta thalassemia. After employing various molecular techniques exploring all regions of both beta and alpha globin genes clusters without finding a convincing explanation for the phenotype of our affected patients, we expanded our search for variants to include other genes. Numerous models and techniques have been developed to speculate on the consequences of single-nucleotide variations. However, these approaches do not generalize well to the affect of combining single nucleotide variations. Evidence from a growing list of diseases shows that the interaction of alleles in several genes can have profound phenotypic consequences. Pathogenicity, genetic disease tendency, and severity may be influenced by the existence of genetic variations. Therefore, the complexity of predicting the impact of genetic variations grows exponentially when a combination of disease-causing variations is being investigated simultaneously. Every specific disease is determined by three primary variables, the number of loci implicated, the amount to which the environment is engaged, and whether or not a prominent locus contributes considerably to the phenotype. Since our several molecular approaches typically failed to yield definitive results, the monogenic model of disease inheritance is often revised to include other considerations, factors, and loci. In this chapter, I have described the details of a family where the

clinical phenotype of the progeny may be explained by the separate segregation of anemia traits. The pathogenesis of the disease in a single patient could involve more than one gene or the severity could be influenced by a gene modifier. Expanding on the Mendelian model can help sort out some conflicting variational and genomic data. There are three types of point mutations that can affect the expression of beta globin, which we need to deeply investigate each of them: 5' UTR and promoter mutations that have defective effect on the transcription of beta-gene, splice-junction and consensus sequence mutations, polyadenylation, and other 3' UTR mutations that alter the processing of mRNA, and non-synonymous variations that alter mRNA translation such as nonsense, frameshift, missense, and initiation codon mutations (Cao & Galanello, 2010).

On a similar note, Giess and colleagues have studied a family in which the mother, son, and daughter all shared the V148G mutation in the superoxide dismutase 1 (*SOD1*) gene, which causes familial amyotrophic lateral sclerosis (FALS), an adult onset neurological degenerative disorder that is transmitted predominantly as a dominant trait. The onset of the disease development presented at the age of 54 in the mother, 25 in the son, and the daughter was asymptomatic at the age of 35 at the time of the study. Both the mother and son died within one year of their diagnosis. This big variability set Giess's group out to identify potential modifiers that may affect age of onset. They discovered that the afflicted son had a homozygous null mutation in the *CNTF* (ciliary neurotrophic factor) gene, suggesting that CNTF may operate as a modulator of FALS. Understanding how often CNTF modifies the FALS phenotype is yet unknown, given these discoveries were based on the data from a single family (Hand & Rouleau, 2002).

A definitive diagnosis is critical for the patient to have confidence regarding the etiology of the condition, for the physician to know the disease course and what to expect and give optimal care to the patient accordingly, and for the patient and family members to have proper genetic counseling. As a result, it is critical to have conclusive evidence regarding the pathogenicity of genetic variations. In our study, results from conventional screening methods did not yield an explanation for the severity of the observed phenotype, which necessitated further investigation.

It is possible that genetic variables, like so-called "modifier genes," are at play in the clinical heterogeneity seen even in disorders with rather straightforward diagnostic criteria. A modifier gene is one gene's ability to alter the phenotypic expression of another gene at a different locus, hence altering the phenotypic expression of that other gene, or a gene's ability to alter the phenotypic manifestation of a mutant gene without altering the normal state (Genin et al., 2008).

DNA sequencing data is being generated at an unprecedented rate worldwide that permits the study of huge cohorts in an effort to better understand the causes of human illness and what affect its severity, which revolutionized the field of medical genetics. Because of the variable clinical manifestations and illness heterogeneity, discovering the underlying causes of a wide range of human diseases remains a formidable challenge. These challenges point to the emergence of more complex genetic models that account for the interplay of several variations and genes. While many bioinformatics resources are available to assist in the identification of Mendelian disease-causing variations, relatively little aid is available for the identification of oligogenic disease-causing variant combinations for patients who have exhausted conventional methods of illness investigation (Renaux et al., 2019). Next generation sequencing has been used extensively in both screening and diagnosing of beta thalassemia (Achour et al., 2021). The Dai people of Yunnan (n=951) underwent a large-scale population carrier screening effort using a targeted NGS technique spanning the globin gene cluster by He et al. (2017). Direct NGS screening for thalassemia found a 49.5% carrier rate, but the conventional method of red cell indices with hemoglobin electrophoresis followed by DNA sequencing only detected a 22% rate in a double-blind trial (He et al., 2017). Using NGS, researchers in Guangxi, China, were able to identify 458 previously undetectable mutations in 57,229 samples, many of which were previously unknown (Munkongdee et al., 2020b).

Since we have started this project, many molecular techniques in genetic diagnosis have emerged and advanced in the last few years. Constant developments and improvements in sequencing techniques have favored the approach towards whole exome sequencing and whole genome sequencing amongst researchers and scientists. Using next-generation sequencing, researchers have been able to pinpoint not just the underlying cause of hereditary anemias, but also the presence of polygenic risk factors and other genetic modifiers. WES enables us to investigate thousands of genes and to get a detailed view of the protein-coding regions of the genome (1-2%) which contains around 60% of identified pathogenic-related genomic variants with the capability to include some intronic and untranslated regions to its targeted content (Ross et al., 2020). Whole genome sequencing will become more useful as the structure and function of promoters and regulator sequences are better understood. Currently, most WGS data will be wasted because it cannot be interpreted. Promoters and enhancers can be identified but the effect of making small changes is usually not usefully understood and so is not predictable. Targeted whole exome sequencing is more affordable and generates more manageable and easier to analyze datasets with high speed and yield which makes it a better alternative to whole genome sequencing (Kedar et al., 2019) (Svidnicki et al., 2020). The main drawback of this approach is

the short read length which can limit genome assembly. Nevertheless, the positive percentage for clinical diagnosis of genetic diseases using exome sequencing or specialized panels is around 40% (Niguidula et al., 2018). There is also the potential for unintended genetic discoveries of variants that are not related to the investigated disease. The challenge of identifying multiple variants located on different genes that are responsible for the disease pathology. The discovery of an increasing number of genetic variations with unclear clinical implications using whole exome sequencing has also resulted in new challenges regarding the diagnostic process (Bertier et al., 2016) (Frésard & Montgomery, 2018) (Ross et al., 2020). Whole-exome sequencing (WES) data interpretation is difficult and time-consuming because of the need to examine all of the variations generated in the VCF file. In this study, we looked at how custom filters may be used to reduce the number of variations of interest in our related family (Figure 112).

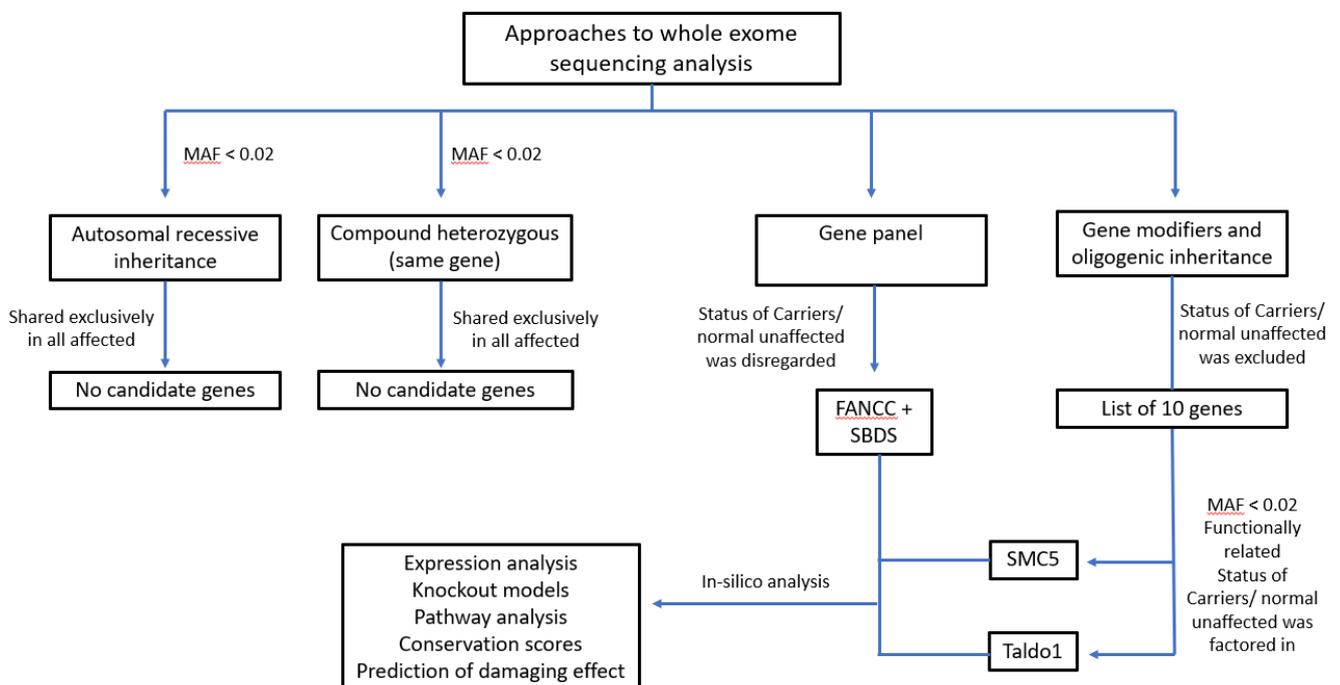


Figure 112: Approaches to whole exome analysis.

In the general population, the allelic frequency of many exome variations is rather low. Protein-truncating variations are frequently linked to illnesses and functional implications in genes with low tolerance for loss-of-function mutations. An obvious relationship may be obscured, however, when illnesses are caused by a collection of alleles with incomplete penetrance that span many genes. This problem becomes exponentially more challenging when the variants in question are reported rare due to lack of observations in the sample cohort. The primary factors to think about while doing such genetic research are defining the phenotype, collecting the genotype, analyzing the data, and verifying the results.

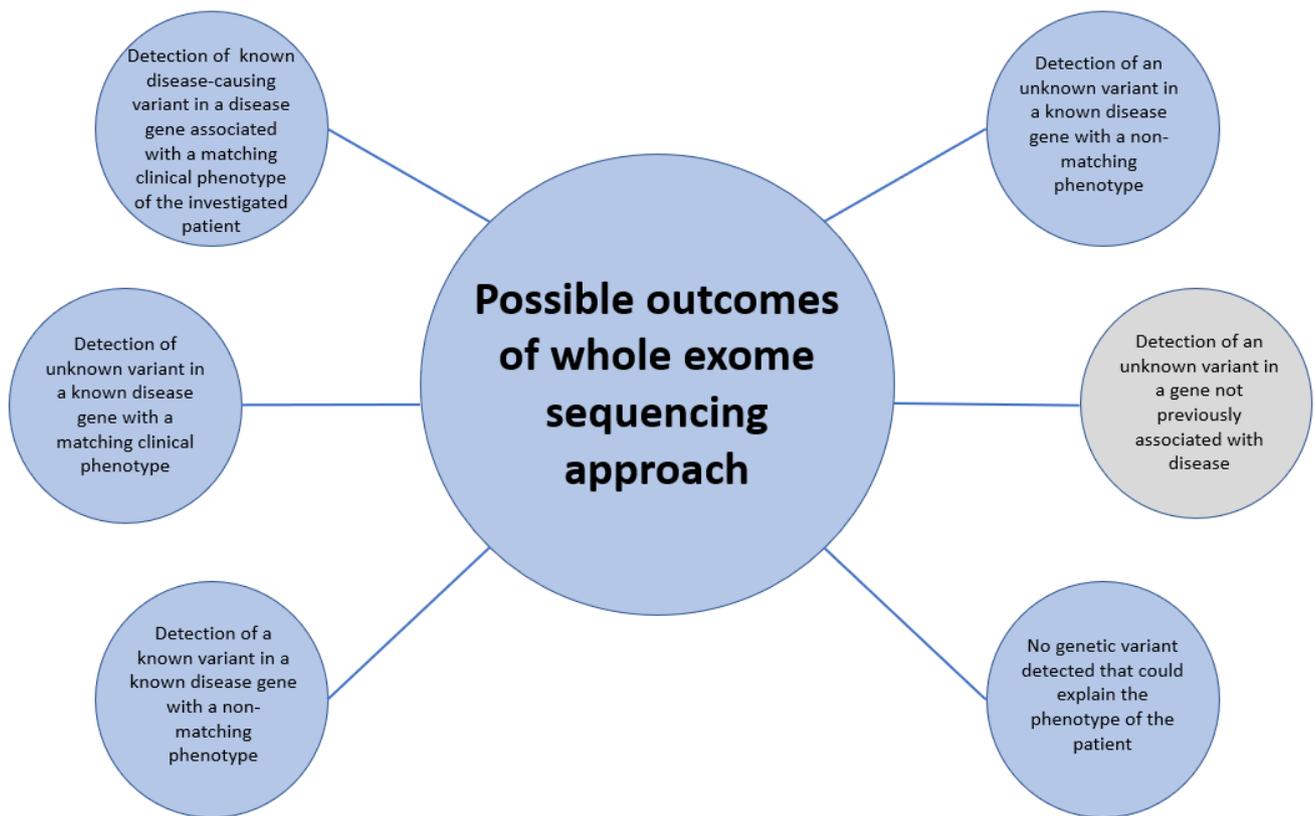


Figure 113: Most frequent possible outcomes of whole exome sequencing approach.

Described by Rodenburg (Rodenburg, 2018).

In 2018, Rodenburg et al have proposed several possible outcomes from analysis of WES data (Figure 113). In this family study, after following various scenarios in the analysis of whole exome data, we have detected variants with unknown significance in genes that have not been associated with beta thalassemia before. We have been able to establish a functional relation that links those genes to our phenotype by exploring gene expression studies, knockout models, genomic pathways associations, conservation scores, and in-silico predictions of damaging effect and pathogenicity of these variants on the produced proteins. The complexity that we found in the genetic make-up of the affected individuals in our study and the difficulty in translating their observed phenotypic outcome has led us to propose two mechanisms that may link the identified candidate genes, *SMC5* and *TALDO1*, with novel associations to beta thalassemia major phenotype. The first mechanism is to consider the *HBB* heterozygous variants (recessive *HBB* variant that is known to be associated with beta thalassemia major), while these candidate genes as genetic modifiers that contribute to phenotypic variation by aggravating the clinical manifestation severity from mild or even intermedia to major (transfusion-dependent). Currently, novel associations of genetic modifiers that can alter the disease manifestation and alleviate the disease severity in patients with the same mutation have been discovered in various

Mendelian diseases (Génin et al., 2008) (Gallati, 2014) (Rahit & Tarailo-Graovac, 2020). In 2018, O’Neal and Knowles have published a review paper on gene modifiers that are associated in the phenotypic variation of the monogenic recessive cystic fibrosis (O’Neal & Knowles, 2018). Another example is the use of whole exome sequencing in the identification of novel modifier genes related to Machado-Joseph disease (Raposo et al., 2022). Also, in another study, *ERFE* gene was identified as a modifier gene that increased the severity in patients with congenital dyserythropoietic anemia. *ERFE* encodes a soluble protein secreted by erythroid precursors, erythroferrone, that has a role in increasing iron overload in the liver when overexpressed by suppression of hepcidin expression (Andolfo et al., 2019). The second proposed mechanism (combined causative genes) is to consider their disease as digenic or polygenic inherited anemia disease with compound heterozygosity in two or more different genes which when combined together contributed to a severe transfusion dependent anemia rather than confining their case in beta thalassemia disorder, hence, consider these candidate genes as causative genes (Table 35).

Table 35: The genetic make-up of both affected and non-affected members of our family cohort.

	I-1	II-3	II-10	II-11	II-12	II-13	I-2	I-3	II-14	II-15	II-16	II-17	II-19	II-20
HBB	Het	Wild type	Het	Wild type	Wild type	Wild type	Wild type	Het	Het	Het	Het	Het	Het	Het
G6PD	Hom	Hom	Hemi	Hom	Hom	Hom	Hemi	wild type	Het	Het	Het	Wild type	Wild type	Het
FANCC	Wild type	Wild type	Wild type	Het	Het	Het	Het	wild type	Het	Het	Het	Het	Wild type	Het
SBDS	Wild type	Het	Wild type	Wild type	Het	Het	Het	Wild type	Het					
SMC5	Wild type	Wild type	Het	Wild type	Wild type	Wild type	Het	Wild type	Het	Het	Het	Het	Het	Wild type
TALDO1	Wild type	Wild type	Het	Wild type	Wild type	Wild type	Het	Wild type	Het	Het	Het	Het	Het	Wild type

The presence of two types of inherited anemias (Hereditary Elliptocytosis and Deletional Hemoglobin H Disease) in an affected patient with moderate anemia have been reported before (Charoenkwan et al., 2017). In addition, the presence of two variations in two different genes, *PIEZO1* (heterozygous) and *SEC23B* (homozygous), resulted in severe iron overload in a proband with Dehydrated hereditary stomatocytosis, a hemolytic anemia disorder (Russo et al., 2018).

In our candidate genes analysis, we used computational tools to investigate the effects of mutations on protein structures and stability. We first used AlphaFold to generate 3D models of the SMC5 and TALDO1 proteins. We then evaluated the models using Ramachandran plots and ProSA-web. The results showed that the AlphaFold models were of high quality, indicating that AlphaFold is a reliable tool for predicting protein structures.

Next, we generated mutant sequences of SMC5 and TALDO1 and modelled them using Swiss Model with the AlphaFold models as templates. We then used DynaMut to predict the impact of the mutations on protein stability. The results showed that the Asp202Asn mutation in TALDO1 was stabilizing, while the Glu226Lys mutation in SMC5 was destabilizing. Finally, we performed MD simulations on the wild-type and mutant SMC5 and TALDO1 proteins. The results showed that the mutant proteins had different stabilities than the wild-type proteins. The Asp202Asn mutation in TALDO1 made the protein more stable, while the Glu226Lys mutation in SMC5 made the protein less stable.

We also further analyzed these variations using different simulation and protein docking tools using I-TASSER and Discovery studio program and showed differences between the wild type protein and the mutated protein in both genes with regards to their 2-dimensional structure, and their active site hydrophobicity, ionizibility, charges, hydrogen bonds, and their solvent accessible surface. We showed differences between the wildtype and mutant genes.

Moreover, using different prediction tools that predict the deleterious effects of single nucleotide polymorphisms and missense mutations, we got different results. Further in-vitro and in-vivo investigations are required to evaluate the functional effect of these variants.

Afterall, my findings suggest that not every monogenic disease case is a straightforward case, some may turn out to be complex. Variable phenotypes in monogenic diseases might be caused by the involvement of other genes which may impact its diagnosis and treatment.

Allele linkage analysis

In a different approach, whole exome analysis have been also used in allele linkage analysis. Allele linkage analysis uses known chromosomal locations of genetic markers that are inherited together with the trait of interest, to pinpoint the gene responsible for a disease (Ott et al., 2015).

In our fourth scenario of data filtering, we looked to see which variants over the whole exome were transmitted from father and not mothers to the affected individuals in both families (in a heterozygous state). As for the rest of the unaffected siblings, if they do carry the father's allele, they would have the same genotype as the father (carrier state). Therefore, looking back at the hb electrophoresis results, which indicated that II-19 is a carrier, while II-3, II-11, II-12, and II-13 are normal, and since II-19 does not have previously identified *HBB* variant, then we can assume that he has inherited the heterozygous pathogenic allele from his father. II-19 was included in while II-3, II-11, II-12, and II-13 were filtered out. Not knowing the hematological status of II-20 led to her sample being disregarded.

We came up with a set of rare variants. We accidently found out that a number of them (*OR5P3*, *OR52B2*, and *OR52I1*) were very close to *HBB* on chromosome 11, in fact flanking *HBB*. They are odorant receptor genes (olfactory genes) which are famous for being wrongly identified as candidates in WES studies. If the variants are all rare and all transmitted together, then they might be a haplotype from one of his Chr11s. They flank one of father's *HBB* loci. They are markers for his *HBB* locus. We seem to have shown that there is a set of markers that flank *HBB* that are being transmitted from one of father's Chr11's to all his affected offspring. Therefore there is a chance that the mutation is in his *HBB* cluster still but we did not find it yet.

6.6 Conclusion

As a speedy and accurate technique for the first screening of beta thalassemia patients, the TaqMan single nucleotide polymorphism (SNP) genotyping assays will reduce the requirement for direct sequencing of the *HBB* gene. This method has successfully confirmed the presence of sickle cell and beta thalassemia variants in 100 out of 154 patients.

There can be no doubt that next-generation sequencing has marked a turning point in the identification and characterization of thalassemia syndromes. In this project, this technique was useful in diagnosing 51 out of 54 patients with beta thalassemia.

Beta thalassemia is a heterogenous disease with a wide range of clinical severity and the steps towards identification of the underlying genetic cause of the phenotype is different from case to case and may require a combination of several molecular techniques. Therefore, the interaction of illness-causing

variations with the rest of an individual's genome is crucial to gaining a complete understanding of the condition.

Excellent detection rates in less time may be achieved with a specialized filtering technique and strategy, making this an option for primary laboratory workflow. This technique helped in the identification of two gene candidates, SMC5 and TALDO1, with possible novel associations in increasing the severity of clinical manifestation in transfusion-dependent patients with heterozygous pathogenic variant of beta thalassemia.

Diagnostic testing based on NGS may now be performed in a time-effective manner and it can also detect polygenic diseases and modifier variations. As our understanding of the genetic aspects of hereditary anemia expands, we will be better able to diagnose these individuals and provide them with individualized clinical therapy.

Documenting such complex cases will aid medical staff in providing appropriate care to similar cases and highlights the importance of following up with the diagnosis investigation process and identifying the genetic causes in inherited diseases to minimize misdiagnosis incidences and to provide the best therapeutic strategy and treatment approach.

6.7 Future Recommendations

Gene allele linkage analysis

Given that *HBB* is the most obvious locus to find a thalassemia allele, we can continue our *HBB* analysis by attempting to confirm that there is indeed a whole continuous linked region by removing the MAF restraint in our WES data filtration and look for variants as markers rather than disease causing variants. Then we can go back into father's WGS data and look for anything wrong with his LCR. Collect new samples to perform WGS to the rest of both families' members (A and B). So far we have looked for known variants in the father's WGS data. By applying our fourth filtration method to the rest of family members, we might be able to assemble paternal haplotypes over the LCR and identify unknown causative variants.

Repeat of MLPA testing

Since we did not have positive controls of alpha gene triplication or quadruplication to include in our previous MLPA test, we can repeat the test with positive controls to be confident of our exclusion of extra alpha globin genes.

Investigation of clinical data and phenotypes

A thorough and detailed investigation of clinical manifestations and phenotypes of all affected members with a hematologist consultation can be very beneficial in our data analysis. An new ethical approval should be applied for to have access to these patients records and medical history.

Candidate genes functional analysis

Investigation of rare genetic variations of undetermined clinical relevance is currently one of the most pressing issues in human molecular genetics. Although several in-silico techniques have been performed for predicting the missense variants structural effects on the produced protein, these proteins interact within the cell with other molecules to perform a biological function, therefore our investigation must be supported with in-vitro and in-vivo functional analysis as these computational methods do not predict the variant effect in actual clinical presentation. To further expand our knowledge of these variations, demonstrate their pathogenic character, and correlate them with the increased severity status of the affected patients, functional consequences of these variants should be investigated experimentally. Scientific research relies heavily on in vitro cultured cell lines as models to investigate a wide range of in vivo events. To get a better understanding of their function in different biological processes, researchers can modify certain molecular targets or cellular signaling pathways using these cellular systems. Researchers studying erythropoiesis in health and illness rely on erythroid cell lines as a tool to manipulate molecular targets and to determine deleterious effect of different variants in vivo. Red blood cell researchers now have access to useful in vitro models of erythroid development because of the recently created human umbilical cord blood-derived erythroid progenitor (HUDEP) cell lines. The HUDEP-2 cell line is the most similar to adult erythroid cells and expresses adult β -globin compared to the other two HUDEP cell lines. Because HUDEP cells are simple to transduce, scientists have found a way to modulate gene expression through the use of CRISPR/Cas9 gene disruption or RNA interference with short hairpin RNAs (shRNAs). There is now a promising new way to use the CRISPR/Cas9 system in conjunction with HUDEP cells to generate a clonal cell population with biallelic disruption of a target gene (Vinjamur & Bauer, 2018) (Gupta et al., 2019) (Daniels et al., 2020).

Furthermore, different experimental designs must be performed to investigate the damaging effect of the candidate genes variants in combination with the beta thalassemia variants (identified in our cohort) and the G6PD variant to find if they inflict severe manifestation together.

