Genetic epidemiology and functional studies of β-thalassaemia in Kilifi, Kenya

Alexander Waiganjo Macharia, MSc

A thesis submitted for the degree of Doctor of Philosophy

The Open University, UK

Affiliated Research Centre

KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya

Collaborating Establishments

University of Cambridge, UK

Karolinska Institutet, Stockholm, Sweden

Submission Date: 22/08/2023

Abstract

It is generally believed that β-thalassaemia is rare in sub-Saharan Africa; however, in studies conducted in Kilifi, we recently observed HbA_2 levels within the diagnostic range for β-thalassaemia (⩾4%) in multiple children and identified two β-thalassaemia mutations. In this thesis, I have investigated this observation further by documenting the prevalence, genetic types, and origins of β-thalassaemia in this region. Additionally, I have explored the malaria protective mechanisms associated with these mutations and examined their health consequences.

I found that in Kilifi, β-thalassaemia is defined by four different mutations, three β^0 (rs33959855, rs33941849, rs193922563) and one a β^+ mutation (rs35004220). I estimated the allele frequency for all mutations to be 0.3% and the birth prevalence of homozygosity to be approximately 1 in 100,000.

Based on evidence from a literature review and ethnolinguistic and haplotype analysis, I hypothesized that rs33959855 was introduced through admixture with the South Asian population during the Asian-Swahili trade in 1200CE. However, the origins of the other mutations remained inconclusive.

Through in vitro studies using β-thalassaemia RBCs and survival analysis, I observed a modest reduction in invasion by *Plasmodium falciparum* 3D7 parasites. However, β-thalassaemia was not associated either with under-5 mortality or protection against hospital admission with clinical malaria. These mutations were, however, associated with several negative outcomes. Specifically, co-inheritance of both β-thalassaemia and the HbS sickle mutation was the cause of 10% of all sickle cell disease cases in the region. Moreover, homozygosity for β-thalassaemia was associated with transfusion dependent β-thalassaemia major, as illustrated by a case report involving rs33941849 mutation. β-thalassaemia carriers were also at a higher risk of admission to hospital with severe anaemia, although the risk was mitigated by coinheritance of α-thalassaemia.

These findings underscore the importance of raising awareness and establishing guidelines for diagnosis and management of β-thalassaemia in Kilifi.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Prof Thomas N. Williams, Prof Julian Rayner and Prof Anna Farnert for their guidance, encouragement and critique of my work throughout this journey. Tom, you have been invaluable as the primary supervisor, I am grateful for your availability and willingness to discuss any challenges that arose.

A special note of thanks goes to members of my progress monitoring committee: Prof George Warimwe, Dr Isabella Oyier and Prof Peter Olupot-Olupot for their invaluable insights and mentorship from the beginning of my PhD.

Special thanks to Dr Gavin Band for his expert guidance on Bioinformatics analysis and to Dr Silvia Kariuki and Dr Sophie Uyoga for their unwavering support and expert guidance in parasite culture assays.

I would like also to thank two exceptional teams from the Human Genetic Factors (HGF) group. Firstly, the field and clinical team, including George Mochamah, Maureen Maoni, Gideon Nyutu, Ruth Mwarabu, Jacob Golijo, Metrine Tendwa, Aphonse Kazungu, Masudi Dzombo and Emmanuel Mabibo. Your assistance in setting up the field study and collecting samples from the participants during the challenging COVID-19 period is greatly appreciated. Secondly, the HGF laboratory team, comprised of Johnstone Makale, Brian Tawa, Patrick Ombati and Wilfred Nyamu, your support with laboratory processes has been indispensable. I also extend my gratitude to my desk mate John Ojal who patiently supported me when I got stuck on statistical analysis using R software.

I wish to express my deep appreciation to the KEMRI-Wellcome Trust Research Programme led by Prof Philip Bejon for providing a conducive research environment. I am also grateful to Prof Sam Kinyanjui, Dr Dorcas Mbuvi and Liz Murabu for support on administrative matters.

Above all, I am grateful to the study participants whose samples were used in this study, I hope the outcome of the analysis will result in a positive impact to your health and that of the broader community.

Dedication

To my wife Monica Mwikamba and my children Tasha Wanjiku and Marcus Macharia. No words can adequately express my gratitude for your unwavering support and understanding throughout this journey. God bless you.

Table of Contents

List of figures

List of tables

List of Publications from this thesis

1. **Macharia AW**, Mochamah G, Makale J et al. Case Report: β-thalassemia major on the East African coast [version 1; peer review: 1 approved]. Wellcome Open Res 2022, 7:188 (https://doi.org/10.12688/wellcomeopenres.17907.1)
2. **Macharia AW**, Mochamah G, Uyoga S, Ndila CM, Nyutu G, Tendwa M, Nyatichi E, Makale J, Ware RE, Williams TN. β-Thalassemia pathogenic variants in a cohort of children from the East African coast. Mol Genet Genomic Med. 2020 Jul;8(7):e1294. doi: 10.1002/mgg3.1294. Epub 2020 May 11. PMID: 32394645; PMCID: PMC7336762.

Chapter 1: Literature Review

1.1 Background of thesis

I was inspired to investigate β-thalassaemia in Kilifi after observing elevated HbA_2 levels, a diagnostic marker for β-thalassaemia, among members of a birth cohort from Kilifi, located on the Kenyan coast (Macharia et al., 2019). This observation was further confirmed in a separate study in which we were investigating the use of hydroxyurea among children with sickle cell disease (REACH trial), where we found out that 10/151 had HbS/β-thalassaemia and 2 β-thalassaemia mutations were identified in this study (McGann et al., 2018), an observation that is discussed in detail in later chapters. This was strange as β-thalassaemia had not been described before in this population. I found it important to investigate this condition further as the outcome of this investigations could inform public health policy on the prevalence, burden of disease associated with the condition, and the affected population, which are important pieces of information for planning interventions such as creating public awareness and in the implementation of carrier and new-born screening programmes.

In this chapter, I perform a literature review to better understand the molecular, cellular, genetics and pathophysiology of β-thalassaemia. Because other haemoglobinopathies are common within this region, I review the consequences of the co-inheritance of β-thalassaemia with α-thalassaemia and sickle cell haemoglobin, the two most common haemoglobinopathies in the Kilifi area. I then look at the global distribution of the β-thalassaemias with a specific focus on sub-Saharan Africa. Lastly, based on observations we have made in Kilifi, that both the sickle mutation and α-thalassaemia are protective against malaria in this region, I review reports on the malaria protective effects of β-thalassaemia and their possible mechanisms.

1.2 Structure and synthesis of normal haemoglobin

Haemoglobin is an important intra-erythrocytic molecule, the main functions of which are to carry oxygen from the lungs to the tissues and carbon dioxide from the tissues to the lungs. Structurally, each haemoglobin molecule is made up of an iron containing molecule known as haem surrounded by four polypeptide globin chains. While the essential role of haem is oxygen transport, the globin molecules protect the haem molecule from oxidation, render it soluble and permit variation in oxygen affinity (Bain, 2020).

In humans, different haemoglobins are expressed during embryonic, foetal, and adult life. The synthesis of these haemoglobins is controlled by two developmentally regulated multigene clusters: the α-globin like cluster located on chromosome 16 and the β-globin like cluster on chromosome 11 (Figure 1). The major forms of haemoglobin that are produced during early embryonic life, consist of α-like Zeta (ζ) chains combined with β-like γ (Hb Portland, ζ2γ2) or epsilon chains (Hb Gower 1, ζ2ε2). Also produced during early embryonic life is Hb Gower 2 consisting of α and ε chains (α2ε2) (Figure 1) (Bain, 2020; Weatherall & Clegg, 2001; Wood, 1976).

Figure 1. Changes in globin genes during prenatal and postnatal age.