REFERENCES

- Abdelgadir, E., Al Sahlawi, M., Al Turki, L., Khamees, K., & Ahmed, W. (2019). Identification of a New Homozygous CEP290 Gene Mutation in a Saudi Family Causing Joubert Syndrome Using Next-Generation Sequencing. *Saudi Journal of Kidney Diseases and Transplantation*, 30(4), 964-968. <https://doi.org/10.4103/1319-2442.265475>
- Abotalib, Z. (2013). Preimplantation genetic diagnosis in Saudi Arabia. *Bioinformation*, 9(8), 388-393. <https://doi.org/10.6026/97320630009388>
- Abuzenadah, A., Hussein, I. M., Damanhour, G. A., FM, A. S., Gari, M. A., Chaudhary, A. G., Zaher, G. F., Al-Attas, A., & Al-Qahtani, M. H. (2011). Molecular basis of beta-thalassemia in the western province of Saudi Arabia: identification of rare beta-thalassemia mutations. *Hemoglobin*, 35(4), 346-357. <https://doi.org/10.3109/03630269.2011.588508>
- Abuzenadah, A. M., Zaher, G. F., Dallol, A., Damanhour, G. A., Chaudhary, A. G., Al-Sayes, F., Gari, M. A., AlZahrani, M., Hindawi, S., & Al-Qahtani, M. H. (2013). Identification of a novel SBF2 missense mutation associated with a rare case of thrombocytopenia using whole-exome sequencing. *Journal of Thrombosis and Thrombolysis*, 36(4), 501-506. <https://doi.org/10.1007/s11239-012-0864-x>
- Achour, A., Koopmann, T. T., Baas, F., & Hartevel, C. L. (2021). The Evolving Role of Next-Generation Sequencing in Screening and Diagnosis of Hemoglobinopathies [Mini Review]. *Frontiers in Physiology*, 12. <https://doi.org/10.3389/fphys.2021.686689>
- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods*, 7(4), 248-249. <https://doi.org/10.1038/nmeth0410-248>
- Agarwal, A. M., Nussenzeig, R. H., Reading, N. S., Patel, J. L., Sangle, N., Salama, M. E., Prchal, J. T., Perkins, S. L., Yaish, H. M., & Christensen, R. D. (2016). Clinical utility of next-generation sequencing in the diagnosis of hereditary haemolytic anaemias. *Br J Haematol*, 174(5), 806-814. <https://doi.org/10.1111/bjh.14131>
- Aggarwal, A., Jamwal, M., Sharma, P., Sachdeva, M. U. S., Bansal, D., Malhotra, P., & Das, R. (2020). Deciphering molecular heterogeneity of Indian families with hereditary spherocytosis using targeted next-generation sequencing: First South Asian study. *Br J Haematol*, 188(5), 784-795. <https://doi.org/10.1111/bjh.16244>
- Ahmed, M. H., Ghatge, M. S., & Safo, M. K. (2020). Hemoglobin: Structure, Function and Allostery. *Subcell Biochem*, 94, 345-382. https://doi.org/10.1007/978-3-030-41769-7_14
- Aizaz, M., Kiani, Y. S., Nisar, M., Shan, S., Paracha, R. Z., & Yang, G. (2023). Genomic Analysis, Evolution and Characterization of E3 Ubiquitin Protein Ligase (TRIM) Gene Family in Common Carp (*Cyprinus carpio*). *Genes (Basel)*, 14(3). <https://doi.org/10.3390/genes14030667>
- Akhtar, M. S., Qaw, F., Borgio, J. F., Albuali, W., Suliman, A., Nasserullah, Z., Al-Jarrash, S., & Al-Ali, A. (2013). Spectrum of α -thalassemia mutations in transfusion-dependent β -thalassemia patients from the Eastern Province of Saudi Arabia. *Hemoglobin*, 37(1), 65-73.
- Al-Awamy, B. H., Niazi, G. A., Al-Mouzan, M. I., Wilson, J. B., Chen, S. S., Webber, B. B., & Huisman, T. H. (1985). Hb F-Dammam or alpha 2A gamma 2(79) (EF3) Asp----Asn. *Hemoglobin*, 9(2), 171-173.
- Al-Jaouni, S. K., Jarullah, J., Azhar, E., & Moradkhani, K. (2011). Molecular characterization of glucose-6-phosphate dehydrogenase deficiency in Jeddah, Kingdom of Saudi Arabia. *BMC Research Notes*, 4(1), 436. <https://doi.org/10.1186/1756-0500-4-436>
- Al-Nafie, A. N., Borgio, J. F., AbdulAzeez, S., Al-Suliman, A. M., Qaw, F. S., Nasserullah, Z. A., Al-Jarrash, S., Al-Madan, M. S., Al-Ali, R. A., & AlKhalifah, M. A. (2015). Co-inheritance of novel ATRX gene

- mutation and globin (α & β) gene mutations in transfusion dependent beta-thalassemia patients. *Blood Cells, Molecules, and Diseases*, 55(1), 27-29.
- Al-Samkari, H., & van Beers, E. J. (2021). Mitapivat, a novel pyruvate kinase activator, for the treatment of hereditary hemolytic anemias. *Ther Adv Hematol*, 12, 20406207211066070. <https://doi.org/10.1177/20406207211066070>
- Al-Samkari, H., van Beers, E. J., Kuo, K. H. M., Barcellini, W., Bianchi, P., Glenthøj, A., del Mar Mañú Pereira, M., van Wijk, R., Glader, B., & Grace, R. F. (2020). The variable manifestations of disease in pyruvate kinase deficiency and their management. *Haematologica*, 105(9), 2229-2239. <https://doi.org/10.3324/haematol.2019.240846>
- Al-Shahrani, M. (2009). Steps toward the prevention of hemoglobinopathies in the kingdom of Saudi Arabia. *Hemoglobin*, 33 Suppl 1, S21-24. <https://doi.org/10.3109/03630260903346437>
- Al Abbar, A., Ngai, S. C., Nogales, N., Alhaji, S. Y., & Abdullah, S. (2020). Induced Pluripotent Stem Cells: Reprogramming Platforms and Applications in Cell Replacement Therapy. *Biores Open Access*, 9(1), 121-136. <https://doi.org/10.1089/biores.2019.0046>
- Al Asoom, L. I., Al Makhaita, M. M., Rafique, N., Al Afandi, D. T., Al Otaibi, W. M., Alsuwat, H. S., Alaithan, M. A., AbdulAzeez, S., & Borgio, J. F. (2020). Effects of (-3.7)alpha Deletion and Sickle-Cell Trait on Ventilatory and Hemodynamic Responses to Maximum Exercise in Young Saudi Females. *J Blood Med*, 11, 371-378. <https://doi.org/10.2147/JBM.S272905>
- Al Zadjali, S., Al-Riyami, A. Z., Gravell, D., Al Haddabi, H., Al Rawahi, M., Al Falahi, K., & Daar, S. (2014). Potential pitfalls in the diagnosis of Hb Handsworth in areas with high prevalence of HbS. *Int J Lab Hematol*, 36(4), 488-492. <https://doi.org/10.1111/ijlh.12157>
- Alabdulaali, M. K. (2007). Sickle cell disease patients in eastern province of Saudi Arabia suffer less severe acute chest syndrome than patients with African haplotypes. *Annals of Thoracic Medicine*, 2(4), 158-162.
- Alaithan, M. A., AbdulAzeez, S., & Borgio, J. F. (2018). A comprehensive review of the prevalence of beta globin gene variations and the co-inheritance of related gene variants in Saudi Arabians with beta-thalassemia. *Saudi Med J*, 39(4), 329-335. <https://doi.org/10.15537/smj.2018.4.21360>
- Albagshi, M. H., Saad, M., Aljasseem, A. M., Bushehab, A. A., Ahmed, N. H., Alabbad, M. M., Omer, N., Alhamad, O. A., Sultan, T. A., & Bahgat, S. (2021). Blood Demand and Challenges for Patients With Beta-Thalassemia Major in Eastern Saudi Arabia. *Cureus*, 13(8), e17470. <https://doi.org/10.7759/cureus.17470>
- Albitar, M., Katsumata, M., & Liebhaber, S. A. (1991). Human alpha-globin genes demonstrate autonomous developmental regulation in transgenic mice. *Mol Cell Biol*, 11(7), 3786-3794. <https://doi.org/10.1128/mcb.11.7.3786-3794.1991>
- Albitar, M., Peschle, C., & Liebhaber, S. A. (1989). Theta, Zeta, and Epsilon Globin Messenger RNAs Are Expressed in Adults. *Blood*, 74(2), 629-637. <https://doi.org/https://doi.org/10.1182/blood.V74.2.629.629>
- Aldakeel, S. A., Ghanem, N. Z., Al-Amodi, A. M., Osman, A. K., Al Asoom, L. I., Ahmed, N. R., Almandil, N. B., Akhtar, M. S., Azeez, S. A., & Borgio, J. F. (2020). Identification of seven novel variants in the beta-globin gene in transfusion-dependent and normal patients. *Arch Med Sci*, 16(2), 453-459. <https://doi.org/10.5114/aoms.2019.84825>
- Alders, M., Ryan, A., Hodges, M., Bliiek, J., Feinberg, A. P., Privitera, O., Westerveld, A., Little, P. F., & Mannens, M. (2000). Disruption of a novel imprinted zinc-finger gene, ZNF215, in Beckwith-Wiedemann syndrome. *Am J Hum Genet*, 66(5), 1473-1484. <https://doi.org/10.1086/302892>
- Alfadhli, S., Kaaba, S., Elshafey, A., Salim, M., AlAwadi, A., & Bastaki, L. (2005). Molecular characterization of glucose-6-phosphate dehydrogenase gene defect in the Kuwaiti population. *Arch Pathol Lab Med*, 129(9), 1144-1147. <https://doi.org/10.5858/2005-129-1144-mcogdg>

- Alfares, A. A. (2018). Applying filtration steps to interpret the results of whole-exome sequencing in a consanguineous population to achieve a high detection rate. *Int J Health Sci (Qassim)*, 12(5), 35-43.
- AlHamdan, N. A., AlMazrou, Y. Y., AlSwaidi, F. M., & Choudhry, A. J. (2007). Premarital screening for thalassemia and sickle cell disease in Saudi Arabia. *Genet Med*, 9(6), 372-377. <http://dx.doi.org/10.1097/GIM.0b013e318065a9e8>
- Ali, S., Mumtaz, S., Shakir, H. A., Khan, M., Tahir, H. M., Mumtaz, S., Mughal, T. A., Hassan, A., Kazmi, S. A. R., Sadia, Irfan, M., & Khan, M. A. (2021). Current status of beta-thalassemia and its treatment strategies. *Mol Genet Genomic Med*, 9(12), e1788. <https://doi.org/10.1002/mgg3.1788>
- Allison, A. C. (1954). Protection afforded by sickle-cell trait against subtertian malarial infection. *Br Med J*, 1(4857), 290-294. <https://doi.org/10.1136/bmj.1.4857.290>
- Alsaeed, Farhat, G. N., Assiri, A. M., Memish, Z., Ahmed, E. M., Saeedi, M. Y., Al-Dossary, M. F., & Bashawri, H. (2018a). Distribution of hemoglobinopathy disorders in Saudi Arabia based on data from the premarital screening and genetic counseling program, 2011-2015. *J Epidemiol Glob Health*, 7 Suppl 1(Suppl 1), S41-S47. <https://doi.org/10.1016/j.jegh.2017.12.001>
- Alsaeed, Farhat, G. N., Assiri, A. M., Memish, Z., Ahmed, E. M., Saeedi, M. Y., Al-Dossary, M. F., & Bashawri, H. (2018b). Distribution of hemoglobinopathy disorders in Saudi Arabia based on data from the premarital screening and genetic counseling program, 2011–2015. *Journal of Epidemiology and Global Health*, 7(S1), S41-S47. <https://doi.org/https://doi.org/10.1016/j.jegh.2017.12.001>
- Alswaidi, F. M., & O'Brien, S. J. (2009). Premarital screening programmes for haemoglobinopathies, HIV and hepatitis viruses: review and factors affecting their success. *Journal of Medical Screening*, 16(1), 22-28. <https://doi.org/10.1258/jms.2008.008029>
- Alwazani, W. A., Zahid, R., Elaimi, A., Bajouh, O., Hindawi, S., Arab, B., Damanhour, G., Saka, M. Y., Turki, R., Khan, J. A., Dallol, A., & Abuzenadah, A. M. (2016). Detection of beta-Thalassemia Mutations Using TaqMan Single Nucleotide Polymorphism Genotyping Assays. *Genet Test Mol Biomarkers*, 20(3), 154-157. <https://doi.org/10.1089/gtmb.2015.0222>
- Amebor, P. M. (2022). Early-life environmental exposures and anaemia among children under age five in Sub-Saharan Africa: An insight from the Demographic & Health Surveys. *Sci Total Environ*, 832, 154957. <https://doi.org/10.1016/j.scitotenv.2022.154957>
- Anderson, H. L., Brodsky, I. E., & Mangalmurti, N. S. (2018). The Evolving Erythrocyte: Red Blood Cells as Modulators of Innate Immunity. *J Immunol*, 201(5), 1343-1351. <https://doi.org/10.4049/jimmunol.1800565>
- Andolfo, I., Rosato, B. E., Marra, R., De Rosa, G., Manna, F., Gambale, A., Iolascon, A., & Russo, R. (2019). The BMP-SMAD pathway mediates the impaired hepatic iron metabolism associated with the ERFE-A260S variant. *Am J Hematol*, 94(11), 1227-1235. <https://doi.org/10.1002/ajh.25613>
- Antwi-Baffour, S., Adjei, J. K., Forson, P. O., Akakpo, S., Kyeremeh, R., & Seidu, M. A. (2019). Comorbidity of Glucose-6-Phosphate Dehydrogenase Deficiency and Sickle Cell Disease Exert Significant Effect on RBC Indices. *Anemia*, 2019, 3179173. <https://doi.org/10.1155/2019/3179173>
- Aoki, T. (2017). A Comprehensive Review of Our Current Understanding of Red Blood Cell (RBC) Glycoproteins. *Membranes (Basel)*, 7(4). <https://doi.org/10.3390/membranes7040056>
- Archer, N. M., Petersen, N., Clark, M. A., Buckee, C. O., Childs, L. M., & Duraisingh, M. T. (2018). Resistance to Plasmodium falciparum in sickle cell trait erythrocytes is driven by oxygen-dependent growth inhibition. *Proc Natl Acad Sci U S A*, 115(28), 7350-7355. <https://doi.org/10.1073/pnas.1804388115>

- Asadov, C., Alimirzoeva, Z., Mammadova, T., Aliyeva, G., Gafarova, S., & Mammadov, J. (2018). β -Thalassemia intermedia: a comprehensive overview and novel approaches. *Int J Hematol*, *108*(1), 5-21. <https://doi.org/10.1007/s12185-018-2411-9>
- Ashiuchi, M., Yagami, T., Willey, R. J., Padovan, J. C., Chait, B. T., Popowicz, A., Manning, L. R., & Manning, J. M. (2005). N-terminal acetylation and protonation of individual hemoglobin subunits: position-dependent effects on tetramer strength and cooperativity. *Protein Sci*, *14*(6), 1458-1471. <https://doi.org/10.1110/ps.041267405>
- Ashley-Koch, A., Yang, Q., & Olney, R. S. (2000). Sickle Hemoglobin (Hb S) Allele and Sickle Cell Disease: A HuGE Review. *American Journal of Epidemiology*, *151*(9), 839-845. <http://aje.oxfordjournals.org/content/151/9/839.abstract>
- Ashorobi, D., Ramsey, A., Yarrarapu, S. N. S., & Bhatt, R. (2022). Sickle Cell Trait. In *StatPearls*. <https://www.ncbi.nlm.nih.gov/pubmed/30725815>
- Atkins, A., Xu, M. J., Li, M., Rogers, N. P., Pryzhkova, M. V., & Jordan, P. W. (2020). SMC5/6 is required for replication fork stability and faithful chromosome segregation during neurogenesis. *eLife*, *9*, e61171. <https://doi.org/10.7554/eLife.61171>
- Atroshi, S. D., Al-Allawi, N., Chui, D. H. K., Najmabadi, H., & Khailany, R. A. (2021). A Novel beta(0)-Thalassemia Mutation, HBB: c.356_357delTT [Codon 118 (-TT)] in an Iraqi Kurd. *Hemoglobin*, *45*(3), 212-214. <https://doi.org/10.1080/03630269.2021.1941082>
- Auerbach, M., & Adamson, J. W. (2016). How we diagnose and treat iron deficiency anemia. *American Journal of Hematology*, *91*(1), 31-38. <https://doi.org/https://doi.org/10.1002/ajh.24201>
- Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurler, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., . . . National Eye Institute, N. I. H. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68-74. <https://doi.org/10.1038/nature15393>
- Azouzi, S., Romana, M., Arashiki, N., Takakuwa, Y., El Nemer, W., Peyrard, T., Colin, Y., Amireault, P., & Le Van Kim, C. (2018). Band 3 phosphorylation induces irreversible alterations of stored red blood cells. *Am J Hematol*, *93*(5), E110-E112. <https://doi.org/10.1002/ajh.25044>
- Babushok, D. V., Duke, J. L., Xie, H. M., Stanley, N., Atienza, J., Perdignes, N., Nicholas, P., Ferriola, D., Li, Y., Huang, H., Ye, W., Morrisette, J. J. D., Kearns, J., Porter, D. L., Podsakoff, G. M., Eisenlohr, L. C., Biegel, J. A., Chou, S. T., Monos, D. S., . . . Olson, T. S. (2017). Somatic HLA Mutations Expose the Role of Class I-Mediated Autoimmunity in Aplastic Anemia and its Clonal Complications. *Blood Adv*, *1*(22), 1900-1910. <https://doi.org/10.1182/bloodadvances.2017010918>
- Badano, J. L., & Katsanis, N. (2002). Beyond Mendel: an evolving view of human genetic disease transmission. *Nat Rev Genet*, *3*(10), 779-789. <https://doi.org/10.1038/nrg910>
- Baird, D. C., Batten, S. H., & Sparks, S. K. (2022). Alpha- and Beta-thalassemia: Rapid Evidence Review. *Am Fam Physician*, *105*(3), 272-280. <https://www.ncbi.nlm.nih.gov/pubmed/35289581>
- Ballas, S. K., Zeidan, A. M., Duong, V. H., DeVaux, M., & Heeney, M. M. (2018). The effect of iron chelation therapy on overall survival in sickle cell disease and beta-thalassemia: A systematic review. *Am J Hematol*, *93*(7), 943-952. <https://doi.org/10.1002/ajh.25103>
- Banki, K., Eddy, R. L., Shows, T. B., Halladay, D. L., Bullrich, F., Croce, C. M., Jurecic, V., Baldini, A., & Perl, A. (1997). The human transaldolase gene (TALDO1) is located on chromosome 11 at p15.4-p15.5. *Genomics*, *45*(1), 233-238. <https://doi.org/10.1006/geno.1997.4932>
- Banne, E., Meiner, V., Shaag, A., Katz-Brull, R., Gamliel, A., Korman, S., Cederboim, S. H., Duvdevani, M. P., Frumkin, A., Zilkha, A., Kapuller, V., Arbell, D., Cohen, E., & Eventov-Friedman, S. (2016). Transaldolase Deficiency: A New Case Expands the Phenotypic Spectrum. *JIMD Rep*, *26*, 31-36. https://doi.org/10.1007/8904_2015_474

- Barbarani, G., Labeledz, A., Stucchi, S., Abbiati, A., & Ronchi, A. E. (2021). Physiological and Aberrant γ -Globin Transcription During Development [Mini Review]. *Frontiers in Cell and Developmental Biology*, 9. <https://doi.org/10.3389/fcell.2021.640060>
- Barrangou, R., & Marraffini, L. A. (2014). CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol Cell*, 54(2), 234-244. <https://doi.org/10.1016/j.molcel.2014.03.011>
- Barrera-Reyes, P. K., & Tejero, M. E. (2019). Genetic variation influencing hemoglobin levels and risk for anemia across populations. *Ann N Y Acad Sci*, 1450(1), 32-46. <https://doi.org/10.1111/nyas.14200>
- Bauer, D. E., Kamran, S. C., Lessard, S., Xu, J., Fujiwara, Y., Lin, C., Shao, Z., Canver, M. C., Smith, E. C., Pinello, L., Sabo, P. J., Vierstra, J., Voit, R. A., Yuan, G. C., Porteus, M. H., Stamatoyannopoulos, J. A., Lettre, G., & Orkin, S. H. (2013). An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science*, 342(6155), 253-257. <https://doi.org/10.1126/science.1242088>
- Bava, K. A., Gromiha, M. M., Uedaira, H., Kitajima, K., & Sarai, A. (2004). ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res*, 32(Database issue), D120-121. <https://doi.org/10.1093/nar/gkh082>
- Belcher, J. D., Marker, P. H., Weber, J. P., Hebbel, R. P., & Vercellotti, G. M. (2000). Activated monocytes in sickle cell disease: potential role in the activation of vascular endothelium and vaso-occlusion. *Blood*, 96(7), 2451-2459. <https://doi.org/10.1182/blood.V96.7.2451>
- Ben, A., Masrati, G., Kessel, A., Narunsky, A., Sprinzak, J., Lahav, S., Ashkenazy, H., & Ben-Tal, N. (2020). ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. *Protein Sci*, 29(1), 258-267. <https://doi.org/10.1002/pro.3779>
- Ben Salah, N., Bou-Fakhredin, R., Mellouli, F., & Taher, A. T. (2017). Revisiting beta thalassemia intermedia: past, present, and future prospects. *Hematology*, 22(10), 607-616. <https://doi.org/10.1080/10245332.2017.1333246>
- Bender, D. A., & Mayes, P. A. (2018). The Pentose Phosphate Pathway & Other Pathways of Hexose Metabolism. In V. W. Rodwell, D. A. Bender, K. M. Botham, P. J. Kennelly, & P. A. Weil (Eds.), *Harper's Illustrated Biochemistry*, 31e. McGraw-Hill Education. accessmedicine.mhmedical.com/content.aspx?aid=1160192322
- Benesch, R., Benesch, R. E., & Enoki, Y. (1968). The interaction of hemoglobin and its subunits with 2,3-diphosphoglycerate. *Proc Natl Acad Sci U S A*, 61(3), 1102-1106. <https://doi.org/10.1073/pnas.61.3.1102>
- Berg JM, T. J., Stryer L. (2002). Hemoglobin Transports Oxygen Efficiently by Binding Oxygen Cooperatively. In *Biochemistry. 5th edition*. W H Freeman.
- Bertier, G., Héту, M., & Joly, Y. (2016). Unsolved challenges of clinical whole-exome sequencing: a systematic literature review of end-users' views. *BMC Med Genomics*, 9(1), 52. <https://doi.org/10.1186/s12920-016-0213-6>
- Bessis, M. (1958). [Erythroblastic island, functional unity of bone marrow]. *Rev Hematol*, 13(1), 8-11. (L'îlot érythroblastique, unité fonctionnelle de la moelle osseuse.)
- Bhandari, J., Thada, P. K., & Puckett, Y. (2023). Fanconi Anemia. In *StatPearls*. <https://www.ncbi.nlm.nih.gov/pubmed/32644559>
- Bhusal, A., Bhandari, S., Seth, T., & Sah, R. P. (2022). Homozygous Lepore Syndrome: A case report. *Ann Med Surg (Lond)*, 80, 104168. <https://doi.org/10.1016/j.amsu.2022.104168>
- Bianchi, Fermo, E., Glader, B., Kanno, H., Agarwal, A., Barcellini, W., Eber, S., Hoyer, J. D., Kuter, D. J., Maia, T. M., Mañu-Pereira, M. d. M., Kalfa, T. A., Pissard, S., Segovia, J.-C., van Beers, E., Gallagher, P. G., Rees, D. C., van Wijk, R., & with the endorsement of EuroBloodNet, t. E. R. N. i. R. H. D. (2019). Addressing the diagnostic gaps in pyruvate kinase deficiency: Consensus recommendations on the diagnosis of pyruvate kinase deficiency. *American Journal of Hematology*, 94(1), 149-161. <https://doi.org/https://doi.org/10.1002/ajh.25325>

- Bianchi, P., & Fermo, E. (2020). Molecular heterogeneity of pyruvate kinase deficiency. *Haematologica*, 105(9), 2218-2228. <https://doi.org/10.3324/haematol.2019.241141>
- Bianchi, P., Fermo, E., Lezon-Geyda, K., van Beers, E. J., Morton, H. D., Barcellini, W., Glader, B., Chonat, S., Ravindranath, Y., Newburger, P. E., Kollmar, N., Despotovic, J. M., Verhovsek, M., Sharma, M., Kwiatkowski, J. L., Kuo, K. H. M., Wlodarski, M. W., Yaish, H. M., Holzhauer, S., . . . Grace, R. F. F. (2020). Genotype-phenotype correlation and molecular heterogeneity in pyruvate kinase deficiency. *American Journal of Hematology*, 95(5), 472-482. <https://doi.org/https://doi.org/10.1002/ajh.25753>
- Bin Zuair, A., Aldossari, S., Alhumaidi, R., Alrabiah, M., & Alshabanat, A. (2023). The Burden of Sickle Cell Disease in Saudi Arabia: A Single-Institution Large Retrospective Study. *Int J Gen Med*, 16, 161-171. <https://doi.org/10.2147/IJGM.S393233>
- Blake, J. A., Baldarelli, R., Kadin, J. A., Richardson, J. E., Smith, C. L., Bult, C. J., & Mouse Genome Database, G. (2021). Mouse Genome Database (MGD): Knowledgebase for mouse-human comparative biology. *Nucleic Acids Res*, 49(D1), D981-D987. <https://doi.org/10.1093/nar/gkaa1083>
- Boehm, C. D., Dowling, C. E., Waber, P. G., Giardina, P. J., & Kazazian Jr, H. H. (1986). Use of oligonucleotide hybridization in the characterization of a β^0 -thalassemia gene ($\beta 37$ TGG \rightarrow TGA) in a Saudi Arabian family. *Blood*, 67(4), 1185-1188.
- Bogdanova, A., Kaestner, L., Simionato, G., Wickrema, A., & Makhro, A. (2020). Heterogeneity of Red Blood Cells: Causes and Consequences. *Front Physiol*, 11, 392. <https://doi.org/10.3389/fphys.2020.00392>
- Boocock, G. R. B., Morrison, J. A., Popovic, M., Richards, N., Ellis, L., Durie, P. R., & Rommens, J. M. (2003). Mutations in SBDS are associated with Shwachman–Diamond syndrome. *Nature Genetics*, 33(1), 97-101. <https://doi.org/10.1038/ng1062>
- Borgio, J. F., Abdulazeez, S., Almandil, N. B., Naserullah, Z. A., Al-Jarrash, S., Al-Suliman, A. M., Elfakharay, H. I., Qaw, F. S., Alabdrabalnabi, F. I., Alkhalifah, M. A., Shakil Akhtar, M., Qutub, H., & Al-Ali, A. K. (2018). The -alpha3.7 deletion in alpha-globin genes increases the concentration of fetal hemoglobin and hemoglobin A2 in a Saudi Arabian population. *Mol Med Rep*, 17(1), 1879-1884. <https://doi.org/10.3892/mmr.2017.8033>
- Borgio, J. F., Al-Madan, M. S., & AbdulAzeez, S. (2016). Mutation near the binding interfaces at α -hemoglobin stabilizing protein is highly pathogenic. *Am J Transl Res*, 8(10), 4224-4232.
- Bou-Fakhredin, R., De Franceschi, L., Motta, I., Eid, A. A., Taher, A. T., & Cappellini, M. D. (2022). Redox Balance in β -Thalassemia and Sickle Cell Disease: A Love and Hate Relationship. *Antioxidants (Basel)*, 11(5). <https://doi.org/10.3390/antiox11050967>
- Bradner, J. E., Mak, R., Tanguturi, S. K., Mazitschek, R., Haggarty, S. J., Ross, K., Chang, C. Y., Bosco, J., West, N., Morse, E., Lin, K., Shen, J. P., Kwiatkowski, N. P., Gheldof, N., Dekker, J., DeAngelo, D. J., Carr, S. A., Schreiber, S. L., Golub, T. R., & Ebert, B. L. (2010). Chemical genetic strategy identifies histone deacetylase 1 (HDAC1) and HDAC2 as therapeutic targets in sickle cell disease. *Proc Natl Acad Sci U S A*, 107(28), 12617-12622. <https://doi.org/10.1073/pnas.1006774107>
- Brancaleoni, V., Di Pierro, E., Motta, I., & Cappellini, M. D. (2016a). Laboratory diagnosis of thalassemia. *Int J Lab Hematol*, 38(S1), 32-40. <https://doi.org/https://doi.org/10.1111/ijlh.12527>
- Brancaleoni, V., Di Pierro, E., Motta, I., & Cappellini, M. D. (2016b). Laboratory diagnosis of thalassemia. *Int J Lab Hematol*, 38 Suppl 1, 32-40. <https://doi.org/10.1111/ijlh.12527>
- Breveglieri, G., Travan, A., D'Aversa, E., Cosenza, L. C., Pellegatti, P., Guerra, G., Gambari, R., & Borgatti, M. (2017). Postnatal and non-invasive prenatal detection of beta-thalassemia mutations based on Taqman genotyping assays. *PLoS One*, 12(2), e0172756. <https://doi.org/10.1371/journal.pone.0172756>

- Brinkman, R., Wildschut, A., & Wittermans, A. (1934). On the occurrence of two kinds of haemoglobin in normal human blood. *J Physiol*, 80(4), 377-387. <https://doi.org/10.1113/jphysiol.1934.sp003098>
- Brittain, T. (2004). Measuring assembly and binding in human embryonic hemoglobins. *Methods Enzymol*, 379, 64-80. [https://doi.org/10.1016/S0076-6879\(04\)79004-7](https://doi.org/10.1016/S0076-6879(04)79004-7)
- Bromberg, Y., & Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res*, 35(11), 3823-3835. <https://doi.org/10.1093/nar/gkm238>
- Brugnara, C. (2018). Sick cell dehydration: Pathophysiology and therapeutic applications. *Clin Hemorheol Microcirc*, 68(2-3), 187-204. <https://doi.org/10.3233/ch-189007>
- Bukau, B., Weissman, J., & Horwich, A. (2006). Molecular chaperones and protein quality control. *Cell*, 125(3), 443-451. <https://doi.org/10.1016/j.cell.2006.04.014>
- Bulger, M., van Doorninck, J. H., Saitoh, N., Telling, A., Farrell, C., Bender, M. A., Felsenfeld, G., Axel, R., & Groudine, M. (1999). Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. *Proc Natl Acad Sci U S A*, 96(9), 5129-5134. <https://doi.org/10.1073/pnas.96.9.5129>
- Bunn, & Forget. (1986). *Hemoglobin: Molecular, Genetic and Clinical Aspects*. W.B. Saunders Company. <http://books.google.com.sa/books?id=hQZsAAAAMAAJ>
- Bunn, F., & Forget, B. (1986). *Hemoglobin--molecular, genetic, and clinical aspects*. WB Saunders Co.
- Bunn, H. F. (1987). Subunit assembly of hemoglobin: an important determinant of hematologic phenotype. *Blood*, 69(1), 1-6. <https://www.ncbi.nlm.nih.gov/pubmed/3539223>
- Bunn, H. F. (1997). Pathogenesis and treatment of sickle cell disease. *N Engl J Med*, 337(11), 762-769. <https://doi.org/10.1056/NEJM199709113371107>
- Burdick, K. J., Cogan, J. D., Rives, L. C., Robertson, A. K., Koziura, M. E., Brokamp, E., Duncan, L., Hannig, V., Pfothenauer, J., Vanzo, R., Paul, M. S., Bican, A., Morgan, T., Duis, J., Newman, J. H., Hamid, R., Phillips, J. A., 3rd, & Undiagnosed Diseases, N. (2020). Limitations of exome sequencing in detecting rare and undiagnosed diseases. *Am J Med Genet A*, 182(6), 1400-1406. <https://doi.org/10.1002/ajmg.a.61558>
- Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L., & Casadio, R. (2009). Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat*, 30(8), 1237-1244. <https://doi.org/10.1002/humu.21047>
- Calado, R. T., Graf, S. A., Wilkerson, K. L., Kajigaya, S., Ancliff, P. J., Dror, Y., Chanock, S. J., Lansdorp, P. M., & Young, N. S. (2007). Mutations in the SBDS gene in acquired aplastic anemia. *Blood*, 110(4), 1141-1146. <https://doi.org/10.1182/blood-2007-03-080044>
- Cano, M., Datta, S., Wang, L., Liu, T., Flores-Bellver, M., Sachdeva, M., Sinha, D., & Handa, J. T. (2021). Nrf2 deficiency decreases NADPH from impaired IDH shuttle and pentose phosphate pathway in retinal pigmented epithelial cells to magnify oxidative stress-induced mitochondrial dysfunction. *Aging Cell*, 20(8), e13444. <https://doi.org/10.1111/acer.13444>
- Canu, G., De Bonis, M., Minucci, A., & Capoluongo, E. (2016). Red blood cell PK deficiency: An update of PK-LR gene mutation database. *Blood Cells Mol Dis*, 57, 100-109. <https://doi.org/10.1016/j.bcmd.2015.12.009>
- Cao, A., & Galanello, R. (2010). Beta-thalassemia. *Genet Med*, 12(2), 61-76. <http://dx.doi.org/10.1097/GIM.0b013e3181cd68ed>
- Cao, A., & Moi, P. (2002). Regulation of the Globin Genes. *Pediatric Research*, 51(4), 415-421. <https://doi.org/10.1203/00006450-200204000-00003>
- Cao, H., Stamatoyannopoulos, G., & Jung, M. (2004). Induction of human gamma globin gene expression by histone deacetylase inhibitors. *Blood*, 103(2), 701-709. <https://doi.org/10.1182/blood-2003-02-0478>

- Cappellini, M. D., Cohen, A., Eleftheriou, A., Piga, A., Porter, J., & Taher, A. (2008). In *Guidelines for the Clinical Management of Thalassaemia* (2nd Revised ed.). <https://www.ncbi.nlm.nih.gov/pubmed/24308075>
- Capriotti, E., & Fariselli, P. (2017). PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. *Nucleic Acids Res*, 45(W1), W247-W252. <https://doi.org/10.1093/nar/gkx369>
- Cario, H. (2018). Hemoglobinopathies—genetically diverse, clinically complex, and globally relevant. *Memo - Magazine of European medical oncology*, 11(3), 235-240. <https://doi.org/10.1007/s12254-018-0402-4>
- Carraway, K. L., Perez, A., Idris, N., Jepson, S., Arango, M., Komatsu, M., Haq, B., Price-Schiavi, S. A., Zhang, J., & Carraway, C. A. (2002). Muc4/sialomucin complex, the intramembrane ErbB2 ligand, in cancer and epithelia: to protect and to survive. *Prog Nucleic Acid Res Mol Biol*, 71, 149-185. [https://doi.org/10.1016/s0079-6603\(02\)71043-x](https://doi.org/10.1016/s0079-6603(02)71043-x)
- Carrocini, G. C., Zamaro, P. J., & Bonini-Domingos, C. R. (2011). What influences Hb fetal production in adulthood? *Rev Bras Hematol Hemoter*, 33(3), 231-236. <https://doi.org/10.5581/1516-8484.20110059>
- Carron, J., Torricelli, C., Silva, J. K., Liu, Y., Pellegrino, R., Lima, C. S. P., & Lourenco, G. J. (2022). Association of Inherited Copy Number Variation in ADAM3A and ADAM5 Pseudogenes with Oropharynx Cancer Risk and Outcome. *Genes (Basel)*, 13(12). <https://doi.org/10.3390/genes13122408>
- Castro, A., Ozturk, K., Pyke, R. M., Xian, S., Zanetti, M., & Carter, H. (2019). Elevated neoantigen levels in tumors with somatic mutations in the HLA-A, HLA-B, HLA-C and B2M genes. *BMC Medical Genomics*, 12(6), 107. <https://doi.org/10.1186/s12920-019-0544-1>
- Cavazzana, M., Antoniani, C., & Miccio, A. (2017). Gene Therapy for β -Hemoglobinopathies. *Molecular therapy : the journal of the American Society of Gene Therapy*, 25. <https://doi.org/10.1016/j.ymthe.2017.03.024>
- Ceccaldi, R., Sarangi, P., & D'Andrea, A. D. (2016). The Fanconi anaemia pathway: new players and new functions. *Nature Reviews Molecular Cell Biology*, 17(6), 337-349. <https://doi.org/10.1038/nrm.2016.48>
- Chai, C. (2019). Principle of Emulsion PCR and Its Applications in Biotechnology. *Journal of Animal Reproduction and Biotechnology*, 34, 259-266. <https://doi.org/10.12750/JARB.34.4.259>
- Chaibunruang, A., Srivorakun, H., Fucharoen, S., Fucharoen, G., Sae-ung, N., & Sanchaisuriya, K. (2010). Interactions of hemoglobin Lepore ($\delta\beta$ hybrid hemoglobin) with various hemoglobinopathies: A molecular and hematological characteristics and differential diagnosis. *Blood Cells, Molecules, and Diseases*, 44(3), 140-145. <https://doi.org/https://doi.org/10.1016/j.bcmed.2009.11.008>
- Chaichompoo, P., Qillah, A., Sirankapracha, P., Kaewchuchuen, J., Rimthong, P., Paiboonsukwong, K., Fucharoen, S., Svasti, S., & Worawichawong, S. (2019). Abnormal red blood cell morphological changes in thalassaemia associated with iron overload and oxidative stress. *J Clin Pathol*, 72(8), 520-524. <https://doi.org/10.1136/ijclinpath-2019-205775>
- Charoenkwan, P., Natesirinilkul, R., Choeyprasert, W., Kulsumritpon, N., & Sangiamporn, O. (2017). Coinheritance of Hereditary Elliptocytosis and Deletional Hemoglobin H Disease. *J Pediatr Hematol Oncol*, 39(2), e69-e70. <https://doi.org/10.1097/mph.0000000000000750>
- Chatterjee, S., & Pal, J. K. (2009). Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biol Cell*, 101(5), 251-262. <https://doi.org/10.1042/BC20080104>
- Chen, H., Lowrey, C. H., & Stamatoyannopoulos, G. (1997). Analysis of enhancer function of the HS-40 core sequence of the human alpha-globin cluster. *Nucleic Acids Res*, 25(14), 2917-2922. <https://doi.org/10.1093/nar/25.14.2917>

- Chen, H., & Pugh, B. F. (2021). What do Transcription Factors Interact With? *J Mol Biol*, 433(14), 166883. <https://doi.org/10.1016/j.jmb.2021.166883>
- Chen, J. K., Xin, X. Q., & Huang, J. G. (2018). A Novel beta-Thalassemia Mutation in a Chinese family: IVS-II-203-205 (TCT>CC) (HBB: c.315+203TCT>CC). *Hemoglobin*, 42(3), 159-160. <https://doi.org/10.1080/03630269.2018.1499524>
- Chen, M., Zhang, J., & Manley, J. L. (2010). Turning on a fuel switch of cancer: hnRNP proteins regulate alternative splicing of pyruvate kinase mRNA. *Cancer Res*, 70(22), 8977-8980. <https://doi.org/10.1158/0008-5472.Can-10-2513>
- Chen, W., Zhang, X., Shang, X., Cai, R., Li, L., Zhou, T., Sun, M., Xiong, F., & Xu, X. (2010). The molecular basis of beta-thalassemia intermedia in southern China: genotypic heterogeneity and phenotypic diversity. *BMC Medical Genetics*, 11(1), 31. <http://www.biomedcentral.com/1471-2350/11/31>
- Cholera, R., Brittain, N. J., Gillrie, M. R., Lopera-Mesa, T. M., Diakite, S. A., Arie, T., Krause, M. A., Guindo, A., Tubman, A., Fujioka, H., Diallo, D. A., Doumbo, O. K., Ho, M., Wellems, T. E., & Fairhurst, R. M. (2008). Impaired cytoadherence of Plasmodium falciparum-infected erythrocytes containing sickle hemoglobin. *Proc Natl Acad Sci U S A*, 105(3), 991-996. <https://doi.org/10.1073/pnas.0711401105>
- Chou, K. C. (1989). Low-frequency resonance and cooperativity of hemoglobin. *Trends Biochem Sci*, 14(6), 212-213. <http://www.ncbi.nlm.nih.gov/pubmed/2763333>
- Cil, B., Ayyildiz, H., & Tuncer, T. (2020). Discrimination of beta-thalassemia and iron deficiency anemia through extreme learning machine and regularized extreme learning machine based decision support system. *Med Hypotheses*, 138, 109611. <https://doi.org/10.1016/j.mehy.2020.109611>
- Cividalli, G., Nathan, D. G., Kan, Y. W., Santamarina, B., & Frigoletto, F. (1974). Relation of beta to gamma synthesis during the first trimester: an approach to prenatal diagnosis of thalassemia. *Pediatr Res*, 8(5), 553-560. <https://doi.org/10.1203/00006450-197405000-00004>
- Clarke, G. M., & Higgins, T. N. (2000). Laboratory investigation of hemoglobinopathies and thalassemias: review and update. *Clin Chem*, 46(8 Pt 2), 1284-1290. <https://www.ncbi.nlm.nih.gov/pubmed/10926923>
- Colaco, S., Colah, R., & Nadkarni, A. (2022). Significance of borderline HbA2 levels in β thalassemia carrier screening. *Scientific Reports*, 12(1), 5414. <https://doi.org/10.1038/s41598-022-09250-5>
- Colombatti, R., Birkegård, C., & Medici, M. (2022). PB2215: GLOBAL EPIDEMIOLOGY OF SICKLE CELL DISEASE: A SYSTEMATIC LITERATURE REVIEW. *HemaSphere*, 6. https://journals.lww.com/hemasphere/fulltext/2022/06003/pb2215_global_epidemiology_of_sickle_cell.2085.aspx
- 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(6121), 819-823. <https://doi.org/10.1126/science.1231143>
- Corrons, Casafont, L. B., & Frasnado, E. F. (2021a). Concise review: how do red blood cells born, live, and die? *Ann Hematol*, 100(10), 2425-2433. <https://doi.org/10.1007/s00277-021-04575-z>
- Corrons, Casafont, L. B., & Frasnado, E. F. (2021b). Concise review: how do red blood cells born, live, and die? *Annals of Hematology*, 100(10), 2425-2433. <https://doi.org/10.1007/s00277-021-04575-z>
- Crosby, W. H. (1957). Siderocytes and the spleen. *Blood*, 12(2), 165-170. <https://www.ncbi.nlm.nih.gov/pubmed/13403981>
- Da Costa, L., Galimand, J., Fenneteau, O., & Mohandas, N. (2013). Hereditary spherocytosis, elliptocytosis, and other red cell membrane disorders. *Blood Rev*, 27(4), 167-178. <https://doi.org/10.1016/j.blre.2013.04.003>
- Dailey, H. A., & Meissner, P. N. (2013). Erythroid heme biosynthesis and its disorders. *Cold Spring Harb Perspect Med*, 3(4), a011676. <https://doi.org/10.1101/cshperspect.a011676>

- Daniels, D. E., Downes, D. J., Ferrer-Vicens, I., Ferguson, D. C. J., Singleton, B. K., Wilson, M. C., Trakarnsanga, K., Kurita, R., Nakamura, Y., Anstee, D. J., & Frayne, J. (2020). Comparing the two leading erythroid lines BEL-A and HUDEP-2. *Haematologica*, *105*(8), e389-e394. <https://doi.org/10.3324/haematol.2019.229211>
- Danjou, F., Anni, F., Perseu, L., Satta, S., Dessi, C., Lai, M. E., Fortina, P., Devoto, M., & Galanello, R. (2012). Genetic modifiers of beta-thalassemia and clinical severity as assessed by age at first transfusion. *Haematologica*, *97*(7), 989-993. <https://doi.org/10.3324/haematol.2011.053504>
- de Back, D. Z., Kostova, E. B., van Kraaij, M., van den Berg, T. K., & van Bruggen, R. (2014). Of macrophages and red blood cells; a complex love story. *Front Physiol*, *5*, 9. <https://doi.org/10.3389/fphys.2014.00009>
- De Gruchy, G. C., Santamaria, J. N., Parsons, I. C., & Crawford, H. (1960). Nonspherocytic congenital hemolytic anemia. *Blood*, *16*, 1371-1397. <https://www.ncbi.nlm.nih.gov/pubmed/13720546>
- De, S. K. (2024). Mitapivat: A Novel Treatment of Hemolytic Anemia in Adults with Pyruvate Kinase Deficiency. *Curr Med Chem*, *31*(6), 683-687. <https://doi.org/10.2174/0929867330666230430002709>
- Debailleul, V., Laine, A., Huet, G., Mathon, P., d'Hooghe, M. C., Aubert, J. P., & Porchet, N. (1998). Human mucin genes MUC2, MUC3, MUC4, MUC5AC, MUC5B, and MUC6 express stable and extremely large mRNAs and exhibit a variable length polymorphism. An improved method to analyze large mRNAs. *J Biol Chem*, *273*(2), 881-890. <https://doi.org/10.1074/jbc.273.2.881>
- Decorsière, A., Mueller, H., van Breugel, P. C., Abdul, F., Gerossier, L., Beran, R. K., Livingston, C. M., Niu, C., Fletcher, S. P., Hantz, O., & Strubin, M. (2016). Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor. *Nature*, *531*(7594), 386-389. <https://doi.org/10.1038/nature17170>
- Defrise-Gussenhoven, E. (1962). Hypothèses de dimérisation et de non-pénétrance. *Human Heredity*, *12*(1), 65-96. <https://doi.org/10.1159/000151184>
- den Hollander, A. I., Koenekoop, R. K., Yzer, S., Lopez, I., Arends, M. L., Voesenek, K. E., Zonneveld, M. N., Strom, T. M., Meitinger, T., Brunner, H. G., Hoyng, C. B., van den Born, L. I., Rohrschneider, K., & Cremers, F. P. (2006). Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *Am J Hum Genet*, *79*(3), 556-561. <https://doi.org/10.1086/507318>
- Deng, Z., Yang, F., Bai, Y., He, L., Li, Q., Wu, Y., Luo, L., Li, H., Ma, L., Yang, Z., He, Y., & Cui, L. (2017). Co-inheritance of glucose-6-phosphate dehydrogenase deficiency mutations and hemoglobin E in a Kachin population in a malaria-endemic region of Southeast Asia. *PLoS One*, *12*(5), e0177917. <https://doi.org/10.1371/journal.pone.0177917>
- Di Maggio, R., & Maggio, A. (2017). The new era of chelation treatments: effectiveness and safety of 10 different regimens for controlling iron overloading in thalassaemia major. *British Journal of Haematology*, *178*(5), 676-688. <https://doi.org/10.1111/bjh.14712>
- Diaz, M., & Pecinka, A. (2018). Scaffolding for Repair: Understanding Molecular Functions of the SMC5/6 Complex. *Genes*, *9*(1), 36. <https://www.mdpi.com/2073-4425/9/1/36>
- DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res*, *41*(7), 4336-4343. <https://doi.org/10.1093/nar/gkt135>
- Duan, Z.-J., Fang, X., Rohde, A., Han, H., Stamatoyannopoulos, G., & Li, Q. (2002). Developmental specificity of recruitment of TBP to the TATA box of the human γ -globin gene. *Proceedings of the National Academy of Sciences*, *99*(8), 5509-5514. <https://doi.org/10.1073/pnas.072084499>
- Dumoulin, A., Padovan, J. C., Manning, L. R., Popowicz, A., Winslow, R. M., Chait, B. T., & Manning, J. M. (1998). The N-terminal sequence affects distant helix interactions in hemoglobin. Implications for mutant proteins from studies on recombinant hemoglobin felix. *J Biol Chem*, *273*(52), 35032-35038. <https://doi.org/10.1074/jbc.273.52.35032>

- Dvořák Tomaščíková, E., Prochazkova, K., Yang, F., Jemelkova, J., Finke, A., Dorn, A., Said, M., Puchta, H., & Pecinka, A. (2023). SMC5/6 complex-mediated SUMOylation stimulates DNA–protein cross-link repair in Arabidopsis. *The Plant Cell*, 35(5), 1532-1547. <https://doi.org/10.1093/plcell/koad020>
- Dzierzak, E., & Philipsen, S. (2013). Erythropoiesis: development and differentiation. *Cold Spring Harb Perspect Med*, 3(4), a011601. <https://doi.org/10.1101/cshperspect.a011601>
- Eaton, W. A. (2020). Hemoglobin S polymerization and sickle cell disease: A retrospective on the occasion of the 70th anniversary of Pauling's Science paper. *Am J Hematol*, 95(2), 205-211. <https://doi.org/10.1002/ajh.25687>
- Edemir, B. (2020). Identification of Prognostic Organic Cation and Anion Transporters in Different Cancer Entities by In Silico Analysis. *Int J Mol Sci*, 21(12). <https://doi.org/10.3390/ijms21124491>
- Efremov, D. G., Dimovski, A. J., & Efremov, G. D. (1991). Detection of beta-thalassemia mutations by ASO hybridization of PCR amplified DNA with digoxigenin ddUTP labeled oligonucleotides. *Hemoglobin*, 15(6), 525-533. <https://doi.org/10.3109/03630269109027900>
- Efremov, G. D. (2007). Dominantly Inherited β -Thalassemia. *Hemoglobin*, 31(2), 193-207. <https://doi.org/10.1080/03630260701290092>
- Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., & Proudfoot, N. J. (1980). The structure and evolution of the human beta-globin gene family. *Cell*, 21(3), 653-668.
- El-Hazmi, M. A. (2004). The natural history and the national pre-marital screening program in Saudi Arabia. *Saudi Med J*, 25(11), 1549-1554. <http://www.ncbi.nlm.nih.gov/pubmed/15573176>
- El-Hazmi, M. A., Al-Hazmi, A. M., & Warsy, A. S. (2011). Sickle cell disease in Middle East Arab countries. *Indian J Med Res*, 134(5), 597-610. <https://doi.org/10.4103/0971-5916.90984>
- El-Hazmi, M. A., & Lehmann, H. (1976). Hemoglobin Riyadh--alpha2beta2 (120(GH3)Lys replaced by Asn). A new variant found in association with alpha-thalassemia and iron deficiency. *Hemoglobin*, 1(1), 59-74.
- Entezari, S., Haghi, S. M., Norouzkhani, N., Sahebazar, B., Vosoughian, F., Akbarzadeh, D., Islampanah, M., Naghsh, N., Abbasalizadeh, M., & Deravi, N. (2022). Iron Chelators in Treatment of Iron Overload. *J Toxicol*, 2022, 4911205. <https://doi.org/10.1155/2022/4911205>
- Etheridge, T. J., Villahermosa, D., Campillo-Funollet, E., Herbert, A. D., Irmisch, A., Watson, A. T., Dang, H. Q., Osborne, M. A., Oliver, A. W., Carr, A. M., & Murray, J. M. (2021). Live-cell single-molecule tracking highlights requirements for stable Smc5/6 chromatin association in vivo. *eLife*, 10, e68579. <https://doi.org/10.7554/eLife.68579>
- Fabry, M. E., Suzuka, S. M., Weinberg, R. S., Lawrence, C., Factor, S. M., Gilman, J. G., Costantini, F., & Nagel, R. L. (2001). Second generation knockout sickle mice: the effect of HbF. *Blood*, 97(2), 410-418. <http://www.ncbi.nlm.nih.gov/pubmed/11154217>
- Fan, A. X., Hossain, M. A., Stees, J., Gavrilo, E., & Bungert, J. (2015). Chapter 11 - Regulation of erythroid cell differentiation by transcription factors, chromatin structure alterations, and noncoding RNA. In S. Huang, M. D. Litt, & C. A. Blakey (Eds.), *Epigenetic Gene Expression and Regulation* (pp. 237-264). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-799958-6.00011-1>
- Farashi, S., Bayat, N., Faramarzi Garous, N., Ashki, M., Montajabi Niat, M., Vakili, S., Imanian, H., Zeinali, S., Najmabadi, H., & Azarkeivan, A. (2015). Interaction of an alpha-Globin Gene Triplication with beta-Globin Gene Mutations in Iranian Patients with beta-Thalassemia Intermedia. *Hemoglobin*, 39(3), 201-206. <https://doi.org/10.3109/03630269.2015.1027914>
- Farashi, S., & Hartevel, C. L. (2018). Molecular basis of alpha-thalassemia. *Blood Cells Mol Dis*, 70, 43-53. <https://doi.org/10.1016/j.bcmd.2017.09.004>

- Farmakis, D., Giakoumis, A., Angastiniotis, M., & Eleftheriou, A. (2020). The changing epidemiology of the ageing thalassaemia populations: A position statement of the Thalassaemia International Federation. *Eur J Haematol*, *105*(1), 16-23. <https://doi.org/10.1111/ejh.13410>
- Farooqui, S. M., Ward, R., & Aziz, M. (2023). Shwachman-Diamond Syndrome. In *StatPearls*. <https://www.ncbi.nlm.nih.gov/pubmed/29939643>
- Fasano, R. M., & Chou, S. T. (2016). Red Blood Cell Antigen Genotyping for Sickle Cell Disease, Thalassemia, and Other Transfusion Complications. *Transfus Med Rev*, *30*(4), 197-201. <https://doi.org/10.1016/j.tmr.2016.05.011>
- Feng, L., Gell, D. A., Zhou, S., Gu, L., Kong, Y., Li, J., Hu, M., Yan, N., Lee, C., Rich, A. M., Armstrong, R. S., Lay, P. A., Gow, A. J., Weiss, M. J., Mackay, J. P., & Shi, Y. (2004). Molecular mechanism of AHSP-mediated stabilization of alpha-hemoglobin. *Cell*, *119*(5), 629-640. <https://doi.org/10.1016/j.cell.2004.11.025>
- Fierro, R., Herrera Mora, P., Zenteno, J. C., & Villarroel Cortes, C. E. (2020). Clinical and molecular evidence of possible digenic inheritance for MFN2/GDAP1 genes in Charcot-Marie-Tooth disease. *Neuromuscul Disord*, *30*(12), 986-990. <https://doi.org/10.1016/j.nmd.2020.10.003>
- Finotti, A., Breda, L., Lederer, C. W., Bianchi, N., Zuccato, C., Kleanthous, M., Rivella, S., & Gambari, R. (2015). Recent trends in the gene therapy of beta-thalassemia. *J Blood Med*, *6*, 69-85. <https://doi.org/10.2147/JBM.S46256>
- Ford, J. (2013). Red blood cell morphology. *Int J Lab Hematol*, *35*(3), 351-357. <https://doi.org/10.1111/ijlh.12082>
- Forget, B. G. (1998). Molecular basis of hereditary persistence of fetal hemoglobin. *Ann N Y Acad Sci*, *850*, 38-44. <http://www.ncbi.nlm.nih.gov/pubmed/9668525>
- Forget, B. G., & Bunn, H. F. (2013). Classification of the disorders of hemoglobin. *Cold Spring Harb Perspect Med*, *3*(2), a011684. <https://doi.org/10.1101/cshperspect.a011684>
- Forth, L. F., & Hoper, D. (2019). Highly efficient library preparation for Ion Torrent sequencing using Y-adapters. *Biotechniques*, *67*(5), 229-237. <https://doi.org/10.2144/btn-2019-0035>
- Fox, A. H., Liew, C., Holmes, M., Kowalski, K., Mackay, J., & Crossley, M. (1999). Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. *Embo j*, *18*(10), 2812-2822. <https://doi.org/10.1093/emboj/18.10.2812>
- Frésard, L., & Montgomery, S. B. (2018). Diagnosing rare diseases after the exome. *Cold Spring Harb Mol Case Stud*, *4*(6). <https://doi.org/10.1101/mcs.a003392>
- Fritsch, E. F., Lawn, R. M., & Maniatis, T. (1980). Molecular cloning and characterization of the human beta-like globin gene cluster. *Cell*, *19*(4), 959-972. [https://doi.org/10.1016/0092-8674\(80\)90087-2](https://doi.org/10.1016/0092-8674(80)90087-2)
- Fucharoen, S., & Weatherall, D. J. (2012). The hemoglobin E thalassemsias. *Cold Spring Harb Perspect Med*, *2*(8). <https://doi.org/10.1101/cshperspect.a011734>
- Gabriel, A. P., J. (2010). Sickle-Cell Anemia: A Look at Global Haplotype Distribution. *Nature Education*, *3*(3):2.
- Gach, A., Pinkier, I., Wysocka, U., Salacinska, K., Salachna, D., Szarras-Czapnik, M., Pietrzyk, A., Sakowicz, A., Nykel, A., Rutkowska, L., Rybak-Krzyszowska, M., Socha, M., Jamsheer, A., & Jakubowski, L. (2022). New findings in oligogenic inheritance of congenital hypogonadotropic hypogonadism. *Arch Med Sci*, *18*(2), 353-364. <https://doi.org/10.5114/aoms.2020.98909>
- Gallati, S. (2014). Disease-modifying genes and monogenic disorders: experience in cystic fibrosis. *App Clin Genet*, *7*, 133-146. <https://doi.org/10.2147/tacg.S18675>
- Gallienne, A. E., Dreau, H. M., Schuh, A., Old, J. M., & Henderson, S. (2012). Ten novel mutations in the erythroid transcription factor KLF1 gene associated with increased fetal hemoglobin levels in adults. *Haematologica*, *97*(3), 340-343. <https://doi.org/10.3324/haematol.2011.055442>