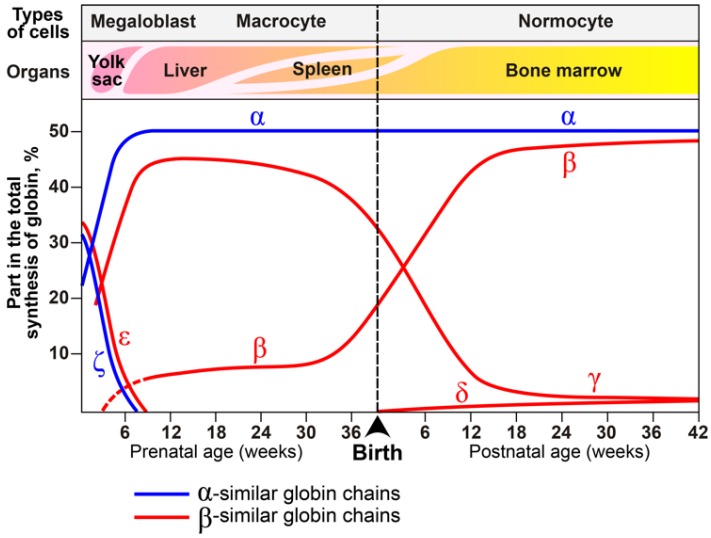


Image shows gene expression of various haemoglobin molecules during prenatal and postnatal period and the organs involved. During the first six weeks of gestational age embryonic genes are expressed resulting in synthesis of ζ-, ε-, α- and γ-globin. The dominant haemoglobins include Portland, $\zeta_2\gamma_2$) Hb Gower 2 ($\zeta_2\varepsilon_2$) HbF ($2\alpha2\gamma$). Six months after birth the production of HbF declines and adult haemoglobin (HbA) consisting of $2\alpha2\beta$ dominates the type of haemoglobin present in adult life. Expression of δ-globin is also activated resulting in synthesis of HbA_2 consisting of $2\alpha2\delta$. Postnatal genetics, original by (Wood, 1976) derivative work by Leonid 2 https://commons.wikimedia.org/wiki/File:Postnatal_genetics.svg under a CC-BY-SA 3.0 licence.

At around 6-8 weeks of gestation a switch occurs within the α-globin gene cluster, whereby the synthesis of ζ chains is silenced at the same time as that of α-chains is increased. During the same period, at the β-globin gene cluster the synthesis of ϵ chains is switched off and the production of γ chains increases steadily. Two types of γ-globin chains are produced namely Gγ and Aγ which differ only at amino acid 136, being glycine in Gγ and alanine in Aγ. The haemoglobin formed between the α and γ chains is referred to as foetal haemoglobin (HbF) (Figure 1). HbF has a higher affinity for oxygen than adult haemoglobin ($\alpha_2\beta_2$; HbA_0), a characteristic that makes it efficient in extracting haemoglobin from the placenta (Bain, 2020; Weatherall & Clegg, 2001; Wood, 1976).

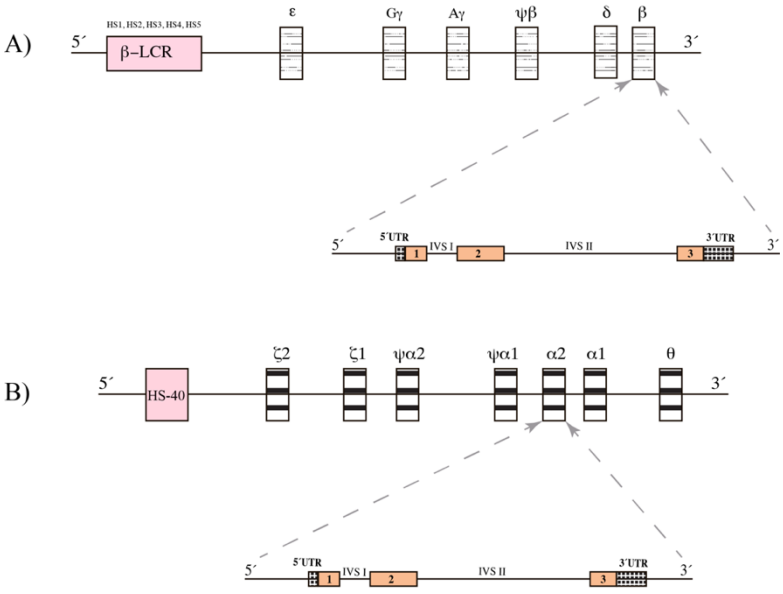
In the first 6 months after birth, the production of HbF steadily declines to levels of 2.3-3.5% and is maintained at these levels throughout adult life. Unlike foetal life, where HbF is produced in all erythrocytes, in adults it is restricted to a subset of cells referred to as F-cells, that produce both HbF and HbA. Concurrent with the reduction in synthesis of γ-globin chains is the increase in synthesis of β-globin chains. Two β-globin chains combine with two α-globin chains to form HbA, which constitutes 96-98% of haemoglobin in adult life (Bain, 2020; Weatherall & Clegg, 2001; Wood, 1976).