- Gao, X.-P., Dong, J.-J., Xie, T., & Guan, X. (2021). Integrative Analysis of MUC4 to Prognosis and Immune Infiltration in Pan-Cancer: Friend or Foe? [Original Research]. *Frontiers in Cell and Developmental Biology*, 9. <https://doi.org/10.3389/fcell.2021.695544>
- Gazzo, A. M., Daneels, D., Cilia, E., Bonduelle, M., Abramowicz, M., Van Dooren, S., Smits, G., & Lenaerts, T. (2016). DIDA: A curated and annotated digenic diseases database. *Nucleic Acids Res*, 44(D1), D900-907. <https://doi.org/10.1093/nar/gkv1068>
- Genin, E., Feingold, J., & Clerget-Darpoux, F. (2008). Identifying modifier genes of monogenic disease: strategies and difficulties. *Hum Genet*, 124(4), 357-368. <https://doi.org/10.1007/s00439-008-0560-2>
- Génin, E., Feingold, J., & Clerget-Darpoux, F. (2008). Identifying modifier genes of monogenic disease: strategies and difficulties. *Hum Genet*, 124(4), 357-368. <https://doi.org/10.1007/s00439-008-0560-2>
- Ghaedi, M., & Niklason, L. E. (2019). Human Pluripotent Stem Cells (iPSC) Generation, Culture, and Differentiation to Lung Progenitor Cells. *Methods Mol Biol*, 1576, 55-92. https://doi.org/10.1007/7651_2016_11
- Ghosh, A., Garee, G., Sweeny, E. A., Nakamura, Y., & Stuehr, D. J. (2018). Hsp90 chaperones hemoglobin maturation in erythroid and nonerythroid cells. *Proceedings of the National Academy of Sciences*, 115(6), E1117-E1126. <https://doi.org/doi:10.1073/pnas.1717993115>
- Giannopoulou, E., Bartsakoulia, M., Tafrali, C., Kourakli, A., Poulas, K., Stavrou, E. F., Papachatzopoulou, A., Georgitsi, M., & Patrinos, G. P. (2012). A single nucleotide polymorphism in the HBBP1 gene in the human beta-globin locus is associated with a mild beta-thalassemia disease phenotype. *Hemoglobin*, 36(5), 433-445. <https://doi.org/10.3109/03630269.2012.717515>
- Gianotten, J., van der Veen, F., Alders, M., Leschot, N. J., Tanck, M. W., Land, J. A., Kremer, J. A., Hoefsloot, L. H., Mannens, M. M., Lombardi, M. P., & Hoffer, M. J. (2003). Chromosomal region 11p15 is associated with male factor subfertility. *Mol Hum Reprod*, 9(10), 587-592. <https://doi.org/10.1093/molehr/gag081>
- Gifford, S. C., Derganc, J., Shevkoplyas, S. S., Yoshida, T., & Bitensky, M. W. (2006). A detailed study of time-dependent changes in human red blood cells: from reticulocyte maturation to erythrocyte senescence. *Br J Haematol*, 135(3), 395-404. <https://doi.org/10.1111/j.1365-2141.2006.06279.x>
- Goh, S. H., Lee, Y. T., Bhanu, N. V., Cam, M. C., Desper, R., Martin, B. M., Moharram, R., Gherman, R. B., & Miller, J. L. (2005). A newly discovered human alpha-globin gene. *Blood*, 106(4), 1466-1472. <https://doi.org/10.1182/blood-2005-03-0948>
- Grace, R. F., & Barcellini, W. (2020). Management of pyruvate kinase deficiency in children and adults. *Blood*, 136(11), 1241-1249. <https://doi.org/https://doi.org/10.1182/blood.2019000945>
- Grace, R. F., Bianchi, P., van Beers, E. J., Eber, S. W., Glader, B., Yaish, H. M., Despotovic, J. M., Rothman, J. A., Sharma, M., McNaull, M. M., Fermo, E., Lezon-Geyda, K., Morton, D. H., Neufeld, E. J., Chonat, S., Kollmar, N., Knoll, C. M., Kuo, K., Kwiatkowski, J. L., . . . Barcellini, W. (2018). Clinical spectrum of pyruvate kinase deficiency: data from the Pyruvate Kinase Deficiency Natural History Study. *Blood*, 131(20), 2183-2192. <https://doi.org/10.1182/blood-2017-10-810796>
- Grammatikopoulos, T., Hadzic, N., Foskett, P., Strautnieks, S., Samyn, M., Vara, R., Dhawan, A., Hertecant, J., Al Jasmi, F., Rahman, O., Deheragoda, M., Bull, L. N., & Thompson, R. J. (2022). Liver Disease and Risk of Hepatocellular Carcinoma in Children With Mutations in TALDO1. *Hepatol Commun*, 6(3), 473-479. <https://doi.org/10.1002/hep4.1824>
- Grange, L. J., Reynolds, J. J., Ullah, F., Isidor, B., Shearer, R. F., Latypova, X., Baxley, R. M., Oliver, A. W., Ganesh, A., Cooke, S. L., Jhujh, S. S., McNee, G. S., Hollingworth, R., Higgs, M. R., Natsume, T., Khan, T., Martos-Moreno, G. Á., Chupp, S., Mathew, C. G., . . . Stewart, G. S. (2022). Pathogenic variants in SLF2 and SMC5 cause segmented chromosomes and mosaic variegated hyperploidy. *Nat Commun*, 13(1), 6664. <https://doi.org/10.1038/s41467-022-34349-8>

- Granick, S., & Levere, R. D. (1964). Heme Synthesis in Erythroid Cells. *Prog Hematol*, 4, 1-47. <https://www.ncbi.nlm.nih.gov/pubmed/14272797>
- Gregory, C. J., & Eaves, A. C. (1977). Human marrow cells capable of erythropoietic differentiation in vitro: definition of three erythroid colony responses. *Blood*, 49(6), 855-864. <https://www.ncbi.nlm.nih.gov/pubmed/861374>
- Grosso, M., Sessa, R., Puzone, S., Storino, M., & Izzo, P. (2012). Molecular Basis of Thalassemia. In. <https://doi.org/10.5772/31362>
- Guitart, J. R., Johnson, J. L., & Chien, W. W. (2016). Research Techniques Made Simple: The Application of CRISPR-Cas9 and Genome Editing in Investigative Dermatology. *Journal of Investigative Dermatology*, 136(9), e87-e93. <https://doi.org/https://doi.org/10.1016/j.jid.2016.06.007>
- Gupta, D., Sturtevant, S., Vieira, B., Nakamura, Y., Krishnamoorthy, S., & Demers, M. (2019). Characterization of a Genetically Engineered HUDEP2 Cell Line Harboring a Sickle Cell Disease Mutation As a Potential Research Tool for Preclinical Sickle Cell Disease Drug Discovery. *Blood*, 134(Supplement_1), 3559-3559. <https://doi.org/10.1182/blood-2019-127330>
- Gutiérrez, L., Caballero, N., Fernández-Calleja, L., Karkoulia, E., & Strouboulis, J. (2020). Regulation of GATA1 levels in erythropoiesis. *IUBMB Life*, 72(1), 89-105. <https://doi.org/https://doi.org/10.1002/iub.2192>
- Gwozdzinski, K., Pieniasek, A., & Gwozdzinski, L. (2021). Reactive Oxygen Species and Their Involvement in Red Blood Cell Damage in Chronic Kidney Disease. *Oxid Med Cell Longev*, 2021, 6639199. <https://doi.org/10.1155/2021/6639199>
- Hahn, E. V., & Gillespie, E. B. (1927). Sickle cell anemia: report of a case greatly improved by splenectomy. Experimental study of sickle cell formation. *Archives of Internal Medicine*, 39(2), 233-254.
- Hamid, M., Keikhaei, B., Galehdari, H., Saberi, A., Sedaghat, A., Shariati, G., & Mohammadi-Anaei, M. (2021). Alpha-globin gene triplication and its effect in beta-thalassemia carrier, sickle cell trait, and healthy individual. *EJHaem*, 2(3), 366-374. <https://doi.org/10.1002/jha2.262>
- Hand, C. K., & Rouleau, G. A. (2002). Familial amyotrophic lateral sclerosis. *Muscle Nerve*, 25(2), 135-159. <https://doi.org/10.1002/mus.10001>
- Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., Wang, L., Lupyan, D., Dahlgren, M. K., Knight, J. L., Kaus, J. W., Cerutti, D. S., Krilov, G., Jorgensen, W. L., Abel, R., & Friesner, R. A. (2016). OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J Chem Theory Comput*, 12(1), 281-296. <https://doi.org/10.1021/acs.jctc.5b00864>
- Hardison. (2012a). Evolution of hemoglobin and its genes. *Cold Spring Harb Perspect Med*, 2(12), a011627. <https://doi.org/10.1101/cshperspect.a011627>
- Hardison. (2012b). Evolution of Hemoglobin and Its Genes. *Cold Spring Harbor Perspectives in Medicine*, 2(12). <https://doi.org/10.1101/cshperspect.a011627>
- Harewood, J., & Azevedo, A. M. (2023). Alpha Thalassemia. In *StatPearls*. <https://www.ncbi.nlm.nih.gov/pubmed/28722856>
- Harrington, A. M., Ward, P. C., & Kroft, S. H. (2008). Iron deficiency anemia, beta-thalassemia minor, and anemia of chronic disease: a morphologic reappraisal. *Am J Clin Pathol*, 129(3), 466-471. <https://doi.org/10.1309/LY7YLUPE7551JYBG>
- Harris, S., Barrie, P. A., Weiss, M. L., & Jeffreys, A. J. (1984). The primate psi beta 1 gene. An ancient beta-globin pseudogene. *J Mol Biol*, 180(4), 785-801. [https://doi.org/10.1016/0022-2836\(84\)90257-2](https://doi.org/10.1016/0022-2836(84)90257-2)
- Harteveld, C. L., Achour, A., Arkesteijn, S. J. G., ter Huurne, J., Verschuren, M., Bhagwandien-Bisoen, S., Schaap, R., Vijfhuizen, L., el Idrissi, H., & Koopmann, T. T. (2022). The hemoglobinopathies, molecular disease mechanisms and diagnostics. *Int J Lab Hematol*, 44(S1), 28-36. <https://doi.org/https://doi.org/10.1111/ijlh.13885>

- Hassan, Vossen, R. H., Chessa, R., den Dunnen, J. T., Bakker, E., Giordano, P. C., & Hartevelde, C. L. (2014). Molecular diagnostics of the HBB gene in an Omani cohort using bench-top DNA Ion Torrent PGM technology. *Blood Cells Mol Dis*, 53(3), 133-137. <https://doi.org/10.1016/j.bcmd.2014.05.002>
- Hassan, AlMuslahi, Al Riyami, M., Bakker, E., Hartevelde, C. L., & Giordano, P. C. (2014). Sickle cell anemia and alpha-thalassemia: a modulating factor in homozygous HbS/S patients in Oman. *Eur J Med Genet*, 57(11-12), 603-606. <https://doi.org/10.1016/j.ejmg.2014.09.005>
- Hattangadi, S. M., Wong, P., Zhang, L., Flygare, J., & Lodish, H. F. (2011). From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. *Blood*, 118(24), 6258-6268. <https://doi.org/10.1182/blood-2011-07-356006>
- Hazelwood, L. F. (2001). *Can't live without it : the story of hemoglobin in sickness and in health*. Nova Science Publishers.
- He, J., Song, W., Yang, J., Lu, S., Yuan, Y., Guo, J., Zhang, J., Ye, K., Yang, F., Long, F., Peng, Z., Yu, H., Cheng, L., & Zhu, B. (2017). Next-generation sequencing improves thalassemia carrier screening among premarital adults in a high prevalence population: the Dai nationality, China. *Genetics in Medicine*, 19(9), 1022-1031. <https://doi.org/10.1038/gim.2016.218>
- He, Z., & Russell, J. E. (2001). Expression, purification, and characterization of human hemoglobins Gower-1 (zeta(2)epsilon(2)), Gower-2 (alpha(2)epsilon(2)), and Portland-2 (zeta(2)beta(2)) assembled in complex transgenic-knockout mice. *Blood*, 97(4), 1099-1105. <https://doi.org/10.1182/blood.v97.4.1099>
- Herrick, J. B. (1910). PECULIAR ELONGATED AND SICKLE-SHAPED RED BLOOD CORPUSCLES IN A CASE OF SEVERE ANEMIA. *Archives of Internal Medicine*, VI(5), 517-521. <https://doi.org/10.1001/archinte.1910.00050330050003>
- Hidayati, N. I., Wijayanti, N., & Handayani, N. S. N. (2020). Detection of HBB:c.92+5G>C and HBB:c.108delC mutations in beta-thalassemia carriers using high-resolution melting analysis. *Mol Biol Rep*, 47(7), 5665-5671. <https://doi.org/10.1007/s11033-020-05625-x>
- Hoban, M. D., Orkin, S. H., & Bauer, D. E. (2016). Genetic treatment of a molecular disorder: gene therapy approaches to sickle cell disease. *Blood*, 127(7), 839-848. <https://doi.org/10.1182/blood-2015-09-618587>
- Hoefele, J., Wolf, M. T., O'Toole, J. F., Otto, E. A., Schultheiss, U., Deschenes, G., Attanasio, M., Utsch, B., Antignac, C., & Hildebrandt, F. (2007). Evidence of oligogenic inheritance in nephronophthisis. *J Am Soc Nephrol*, 18(10), 2789-2795. <https://doi.org/10.1681/ASN.2007020243>
- Holland, P. M., Abramson, R. D., Watson, R., & Gelfand, D. H. (1991). Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc Natl Acad Sci U S A*, 88(16), 7276-7280. <https://doi.org/10.1073/pnas.88.16.7276>
- Holwerda, S., & De Laat, W. (2012). Chromatin loops, gene positioning, and gene expression [Review]. *Frontiers in Genetics*, 3. <https://doi.org/10.3389/fgene.2012.00217>
- Hossain, M. S., Roy, A. S., & Islam, M. S. (2020). In silico analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions. *Scientific Reports*, 10(1), 14542. <https://doi.org/10.1038/s41598-020-71457-1>
- Hou, H., Lu, Q., Li, D., Wu, S., Chu, X., Xu, G., & Hu, S. (2018). Quantitative Proteomic Profiling Identifies Differentially Regulated Pathways in Fanconi Anemia and Aplastic Anemia. *Blood*, 132, 5100. <https://doi.org/https://doi.org/10.1182/blood-2018-99-116347>
- Hsiao, V., & Jim, V. (2008). Controlling α -globin: a review of α -globin expression and its impact on β -thalassemia. *Haematologica*, 93(12), 1868-1876. <https://doi.org/10.3324/haematol.13490>
- Hu, X. (2016). CRISPR/Cas9 system and its applications in human hematopoietic cells. *Blood Cells Mol Dis*, 62, 6-12. <https://doi.org/10.1016/j.bcmd.2016.09.003>

- Huang, Z., Hearne, L., Irby, C. E., King, S. B., Ballas, S. K., & Kim-Shapiro, D. B. (2003). Kinetics of Increased Deformability of Deoxygenated Sickle Cells upon Oxygenation. *Biophysical Journal*, 85(4), 2374-2383. [https://doi.org/http://dx.doi.org/10.1016/S0006-3495\(03\)74661-X](https://doi.org/http://dx.doi.org/10.1016/S0006-3495(03)74661-X)
- Hung, S. Y., Lin, C. C., Hsu, C. L., Yao, C. Y., Wang, Y. H., Tsai, C. H., Hou, H. A., Chou, W. C., & Tien, H. F. (2021). The expression levels of long non-coding RNA KIAA0125 are associated with distinct clinical and biological features in myelodysplastic syndromes. *Br J Haematol*, 192(3), 589-598. <https://doi.org/10.1111/bjh.17231>
- Hwang, W. Y. F., Yanfang; Reyon, Deepak; Maeder, Morgan L.; Tsai, Shengdar Q.; Sander, Jeffrey D.; Peterson, Randall T.; Yeh, J.-R. Joanna; Joung, J. Keith. (2013). Efficient In Vivo Genome Editing Using RNA-Guided Nucleases. *Nature biotechnology*, 31((3): 227-229.). <https://doi.org/doi:10.1038/nbt.2501>
- Ingram, V. M. (1956). A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature*, 178(4537), 792-794.
- Ingram, V. M. (1957). The sulphhydryl groups of sickle-cell haemoglobin. *Biochem J*, 65(4), 760-763. <https://doi.org/10.1042/bj0650760>
- Iolascon, A., Andolfo, I., & Russo, R. (2019). Advances in understanding the pathogenesis of red cell membrane disorders. *Br J Haematol*, 187(1), 13-24. <https://doi.org/10.1111/bjh.16126>
- Jackson, D. A., McDowell, J. C., & Dean, A. (2003). Beta-globin locus control region HS2 and HS3 interact structurally and functionally. *Nucleic Acids Res*, 31(4), 1180-1190. <https://doi.org/10.1093/nar/gkg217>
- Jacq, C., Miller, J. R., & Brownlee, G. G. (1977). A pseudogene structure in 5S DNA of *Xenopus laevis*. *Cell*, 12(1), 109-120. [https://doi.org/10.1016/0092-8674\(77\)90189-1](https://doi.org/10.1016/0092-8674(77)90189-1)
- Jaing, T.-H., Chang, T.-Y., Chen, S.-H., Lin, C.-W., Wen, Y.-C., & Chiu, C.-C. (2021). Molecular genetics of β -thalassemia: A narrative review. *Medicine*, 100(45). https://journals.lww.com/md-journal/fulltext/2021/11120/molecular_genetics_of_thalassemia_a_narrative.4.aspx
- Jankovic, L., Efremov, G. D., Josifovska, O., Juricic, D., Stoming, T. A., Kutlar, A., & Huisman, T. H. (1990). An initiation codon mutation as a cause of a beta-thalassemia. *Hemoglobin*, 14(2), 169-176. <https://doi.org/10.3109/03630269009046958>
- Jaskiewicz, E., Jodłowska, M., Kaczmarek, R., & Zerka, A. (2019). Erythrocyte glycoporphins as receptors for Plasmodium merozoites. *Parasit Vectors*, 12(1), 317. <https://doi.org/10.1186/s13071-019-3575-8>
- Jastaniah, W. (2011). Epidemiology of sickle cell disease in Saudi Arabia. *Ann Saudi Med*, 31(3), 289-293. <https://doi.org/10.4103/0256-4947.81540>
- Jatavan, P., Chattipakorn, N., & Tongsong, T. (2018). Fetal hemoglobin Bart's hydrops fetalis: pathophysiology, prenatal diagnosis and possibility of intrauterine treatment. *J Matern Fetal Neonatal Med*, 31(7), 946-957. <https://doi.org/10.1080/14767058.2017.1301423>
- Jensen, W. N., Bromberg, P. A., & Bessis, M. C. (1967). Microincision of sickled erythrocytes by a laser beam. *Science*, 155(3763), 704-707. <https://doi.org/10.1126/science.155.3763.704>
- Ji, P., Murata-Hori, M., & Lodish, H. F. (2011). Formation of mammalian erythrocytes: chromatin condensation and enucleation. *Trends Cell Biol*, 21(7), 409-415. <https://doi.org/10.1016/j.tcb.2011.04.003>
- Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., & Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res*, 41(20), e188. <https://doi.org/10.1093/nar/gkt780>
- Jiao, B., Johnson, K. M., Ramsey, S. D., Bender, M. A., Devine, B., & Basu, A. (2023). Long-term survival with sickle cell disease: a nationwide cohort study of Medicare and Medicaid beneficiaries. *Blood Advances*, 7(13), 3276-3283. <https://doi.org/10.1182/bloodadvances.2022009202>

- 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(6096), 816-821. <https://doi.org/10.1126/science.1225829>
- Joenje, H., & Patel, K. J. (2001). The emerging genetic and molecular basis of Fanconi anaemia. *Nature Reviews Genetics*, 2(6), 446-458. <https://doi.org/10.1038/35076590>
- Joshi, D. D., Nickerson, H. J., & McManus, M. J. (2004). Hydrops fetalis caused by homozygous alpha-thalassemia and Rh antigen alloimmunization: report of a survivor and literature review. *Clin Med Res*, 2(4), 228-232. <https://doi.org/10.3121/cmr.2.4.228>
- Ju, A. P., Jiang, F., Li, J., Tang, X. W., & Li, D. Z. (2020). Detection of an alpha-Globin Fusion Gene Using Real-Time Polymerase Chain Reaction-Based Multicolor Melting Curve. *Hemoglobin*, 44(6), 427-431. <https://doi.org/10.1080/03630269.2020.1838923>
- Kajiwarra, K., Berson, E. L., & Dryja, T. P. (1994). Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science*, 264(5165), 1604-1608. <https://doi.org/10.1126/science.8202715>
- Kanagal-Shamanna, R. (2016). Emulsion PCR: Techniques and Applications. *Methods Mol Biol*, 1392, 33-42. https://doi.org/10.1007/978-1-4939-3360-0_4
- Karamzade, A., Mirzapour, H., Hoseinzade, M., Asadi, S., Gholamrezapour, T., Tavakoli, P., & Salehi, M. (2014). alpha-Globin gene mutations in Isfahan Province, Iran. *Hemoglobin*, 38(3), 161-164. <https://doi.org/10.3109/03630269.2014.893531>
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., . . . Genome Aggregation Database, C. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809), 434-443. <https://doi.org/10.1038/s41586-020-2308-7>
- Karczewski, K. J., Weisburd, B., Thomas, B., Solomonson, M., Ruderfer, D. M., Kavanagh, D., Hamamsy, T., Lek, M., Samocha, K. E., Cummings, B. B., Birnbaum, D., Daly, M. J., & MacArthur, D. G. (2017). The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res*, 45(D1), D840-d845. <https://doi.org/10.1093/nar/gkw971>
- Karg, A., Dinç, Z. A., Başok, O., & Uçvet, A. (2006). MUC4 expression and its relation to ErbB2 expression, apoptosis, proliferation, differentiation, and tumor stage in non-small cell lung cancer (NSCLC). *Pathol Res Pract*, 202(8), 577-583. <https://doi.org/10.1016/j.prp.2006.04.002>
- Kasai, S., Mimura, J., Ozaki, T., & Itoh, K. (2018). Emerging Regulatory Role of Nrf2 in Iron, Heme, and Hemoglobin Metabolism in Physiology and Disease. *Front Vet Sci*, 5, 242. <https://doi.org/10.3389/fvets.2018.00242>
- Katsumura, K. R., & Bresnick, E. H. (2017). The GATA factor revolution in hematology. *Blood*, 129(15), 2092-2102. <https://doi.org/10.1182/blood-2016-09-687871>
- 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(9), a015412. <https://doi.org/10.1101/cshperspect.a015412>
- Kattamis, A., Forni, G. L., Aydinok, Y., & Viprakasit, V. (2020). Changing patterns in the epidemiology of β -thalassemia. *Eur J Haematol*, 105(6), 692-703. <https://doi.org/10.1111/ejh.13512>
- Kaufman, D. P., Khattar, J., & Lappin, S. L. (2023). Physiology, Fetal Hemoglobin. In *StatPearls*. <https://www.ncbi.nlm.nih.gov/pubmed/29763187>
- Kaul, D. K., Fabry, M. E., Suzuka, S. M., & Zhang, X. (2013). Antisickling fetal hemoglobin reduces hypoxia-inducible factor-1alpha expression in normoxic sickle mice: microvascular implications. *Am J Physiol Heart Circ Physiol*, 304(1), H42-50. <https://doi.org/10.1152/ajpheart.00296.2012>
- Kaur Gharial, P. (2022). Haemoglobinopathies. *InnovAiT*, 15(5), 294-300. <https://doi.org/10.1177/17557380221078615>

- Kawasaki, E., Saiki, R., & Erlich, H. (1993). Genetic analysis using polymerase chain reaction-amplified DNA and immobilized oligonucleotide probes: reverse dot-blot typing. *Methods Enzymol*, *218*, 369-381. [https://doi.org/10.1016/0076-6879\(93\)18029-c](https://doi.org/10.1016/0076-6879(93)18029-c)
- Kay, J. G., & Fairn, G. D. (2019). Distribution, dynamics and functional roles of phosphatidylserine within the cell. *Cell Communication and Signaling*, *17*(1), 126. <https://doi.org/10.1186/s12964-019-0438-z>
- Kazazian, H. H., Jr., & Woodhead, A. P. (1973). Hemoglobin A synthesis in the developing fetus. *N Engl J Med*, *289*(2), 58-62. <https://doi.org/10.1056/NEJM197307122890202>
- Kedar, P. S., Harigae, H., Ito, E., Muramatsu, H., Kojima, S., Okuno, Y., Fujiwara, T., Dongerdiye, R., Warang, P. P., & Madkaikar, M. R. (2019). Study of pathophysiology and molecular characterization of congenital anemia in India using targeted next-generation sequencing approach. *Int J Hematol*, *110*(5), 618-626. <https://doi.org/10.1007/s12185-019-02716-9>
- Keller, A., Zhuang, H., Chi, Q., Vosshall, L. B., & Matsunami, H. (2007). Genetic variation in a human odorant receptor alters odour perception. *Nature*, *449*(7161), 468-472. <https://doi.org/10.1038/nature06162>
- Khan, A., & Rehman, A. U. (2022). Laboratory Evaluation of Beta Thalassemia. In *StatPearls*. <https://www.ncbi.nlm.nih.gov/pubmed/36251822>
- Kho, S. L., Chua, K. H., George, E., & Tan, J. A. (2013). High throughput molecular confirmation of beta-thalassemia mutations using novel TaqMan probes. *Sensors (Basel)*, *13*(2), 2506-2514. <https://doi.org/10.3390/s130202506>
- 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*(6890), 758-763. <https://doi.org/10.1038/nature00803>
- Kilander, M. B. C., Wang, C. H., Chang, C. H., Nestor, J. E., Herold, K., Tsai, J. W., Nestor, M. W., & Lin, Y. C. (2018). A rare human CEP290 variant disrupts the molecular integrity of the primary cilium and impairs Sonic Hedgehog machinery. *Sci Rep*, *8*(1), 17335. <https://doi.org/10.1038/s41598-018-35614-x>
- Kim, A., & Dean, A. (2012). Chromatin loop formation in the β -globin locus and its role in globin gene transcription. *Mol Cells*, *34*(1), 1-5. <https://doi.org/10.1007/s10059-012-0048-8>
- Kim, S. Y., Lee, S. H., Cho, S. I., Song, S. H., Hattori, Y., Park, S.-K., Song, J., Choi, Y., Yang, M., Park, H., Kim, S.-R., Seong, M.-W., Kim, J. Y., Cho, H.-I., & Park, S. S. (2010). Molecular identification of the novel G γ - β hybrid hemoglobin: Hb G γ - β Ulsan (G γ through 13; β from 19). *Blood Cells, Molecules, and Diseases*, *45*(4), 276-279. <https://doi.org/https://doi.org/10.1016/j.bcmed.2010.10.001>
- Kim, Y., Park, J., & Kim, M. (2017). Diagnostic approaches for inherited hemolytic anemia in the genetic era. *Blood Res*, *52*(2), 84-94. <https://doi.org/10.5045/br.2017.52.2.84>
- King, A. J., Songdej, D., Downes, D. J., Beagrie, R. A., Liu, S., Buckley, M., Hua, P., Suci, M. C., Marieke Oudelaar, A., Hanssen, L. L. P., Jeziorska, D., Roberts, N., Carpenter, S. J., Francis, H., Telenius, J., Olijnik, A.-A., Sharpe, J. A., Sloane-Stanley, J., Eglinton, J., . . . Babbs, C. (2021). Reactivation of a developmentally silenced embryonic globin gene. *Nat Commun*, *12*(1), 4439. <https://doi.org/10.1038/s41467-021-24402-3>
- Kingsley, P. D., Malik, J., Emerson, R. L., Bushnell, T. P., McGrath, K. E., Bloedorn, L. A., Bulger, M., & Palis, J. (2006). "Maturation" globin switching in primary primitive erythroid cells. *Blood*, *107*(4), 1665-1672. <https://doi.org/10.1182/blood-2005-08-3097>
- Kirkham, J. K., Estepp, J. H., Weiss, M. J., & Rashkin, S. R. (2023). Genetic Variation and Sickle Cell Disease Severity: A Systematic Review and Meta-Analysis. *JAMA Network Open*, *6*(10), e2337484-e2337484. <https://doi.org/10.1001/jamanetworkopen.2023.37484>
- Kohne, E. (2011). Hemoglobinopathies: clinical manifestations, diagnosis, and treatment. *Dtsch Arztebl Int*, *108*(31-32), 532-540. <https://doi.org/10.3238/arztebl.2011.0532>

- Komar, A. A., Kommer, A., Krashennikov, I. A., & Spirin, A. S. (1997). Cotranslational folding of globin. *J Biol Chem*, 272(16), 10646-10651. <https://doi.org/10.1074/jbc.272.16.10646>
- Kong, Y., Zhou, S., Kihm, A. J., Katein, A. M., Yu, X., Gell, D. A., Mackay, J. P., Adachi, K., Foster-Brown, L., Loudon, C. S., Gow, A. J., & Weiss, M. J. (2004). Loss of alpha-hemoglobin-stabilizing protein impairs erythropoiesis and exacerbates beta-thalassemia. *J Clin Invest*, 114(10), 1457-1466. <https://doi.org/10.1172/jci21982>
- Koonin, E. V., & Makarova, K. S. (2009). CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000 Biol Rep*, 1, 95. <https://doi.org/10.3410/B1-95>
- Koralkova, P., van Solinge, W. W., & van Wijk, R. (2014). Rare hereditary red blood cell enzymopathies associated with hemolytic anemia - pathophysiology, clinical aspects, and laboratory diagnosis. *Int J Lab Hematol*, 36(3), 388-397. <https://doi.org/10.1111/ijlh.12223>
- Körber, C., Wölfler, A., Neubauer, M., & Robier, C. (2017). Red blood cell morphology in patients with β -thalassemia minor. *LaboratoriumsMedizin*, 41(1), 49-52. <https://doi.org/doi:10.1515/labmed-2016-0052>
- Korber, E. (1866). Über Differenzen des Blufarbstoffs. *Inaugural dissertation. Universität Dorpat*.
- Kozłowski, P., Jasinska, A. J., & Kwiatkowski, D. J. (2008). New applications and developments in the use of multiplex ligation-dependent probe amplification. *Electrophoresis*, 29(23), 4627-4636. <https://doi.org/10.1002/elps.200800126>
- Krüger, F. (1888). Über die ungleiche Resistenz des Blutfarbstoffes verschiedener Tiere gegen zersetzende Agentien. *Zeitschr. f. Biol*, 24.
- Kuijpers, T. W., Alders, M., Tool, A. T., Mellink, C., Roos, D., & Hennekam, R. C. (2005). Hematologic abnormalities in Shwachman Diamond syndrome: lack of genotype-phenotype relationship. *Blood*, 106(1), 356-361. <https://doi.org/10.1182/blood-2004-11-4371>
- Kulozik, A. E., Thein, S. L., Wainscoat, J. S., Gale, R., Kay, L. A., Wood, J. K., Weatherall, D. J., & Huehns, E. R. (1987). Thalassaemia intermedia: interaction of the triple alpha-globin gene arrangement and heterozygous beta-thalassaemia. *Br J Haematol*, 66(1), 109-112. <https://doi.org/10.1111/j.1365-2141.1987.tb06898.x>
- Kumar, V., & Robbins, S. L. (2007). *Robbins basic pathology* (8th ed.). Saunders/Elsevier.
- Kurosaki, T., Popp, M. W., & Maquat, L. E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat Rev Mol Cell Biol*, 20(7), 406-420. <https://doi.org/10.1038/s41580-019-0126-2>
- Kuypers, F. A., Yuan, J., Lewis, R. A., Snyder, L. M., Kiefer, C. R., Bunyaratvej, A., Fucharoen, S., Ma, L., Styles, L., de Jong, K., & Schrier, S. L. (1998). Membrane phospholipid asymmetry in human thalassemia. *Blood*, 91(8), 3044-3051. <https://www.ncbi.nlm.nih.gov/pubmed/9531618>
- Kwaifa, I., Lai, M. I., & Md Noor, S. (2020). Non-deletional alpha thalassaemia: a review. *Orphanet J Rare Dis*, 15(1), 166. <https://doi.org/10.1186/s13023-020-01429-1>
- Labie, D., Pagnier, J., Lapoumeroulie, C., Rouabhi, F., Dunda-Belkhdja, O., Chardin, P., Beldjord, C., Wajcman, H., Fabry, M. E., & Nagel, R. L. (1985). Common haplotype dependency of high G gamma-globin gene expression and high Hb F levels in beta-thalassemia and sickle cell anemia patients. *Proc Natl Acad Sci U S A*, 82(7), 2111-2114. <https://doi.org/10.1073/pnas.82.7.2111>
- Lafuente-Barquero, J., Luke-Glaser, S., Graf, M., Silva, S., Gomez-Gonzalez, B., Lockhart, A., Lisby, M., Aguilera, A., & Luke, B. (2017). The Smc5/6 complex regulates the yeast Mph1 helicase at RNA-DNA hybrid-mediated DNA damage. *PLoS Genet*, 13(12), e1007136. <https://doi.org/10.1371/journal.pgen.1007136>
- Lam, V. M. S., Xie, S. S., Tam, J. W. O., Woo, Y. K., Gu, Y. L., & Li, A. M. C. (1990). A NEW SINGLE NUCLEOTIDE CHANGE AT THE INITIATION CODON (ATG → AGG) IDENTIFIED IN AMPLIFIED GENOMIC DNA OF A CHINESE β -THALASSEMIC PATIENT. *Blood*, 75(5), 1207-1208. <https://doi.org/https://doi.org/10.1182/blood.V75.5.1207.1207>

- Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A., McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., . . . Maglott, D. R. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*, *46*(D1), D1062-d1067. <https://doi.org/10.1093/nar/gkx1153>
- Lang, E., & Lang, F. (2015). Mechanisms and pathophysiological significance of eryptosis, the suicidal erythrocyte death. *Semin Cell Dev Biol*, *39*, 35-42. <https://doi.org/10.1016/j.semcdb.2015.01.009>
- Lang, F., Lang, E., & Foller, M. (2012). Physiology and pathophysiology of eryptosis. *Transfus Med Hemother*, *39*(5), 308-314. <https://doi.org/10.1159/000342534>
- Langer, A. L. (1993). Beta-Thalassemia. In M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. W. Gripp, & A. Amemiya (Eds.), *GeneReviews*((R)). <https://www.ncbi.nlm.nih.gov/pubmed/20301599>
- Langer, A. L., & Esrick, E. B. (2021). β -Thalassemia: evolving treatment options beyond transfusion and iron chelation. *Hematology*, *2021*(1), 600-606. <https://doi.org/10.1182/hematology.2021000313>
- Laššuthová, P., Vill, K., Erdem-Ozdamar, S., Schröder, J. M., Topaloglu, H., Horvath, R., Müller-Felber, W., Bansagi, B., Schlotter-Weigel, B., Gläser, D., Neupauerová, J., Sedláčková, L., Staněk, D., Mazanec, R., Weis, J., Seeman, P., & Senderek, J. (2018). Novel SBF2 mutations and clinical spectrum of Charcot-Marie-Tooth neuropathy type 4B2. *Clin Genet*, *94*(5), 467-472. <https://doi.org/10.1111/cge.13417>
- Laurentino, M. R., Almeida, T. P., Maia, P. A., Nascimento, F. O. F., Advincula, A. F., Machado, C. M. G., & Lemes, R. P. G. (2019). Clinical difference between identical twins with sickle cell anemia. *Jornal Brasileiro de Patologia e Medicina Laboratorial*, *55*.
- Lawn, R. M., Efstratiadis, A., O'Connell, C., & Maniatis, T. (1980). The nucleotide sequence of the human beta-globin gene. *Cell*, *21*(3), 647-651. <http://www.ncbi.nlm.nih.gov/pubmed/6254664>
- Lecine, P., Blank, V., & Shivdasani, R. (1998). Characterization of the Hematopoietic Transcription Factor NF-E2 in Primary Murine Megakaryocytes*. *Journal of Biological Chemistry*, *273*(13), 7572-7578. <https://doi.org/https://doi.org/10.1074/jbc.273.13.7572>
- Lee, J. S., Cho, S. I., Park, S. S., & Seong, M. W. (2021). Molecular basis and diagnosis of thalassemia. *Blood Res*, *56*(S1), S39-S43. <https://doi.org/10.5045/br.2021.2020332>
- Lee, J. S., Rhee, T. M., Jeon, K., Cho, Y., Lee, S. W., Han, K. D., Seong, M. W., Park, S. S., & Lee, Y. K. (2022). Epidemiologic Trends of Thalassemia, 2006-2018: A Nationwide Population-Based Study. *J Clin Med*, *11*(9). <https://doi.org/10.3390/jcm11092289>
- Lee, S. H., Crocker, P. R., Westaby, S., Key, N., Mason, D. Y., Gordon, S., & Weatherall, D. J. (1988). Isolation and immunocytochemical characterization of human bone marrow stromal macrophages in hemopoietic clusters. *J Exp Med*, *168*(3), 1193-1198. <https://doi.org/10.1084/jem.168.3.1193>
- Lehmann, H., Maranjian, G., & Mourant, A. E. (1963). Distribution of Sickle-cell Hæmoglobin in Saudi Arabia. *Nature*, *198*(4879), 492-493. <https://doi.org/10.1038/198492b0>
- Leimberg, M. J., Prus, E., Konijn, A. M., & Fibach, E. (2008). Macrophages function as a ferritin iron source for cultured human erythroid precursors. *J Cell Biochem*, *103*(4), 1211-1218. <https://doi.org/10.1002/jcb.21499>
- Lenglet, H., Schmitt, C., Grange, T., Manceau, H., Karboul, N., Bouchet-Crivat, F., Robreau, A. M., Nicolas, G., Lamoril, J., Simonin, S., Mirmiran, A., Karim, Z., Casalino, E., Deybach, J. C., Puy, H., Peoc'h, K., & Gouya, L. (2018). From a dominant to an oligogenic model of inheritance with environmental modifiers in acute intermittent porphyria. *Hum Mol Genet*, *27*(7), 1164-1173. <https://doi.org/10.1093/hmg/ddy030>

- Lerer, I., Sagi, M., Ben-Neriah, Z., Wang, T., Levi, H., & Abeliovich, D. (2001). A deletion mutation in GJB6 cooperating with a GJB2 mutation in trans in non-syndromic deafness: A novel founder mutation in Ashkenazi Jews. *Hum Mutat*, 18(5), 460. <https://doi.org/10.1002/humu.1222>
- Lettre, G. (2012). The search for genetic modifiers of disease severity in the beta-hemoglobinopathies. *Cold Spring Harb Perspect Med*, 2(10). <https://doi.org/10.1101/cshperspect.a015032>
- 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(33), 11869-11874. <https://doi.org/10.1073/pnas.0804799105>
- Levings, P. P., & Bungert, J. (2002). The human β -globin locus control region. *European Journal of Biochemistry*, 269(6), 1589-1599. <https://doi.org/10.1046/j.1432-1327.2002.02797.x>
- Lewis, B. A., & Orkin, S. H. (1995). A functional initiator element in the human beta-globin promoter. *J Biol Chem*, 270(47), 28139-28144. <https://doi.org/10.1074/jbc.270.47.28139>
- Li, B., Kamarck, M. L., Peng, Q., Lim, F. L., Keller, A., Smeets, M. A. M., Mainland, J. D., & Wang, S. (2022). From musk to body odor: Decoding olfaction through genetic variation. *PLoS Genet*, 18(2), e1009564. <https://doi.org/10.1371/journal.pgen.1009564>
- Li, H., & Lykotrafitis, G. (2014). Erythrocyte membrane model with explicit description of the lipid bilayer and the spectrin network. *Biophys J*, 107(3), 642-653. <https://doi.org/10.1016/j.bpj.2014.06.031>
- Li, M., Nishio, S. Y., Naruse, C., Riddell, M., Sapski, S., Katsuno, T., Hikita, T., Mizapourshafiyi, F., Smith, F. M., Cooper, L. T., Lee, M. G., Asano, M., Boettger, T., Krueger, M., Wietelmann, A., Graumann, J., Day, B. W., Boyd, A. W., Offermanns, S., . . . Nakayama, M. (2020). Digenic inheritance of mutations in EPHA2 and SLC26A4 in Pendred syndrome. *Nat Commun*, 11(1), 1343. <https://doi.org/10.1038/s41467-020-15198-9>
- Li, W., Guo, R., Song, Y., & Jiang, Z. (2021). Erythroblastic Island Macrophages Shape Normal Erythropoiesis and Drive Associated Disorders in Erythroid Hematopoietic Diseases [Review]. *Frontiers in Cell and Developmental Biology*, 8. <https://doi.org/10.3389/fcell.2020.613885>
- Li, W., Wang, Y., Chen, L., & An, X. (2019). Erythroblast island macrophages: recent discovery and future perspectives. *Blood Sci*, 1(1), 61-64. <https://doi.org/10.1097/bs9.0000000000000017>
- Li, W., Wang, Y., Zhao, H., Zhang, H., Xu, Y., Wang, S., Guo, X., Huang, Y., Zhang, S., Han, Y., Wu, X., Rice, C. M., Huang, G., Gallagher, P. G., Mendelson, A., Yazdanbakhsh, K., Liu, J., Chen, L., & An, X. (2019). Identification and transcriptome analysis of erythroblastic island macrophages. *Blood*, 134(5), 480-491. <https://doi.org/10.1182/blood.2019000430>
- Liang, P., Xu, Y., Zhang, X., Ding, C., Huang, R., Zhang, Z., Lv, J., Xie, X., Chen, Y., Li, Y., Sun, Y., Bai, Y., Songyang, Z., Ma, W., Zhou, C., & Huang, J. (2015). CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein Cell*, 6(5), 363-372. <https://doi.org/10.1007/s13238-015-0153-5>
- Liebhaber, S. A., Cash, F. E., & Ballas, S. K. (1986). Human alpha-globin gene expression. The dominant role of the alpha 2-locus in mRNA and protein synthesis. *J Biol Chem*, 261(32), 15327-15333. <https://www.ncbi.nlm.nih.gov/pubmed/3771577>
- Liebhaber, S. A., Goossens, M. J., & Kan, Y. W. (1980). Cloning and complete nucleotide sequence of human 5'-alpha-globin gene. *Proceedings of the National Academy of Sciences*, 77(12), 7054-7058. <http://www.pnas.org/content/77/12/7054.abstract>
- Liebhaber, S. A., Wang, Z., Cash, F. E., Monks, B., & Russell, J. E. (1996). Developmental silencing of the embryonic zeta-globin gene: concerted action of the promoter and the 3'-flanking region combined with stage-specific silencing by the transcribed segment. *Mol Cell Biol*, 16(6), 2637-2646. <https://doi.org/10.1128/mcb.16.6.2637>

- Lim, W. F., Muniandi, L., George, E., Sathar, J., Teh, L. K., & Lai, M. I. (2015). HbF in HbE/ β -thalassemia: A clinical and laboratory correlation. *Hematology*, 20(6), 349-353. <https://doi.org/10.1179/1607845414Y.0000000203>
- Lin, X.-M., Jiang, F., Li, J., & Li, D.-Z. (2022). Dominantly Inherited β -Thalassemia Caused by a Single Nucleotide Deletion in Exon 3 of the β -Globin Gene: Hb Xiangyang (HBB: c.393delT). *Hemoglobin*, 46(4), 253-255. <https://doi.org/10.1080/03630269.2022.2072325>
- Lipsick, J. S. (2010). The *C-MYB* story; is it definitive? *Proceedings of the National Academy of Sciences*, 107(40), 17067-17068. <https://doi.org/doi:10.1073/pnas.1012402107>
- Liu, W., Palovcak, A., Li, F., Zafar, A., Yuan, F., & Zhang, Y. (2020). Fanconi anemia pathway as a prospective target for cancer intervention. *Cell & Bioscience*, 10(1), 39. <https://doi.org/10.1186/s13578-020-00401-7>
- Loggetto, S. R. (2013). Sickle cell anemia: clinical diversity and beta S-globin haplotypes. *Rev Bras Hematol Hemoter*, 35(3), 155-157. <https://doi.org/10.5581/1516-8484.20130048>
- Lord, J., & Baralle, D. (2021). Splicing in the Diagnosis of Rare Disease: Advances and Challenges. *Front Genet*, 12, 689892. <https://doi.org/10.3389/fgene.2021.689892>
- Lowrey, C. H., Bodine, D. M., & Nienhuis, A. W. (1992). Mechanism of DNase I hypersensitive site formation within the human globin locus control region. *Proc Natl Acad Sci U S A*, 89(3), 1143-1147. <https://doi.org/10.1073/pnas.89.3.1143>
- Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*, 48(D1), D265-d268. <https://doi.org/10.1093/nar/gkz991>
- Lubeck, D., Agodoa, I., Bhakta, N., Danese, M., Pappu, K., Howard, R., Gleeson, M., Halperin, M., & Lanzkron, S. (2019). Estimated Life Expectancy and Income of Patients With Sickle Cell Disease Compared With Those Without Sickle Cell Disease. *JAMA Netw Open*, 2(11), e1915374. <https://doi.org/10.1001/jamanetworkopen.2019.15374>
- Luo, H., Zou, Y., & Liu, Y. (2020). A Novel beta-Thalassemia Mutation [IVS-I-6 (T>G), HBB: c.92+6T>G] in a Chinese Family. *Hemoglobin*, 44(1), 55-57. <https://doi.org/10.1080/03630269.2020.1714648>
- Luo, X., Zhang, X. M., Wu, L. S., Chen, J., & Chen, Y. (2021). Prevalence and clinical phenotype of the triplicated alpha-globin genes and its ethnic and geographical distribution in Guizhou of China. *BMC Med Genomics*, 14(1), 97. <https://doi.org/10.1186/s12920-021-00944-9>
- Lux, S. E. (1979). Dissecting the red cell membrane skeleton. *Nature*, 281(5731), 426-428. <https://doi.org/10.1038/281426a0>
- Lux, S. E., IV. (2016). Anatomy of the red cell membrane skeleton: unanswered questions. *Blood*, 127(2), 187-199. <https://doi.org/10.1182/blood-2014-12-512772>
- Ma, Liu, S., Gao, J., Chen, C., Zhang, X., Yuan, H., Chen, Z., Yin, X., Sun, C., Mao, Y., Zhou, F., Shao, Y., Liu, Q., Xu, J., Cheng, L., Yu, D., Li, P., Yi, P., He, J., . . . Yu, J. (2021). Genome-wide analysis of pseudogenes reveals HBBP1's human-specific essentiality in erythropoiesis and implication in beta-thalassemia. *Dev Cell*, 56(4), 478-493 e411. <https://doi.org/10.1016/j.devcel.2020.12.019>
- Ma, S., Deng, X., Liao, L., Deng, Z., Qiu, Y., Wei, H., & Lin, F. (2018). Analysis of the causes of the misdiagnosis of hereditary spherocytosis. *Oncol Rep*, 40(3), 1451-1458. <https://doi.org/10.3892/or.2018.6578>
- Ma, Xi, H. R., Gao, X. X., Yang, J. M., Kurita, R., Nakamura, Y., Song, X. M., Chen, H. Y., & Lu, D. R. (2021). Long noncoding RNA HBBP1 enhances gamma-globin expression through the ETS transcription factor ELK1. *Biochem Biophys Res Commun*, 552, 157-163. <https://doi.org/10.1016/j.bbrc.2021.03.051>
- MacDonald, R. (1977). Red cell 2,3-diphosphoglycerate and oxygen affinity. *Anaesthesia*, 32(6), 544-553. <https://doi.org/10.1111/j.1365-2044.1977.tb10002.x>

- Maggio, A., Giambona, A., Cai, S. P., Wall, J., Kan, Y. W., & Chehab, F. F. (1993). Rapid and simultaneous typing of hemoglobin S, hemoglobin C, and seven Mediterranean beta-thalassemia mutations by covalent reverse dot-blot analysis: application to prenatal diagnosis in Sicily. *Blood*, *81*(1), 239-242. <https://www.ncbi.nlm.nih.gov/pubmed/8417793>
- Magrinelli, F., & Lohmann, K. (2022). PRKRAP1 and Other Pseudogenes in Movement Disorders: The Troublemakers in Genetic Analyses Are More Than Genomic Fossils. *Mov Disord Clin Pract*, *9*(5), 698-702. <https://doi.org/10.1002/mdc3.13499>
- Makkawi, M., Alasmari, S., Hawan, A. A., Shahrani, M. M. A., & Dera, A. A. (2021). Hemoglobinopathies: An update on the prevalence trends in Southern Saudi Arabia. *Saudi Med J*, *42*(7), 784-789. <https://doi.org/10.15537/smj.2021.42.7.20210273>
- Malik, J., Kim, A. R., Tyre, K. A., Cherukuri, A. R., & Palis, J. (2013). Erythropoietin critically regulates the terminal maturation of murine and human primitive erythroblasts. *Haematologica*, *98*(11), 1778-1787. <https://doi.org/10.3324/haematol.2013.087361>
- Malli, T., Rammer, M., Haslinger, S., Burghofer, J., Burgstaller, S., Boesmueller, H.-C., Marschon, R., Kranewitter, W., Erdel, M., Deutschbauer, S., & Webersinke, G. (2018). Overexpression of the proneural transcription factor ASCL1 in chronic lymphocytic leukemia with a t(12;14)(q23.2;q32.3). *Molecular Cytogenetics*, *11*(1), 3. <https://doi.org/10.1186/s13039-018-0355-7>
- Mallik, N., Das, R., Malhotra, P., & Sharma, P. (2021). Congenital erythrocytosis. *Eur J Haematol*, *107*(1), 29-37. <https://doi.org/10.1111/ejh.13632>
- Malnic, B., Godfrey, P. A., & Buck, L. B. (2004). The human olfactory receptor gene family. *Proc Natl Acad Sci U S A*, *101*(8), 2584-2589. <https://doi.org/10.1073/pnas.0307882100>
- Manchinu, M. F., Simbula, M., Caria, C. A., Musu, E., Perseu, L., Porcu, S., Steri, M., Poddie, D., Frau, J., Cocco, E., Manunza, L., Barella, S., & Ristaldi, M. S. (2020). Delta-Globin Gene Expression Is Enhanced in vivo by Interferon Type I [Brief Research Report]. *Frontiers in Medicine*, *7*. <https://doi.org/10.3389/fmed.2020.00163>
- Manco, L., Ribeiro, M. L., Maximo, V., Almeida, H., Costa, A., Freitas, O., Barbot, J., Abade, A., & Tamagnini, G. (2000). A new PKLR gene mutation in the R-type promoter region affects the gene transcription causing pyruvate kinase deficiency. *Br J Haematol*, *110*(4), 993-997. <https://doi.org/10.1046/j.1365-2141.2000.02283.x>
- Maniatis, T., Fritsch, E. F., Lauer, J., & Lawn, R. M. (1980). The molecular genetics of human hemoglobins. *Annu Rev Genet*, *14*, 145-178. <https://doi.org/10.1146/annurev.ge.14.120180.001045>
- Manning, Russell, J. E., Padovan, J. C., Chait, B. T., Popowicz, A., Manning, R. S., & Manning, J. M. (2007a). Human embryonic, fetal, and adult hemoglobins have different subunit interface strengths. Correlation with lifespan in the red cell. *Protein Science*, *16*(8), 1641-1658. <https://doi.org/10.1110/ps.072891007>
- Manning, Russell, J. E., Padovan, J. C., Chait, B. T., Popowicz, A., Manning, R. S., & Manning, J. M. (2007b). Human embryonic, fetal, and adult hemoglobins have different subunit interface strengths. Correlation with lifespan in the red cell. *Protein Sci*, *16*(8), 1641-1658. <https://doi.org/10.1110/ps.072891007>
- Manning, J. M., Dumoulin, A., Li, X., & Manning, L. R. (1998). Normal and abnormal protein subunit interactions in hemoglobins. *J Biol Chem*, *273*(31), 19359-19362. <https://doi.org/10.1074/jbc.273.31.19359>
- Manning, L. R., Russell, J. E., Popowicz, A. M., Manning, R. S., Padovan, J. C., & Manning, J. M. (2009). Energetic differences at the subunit interfaces of normal human hemoglobins correlate with their developmental profile. *Biochemistry*, *48*(32), 7568-7574. <https://doi.org/10.1021/bi900857r>