A smaller number of β-globin like delta (δ) chains are also produced, which combine with α-globin chains to form haemoglobin A_2 (HbA_2). The percentage of HbA_2 at birth is usually between 0.2% and 0.3%, which subsequently rises to adult levels of 2%-3.5% during the first two years of life. Unlike HbF, HbA_2 has a pan-cellular distribution (Figure 1). Although the physiological functions of HbA_2 are unknown, it is a haemoglobin of interest to the medical field due to its ability to inhibit polymerization of HbS and its importance in the diagnosis of β-thalassaemia carriers, in whom it is elevated (M. H. Steinberg & Rodgers, 2015).

Genetics of haemoglobin synthesis

The genes encoding α-like and β-like globins are arranged in clusters from 5´ to 3´ on chromosomes 16 and 11, respectively (Figure 2). The two clusters share some similarities, including the fact that they both include pseudogenes which are non-functional homologues that are transcribed but not translated to protein. Similarly, all the coding genes are organised in a similar format consisting of three exons interspersed by two introns and flanked on both ends by non-coding sequences known as untranslated regions (UTRs) (Figure 2). Next to the 5´ UTR of each gene is a promoter sequence, that RNA polymerase and other transcription factors bind to start transcription. Several intergenic elements, referred to as enhancers, also play the important role of increasing transcriptional activity of certain promoters. In addition to these promoter and enhancer sequences, each gene cluster has an upstream regulatory region that plays an important role of gene expression at various stages of erythrocyte development. This region is known as the β-locus control region (β-LCR) in the β-globin cluster, and as HS-40 in α-globin gene cluster (Bain, 2020; Weatherall, 2001).

Figure 2. Organisation of β-globin like and α-globin like gene clusters on chromosomes 11 (A) and 16 (B).



Images A and B show the organisation of the β-globin like and α-globin like gene clusters on chromosomes 11 and 16 respectively. From the 5´ to 3´ end of each cluster is the locus control region followed by the genes as organised on each chromosome. The expanded region shows the location the 3 exons and 2 introns present in each gene. Also highlighted in the same region is the 5´ and 3´ untranslated regions (UTRs). Figure based on a description by (Clark & Thein, 2004).

Synthesis of the α- and β-globins takes place in erythrocyte precursors between pro-erythroblast and reticulocyte stages. Activation and expression of the genes that encode them follows the order in which they are arranged, starting from the 5´ to 3´ end (Michela, Raffaele, Stella, Maria Rosaria, & Paola, 2012; Weatherall, 2001). This process can be divided into three main stages: transcription, RNA splicing and translation. I discuss the synthesis of β-globin below, but a similar process occurs in α-globin synthesis.

- Transcription: This is the process by which DNA is translated to RNA, a process that involves RNA polymerase. Transcription is initiated and controlled through the interaction of genes with transcription factors, promoters and β-LCR. β-LCR contains 5 DNAse I hypersensitive erythroid specific sites (HS1, HS2, HS3, HS4, HS5), of which HS3 is the most important as it opens the chromatin to allow transcription factors to access the site and to interact with specific globin gene promoters that are selected based on development stage. Many erythroid-specific transcriptional factors interact with LCR and gene promoters to activate or repress transcription of genes, but the main ones are EKLF, GATA-1, GATA-2 and NF-E2. Enhancers also play a role by increasing transcription of selected promoters. Once a gene is transcribed, a nascent RNA is formed which is stabilised by the addition of 7-methylguanosine cap (CAP) at the 5´ end and a polyadenylate tail at the 3´ end (Alberts, 2008; Bain, 2020; Michela et al., 2012).
- RNA splicing: After synthesis, the nascent RNA undergoes splicing, which is the process through which introns are removed from the sequence. Splicing requires recognition of intron junctions of which vast majority are conserved and start with GT and end with an AG (also referred to as the GT-AG rule). Mutation in these sequences causes abnormal splicing, which results in the synthesis of abnormal RNA and consequently reduced or absence synthesis of β-globin (Alberts, 2008; Bain, 2020; Kaminsky, Kruger, Hempelmann, & Bommer, 1986).