- Mardis, E. R. (2013). Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif)*, 6, 287-303. <https://doi.org/10.1146/annurev-anchem-062012-092628>
- Marengo-Rowe, A. J. (2006). Structure-function relations of human hemoglobins. *Proc (Bayl Univ Med Cent)*, 19(3), 239-245. <http://www.ncbi.nlm.nih.gov/pubmed/17252042>
- Marengo-Rowe, A. J. (2007). The thalassemias and related disorders. *Proc (Bayl Univ Med Cent)*, 20(1), 27-31. <http://www.ncbi.nlm.nih.gov/pubmed/17256039>
- Mariani, R., Trombini, P., Pozzi, M., & Piperno, A. (2009). Iron metabolism in thalassemia and sickle cell disease. *Mediterr J Hematol Infect Dis*, 1(1), e2009006. <https://doi.org/10.4084/MJHID.2009.006>
- Marino, K., Mazzarella, L., & Cocc, E. (2023). Expression, isolation and characterisation of tetrameric haemoglobin as a model to study the Root effect.
- Marraffini, L. A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature*, 526(7571), 55-61. <https://doi.org/10.1038/nature15386>
- Martyn, G. E., Quinlan, K. G. R., & Crossley, M. (2017). The regulation of human globin promoters by CCAAT box elements and the recruitment of NF-Y. *Biochim Biophys Acta Gene Regul Mech*, 1860(5), 525-536. <https://doi.org/10.1016/j.bbaggm.2016.10.002>
- Massart, F., & Saggese, G. (2010). Morphogenetic targets and genetics of undescended testis. *Sex Dev*, 4(6), 326-335. <https://doi.org/10.1159/000321006>
- Masuda, T., Wang, X., Maeda, M., Canver, M. C., Sher, F., Funnell, A. P., Fisher, C., Suci, M., Martyn, G. E., Norton, L. J., Zhu, C., Kurita, R., Nakamura, Y., Xu, J., Higgs, D. R., Crossley, M., Bauer, D. E., Orkin, S. H., Kharchenko, P. V., & Maeda, T. (2016). Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. *Science*, 351(6270), 285-289. <https://doi.org/10.1126/science.aad3312>
- Mathias, L. A., Fisher, T. C., Zeng, L., Meiselman, H. J., Weinberg, K. I., Hiti, A. L., & Malik, P. (2000). Ineffective erythropoiesis in β -thalassemia major is due to apoptosis at the polychromatophilic normoblast stage. *Experimental Hematology*, 28(12), 1343-1353. [https://doi.org/https://doi.org/10.1016/S0301-472X\(00\)00555-5](https://doi.org/https://doi.org/10.1016/S0301-472X(00)00555-5)
- Mattevi, A., Valentini, G., Rizzi, M., Speranza, M. L., Bolognesi, M., & Coda, A. (1995). Crystal structure of Escherichia coli pyruvate kinase type I: molecular basis of the allosteric transition. *Structure*, 3(7), 729-741. [https://doi.org/10.1016/S0969-2126\(01\)00207-6](https://doi.org/10.1016/S0969-2126(01)00207-6)
- Maurer, M. H., Kohler, A., Hudemann, M., Jungling, J., Biskup, S., & Menzel, M. (2022). Case Report of a Juvenile Patient with Autism Spectrum Disorder with a Novel Combination of Copy Number Variants in ADGRL3 (LPHN3) and Two Pseudogenes. *Appl Clin Genet*, 15, 125-131. <https://doi.org/10.2147/TACG.S361239>
- May, A., & Forrester, L. M. (2020). The erythroblastic island niche: modeling in health, stress, and disease. *Experimental Hematology*, 91, 10-21. <https://doi.org/https://doi.org/10.1016/j.exphem.2020.09.185>
- McMahon, T. J., Shan, S., Riccio, D. A., Batchvarova, M., Zhu, H., Telen, M. J., & Zennadi, R. (2019). Nitric oxide loading reduces sickle red cell adhesion and vaso-occlusion in vivo. *Blood Adv*, 3(17), 2586-2597. <https://doi.org/10.1182/bloodadvances.2019031633>
- Memish, Z. A., Owaidah, T. M., & Saeedi, M. Y. (2011). Marked regional variations in the prevalence of sickle cell disease and beta-thalassemia in Saudi Arabia: findings from the premarital screening and genetic counseling program. *J Epidemiol Glob Health*, 1(1), 61-68. <https://doi.org/10.1016/j.jegh.2011.06.002>
- Memish, Z. A., & Saeedi, M. Y. (2011). Six-year outcome of the national premarital screening and genetic counseling program for sickle cell disease and beta-thalassemia in Saudi Arabia. *Ann Saudi Med*, 31(3), 229-235. <https://doi.org/10.4103/0256-4947.81527>
- 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(10), 1197-1199. <https://doi.org/10.1038/ng2108>
- Migliaccio, A. R. (2010). Erythroblast enucleation. *Haematologica*, 95(12), 1985-1988. <https://doi.org/10.3324/haematol.2010.033225>
- Migliaccio, A. R. (2020). GATA1 gets personal. *Haematologica*, 105(4), 852-854. <https://doi.org/10.3324/haematol.2019.246355>
- 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(5), 2776-2786. <https://doi.org/10.1128/mcb.13.5.2776-2786.1993>
- Minton, K. (2021). Red blood cells join the ranks as immune sentinels. *Nature Reviews Immunology*, 21(12), 760-761. <https://doi.org/10.1038/s41577-021-00648-2>
- Mirasena, S., Shimbhu, D., Sanguansersri, M., & Sanguansersri, T. (2008). Detection of beta-thalassemia mutations using a multiplex amplification refractory mutation system assay. *Hemoglobin*, 32(4), 403-409. <https://doi.org/10.1080/03630260701798391>
- Mobarra, N., Shanaki, M., Ehteram, H., Nasiri, H., Sahmani, M., Saeidi, M., Goudarzi, M., Pourkarim, H., & Azad, M. (2016). A Review on Iron Chelators in Treatment of Iron Overload Syndromes. *Int J Hematol Oncol Stem Cell Res*, 10(4), 239-247.
- Mohamed, D. S., Shaban, N. S., Labib, M. M., & Shehata, O. (2022). Sesame oil ameliorates valproic acid-induced hepatotoxicity in mice: integrated in vivo-in silico study. *J Biomol Struct Dyn*, 1-21. <https://doi.org/10.1080/07391102.2022.2135593>
- Mohandas, N., & An, X. (2012). Malaria and human red blood cells. *Med Microbiol Immunol*, 201(4), 593-598. <https://doi.org/10.1007/s00430-012-0272-z>
- Moiz, B., Nasir, A., Khan, S. A., Kherani, S. A., & Qadir, M. (2012). Neonatal Hyperbilirubinemia in infants with G6PD c.563C > T Variant. *BMC Pediatrics*, 12(1), 708. <https://doi.org/10.1186/1471-2431-12-126>
- Molchanova, T. P., Pobedimskaya, D. D., & Huisman, T. H. (1994). The differences in quantities of alpha 2- and alpha 1-globin gene variants in heterozygotes. *Br J Haematol*, 88(2), 300-306. <https://doi.org/10.1111/j.1365-2141.1994.tb05022.x>
- Mollan, T. L., Yu, X., Weiss, M. J., & Olson, J. S. (2010). The role of alpha-hemoglobin stabilizing protein in redox chemistry, denaturation, and hemoglobin assembly. *Antioxid Redox Signal*, 12(2), 219-231. <https://doi.org/10.1089/ars.2009.2780>
- Mondesert, E., GIANILY-BLAIZOT, M., DELAGE, J., DANTON, F., & AGUILAR-MARTINEZ, P. (2022). PB2235: COMPLETE DELETION OF THE CACCC BOX OF THE HBB GENE PROMOTER: A NEW MECHANISM LEADING TO A SILENT PHENOTYPE OF BETA THALASSEMIA. *HemaSphere*, 6, 2105-2106. <https://doi.org/10.1097/01.HS9.0000851768.05486.cf>
- Moniaux, N., Escande, F., Batra, S. K., Porchet, N., Laine, A., & Aubert, J. P. (2000). Alternative splicing generates a family of putative secreted and membrane-associated MUC4 mucins. *Eur J Biochem*, 267(14), 4536-4544. <https://doi.org/10.1046/j.1432-1327.2000.01504.x>
- Monod, J., Wyman, J., & Changeux, J. P. (1965). On the Nature of Allosteric Transitions: A Plausible Model. *J Mol Biol*, 12, 88-118. [https://doi.org/10.1016/s0022-2836\(65\)80285-6](https://doi.org/10.1016/s0022-2836(65)80285-6)
- Montellano, P. R. (2018). A New Step in the Treatment of Sickle Cell Disease(Published as part of the Biochemistry series "Biochemistry to Bedside"). *Biochemistry*, 57(5), 470-471. <https://doi.org/10.1021/acs.biochem.7b00785>
- Month, S. R., Wood, R. W., Trifillis, P. T., Orchowski, P. J., Sharon, B., Ballas, S. K., Surrey, S., & Schwartz, E. (1990). Analysis of 5' flanking regions of the gamma globin genes from major African haplotype backgrounds associated with sickle cell disease. *J Clin Invest*, 85(2), 364-370. <https://doi.org/10.1172/JCI114447>
- Moon, S. J., Dong, W., Stephanopoulos, G. N., & Sikes, H. D. (2020). Oxidative pentose phosphate pathway and glucose anaplerosis support maintenance of mitochondrial NADPH pool under

- mitochondrial oxidative stress. *Bioengineering & Translational Medicine*, 5(3), e10184. <https://doi.org/https://doi.org/10.1002/btm2.10184>
- Moras, M., Lefevre, S. D., & Ostuni, M. A. (2017). From Erythroblasts to Mature Red Blood Cells: Organelle Clearance in Mammals [Mini Review]. *Frontiers in Physiology*, 8. <https://doi.org/10.3389/fphys.2017.01076>
- Mukherjee, M., Rahaman, M., Ray, S. K., Shukla, P. C., Dolai, T. K., & Chakravorty, N. (2022). Revisiting fetal hemoglobin inducers in beta-hemoglobinopathies: a review of natural products, conventional and combinatorial therapies. *Mol Biol Rep*, 49(3), 2359-2373. <https://doi.org/10.1007/s11033-021-06977-8>
- Muncie, H. L., Jr., & Campbell, J. (2009). Alpha and beta thalassemia. *Am Fam Physician*, 80(4), 339-344.
- Munkongdee, Chen, P., Winichagoon, P., Fucharoen, S., & Paiboonsukwong, K. (2020a). Update in Laboratory Diagnosis of Thalassemia. *Front Mol Biosci*, 7, 74. <https://doi.org/10.3389/fmolb.2020.00074>
- Munkongdee, Chen, P., Winichagoon, P., Fucharoen, S., & Paiboonsukwong, K. (2020b). Update in Laboratory Diagnosis of Thalassemia [Review]. *Frontiers in Molecular Biosciences*, 7. <https://doi.org/10.3389/fmolb.2020.00074>
- Munshi, N. V. (2016). CRISPR (Clustered Regularly Interspaced Palindromic Repeat)/Cas9 System: A Revolutionary Disease-Modifying Technology. *Circulation*, 134(11), 777-779. <https://doi.org/10.1161/CIRCULATIONAHA.116.024007>
- Murakami, K., & Yoshino, M. (2017). Zinc inhibition of pyruvate kinase of M-type isozyme. *Biometals*, 30(3), 335-340. <https://doi.org/10.1007/s10534-017-0009-y>
- Murru, S., Poddie, D., Sciaratta, G. V., Agosti, S., Baffico, M., Melevendi, C., Pirastu, M., & Cao, A. (1992). A novel beta-globin structural mutant, Hb Brescia (beta 114 Leu-Pro), causing a severe beta-thalassemia intermedia phenotype. *Hum Mutat*, 1(2), 124-128. <https://doi.org/10.1002/humu.1380010207>
- Musallam, K. M., Cappellini, M. D., & Taher, A. T. (2013). Iron overload in beta-thalassemia intermedia: an emerging concern. *Curr Opin Hematol*, 20(3), 187-192. <https://doi.org/10.1097/MOH.0b013e32835f5a5c>
- Musallam, K. M., Taher, A. T., & Rachmilewitz, E. A. (2012). beta-thalassemia intermedia: a clinical perspective. *Cold Spring Harb Perspect Med*, 2(7), a013482. <https://doi.org/10.1101/cshperspect.a013482>
- Myers, R. M., Tilly, K., & Maniatis, T. (1986). Fine structure genetic analysis of a beta-globin promoter. *Science*, 232(4750), 613-618. <https://doi.org/10.1126/science.3457470>
- Myers, T. D., Ferguson, C., Gliniak, E., Homanics, G. E., & Palladino, M. J. (2022). Murine model of triosephosphate isomerase deficiency with anemia and severe neuromuscular dysfunction. *Curr Res Neurobiol*, 3, 100062. <https://doi.org/10.1016/j.crneur.2022.100062>
- 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(1), 60-66. <https://doi.org/10.1016/j.bcmd.2007.12.003>
- Narayanan, A., Gagliardi, F., Gallotti, A. L., Mazzoleni, S., Cominelli, M., Fagnocchi, L., Pala, M., Piras, I. S., Zordan, P., Moretta, N., Tratta, E., Brugnara, G., Altabella, L., Bozzuto, G., Gorombe, P., Molinari, A., Padua, R.-A., Bulfone, A., Politi, L. S., . . . Galli, R. (2019). The proneural gene ASCL1 governs the transcriptional subgroup affiliation in glioblastoma stem cells by directly repressing the mesenchymal gene NDRG1. *Cell Death & Differentiation*, 26(9), 1813-1831. <https://doi.org/10.1038/s41418-018-0248-7>
- Ndefo, U. A., Maxwell, A. E., Nguyen, H., & Chiobi, T. L. (2008). Pharmacological management of sickle cell disease. *P t*, 33(4), 238-243.

- Needs, T., Gonzalez-Mosquera, L. F., & Lynch, D. T. (2022). Beta Thalassemia. In *StatPearls*. <https://www.ncbi.nlm.nih.gov/pubmed/30285376>
- Needs, T., Gonzalez-Mosquera, L. F., & Lynch, D. T. (2023). Beta Thalassemia. In *StatPearls*. <https://www.ncbi.nlm.nih.gov/pubmed/30285376>
- Neel, J. V. (1951). The inheritance of the sickling phenomenon, with particular reference to sickle cell disease. *Blood*, 6(5), 389-412. <http://www.ncbi.nlm.nih.gov/pubmed/14830636>
- Nelson, A. S., & Myers, K. C. (2018). Diagnosis, Treatment, and Molecular Pathology of Shwachman-Diamond Syndrome. *Hematol Oncol Clin North Am*, 32(4), 687-700. <https://doi.org/10.1016/j.hoc.2018.04.006>
- Neveling, K., Feenstra, I., Gilissen, C., Hoefsloot, L. H., Kamsteeg, E.-J., Mensenkamp, A. R., Rodenburg, R. J. T., Yntema, H. G., Spruijt, L., Vermeer, S., Rinne, T., van Gassen, K. L., Bodmer, D., Lugtenberg, D., de Reuver, R., Buijsman, W., Derks, R. C., Wieskamp, N., van den Heuvel, B., . . . Nelen, M. R. (2013). A Post-Hoc Comparison of the Utility of Sanger Sequencing and Exome Sequencing for the Diagnosis of Heterogeneous Diseases. *Human Mutation*, 34(12), 1721-1726. <https://doi.org/https://doi.org/10.1002/humu.22450>
- Newton, C. R., Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., Smith, J. C., & Markham, A. F. (1989). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res*, 17(7), 2503-2516. <https://doi.org/10.1093/nar/17.7.2503>
- Ng, A. Y. H., Li, Z., Jones, M. M., Yang, S., Li, C., Fu, C., Tu, C., Oursler, M. J., Qu, J., & Yang, S. (2019). Regulator of G protein signaling 12 enhances osteoclastogenesis by suppressing Nrf2-dependent antioxidant proteins to promote the generation of reactive oxygen species. *eLife*, 8, e42951. <https://doi.org/10.7554/eLife.42951>
- Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*, 31(13), 3812-3814. <https://doi.org/10.1093/nar/gkg509>
- Nickless, A., Bailis, J. M., & You, Z. (2017). Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell Biosci*, 7, 26. <https://doi.org/10.1186/s13578-017-0153-7>
- Niguidula, N., Alamillo, C., Shahmirzadi Mowlavi, L., Powis, Z., Cohen, J. S., & Farwell Hagman, K. D. (2018). Clinical whole-exome sequencing results impact medical management. *Mol Genet Genomic Med*, 6(6), 1068-1078. <https://doi.org/10.1002/mgg3.484>
- Niu, X., He, W., Song, B., Ou, Z., Fan, D., Chen, Y., Fan, Y., & Sun, X. (2016). Combining Single Strand Oligodeoxynucleotides and CRISPR/Cas9 to Correct Gene Mutations in beta-Thalassemia-induced Pluripotent Stem Cells. *J Biol Chem*, 291(32), 16576-16585. <https://doi.org/10.1074/jbc.M116.719237>
- Noguchi, C. T., & Schechter, A. N. (1981). The intracellular polymerization of sickle hemoglobin and its relevance to sickle cell disease. *Blood*, 58(6), 1057-1068. <https://www.ncbi.nlm.nih.gov/pubmed/7030432>
- Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R., & Grosveld, F. (1995). Defective haematopoiesis in fetal liver resulting from inactivation of the EKLf gene. *Nature*, 375(6529), 316-318. <https://doi.org/10.1038/375316a0>
- O'Neal, W. K., & Knowles, M. R. (2018). Cystic Fibrosis Disease Modifiers: Complex Genetics Defines the Phenotypic Diversity in a Monogenic Disease. *Annu Rev Genomics Hum Genet*, 19, 201-222. <https://doi.org/10.1146/annurev-genom-083117-021329>
- Olender, T., Lancet, D., & Nebert, D. W. (2008). Update on the olfactory receptor (OR) gene superfamily. *Hum Genomics*, 3(1), 87-97. <https://doi.org/10.1186/1479-7364-3-1-87>
- Olivieri, N. F., Pakbaz, Z., & Vichinsky, E. (2011). Hb E/beta-thalassaemia: a common & clinically diverse disorder. *Indian J Med Res*, 134(4), 522-531. <https://www.ncbi.nlm.nih.gov/pubmed/22089616>
- Olson, T. S., Frost, B. F., Duke, J. L., Dribus, M., Xie, H. M., Prudowsky, Z. D., Furutani, E., Gudera, J., Shah, Y. B., Ferriola, D., Dinou, A., Pagkrati, I., Kim, S., Xu, Y., He, M., Zheng, S., Nijim, S., Lin, P.,

- Xu, C., . . . Babushok, D. V. (2022). Pathogenicity and impact of HLA class I alleles in aplastic anemia patients of different ethnicities. *JCI Insight*, 7(22). <https://doi.org/10.1172/jci.insight.163040>
- Olwi, D. I., Merdad, L. A., & Ramadan, E. K. (2018). Thalassemia: a prevalent disease yet unknown term among college students in Saudi Arabia. *J Community Genet*, 9(3), 277-282. <https://doi.org/10.1007/s12687-017-0351-3>
- Origa, R., & Moi, P. (1993). Alpha-Thalassemia. In R. A. Pagon, M. P. Adam, H. H. Ardinger, T. D. Bird, C. R. Dolan, C. T. Fong, R. J. H. Smith, & K. Stephens (Eds.), *GeneReviews(R)*. <http://www.ncbi.nlm.nih.gov/pubmed/20301608>
- Ott, J., Wang, J., & Leal, S. M. (2015). Genetic linkage analysis in the age of whole-genome sequencing. *Nature Reviews Genetics*, 16(5), 275-284. <https://doi.org/10.1038/nrg3908>
- Ouellette, R. J., & Rawn, J. D. (2015). *Principles of organic chemistry*. Elsevier.
- P., B., Fermo, E., Glader, B., Kanno, H., Agarwal, A., Barcellini, W., Eber, S., Hoyer, J. D., Kuter, D. J., Maia, T. M., Mañu-Pereira, M. D. M., Kalfa, T. A., Pissard, S., Segovia, J. C., van Beers, E., Gallagher, P. G., Rees, D. C., & van Wijk, R. (2019). Addressing the diagnostic gaps in pyruvate kinase deficiency: Consensus recommendations on the diagnosis of pyruvate kinase deficiency. *Am J Hematol*, 94(1), 149-161. <https://doi.org/10.1002/ajh.25325>
- Paikari, A., & Sheehan, V. A. (2018). Fetal haemoglobin induction in sickle cell disease. *Br J Haematol*, 180(2), 189-200. <https://doi.org/10.1111/bjh.15021>
- Palis, J. (2014). Primitive and definitive erythropoiesis in mammals. *Front Physiol*, 5, 3. <https://doi.org/10.3389/fphys.2014.00003>
- Pang, Q., Christianson, T. A., Keeble, W., Diaz, J., Faulkner, G. R., Reifsteck, C., Olson, S., & Bagby, G. C. (2001). The Fanconi anemia complementation group C gene product: structural evidence of multifunctionality. *Blood*, 98(5), 1392-1401. <https://doi.org/10.1182/blood.V98.5.1392>
- Panigrahi, I., & Agarwal, S. (2008). Genetic determinants of phenotype in beta-thalassemia. *Hematology*, 13(4), 247-252. <https://doi.org/10.1179/102453308X316031>
- Paradkar, S., & Gambhire, P. (2021). The Role of Cytoskeleton of a Red Blood Cell in Its Deformability. *Journal of the Indian Institute of Science*, 101(1), 39-46. <https://doi.org/10.1007/s41745-020-00221-1>
- Pasvol, G., Weatherall, D. J., Wilson, R. J., Smith, D. H., & Gilles, H. M. (1976). Fetal haemoglobin and malaria. *Lancet*, 1(7972), 1269-1272. [https://doi.org/10.1016/s0140-6736\(76\)91738-4](https://doi.org/10.1016/s0140-6736(76)91738-4)
- Pauling, L., Itano, H. A., & et al. (1949). Sickle cell anemia a molecular disease. *Science*, 110(2865), 543-548. <https://doi.org/10.1126/science.110.2865.543>
- Pauling, L., Itano, H. A., Wells, I. C., Schroeder, W. A., Kay, L. M., Singer, S., & Corey, R. (1950). Sickle cell anemia hemoglobin. *Science*,
- Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K. A., Lin, G. N., Nam, H. J., Mort, M., Cooper, D. N., Sebat, J., Iakoucheva, L. M., Mooney, S. D., & Radivojac, P. (2020). Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun*, 11(1), 5918. <https://doi.org/10.1038/s41467-020-19669-x>
- Peng, M., Han, S., Sun, J., He, X., Lv, Y., & Yang, L. (2022). Evaluation of novel compound variants of CEP290 in prenatally suspected case of Meckel syndrome through whole exome sequencing. *Molecular Genetics & Genomic Medicine*, 10(5), e1935. <https://doi.org/https://doi.org/10.1002/mgg3.1935>
- Pennisi, E. (2010). Genomics. Semiconductors inspire new sequencing technologies. *Science*, 327(5970), 1190. <https://doi.org/10.1126/science.327.5970.1190>
- Perkins, A. C., Sharpe, A. H., & Orkin, S. H. (1995). Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature*, 375(6529), 318-322. <https://doi.org/10.1038/375318a0>

- Perrine, S. P. (2005). Fetal globin induction--can it cure beta thalassemia? *Hematology Am Soc Hematol Educ Program*, 38-44. <https://doi.org/10.1182/asheducation-2005.1.38>
- Peschle, C., Migliaccio, A. R., Migliaccio, G., Petrini, M., Calandrini, M., Russo, G., Mastroberardino, G., Presta, M., Gianni, A. M., Comi, P., & et al. (1984). Embryonic---Fetal Hb switch in humans: studies on erythroid bursts generated by embryonic progenitors from yolk sac and liver. *Proc Natl Acad Sci U S A*, 81(8), 2416-2420. <https://doi.org/10.1073/pnas.81.8.2416>
- Petersen, B.-S., Fredrich, B., Hoepfner, M. P., Ellinghaus, D., & Franke, A. (2017). Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genetics*, 18(1), 14. <https://doi.org/10.1186/s12863-017-0479-5>
- Peterson, K. R., Fedosyuk, H., & Costa, F. C. (2008). Hereditary Persistence of Fetal Hemoglobin: Old, New and Future Mutations in the $\alpha\gamma$ -Globin Gene-Proximal Region. *Blood*, 112(11), 492-492. <https://doi.org/10.1182/blood.V112.11.492.492>
- Pettersson, E., Lundeberg, J., & Ahmadian, A. (2009). Generations of sequencing technologies. *Genomics*, 93(2), 105-111. <https://doi.org/10.1016/j.ygeno.2008.10.003>
- Pfeffer, D. A., Satyagraha, A. W., Sadhewa, A., Alam, M. S., Bancone, G., Boum, Y., 2nd, Brito, M., Cui, L., Deng, Z., Domingo, G. J., He, Y., Khan, W. A., Kibria, M. G., Lacerda, M., Menard, D., Monteiro, W., Pal, S., Parikh, S., Roca-Feltrer, A., . . . Ley, B. (2022). Genetic Variants of Glucose-6-Phosphate Dehydrogenase and Their Associated Enzyme Activity: A Systematic Review and Meta-Analysis. *Pathogens*, 11(9). <https://doi.org/10.3390/pathogens11091045>
- Piel, F. B., Patil, A. P., Howes, R. E., Nyangiri, O. A., Gething, P. W., Williams, T. N., Weatherall, D. J., & Hay, S. I. (2010). Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nat Commun*, 1, 104. <https://doi.org/10.1038/ncomms1104>
- Piel, F. B., Steinberg, M. H., & Rees, D. C. (2017). Sickle Cell Disease. *The New England Journal of Medicine*, 376(16), 1561-1573. <https://doi.org/https://doi.org/10.1056/NEJMra1510865>
- Piety, N. Z., Reinhart, W. H., Purreau, P. H., Abidi, R., & Shevkopyas, S. S. (2016). Shape matters: the effect of red blood cell shape on perfusion of an artificial microvascular network. *Transfusion*, 56(4), 844-851. <https://doi.org/10.1111/trf.13449>
- Pimpakan, T., Mungkalasut, P., Tansakul, P., Chanda, M., Jugnam-Ang, W., Charucharana, S., Cheepsunthorn, P., Fucharoen, S., Punnahitananda, S., & Cheepsunthorn, C. L. (2022). Effect of neonatal reticulocytosis on glucose 6-phosphate dehydrogenase (G6PD) activity and G6PD deficiency detection: a cross-sectional study. *BMC Pediatrics*, 22(1), 678. <https://doi.org/10.1186/s12887-022-03740-1>
- Platt, O. S., Brambilla, D. J., Rosse, W. F., Milner, P. F., Castro, O., Steinberg, M. H., & Klug, P. P. (1994). Mortality in sickle cell disease. Life expectancy and risk factors for early death. *N Engl J Med*, 330(23), 1639-1644. <https://doi.org/10.1056/NEJM199406093302303>
- Platt, O. S., Thorington, B. D., Brambilla, D. J., Milner, P. F., Rosse, W. F., Vichinsky, E., & Kinney, T. R. (1991). Pain in sickle cell disease. Rates and risk factors. *N Engl J Med*, 325(1), 11-16. <https://doi.org/10.1056/NEJM199107043250103>
- Popay, T. M., & Dixon, J. R. (2022). Coming full circle: On the origin and evolution of the looping model for enhancer-promoter communication. *J Biol Chem*, 298(8), 102117. <https://doi.org/10.1016/j.jbc.2022.102117>
- Porcu, S., Simbula, M., Marongiu, M. F., Perra, A., Poddie, D., Perseu, L., Kowalik, M. A., Littera, R., Barella, S., Caria, C. A., Demartis, F. R., & Ristaldi, M. S. (2021). Delta-globin gene expression improves sickle cell disease in a humanised mouse model. *British Journal of Haematology*, 193(6), 1228-1237. <https://doi.org/https://doi.org/10.1111/bjh.17561>
- Pornprasert, S., & Phanthong, S. (2013). Anemia in patients with coinherited thalassemia and glucose-6-phosphate dehydrogenase deficiency. *Hemoglobin*, 37(6), 536-543. <https://doi.org/10.3109/03630269.2013.819558>

- Port, F., Chen, H. M., Lee, T., & Bullock, S. L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc Natl Acad Sci U S A*, *111*(29), E2967-2976. <https://doi.org/10.1073/pnas.1405500111>
- Porter, M. L., & Dennis, B. L. (2002). Hyperbilirubinemia in the term newborn. *Am Fam Physician*, *65*(4), 599-606. <https://www.ncbi.nlm.nih.gov/pubmed/11871676>
- Pretini, V., Koenen, M. H., Kaestner, L., Fens, M., Schiffelers, R. M., Bartels, M., & Van Wijk, R. (2019). Red Blood Cells: Chasing Interactions. *Front Physiol*, *10*, 945. <https://doi.org/10.3389/fphys.2019.00945>
- Qin, K., Huang, P., Feng, R., Keller, C. A., Peslak, S. A., Khandros, E., Saari, M. S., Lan, X., Mayuranathan, T., Doerfler, P. A., Abdulmalik, O., Giardine, B., Chou, S. T., Shi, J., Hardison, R. C., Weiss, M. J., & Blobel, G. A. (2022). Dual function NFI factors control fetal hemoglobin silencing in adult erythroid cells. *Nature Genetics*, *54*(6), 874-884. <https://doi.org/10.1038/s41588-022-01076-1>
- Qiu, C., Olivier, E. N., Velho, M., & Bouhassira, E. E. (2008). Globin switches in yolk sac-like primitive and fetal-like definitive red blood cells produced from human embryonic stem cells. *Blood*, *111*(4), 2400-2408. <https://doi.org/10.1182/blood-2007-07-102087>
- Quinn, C. T. (2016). Minireview: Clinical severity in sickle cell disease: the challenges of definition and prognostication. *Exp Biol Med (Maywood)*, *241*(7), 679-688. <https://doi.org/10.1177/1535370216640385>
- Rahimmanesh, I., Boshtam, M., Kouhpayeh, S., Khanahmad, H., Dabiri, A., Ahangarzadeh, S., Esmaili, Y., Bidram, E., Vaseghi, G., Haghjooy Javanmard, S., Shariati, L., Zarrabi, A., & Varma, R. S. (2022). Gene Editing-Based Technologies for Beta-hemoglobinopathies Treatment. *Biology*, *11*(6), 862. <https://www.mdpi.com/2079-7737/11/6/862>
- Rahit, K. M. T. H., & Tarailo-Graovac, M. (2020). Genetic Modifiers and Rare Mendelian Disease. *Genes*, *11*(3), 239. <https://www.mdpi.com/2073-4425/11/3/239>
- Ramasamy, I. (2020). Atypical hereditary spherocytosis phenotype associated with pseudohypokalaemia and a new variant in the band 3 protein. *BMJ Case Rep*, *13*(12). <https://doi.org/10.1136/bcr-2020-238428>
- Randhawa, Z. I., Jones, R. T., & Lie-Injo, L. E. (1984). Human hemoglobin Portland II (zeta 2 beta 2). Isolation and characterization of Portland hemoglobin components and their constituent globin chains. *J Biol Chem*, *259*(11), 7325-7330. <https://www.ncbi.nlm.nih.gov/pubmed/6539334>
- Raposo, M., Bettencourt, C., Melo, A. R. V., Ferreira, A. F., Alonso, I., Silva, P., Vasconcelos, J., Kay, T., Saraiva-Pereira, M. L., Costa, M. D., Vilasboas-Campos, D., Bettencourt, B. F., Bruges-Armas, J., Houlden, H., Heutink, P., Jardim, L. B., Sequeiros, J., Maciel, P., & Lima, M. (2022). Novel Machado-Joseph disease-modifying genes and pathways identified by whole-exome sequencing. *Neurobiology of Disease*, *162*, 105578. <https://doi.org/https://doi.org/10.1016/j.nbd.2021.105578>
- Rattananon, P., Anurathapan, U., Bhukhai, K., & Hongeng, S. (2021). The Future of Gene Therapy for Transfusion-Dependent Beta-Thalassemia: The Power of the Lentiviral Vector for Genetically Modified Hematopoietic Stem Cells [Mini Review]. *Frontiers in Pharmacology*, *12*. <https://doi.org/10.3389/fphar.2021.730873>
- Raya, A., Rodríguez-Pizà, I., Guenechea, G., Vassena, R., Navarro, S., Barrero, M. J., Consiglio, A., Castellà, M., Río, P., Sleep, E., González, F., Tiscornia, G., Garreta, E., Aasen, T., Veiga, A., Verma, I. M., Surrallés, J., Bueren, J., & Izpisua Belmonte, J. C. (2009). Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature*, *460*(7251), 53-59. <https://doi.org/10.1038/nature08129>
- Rees, D. C., Brousse, V. A. M., & Brewin, J. N. (2022). Determinants of severity in sickle cell disease. *Blood Rev*, *56*, 100983. <https://doi.org/https://doi.org/10.1016/j.blre.2022.100983>