- Translation: After splicing, the mRNA moves to the cytoplasm where it is used as a template for synthesis of β-globin polypeptide, a process known as translation. During translation the 80S ribosome attaches to the mRNA and scans for the initiation codon which is a triplet codon AUG that encodes for amino acid methionine and is usually the first codon at the 5´ end. Mutations that result in a change at this codon to a different amino acid sequence usually result in a complete abrogation of β-globin synthesis (β^0-thalassaemia). Addition of amino acids to a growing β-polypeptide chain is done by transfer RNA (tRNA). Each tRNA is specific to a nucleotide codon and an amino acid. tRNA recognizes the mRNA in a triplet codon manner and attaches to the ribosome once it recognises a matching triplet codon on the mRNA. The ribosomes then bind the amino acid to the growing polypeptide chain. Translation continues until the ribosomes encounter codons UAA, UAG or UGA that encodes for a stop codon, at which point the ribosomes terminate protein synthesis. The termination codon is usually found after the 3´ UTR. Some β-thalassaemia mutations create a stop codon prematurely resulting in premature termination of protein synthesis. Depending on the positioning of the stop codon it may result in synthesis of an abnormal/non-functional β-globin polypeptide chain, causing β-thalassaemia (Alberts, 2008; Bain, 2020; Michela et al., 2012).

At the end of translation, the β-globin chains form a dimer with the α-globin chains after which haem is in cooperated by mitochondria to form a haemoglobin molecule.

1.3 Inherited haemoglobin disorders

Haemoglobin disorders are among the most common inherited conditions in the world, occurring at an estimated combined carrier frequency of 4.5% and with a birth rate of ~2/1000. Approximately 75% of all affected children are born in Africa (WHO, 1989). These disorders can broadly be classified into two groups:

1. Structural disorders, which result in a change in the structure of the individual globin chains of haemoglobin.
2. The thalassaemias, which result from a quantitative reduction in the synthesis of normal haemoglobin chains.

Structural disorders

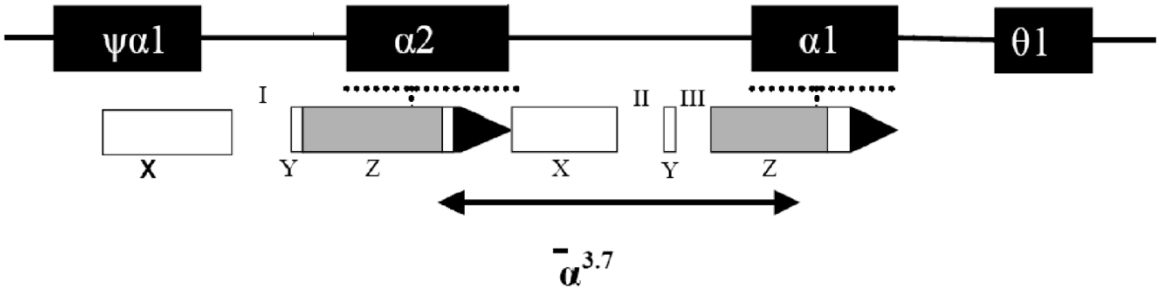
Globally, the most common structural disorders of haemoglobin are haemoglobin S (HbS), haemoglobin C (HbC) and haemoglobin E (HbE), all of which result from single nucleotide polymorphisms (SNPs) within the β-globin gene. HbS, the most common structural haemoglobin defect in sub-Saharan Africa, results from a SNP change of A>T at position 5248232 (GRCh37 reference genome) of chromosome 11, which results in the substitution of the amino acid valine for glutamic acid at position 6 of the β-globin polypeptide chain (M. H. Steinberg, Forget, B. G., Higgs, D. R., & Weatherall, D. J., 2009). The hydrophobic interactions between β^s-valine on one haemoglobin tetramer and β-85 phenylalanine and β-88 leucine of an adjacent tetramer results in HbS polymerization. Once these polymers are formed, they cause the RBC to become rigid and, through a cascade of subsequent events, result in vaso-occlusion and end-organ damage (Stuart & Nagel, 2004).