- Rehman, A. U., Rashid, A., Hussain, Z., & Shah, K. (2022). A novel homozygous missense variant p.D339N in the PKLR gene correlates with pyruvate kinase deficiency in a Pakistani family: a case report. *J Med Case Rep*, 16(1), 66. <https://doi.org/10.1186/s13256-022-03292-z>
- Renaux, A., Papadimitriou, S., Versbraegen, N., Nachtegael, C., Boutry, S., Nowe, A., Smits, G., & Lenaerts, T. (2019). ORVAL: a novel platform for the prediction and exploration of disease-causing oligogenic variant combinations. *Nucleic Acids Res*, 47(W1), W93-W98. <https://doi.org/10.1093/nar/gkz437>
- Repsold, L., & Joubert, A. M. (2018). Eryptosis: An Erythrocyte's Suicidal Type of Cell Death. *Biomed Res Int*, 2018, 9405617. <https://doi.org/10.1155/2018/9405617>
- Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*, 39(17), e118. <https://doi.org/10.1093/nar/gkr407>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, 17(5), 405-424. <https://doi.org/10.1038/gim.2015.30>
- Richardson, S. R., & O'Malley, G. F. (2023). Glucose 6 Phosphate Dehydrogenase Deficiency. In *StatPearls*. <https://www.ncbi.nlm.nih.gov/pubmed/29262208>
- Rieu, S., Géminard, C., Rabesandratana, H., Sainte-Marie, J., & Vidal, M. (2000). Exosomes released during reticulocyte maturation bind to fibronectin via integrin $\alpha\beta 1$. *European Journal of Biochemistry*, 267(2), 583-590.
- Rizzuto, V., Koopmann, T. T., Blanco-Álvarez, A., Tazón-Vega, B., Idrizovic, A., Díaz de Heredia, C., Del Orbe, R., Pampliega, M. V., Velasco, P., Beneitez, D., Santen, G. W. E., Waisfisz, Q., Elting, M., Smiers, F. J. W., de Pagter, A. J., Kerkhoffs, J.-L. H., Hartevelde, C. L., & Mañú-Pereira, M. d. M. (2021). Usefulness of NGS for Diagnosis of Dominant Beta-Thalassemia and Unstable Hemoglobinopathies in Five Clinical Cases [Original Research]. *Frontiers in Physiology*, 12. <https://doi.org/10.3389/fphys.2021.628236>
- Robledo, R., Beggs, W. R., & Bender, P. K. (2005). TaqMan genotyping of insertion/deletion polymorphisms. *Methods Mol Biol*, 311, 165-176. <https://doi.org/10.1385/1-59259-957-5:165>
- Rodenburg, R. J. (2018). The functional genomics laboratory: functional validation of genetic variants. *J Inherit Metab Dis*, 41(3), 297-307. <https://doi.org/10.1007/s10545-018-0146-7>
- Rodrigo, L. (2019). Iron Deficiency Anemia. In Rijeka: IntechOpen.
- Rodrigues, C. H., Pires, D. E., & Ascher, D. B. (2018). DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res*, 46(W1), W350-W355. <https://doi.org/10.1093/nar/gky300>
- Roopa, D. L., Shyamsunder, K., Karunakar, P., Rajabathar, J. R., Venkatesulu, A., Karnan, M., Kiran, K. S., Selvaraj, M., & Basavarajaiah, S. M. (2023). Naphtho[2,1-b]furan derived triazole-pyrimidines as highly potential InhA and Cytochrome c peroxidase inhibitors: Synthesis, DFT calculations, drug-likeness profile, molecular docking and dynamic studies. *Journal of Molecular Structure*, 1287, 135685. <https://doi.org/https://doi.org/10.1016/j.molstruc.2023.135685>
- Ropero, P., Gonzalez Fernandez, F. A., Nieto, J. M., Torres-Jimenez, W. M., & Benavente, C. (2022). beta-Thalassemia Intermedia: Interaction of alpha-Globin Gene Triplication With beta-thalassemia Heterozygous in Spain. *Front Med (Lausanne)*, 9, 866396. <https://doi.org/10.3389/fmed.2022.866396>
- Ross, J. P., Dion, P. A., & Rouleau, G. A. (2020). Exome sequencing in genetic disease: recent advances and considerations. *F1000Res*, 9. <https://doi.org/10.12688/f1000research.19444.1>

- Rossi, F., Helbling-Leclerc, A., Kawasumi, R., Jegadesan, N. K., Xu, X., Devulder, P., Abe, T., Takata, M., Xu, D., Rosselli, F., & Branzei, D. (2020). SMC5/6 acts jointly with Fanconi anemia factors to support DNA repair and genome stability. *EMBO Rep*, 21(2), e48222. <https://doi.org/10.15252/embr.201948222>
- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., . . . Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356), 348-352. <https://doi.org/10.1038/nature10242>
- Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*, 5(4), 725-738. <https://doi.org/10.1038/nprot.2010.5>
- Roy, M. K., Cendali, F., Ooyama, G., Gamboni, F., Morton, H., & D'Alessandro, A. (2021). Red Blood Cell Metabolism in Pyruvate Kinase Deficient Patients [Original Research]. *Frontiers in Physiology*, 12. <https://doi.org/10.3389/fphys.2021.735543>
- Roy, N. B., Wilson, E. A., Henderson, S., Wray, K., Babbs, C., Okoli, S., Atoyebi, W., Mixon, A., Cahill, M. R., Carey, P., Cullis, J., Curtin, J., Dreau, H., Ferguson, D. J., Gibson, B., Hall, G., Mason, J., Morgan, M., Proven, M., . . . Schuh, A. (2016). A novel 33-Gene targeted resequencing panel provides accurate, clinical-grade diagnosis and improves patient management for rare inherited anaemias. *Br J Haematol*, 175(2), 318-330. <https://doi.org/10.1111/bjh.14221>
- Roy, N. B. A., & Babbs, C. (2019). The pathogenesis, diagnosis and management of congenital dyserythropoietic anaemia type I. *Br J Haematol*, 185(3), 436-449. <https://doi.org/10.1111/bjh.15817>
- Rund, D., Cohen, T., Filon, D., Dowling, C. E., Warren, T. C., Barak, I., Rachmilewitz, E., Kazazian, H. H., Jr., & Oppenheim, A. (1991). Evolution of a genetic disease in an ethnic isolate: beta-thalassemia in the Jews of Kurdistan. *Proc Natl Acad Sci U S A*, 88(1), 310-314. <https://doi.org/10.1073/pnas.88.1.310>
- Rusk, N. (2011). Torrents of sequence. *Nature Methods*, 8(1), 44-44. <https://doi.org/10.1038/nmeth.f.330>
- Russo, R., Andolfo, I., Manna, F., Gambale, A., Marra, R., Rosato, B. E., Caforio, P., Pinto, V., Pignataro, P., Radhakrishnan, K., Unal, S., Tomaiuolo, G., Forni, G. L., & Iolascon, A. (2018). Multi-gene panel testing improves diagnosis and management of patients with hereditary anemias. *Am J Hematol*, 93(5), 672-682. <https://doi.org/10.1002/ajh.25058>
- Russo, R., Marra, R., Rosato, B. E., Iolascon, A., & Andolfo, I. (2020). Genetics and Genomics Approaches for Diagnosis and Research Into Hereditary Anemias. *Front Physiol*, 11, 613559. <https://doi.org/10.3389/fphys.2020.613559>
- Sabiha, B., Haider, S. A., Jan, H., Yousafzai, Y. M., Afridi, O. K., Khan, A. A., & Ali, J. (2020). Development of the Next Generation Sequencing-Based Diagnostic Test for beta-Thalassemia and its Validation in a Pashtun Family. *Hemoglobin*, 44(4), 254-258. <https://doi.org/10.1080/03630269.2020.1793773>
- Sadahira, Y., Yoshino, T., & Monobe, Y. (1995). Very late activation antigen 4-vascular cell adhesion molecule 1 interaction is involved in the formation of erythroblastic islands. *J Exp Med*, 181(1), 411-415. <https://doi.org/10.1084/jem.181.1.411>
- Sahli, C. A., Ben Salem, I., Jouini, L., Laouini, N., Dabboubi, R., Hadj Fredj, S., Siala, H., Othmeni, R., Dakhlaoui, B., Fattoum, S., Bibi, A., & Messaoud, T. (2016). Setup of a Protocol of Molecular Diagnosis of beta-Thalassemia Mutations in Tunisia using Denaturing High-Performance Liquid Chromatography (DHPLC). *J Clin Lab Anal*, 30(5), 392-398. <https://doi.org/10.1002/jcla.21867>
- Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*, 32(4), 347-355. <https://doi.org/10.1038/nbt.2842>

- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 74(12), 5463-5467. <https://doi.org/10.1073/pnas.74.12.5463>
- Sankaran, Menne, T. F., Scepanovic, D., Vergilio, J. A., Ji, P., Kim, J., Thiru, P., Orkin, S. H., Lander, E. S., & Lodish, H. F. (2011). MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. *Proc Natl Acad Sci U S A*, 108(4), 1519-1524. <https://doi.org/10.1073/pnas.1018384108>
- Sankaran, V. G., Menne, T. F., Xu, J., Akie, T. E., Lettre, G., Van Handel, B., Mikkola, H. K. A., 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(5909), 1839-1842. <https://doi.org/doi:10.1126/science.1165409>
- Sankaran, V. G., & Orkin, S. H. (2013). The switch from fetal to adult hemoglobin. *Cold Spring Harb Perspect Med*, 3(1), a011643. <https://doi.org/10.1101/cshperspect.a011643>
- Sankaran, V. G., & Weiss, M. J. (2015). Anemia: progress in molecular mechanisms and therapies. *Nat Med*, 21(3), 221-230. <https://doi.org/10.1038/nm.3814>
- Sankaran, V. G., Xu, J., & Orkin, S. H. (2010). Advances in the understanding of haemoglobin switching. *Br J Haematol*, 149(2), 181-194. <https://doi.org/10.1111/j.1365-2141.2010.08105.x>
- Sathupak, S., Leecharoenkiat, K., & Kampuansai, J. (2021). Prevalence and molecular characterization of glucose-6-phosphate dehydrogenase deficiency in the Lue ethnic group of northern Thailand. *Sci Rep*, 11(1), 2956. <https://doi.org/10.1038/s41598-021-82477-w>
- Savić, N., & Schwank, G. (2016). Advances in therapeutic CRISPR/Cas9 genome editing. *Translational Research*, 168, 15-21. <https://doi.org/https://doi.org/10.1016/j.trsl.2015.09.008>
- Savitt, T. L. (2014). Learning About Sickle Cell: The Patient in Early Sickle Cell Disease Case Reports, 1910-1933. *J Natl Med Assoc*, 106(1), 31-41. [https://doi.org/10.1016/s0027-9684\(15\)30068-7](https://doi.org/10.1016/s0027-9684(15)30068-7)
- Sayer, J. A., Otto, E. A., O'Toole, J. F., Nurnberg, G., Kennedy, M. A., Becker, C., Hennies, H. C., Helou, J., Attanasio, M., Fausett, B. V., Utsch, B., Khanna, H., Liu, Y., Drummond, I., Kawakami, I., Kusakabe, T., Tsuda, M., Ma, L., Lee, H., . . . Hildebrandt, F. (2006). The centrosomal protein nephrocystin-6 is mutated in Joubert syndrome and activates transcription factor ATF4. *Nat Genet*, 38(6), 674-681. <https://doi.org/10.1038/ng1786>
- Schaffer, A. A. (2013). Digenic inheritance in medical genetics. *J Med Genet*, 50(10), 641-652. <https://doi.org/10.1136/jmedgenet-2013-101713>
- Schleinitz, D., Distefano, J. K., & Kovacs, P. (2011). Targeted SNP genotyping using the TaqMan(R) assay. *Methods Mol Biol*, 700, 77-87. https://doi.org/10.1007/978-1-61737-954-3_6
- Schormann, N., Hayden, K. L., Lee, P., Banerjee, S., & Chattopadhyay, D. (2019). An overview of structure, function, and regulation of pyruvate kinases. *Protein Sci*, 28(10), 1771-1784. <https://doi.org/10.1002/pro.3691>
- Schrier, S. L., Centis, F., Verneris, M., Ma, L., & Angelucci, E. (2003). The role of oxidant injury in the pathophysiology of human thalassemias. *Redox Rep*, 8(5), 241-245. <https://doi.org/10.1179/135100003225002835>
- Schroer, A. B., Mohamed, J. S., Willard, M. D., Setola, V., Oestreich, E., & Siderovski, D. P. (2019). A role for Regulator of G protein Signaling-12 (RGS12) in the balance between myoblast proliferation and differentiation. *PLoS One*, 14(8), e0216167. <https://doi.org/10.1371/journal.pone.0216167>
- Schroit, A. J., Madsen, J. W., & Tanaka, Y. (1985). In vivo recognition and clearance of red blood cells containing phosphatidylserine in their plasma membranes. *J Biol Chem*, 260(8), 5131-5138. <https://www.ncbi.nlm.nih.gov/pubmed/3988747>
- Schubert, L., Hendriks, I. A., Hertz, E. P. T., Wu, W., Sellés-Baiget, S., Hoffmann, S., Viswalingam, K. S., Gallina, I., Pentakota, S., Benedict, B., Johansen, J., Apelt, K., Luijsterburg, M. S., Rasmussen, S., Lisby, M., Liu, Y., Nielsen, M. L., Mailand, N., & Duxin, J. P. (2022). SCAI promotes error-free

- repair of DNA interstrand crosslinks via the Fanconi anemia pathway. *EMBO reports*, 23(4), e53639. <https://doi.org/https://doi.org/10.15252/embr.202153639>
- Scionti, F., Di Martino, M. T., Pensabene, L., Bruni, V., & Concolino, D. (2018). The Cytoscan HD Array in the Diagnosis of Neurodevelopmental Disorders. *High Throughput*, 7(3). <https://doi.org/10.3390/ht7030028>
- Scriver, J. B., & Waugh, T. R. (1930). STUDIES ON A CASE OF SICKLE-CELL ANAEMIA. *Can Med Assoc J*, 23(3), 375-380.
- Secrest, M. H., Storm, M., Carrington, C., Casso, D., Gilroy, K., Pladson, L., & Boscoe, A. N. (2020). Prevalence of pyruvate kinase deficiency: A systematic literature review. *Eur J Haematol*, 105(2), 173-184. <https://doi.org/10.1111/ejh.13424>
- Seki, M., & Shirasawa, H. (1965). Role of the reticular cells during maturation process of the erythroblast. 3. The fate of phagocytized nucleus. *Acta Pathol Jpn*, 15(4), 387-405. <https://doi.org/10.1111/j.1440-1827.1965.tb01930.x>
- Selwyn, J. G., & Dacie, J. V. (1954). Autohemolysis and Other Changes Resulting from the Incubation in Vitro of Red Cells from Patients with Congenital Hemolytic Anemia. *Blood*, 9(5), 414-438. <https://doi.org/10.1182/blood.V9.5.414.414>
- Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol*, 14(8), e1002533. <https://doi.org/10.1371/journal.pbio.1002533>
- Serjeant. (2013a). The Natural History of Sickle Cell Disease. *Cold Spring Harbor Perspectives in Medicine*, 3(10). <https://doi.org/10.1101/cshperspect.a011783>
- Serjeant. (2013b). The natural history of sickle cell disease. *Cold Spring Harb Perspect Med*, 3(10), a011783. <https://doi.org/10.1101/cshperspect.a011783>
- Shaeffer, J. R., & Kania, M. A. (1995). Degradation of monoubiquitinated alpha-globin by 26S proteasomes. *Biochemistry*, 34(12), 4015-4021. <https://doi.org/10.1021/bi00012a020>
- Shaikh, T. H. (2017). Copy Number Variation Disorders. *Curr Genet Med Rep*, 5(4), 183-190. <https://doi.org/10.1007/s40142-017-0129-2>
- Shakiba, M., & Keramatipour, M. (2018). Effect of Whole Exome Sequencing in Diagnosis of Inborn Errors of Metabolism and Neurogenetic Disorders. *Iran J Child Neurol*, 12(1), 7-15. <https://www.ncbi.nlm.nih.gov/pubmed/29379558>
- Shakin, S. H., & Liehaber, S. A. (1986). Translational profiles of alpha 1-, alpha 2-, and beta-globin messenger ribonucleic acids in human reticulocytes. *J Clin Invest*, 78(4), 1125-1129. <https://doi.org/10.1172/JCI112670>
- Sharma, D. C., Singhal, S., Woike, P., Rai, S., Yadav, M., & Gaur, R. (2020). Hereditary persistence of fetal hemoglobin. *Asian J Transfus Sci*, 14(2), 185-186. https://doi.org/10.4103/ajts.AJTS_71_16
- Shen, G. Q., Abdullah, K. G., & Wang, Q. K. (2009). The TaqMan method for SNP genotyping. *Methods Mol Biol*, 578, 293-306. https://doi.org/10.1007/978-1-60327-411-1_19
- Shen, S. C., Fleming, E. M., & Castle, W. B. (1949). Studies on the destruction of red blood cells; irreversibly sickled erythrocytes; their experimental production in vitro. *Blood*, 4(5), 498-504. <https://www.ncbi.nlm.nih.gov/pubmed/18126358>
- Shen, Y., Verboon, J. M., Zhang, Y., Liu, N., Kim, Y. J., Marglous, S., Nandakumar, S. K., Voit, R. A., Fiorini, C., Ejaz, A., Basak, A., Orkin, S. H., Xu, J., & Sankaran, V. G. (2021). A unified model of human hemoglobin switching through single-cell genome editing. *Nat Commun*, 12(1), 4991. <https://doi.org/10.1038/s41467-021-25298-9>
- Siegel, R. J., Bridges Jr, S. L., & Ahmed, S. (2019). HLA-C: An Accomplice in Rheumatic Diseases. *ACR Open Rheumatology*, 1(9), 571-579. <https://doi.org/https://doi.org/10.1002/acr2.11065>
- Slimane, A., Thiago Trovati, M., Olivier, H., & Mariane de, M. (2020). Innate immune cells, major protagonists of sickle cell disease pathophysiology. *Haematologica*, 105(2), 273-283. <https://doi.org/10.3324/haematol.2019.229989>

- Smith, D. W. (1980). The molecular biology of mammalian hemoglobin synthesis. *Ann Clin Lab Sci*, 10(2), 116-122. <https://www.ncbi.nlm.nih.gov/pubmed/6992694>
- Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B., & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature*, 321(6071), 674-679. <https://doi.org/10.1038/321674a0>
- Soares, D. S., Homem, C. C. F., & Castro, D. S. (2022). Function of Proneural Genes *Ascl1* and *Asense* in Neurogenesis: How Similar Are They? [Mini Review]. *Frontiers in Cell and Developmental Biology*, 10. <https://doi.org/10.3389/fcell.2022.838431>
- Socolovsky, M., Nam, H., Fleming, M. D., Haase, V. H., Brugnara, C., & Lodish, H. F. (2001). Ineffective erythropoiesis in *Stat5a(-/-)5b(-/-)* mice due to decreased survival of early erythroblasts. *Blood*, 98(12), 3261-3273. <https://doi.org/10.1182/blood.v98.12.3261>
- Solovieff, N., Milton, J. N., Hartley, S. W., Sherva, R., Sebastiani, P., Dworkis, D. A., Klings, E. S., Farrer, L. A., Garrett, M. E., Ashley-Koch, A., Telen, M. J., Fucharoen, S., Ha, S. Y., Li, C.-K., Chui, D. H. K., Baldwin, C. T., & Steinberg, M. H. (2010). Fetal hemoglobin in sickle cell anemia: genome-wide association studies suggest a regulatory region in the 5' olfactory receptor gene cluster. *Blood*, 115(9), 1815-1822. <https://doi.org/10.1182/blood-2009-08-239517>
- 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(9), 1053-1065. <https://doi.org/10.1089/scd.2014.0347>
- Spangenberg, M. N., Grille, S., Simoes, C., Dell'Oca, N., Boada, M., Guillermo, C., Raggio, V., & Spangenberg, L. (2022). Two mutations in the *SBDS* gene reveal a diagnosis of Shwachman-Diamond syndrome in a patient with atypical symptoms. *Cold Spring Harb Mol Case Stud*, 8(7). <https://doi.org/10.1101/mcs.a006237>
- Spritz, R. A., Jagadeeswaran, P., Choudary, P. V., Biro, P. A., Elder, J. T., deRiel, J. K., Manley, J. L., Gefter, M. L., Forget, B. G., & Weissman, S. M. (1981). Base substitution in an intervening sequence of a beta⁺-thalassemic human globin gene. *Proc Natl Acad Sci U S A*, 78(4), 2455-2459. <https://doi.org/10.1073/pnas.78.4.2455>
- Stanchfield, M. L., Webster, S. E., Webster, M. K., & Linn, C. L. (2020). Involvement of HB-EGF/*Ascl1*/*Lin28a* Genes in Dedifferentiation of Adult Mammalian Müller Glia. *Front Mol Biosci*, 7, 200. <https://doi.org/10.3389/fmolb.2020.00200>
- Steinberg, M. H. (2020). Fetal hemoglobin in sickle cell anemia. *Blood*, 136(21), 2392-2400. <https://doi.org/https://doi.org/10.1182/blood.2020007645>
- Steinberg, M. H., Forget, B. G., Higgs, D. R., & Weatherall, D. J. (2009). *Disorders of hemoglobin: genetics, pathophysiology, and clinical management*. Cambridge University Press.
- Steinberg, M. H., & Rodgers, G. P. (2015). HbA2: biology, clinical relevance and a possible target for ameliorating sickle cell disease. *British Journal of Haematology*, 170(6), 781-787. <https://doi.org/https://doi.org/10.1111/bjh.13570>
- Sterling, D., & Casey, J. R. (2002). Bicarbonate transport proteins. *Biochem Cell Biol*, 80(5), 483-497. <https://doi.org/10.1139/o02-152>
- Stuppia, L., Antonucci, I., Palka, G., & Gatta, V. (2012). Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int J Mol Sci*, 13(3), 3245-3276. <https://doi.org/10.3390/ijms13033245>
- Sugiyama, M., Hamanoue, S., Nagai, J. I., Tsurusaki, Y., Kurosawa, K., Tanaka, M., Tanaka, Y., & Goto, H. (2019). Hemoglobin beta Kanagawa [c.443A>C; p.(Ter148Serext*21)]: A novel beta-globin gene mutation causing dominantly inherited beta-thalassemia. *Pediatr Blood Cancer*, 66(9), e27871. <https://doi.org/10.1002/pbc.27871>

- Suhaimi, S. A., Zulkipli, I. N., Ghani, H., & Abdul-Hamid, M. R. W. (2022). Applications of next generation sequencing in the screening and diagnosis of thalassemia: A mini-review [Mini Review]. *Frontiers in Pediatrics*, 10. <https://doi.org/10.3389/fped.2022.1015769>
- Sukumar, S., Mukherjee, M. B., Colah, R. B., & Mohanty, D. (2004). Molecular basis of G6PD deficiency in India. *Blood Cells Mol Dis*, 33(2), 141-145. <https://doi.org/10.1016/j.bcmd.2004.06.003>
- Sun, N., & Zhao, H. (2014). Seamless correction of the sickle cell disease mutation of the HBB gene in human induced pluripotent stem cells using TALENs. *Biotechnol Bioeng*, 111(5), 1048-1053. <https://doi.org/10.1002/bit.25018>
- Sun, Y., Ruivenkamp, C. A., Hoffer, M. J., Vrijenhoek, T., Kriek, M., van Asperen, C. J., den Dunnen, J. T., & Santen, G. W. (2015). Next-generation diagnostics: gene panel, exome, or whole genome? *Hum Mutat*, 36(6), 648-655. <https://doi.org/10.1002/humu.22783>
- Suwannakhon, N., Pangeson, T., Seeratanachot, T., Mahingsa, K., Pingyod, A., Bumrungrakdee, W., & Sanguanserm Sri, T. (2019). Noninvasive prenatal screening test for compound heterozygous beta thalassemia using an amplification refractory mutation system real-time polymerase chain reaction technique. *Hematol Rep*, 11(3), 8124. <https://doi.org/10.4081/hr.2019.8124>
- Svidnicki, M., Zanetta, G. K., Congrains-Castillo, A., Costa, F. F., & Saad, S. T. O. (2020). Targeted next-generation sequencing identified novel mutations associated with hereditary anemias in Brazil. *Ann Hematol*, 99(5), 955-962. <https://doi.org/10.1007/s00277-020-03986-8>
- Sydenstricker, V. P. (1924). FURTHER OBSERVATIONS ON SICKLE CELL ANEMIA. *Journal of the American Medical Association*, 83(1), 12-17. <https://doi.org/10.1001/jama.1924.02660010016004>
- Taha, I., Foroni, S., Valli, R., Frattini, A., Rocchia, P., Porta, G., Zecca, M., Bergami, E., Cipolli, M., Pasquali, F., Danesino, C., Scotti, C., & Minelli, A. (2022). Case Report: Heterozygous Germline Variant in EIF6 Additional to Biallelic SBDS Pathogenic Variants in a Patient With Ribosomopathy Shwachman–Diamond Syndrome [Case Report]. *Frontiers in Genetics*, 13. <https://doi.org/10.3389/fgene.2022.896749>
- Taher, A., Vichinsky, E., Musallam, K., Cappellini, M. D., & Viprakasit, V. (2013). In D. Weatherall (Ed.), *Guidelines for the Management of Non Transfusion Dependent Thalassemia (NTDT)*. <https://www.ncbi.nlm.nih.gov/pubmed/24672826>
- Taliaferro, W. H., & Huck, J. G. (1923). The Inheritance of Sickle-Cell Anaemia in Man. *Genetics*, 8(6), 594-598. <http://www.ncbi.nlm.nih.gov/pubmed/17246028>
- Tanaka, K. R., Valentine, W. N., & Miwa, S. (1962). Pyruvate kinase (PK) deficiency hereditary nonspherocytic hemolytic anemia. *Blood*, 19, 267-295.
- Tanasie, N. L., Gutierrez-Escribano, P., Jaklin, S., Aragon, L., & Stigler, J. (2022). Stabilization of DNA fork junctions by Smc5/6 complexes revealed by single-molecule imaging. *Cell Rep*, 41(10), 111778. <https://doi.org/10.1016/j.celrep.2022.111778>
- Tanimoto, K., Liu, Q., Bungert, J., & Engel, J. D. (1999). Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. *Nature*, 398(6725), 344-348. <https://doi.org/10.1038/18698>
- Tasan, I., Jain, S., & Zhao, H. (2016). Use of genome-editing tools to treat sickle cell disease. *Hum Genet*, 135(9), 1011-1028. <https://doi.org/10.1007/s00439-016-1688-0>
- Tavazzi, D., Taher, A., & Cappellini, M. D. (2008). Red blood cell enzyme disorders: an overview. *Pediatr Ann*, 37(5), 303-310. <https://doi.org/10.3928/00904481-20080501-08>
- Taylor, S. M., Parobek, C. M., & Fairhurst, R. M. (2012). Haemoglobinopathies and the clinical epidemiology of malaria: a systematic review and meta-analysis. *Lancet Infect Dis*, 12(6), 457-468. [https://doi.org/10.1016/s1473-3099\(12\)70055-5](https://doi.org/10.1016/s1473-3099(12)70055-5)
- Teh, Lee, T.-Y., Tan, J. A. M. A., Lai, M.-I., & George, E. (2015). The use of Taqman genotyping assays for rapid confirmation of β -thalassaemia mutations in the Malays: accurate diagnosis with low DNA concentrations. *Int J Lab Hematol*, 37(1), 79-89. <https://doi.org/https://doi.org/10.1111/ijlh.12240>

- Thein, S. L. (1992). Dominant beta thalassaemia: molecular basis and pathophysiology. *Br J Haematol*, 80(3), 273-277. <https://doi.org/10.1111/j.1365-2141.1992.tb08132.x>
- Thein, S. L. (2013). The molecular basis of beta-thalassemia. *Cold Spring Harb Perspect Med*, 3(5), a011700. <https://doi.org/10.1101/cshperspect.a011700>
- Thein, S. L. (2018). Molecular basis of beta thalassemia and potential therapeutic targets. *Blood Cells Mol Dis*, 70, 54-65. <https://doi.org/10.1016/j.bcmd.2017.06.001>
- Thein, S. L., Menzel, S., Lathrop, M., & Garner, C. (2009). Control of fetal hemoglobin: new insights emerging from genomics and clinical implications. *Hum Mol Genet*, 18(R2), R216-223. <https://doi.org/10.1093/hmg/ddp401>
- 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(27), 11346-11351. <https://doi.org/10.1073/pnas.0611393104>
- Thiam, F., Diop, G., Coulonges, C., Derbois, C., Mbengue, B., Thiam, A., Nguer, C. M., Zagury, J. F., Deleuze, J. F., & Dieye, A. (2022). G6PD and HBB polymorphisms in the Senegalese population: prevalence, correlation with clinical malaria. *PeerJ*, 10, e13487. <https://doi.org/10.7717/peerj.13487>
- Thierry-Mieg, D., & Thierry-Mieg, J. (2006). AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol*, 7 Suppl 1(Suppl 1), S12 11-14. <https://doi.org/10.1186/gb-2006-7-s1-s12>
- Thom, C. S., Dickson, C. F., Gell, D. A., & Weiss, M. J. (2013). Hemoglobin variants: biochemical properties and clinical correlates. *Cold Spring Harb Perspect Med*, 3(3), a011858. <https://doi.org/10.1101/cshperspect.a011858>
- Thomas, C., & Lumb, A. B. (2012). Physiology of haemoglobin. *Continuing Education in Anaesthesia Critical Care & Pain*, 12(5), 251-256. <https://doi.org/10.1093/bjaceaccp/mks025>
- Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., & Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*, 13(9), 2129-2141. <https://doi.org/10.1101/gr.772403>
- Thornley, I., Dror, Y., Sung, L., Wynn, R. F., & Freedman, M. H. (2002). Abnormal telomere shortening in leucocytes of children with Shwachman-Diamond syndrome. *Br J Haematol*, 117(1), 189-192. <https://doi.org/10.1046/j.1365-2141.2002.03371.x>
- Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*, 10(6), 1453-1465. [https://doi.org/10.1016/s1097-2765\(02\)00781-5](https://doi.org/10.1016/s1097-2765(02)00781-5)
- Tripathi, P., Agarwal, S., Mandal, K., Gupta, A., & Sarangi, A. N. (2023). Impact of Genetic Polymorphisms in Modifier Genes in Determining Fetal Hemoglobin Levels in Beta-Thalassemia. *Thalassemia Reports*, 13(1), 85-112. <https://www.mdpi.com/2039-4365/13/1/9>
- Tutar, L., Ozgur, A., & Tutar, Y. (2018). Involvement of miRNAs and Pseudogenes in Cancer. *Methods Mol Biol*, 1699, 45-66. https://doi.org/10.1007/978-1-4939-7435-1_3
- Tutar, Y. (2012). Pseudogenes. *Comp Funct Genomics*, 2012, 424526. <https://doi.org/10.1155/2012/424526>
- Tylki-Szymańska, A., & Stradomska, T. J. (2011). [A newly discovered metabolic diseases due to defects in the pentose pathway]. *Postepy Biochem*, 57(2), 168-171. (Nowo opisane choroby metaboliczne związane z błędami metabolizmu na szlaku przemiany pentoz.)
- Tylki-Szymańska, A., Stradomska, T. J., Wamelink, M. M., Salomons, G. S., Taybert, J., Pawłowska, J., & Jakobs, C. (2009). Transaldolase deficiency in two new patients with a relative mild phenotype. *Mol Genet Metab*, 97(1), 15-17. <https://doi.org/10.1016/j.ymgme.2009.01.016>

- 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., . . . 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*(5), 1620-1625. <https://doi.org/10.1073/pnas.0711566105>
- Udy, D. B., & Bradley, R. K. (2022). Nonsense-mediated mRNA decay uses complementary mechanisms to suppress mRNA and protein accumulation. *Life Sci Alliance*, *5*(3). <https://doi.org/10.26508/lsa.202101217>
- Valayannopoulos, V., Verhoeven, N. M., Mention, K., Salomons, G. S., Sommelet, D., Gonzales, M., Touati, G., de Lonlay, P., Jakobs, C., & Saudubray, J. M. (2006). Transaldolase deficiency: a new cause of hydrops fetalis and neonatal multi-organ disease. *J Pediatr*, *149*(5), 713-717. <https://doi.org/10.1016/j.jpeds.2006.08.016>
- Valentine, W. N. (1961). A specific erythrocyte glycolytic enzymedefect (pyruvate kinase) in three subjects with congenital nonspherocytic hemolytic anemia. *Trans Ass Amer Physicians*, *74*, 100-110.
- Valentini, G., Chiarelli, L., Fortin, R., Speranza, M. L., Galizzi, A., & Mattevi, A. (2000). The allosteric regulation of pyruvate kinase. *J Biol Chem*, *275*(24), 18145-18152. <https://doi.org/10.1074/jbc.M001870200>
- Valentini, G., Chiarelli, L. R., Fortin, R., Dolzan, M., Galizzi, A., Abraham, D. J., Wang, C., Bianchi, P., Zanella, A., & Mattevi, A. (2002). Structure and function of human erythrocyte pyruvate kinase. Molecular basis of nonspherocytic hemolytic anemia. *J Biol Chem*, *277*(26), 23807-23814. <https://doi.org/10.1074/jbc.M202107200>
- van der Crabben, S. N., Hennis, M. P., McGregor, G. A., Ritter, D. I., Nagamani, S. C., Wells, O. S., Harakalova, M., Chinn, I. K., Alt, A., Vondrova, L., Hochstenbach, R., van Montfrans, J. M., Terheggen-Lagro, S. W., van Lieshout, S., van Roosmalen, M. J., Renkens, I., Duran, K., Nijman, I. J., Kloosterman, W. P., . . . van Haften, G. (2016). Destabilized SMC5/6 complex leads to chromosome breakage syndrome with severe lung disease. *J Clin Invest*, *126*(8), 2881-2892. <https://doi.org/10.1172/jci82890>
- van Dijk, T. B., Gillemans, N., Pourfarzad, F., van Lom, K., von Lindern, M., Grosveld, F., & Philipsen, S. (2010). Fetal globin expression is regulated by Friend of Prmt1. *Blood*, *116*(20), 4349-4352. <https://doi.org/10.1182/blood-2010-03-274399>
- Van Ziffle, J., Yang, W., & Chehab, F. F. (2011). Homozygous deletion of six olfactory receptor genes in a subset of individuals with Beta-thalassemia. *PLoS One*, *6*(2), e17327. <https://doi.org/10.1371/journal.pone.0017327>
- Vandorpe, D. H., Rivera, A., Ganter, M., Dankwa, S., Wohlgemuth, J. G., Dlott, J. S., Snyder, L. M., Brugnara, C., Duraisingh, M., & Alper, S. L. (2022). Purinergic signaling is essential for full Psickle activation by hypoxia and by normoxic acid pH in mature human sickle red cells and in vitro-differentiated cultured human sickle reticulocytes. *Pflügers Archiv - European Journal of Physiology*, *474*(5), 553-565. <https://doi.org/10.1007/s00424-022-02665-z>
- Varawalla, Old, J. M., Sarkar, R., Venkatesan, R., & Weatherall, D. J. (1991). The spectrum of beta-thalassaemia mutations on the Indian subcontinent: the basis for prenatal diagnosis. *Br J Haematol*, *78*(2), 242-247. <https://doi.org/10.1111/j.1365-2141.1991.tb04423.x>
- Venegas, A. B., Natsume, T., Kanemaki, M., & Hickson, I. D. (2020). Inducible Degradation of the Human SMC5/6 Complex Reveals an Essential Role Only during Interphase. *Cell Reports*, *31*(3), 107533. <https://doi.org/https://doi.org/10.1016/j.celrep.2020.107533>
- Verhoeven, N. M., Huck, J. H. J., Roos, B., Struys, E. A., Salomons, G. S., Douwes, A. C., van der Knaap, M. S., & Jakobs, C. (2001). Transaldolase Deficiency: Liver Cirrhosis Associated with a New Inborn Error in the Pentose Phosphate Pathway. *The American Journal of Human Genetics*, *68*(5), 1086-1092. <https://doi.org/https://doi.org/10.1086/320108>

- Vichinsky, E. P. (2009). Alpha thalassemia major--new mutations, intrauterine management, and outcomes. *Hematology Am Soc Hematol Educ Program*, 35-41. <https://doi.org/10.1182/asheducation-2009.1.35>
- Vichinsky, E. P. (2013). Clinical manifestations of alpha-thalassemia. *Cold Spring Harb Perspect Med*, 3(5), a011742. <https://doi.org/10.1101/cshperspect.a011742>
- Vinjamur, D. S., & Bauer, D. E. (2018). Growing and Genetically Manipulating Human Umbilical Cord Blood-Derived Erythroid Progenitor (HUDEP) Cell Lines. *Methods Mol Biol*, 1698, 275-284. https://doi.org/10.1007/978-1-4939-7428-3_17
- 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(24), 2797-2802. <https://doi.org/10.1093/hmg/10.24.2797>
- Vo, L. T. T., Nguyen, T. T., Le, H. X., & Le, H. T. T. (2018). Analysis of Common beta-Thalassemia Mutations in North Vietnam. *Hemoglobin*, 42(1), 16-22. <https://doi.org/10.1080/03630269.2018.1428621>
- Voon, H. P., & Vadolas, J. (2008). Controlling alpha-globin: a review of alpha-globin expression and its impact on beta-thalassemia. *Haematologica*, 93(12), 1868-1876. <https://doi.org/10.3324/haematol.13490>
- Voskou, S., Aslan, M., Fanis, P., Phylactides, M., & Kleanthous, M. (2015). Oxidative stress in β -thalassaemia and sickle cell disease. *Redox Biology*, 6, 226-239. <https://doi.org/https://doi.org/10.1016/j.redox.2015.07.018>
- Vue, T. Y., Kollipara, R. K., Borromeo, M. D., Smith, T., Mashimo, T., Burns, D. K., Bachoo, R. M., & Johnson, J. E. (2020). ASCL1 regulates neurodevelopmental transcription factors and cell cycle genes in brain tumors of glioma mouse models. *Glia*, 68(12), 2613-2630. <https://doi.org/10.1002/glia.23873>
- Vulliamy, T. J., D'Urso, M., Battistuzzi, G., Estrada, M., Foulkes, N. S., Martini, G., Calabro, V., Poggi, V., Giordano, R., Town, M., & et al. (1988). Diverse point mutations in the human glucose-6-phosphate dehydrogenase gene cause enzyme deficiency and mild or severe hemolytic anemia. *Proc Natl Acad Sci U S A*, 85(14), 5171-5175. <https://doi.org/10.1073/pnas.85.14.5171>
- Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*, 8(2), 127-134. <https://doi.org/10.1093/protein/8.2.127>
- Wang, Qiao, J. D., Liu, X. R., Liu, D. T., Chen, Y. H., Wu, Y., Sun, Y., Yu, J., Ren, R. N., Mei, Z., Liu, Y. X., Shi, Y. W., Jiang, M., Lin, S. M., He, N., Li, B., Bian, W. J., Li, B. M., Yi, Y. H., . . . Liao, W. P. (2021). UNC13B variants associated with partial epilepsy with favourable outcome. *Brain*, 144(10), 3050-3060. <https://doi.org/10.1093/brain/awab164>
- Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16), e164-e164. <https://doi.org/10.1093/nar/gkq603>
- Wang, L. C., & Gautier, J. (2010). The Fanconi anemia pathway and ICL repair: implications for cancer therapy. *Crit Rev Biochem Mol Biol*, 45(5), 424-439. <https://doi.org/10.3109/10409238.2010.502166>
- Wang, W., Ma, E. S. K., Chan, A. Y. Y., Prior, J., Erber, W. N., Chan, L. C., Chui, D. H. K., & Chong, S. S. (2003). Single-Tube Multiplex-PCR Screen for Anti-3.7 and Anti-4.2 α -Globin Gene Triplications. *Clinical Chemistry*, 49(10), 1679-1682. <https://doi.org/10.1373/49.10.1679>
- Wang, X., Angelis, N., & Thein, S. L. (2018). MYB – A regulatory factor in hematopoiesis. *Gene*, 665, 6-17. <https://doi.org/https://doi.org/10.1016/j.gene.2018.04.065>
- Wang, X., & Thein, S. L. (2018). Switching from fetal to adult hemoglobin. *Nat Genet*, 50(4), 478-480. <https://doi.org/10.1038/s41588-018-0094-z>

- Wang, X., Yuan, L., Ye, Y., Lu, B., & Xu, X. (2022). Apoptosis inhibition in T-cell acute lymphoblastic leukemia by UNC13B. *Materials Express*, 12(5), 675-682.
- Wang, Y., Tan, J., Liu, D., Yang, Y., & Wu, H. (2019). The Association of UNC13B Gene Polymorphisms and Diabetic Kidney Disease in a Chinese Han Population. *Med Sci Monit*, 25, 8527-8533. <https://doi.org/10.12659/msm.919930>
- Wang, Lin, C. C., Hsu, C. L., Hung, S. Y., Yao, C. Y., Lee, S. H., Tsai, C. H., Hou, H. A., Chou, W. C., & Tien, H. F. (2021). Distinct clinical and biological characteristics of acute myeloid leukemia with higher expression of long noncoding RNA KIAA0125. *Ann Hematol*, 100(2), 487-498. <https://doi.org/10.1007/s00277-020-04358-y>
- Weatherall, D. J. (1980). The thalassemia syndromes. *Tex Rep Biol Med*, 40, 323-333. <http://www.ncbi.nlm.nih.gov/pubmed/7034274>
- Weatherall, M. W., Higgs, D. R., Weiss, H., Weatherall, D. J., & Serjeant, G. R. (2005). Phenotype/genotype relationships in sickle cell disease: a pilot twin study. *Clin Lab Haematol*, 27(6), 384-390. <https://doi.org/10.1111/j.1365-2257.2005.00731.x>
- Webb, B., & Sali, A. (2021). Protein Structure Modeling with MODELLER. *Methods Mol Biol*, 2199, 239-255. https://doi.org/10.1007/978-1-0716-0892-0_14
- Wickner, S., Maurizi, M. R., & Gottesman, S. (1999). Posttranslational quality control: folding, refolding, and degrading proteins. *Science*, 286(5446), 1888-1893. <https://doi.org/10.1126/science.286.5446.1888>
- Wickramasinghe, S. N., & Bush, V. (1975). Observations on the Ultrastructure of Erythropoietic Cells and Reticulum Cells in the Bone Marrow of Patients with Homozygous β -Thalassaemia. *British Journal of Haematology*, 30(4), 395-399. <https://doi.org/https://doi.org/10.1111/j.1365-2141.1975.tb01853.x>
- Wijgerde, M., Grosveld, F., & Fraser, P. (1995). Transcription complex stability and chromatin dynamics in vivo. *Nature*, 377(6546), 209-213. <https://doi.org/10.1038/377209a0>
- Wijk, & Solinge. (2005a). The energy-less red blood cell is lost: erythrocyte enzyme abnormalities of glycolysis. *Blood*, 106(13), 4034-4042. <https://doi.org/https://doi.org/10.1182/blood-2005-04-1622>
- Wijk, & Solinge. (2005b). The energy-less red blood cell is lost: erythrocyte enzyme abnormalities of glycolysis. *Blood*, 106(13), 4034-4042. <https://doi.org/10.1182/blood-2005-04-1622>
- Wilber, A., Nienhuis, A. W., & Persons, D. A. (2011). Transcriptional regulation of fetal to adult hemoglobin switching: new therapeutic opportunities. *Blood*, 117(15), 3945-3953. <https://doi.org/10.1182/blood-2010-11-316893>
- Wildmann, C., Larondelle, Y., Vaerman, J. L., Eeckels, R., Martiat, P., & Philippel, M. (1993). An Initiation Codon Mutation as a Cause of β -Thalassaemia in a Belgian Family. *Hemoglobin*, 17(1), 19-30. <https://doi.org/10.3109/03630269308998882>
- Williams, M., Valayannopoulos, V., Altassan, R., Chung, W. K., Heijboer, A. C., Keng, W. T., Lapatto, R., McClean, P., Mulder, M. F., Tylki-Szymańska, A., Walenkamp, M.-J. E., Alfadhel, M., Alakeel, H., Salomons, G. S., Eyaid, W., & Wamelink, M. M. C. (2019). Clinical, biochemical, and molecular overview of transaldolase deficiency and evaluation of the endocrine function: Update of 34 patients. *Journal of Inherited Metabolic Disease*, 42(1), 147-158. <https://doi.org/https://doi.org/10.1002/jimd.12036>
- Wong, T. E., Calicchio, M. L., Fleming, M. D., Shimamura, A., & Harris, M. H. (2010). SBDS protein expression patterns in the bone marrow. *Pediatr Blood Cancer*, 55(3), 546-549. <https://doi.org/10.1002/pbc.22573>
- Wongtong, N., Jones, S., Deng, Y., Cai, J., & Ataga, K. I. (2015). Monocytosis is associated with hemolysis in sickle cell disease. *Hematology*, 20(10), 593-597. <https://doi.org/10.1179/1607845415y.0000000011>

- Wood, B. L., Gibson, D. F., & Tait, J. F. (1996). Increased erythrocyte phosphatidylserine exposure in sickle cell disease: flow-cytometric measurement and clinical associations. *Blood*, *88*(5), 1873-1880. <https://www.ncbi.nlm.nih.gov/pubmed/8781447>
- Wooderchak-Donahue, W. L., O'Fallon, B., Furtado, L. V., Durtschi, J. D., Plant, P., Ridge, P. G., Rope, A. F., Yetman, A. T., & Bayrak-Toydemir, P. (2012). A direct comparison of next generation sequencing enrichment methods using an aortopathy gene panel- clinical diagnostics perspective. *BMC Med Genomics*, *5*, 50. <https://doi.org/10.1186/1755-8794-5-50>
- Wu, D., Zhang, L., Qiang, Y., & Wang, K. (2022). Improved detection of SBDS gene mutation by a new method of next-generation sequencing analysis based on the Chinese mutation spectrum. *PLoS One*, *17*(12), e0269029. <https://doi.org/10.1371/journal.pone.0269029>
- Wu, K. C., Cui, J. Y., & Klaassen, C. D. (2011). Beneficial role of Nrf2 in regulating NADPH generation and consumption. *Toxicol Sci*, *123*(2), 590-600. <https://doi.org/10.1093/toxsci/kfr183>
- 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*(9), 1526-1533. <https://doi.org/10.1101/gr.173427.114>
- Xu, J., Sankaran, V. G., Ni, M., Menne, T. F., Puram, R. V., Kim, W., & Orkin, S. H. (2010). Transcriptional silencing of gamma-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev*, *24*(8), 783-798. <https://doi.org/10.1101/gad.1897310>
- Xu, X. M., Zhou, Y. Q., Luo, G. X., Liao, C., Zhou, M., Chen, P. Y., Lu, J. P., Jia, S. Q., Xiao, G. F., Shen, X., Li, J., Chen, H. P., Xia, Y. Y., Wen, Y. X., Mo, Q. H., Li, W. D., Li, Y. Y., Zhuo, L. W., Wang, Z. Q., . . . Zhong, M. (2004). The prevalence and spectrum of α and β thalassaemia in Guangdong Province: implications for the future health burden and population screening. *Journal of Clinical Pathology*, *57*(5), 517-522. <https://doi.org/10.1136/jcp.2003.014456>
- Xue, J., Han, J., Zhao, X., Zhen, L., Mei, S., Hu, Z., & Li, X. (2022). Prenatal Diagnosis of Fetus With Transaldolase Deficiency Identifies Compound Heterozygous Variants: A Case Report [Case Report]. *Frontiers in Genetics*, *12*. <https://doi.org/10.3389/fgene.2021.752272>
- Yagami, T., Ballard, B. T., Padovan, J. C., Chait, B. T., Popowicz, A. M., & Manning, J. M. (2002). N-terminal contributions of the gamma-subunit of fetal hemoglobin to its tetramer strength: remote effects at subunit contacts. *Protein Sci*, *11*(1), 27-35. <https://doi.org/10.1110/ps.30602>
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat Methods*, *12*(1), 7-8. <https://doi.org/10.1038/nmeth.3213>
- Yang, K., Liu, X., Chen, K., Luo, S., Kong, W., Huang, W., & Xiao, J. (2022). A case of G6PD Utrecht associated with beta-thalassemia responding to splenectomy. *Pediatr Blood Cancer*, *69*(9), e29837. <https://doi.org/10.1002/pbc.29837>
- Yang, M. Y., Lin, P. M., Yang, C. H., Hu, M. L., Chen, I. Y., Lin, S. F., & Hsu, C. M. (2021). Loss of ZNF215 imprinting is associated with poor five-year survival in patients with cytogenetically abnormal-acute myeloid leukemia. *Blood Cells Mol Dis*, *90*, 102577. <https://doi.org/10.1016/j.bcmed.2021.102577>
- Ye, L., Chang, J. C., Lin, C., Sun, X., Yu, J., & Kan, Y. W. (2009). Induced pluripotent stem cells offer new approach to therapy in thalassemia and sickle cell anemia and option in prenatal diagnosis in genetic diseases. *Proc Natl Acad Sci U S A*, *106*(24), 9826-9830. <https://doi.org/10.1073/pnas.0904689106>
- Youssofian, H. (1996). Cytoplasmic localization of FAC is essential for the correction of a prerepair defect in Fanconi anemia group C cells. *J Clin Invest*, *97*(9), 2003-2010. <https://doi.org/10.1172/jci118635>
- 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*(6), 2040-2045. <https://doi.org/10.1182/blood-2002-02-0387>

- 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(7), 1856-1865. <https://doi.org/10.1172/jci31664>
- Yuan, G., Yang, S., Ng, A., Fu, C., Oursler, M. J., Xing, L., & Yang, S. (2020). RGS12 Is a Novel Critical NF- κ B Activator in Inflammatory Arthritis. *iScience*, 23(6), 101172. <https://doi.org/10.1016/j.isci.2020.101172>
- Yuan, G., Yang, S., & Yang, S. (2022). Macrophage RGS12 contributes to osteoarthritis pathogenesis through enhancing the ubiquitination. *Genes & Diseases*, 9(5), 1357-1367. <https://doi.org/https://doi.org/10.1016/j.gendis.2021.08.005>
- Zakaria, N. A., Bahar, R., Abdullah, W. Z., Mohamed Yusoff, A. A., Shamsuddin, S., Abdul Wahab, R., & Johan, M. F. (2022). Genetic Manipulation Strategies for β -Thalassemia: A Review. *Front Pediatr*, 10, 901605. <https://doi.org/10.3389/fped.2022.901605>
- Zakaria, N. A., Islam, M. A., Abdullah, W. Z., Bahar, R., Mohamed Yusoff, A. A., Abdul Wahab, R., Shamsuddin, S., & Johan, M. F. (2021). Epigenetic Insights and Potential Modifiers as Therapeutic Targets in β -Thalassemia. *Biomolecules*, 11(5), 755. <https://www.mdpi.com/2218-273X/11/5/755>
- Zamora, E. A., & Schaefer, C. A. (2022). Hereditary Spherocytosis. In *StatPearls*. <https://www.ncbi.nlm.nih.gov/pubmed/30969619>
- Zanella, A., Bianchi, P., Baronciani, L., Zappa, M., Bredi, E., Vercellati, C., Alfinito, F., Pelissero, G., & Sirchia, G. (1997). Molecular Characterization of PK-LR Gene in Pyruvate Kinase-Deficient Italian Patients. *Blood*, 89(10), 3847-3852. <https://doi.org/https://doi.org/10.1182/blood.V89.10.3847>
- Zanella, A., Fermo, E., Bianchi, P., Chiarelli, L. R., & Valentini, G. (2007). Pyruvate kinase deficiency: The genotype-phenotype association. *Blood Rev*, 21(4), 217-231. <https://doi.org/https://doi.org/10.1016/j.blre.2007.01.001>
- Zanella, A., Fermo, E., Bianchi, P., & Valentini, G. (2005). Red cell pyruvate kinase deficiency: molecular and clinical aspects. *Br J Haematol*, 130(1), 11-25. <https://doi.org/10.1111/j.1365-2141.2005.05527.x>
- Zhang, H. B., Liu, D. P., & Liang, C. C. (2002). The control of expression of the alpha-globin gene cluster. *Int J Hematol*, 76(5), 420-426. <https://doi.org/10.1007/BF02982807>
- Zhang, J., & Lodish, H. F. (2007). Endogenous K-ras signaling in erythroid differentiation. *Cell Cycle*, 6(16), 1970-1973. <https://doi.org/10.4161/cc.6.16.4577>
- Zhang, X., Li, J., Sejas, D. P., Rathbun, K. R., Bagby, G. C., & Pang, Q. (2004). The Fanconi anemia proteins functionally interact with the protein kinase regulated by RNA (PKR). *J Biol Chem*, 279(42), 43910-43919. <https://doi.org/10.1074/jbc.M403884200>
- Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9, 40. <https://doi.org/10.1186/1471-2105-9-40>
- Zhang, Y., Xu, Y., Zhang, S., Lu, Z., Li, Y., & Zhao, B. (2022). The regulation roles of Ca(2+) in erythropoiesis: What have we learned? *Exp Hematol*, 106, 19-30. <https://doi.org/10.1016/j.exphem.2021.12.192>
- Zhang, Z., Wang, J., He, J., Zheng, Z., Zeng, X., Zhang, C., Ye, J., Zhang, Y., Zhong, N., & Lu, W. (2013). Genetic variants in MUC4 gene are associated with lung cancer risk in a Chinese population. *PLoS One*, 8(10), e77723. <https://doi.org/10.1371/journal.pone.0077723>
- Zhou, D., Liu, K., Sun, C. W., Pawlik, K. M., & Townes, T. M. (2010). KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching. *Nat Genet*, 42(9), 742-744. <https://doi.org/10.1038/ng.637>
- Zhou, J. Y., Jiang, F., Li, J., Chen, G. L., & Li, D. Z. (2019). Coinheritance of Hb City of Hope (HBB: c.208G>A) and beta-Thalassemia: Compromising the Molecular Diagnosis of the Codons 71/72

- (+A) (HBB: c.216_217insA) Mutation by Reverse Dot-Blot Hybridization. *Hemoglobin*, 43(2), 145-147. <https://doi.org/10.1080/03630269.2019.1626741>
- Zhu, X., Wang, Y., Pi, W., Liu, H., Wickrema, A., & Tuan, D. (2012). NF- κ B Recruits Both Transcription Activator and Repressor to Modulate Tissue- and Developmental Stage-Specific Expression of Human γ -Globin Gene. *PLoS One*, 7(10), e47175. <https://doi.org/10.1371/journal.pone.0047175>
- Zivot, A., Lipton, J. M., Narla, A., & Blanc, L. (2018). Erythropoiesis: insights into pathophysiology and treatments in 2017. *Mol Med*, 24(1), 11. <https://doi.org/10.1186/s10020-018-0011-z>