The thalassaemias

The thalassaemias are classified according to which globin chain is affected. The most common types are those that affect the α- and β-globin chains, that are referred to as the α- and β-thalassaemias, respectively (Harteveld & Higgs, 2010; Weatherall & Clegg, 2001). The α-thalassaemias are caused by the deletion of one or more of the two α-globin genes located on each copy of chromosome 16. Less frequently observed are the non-deletion types of α-thalassaemia, that result from point mutations in critical regions such as the promoter region. The α-thalassaemias can broadly be classified into two groups: the α^+-thalassaemias in which

the mutations result in reduced expression of α-globin chains and α^0-thalassaemias in which the mutations completely abolish synthesis of α-globin chains. One of the most common α^+-thalassaemias results from a 3.7kb deletion ($\alpha^{-3.7}$) caused by an unequal recombination of homologous segments (Z boxes) that are 3.7kb apart (Figure 3). This unequal crossover produces a 3.7kb deletion on one of the chromosomes and an α-triplicated segment on the alternative chromosome. A similar misalignment of the X boxes which are 4.2kb apart results in a 4.2kb deletion ($\alpha^{-4.2}$) on one chromosome and an α-triplicated segment on the alternative chromosomes (Harteveld & Higgs, 2010; M. H. Steinberg, Forget, B. G., Higgs, D. R., & Weatherall, D. J., 2009).

Figure 3. A section of chromosome 16 showing the location of 3.7kb deletion resulting in α^+-thalassaemia.



The region marked XYZ are homologous regions located along the chromosome. It is the misalignment of these homologous regions during meiosis that results in unequal crossover of chromosomal segments. The arrow below the XYZ boxes shows the span of the 3.7kb deletion that results from misalignment of the Z boxes. I drew this image based on the description by (Weatherall & Clegg, 2001)

The α^0-thalassaemias are caused by the partial or complete deletion of both of the α-globin genes on a single chromosome in cis, resulting in no synthesis of α-globin chains. The length of these deletions, which are geographically diverse, are usually named by the region where the mutation was first discovered. The two most common are the –SEA and –MED that are found in south-east Asia and the Mediterranean, respectively (Kalle Kwaifa, Lai, & Md Noor, 2020).

In general terms, the severity of the various forms of α-thalassaemia depends on the number of α-globin genes that are deleted and can range from mild and asymptomatic anaemia when only one copy is deleted to severe and fatal forms when 3 (HbH disease) or 4 copies (Hb Barts

Hydrops Fetalis) are deleted. The 3.7kb mutation is the most common mutation seen in sub-Saharan Africa, where it is found at frequencies of between 5% and 40% (Piel & Weatherall, 2014). In our previous studies we have identified the 3.7kb mutation as the main cause of α-thalassaemia in Kilifi where it is found at an estimated frequency of 40% (Ndila et al., 2020; M. H. Steinberg, Forget, B. G., Higgs, D. R., & Weatherall, D. J., 2009; Williams, Wambua, et al., 2005).

The second most common type of thalassaemia, the β-thalassaemias, are caused by mutations in *HBB* that result in either reduced (β^+) or absent (β^0) synthesis of normal β-globin. The pathophysiological consequences of the different forms of β-thalassaemia are largely dependent on the degree by which they reduce the amount of β-globin that is synthesized (Weatherall & Clegg, 2001), as discussed in detail below.

1.4 Pathophysiology of the β-thalassaemias

Any reduction in β-globin chain production results in an excess in unpaired α-globin chains, which precipitate in both RBCs and RBC precursors to result in ineffective erythropoiesis and premature RBC destruction. The β-thalassaemias can be classified both genetically and based on their associated clinical syndromes (Table 1). The genetic classification is based on the type of mutation inherited, with β^+ representing mutations that result in the reduced synthesis and β^0 in absent synthesis of normal β-globin molecules. The clinical classification is based on disease severity. The most severe form of β-thalassaemia occurs when two β^0 alleles are inherited together, resulting in a complete absence of β-globin chain production, a condition referred to as β-thalassaemia major. Such individuals are transfusion dependent from the time HbF production has completely waned – usually at around two years of age. Patients with β-thalassaemia intermedia experience mild to moderate disease, although in some cases it may result in severe anaemia and a need for regular blood transfusions. Conversely, individuals with β-thalassaemia trait (β/β^0 or β/β^+) are generally asymptomatic, although some may manifest with mild, hypochromic, microcytic anaemia, elevated HbA_2 levels and varying HbF levels

(Chonat & Quinn, 2017; Dharmawardena, Premaratne, Wickremasinghe, Mendis, & Fernando, 2022; Weatherall & Clegg, 2001). Some, studies, however, have reported a connection between β-thalassaemia trait and specific clinical outcomes. These include increased hospital admission with cholelithiasis, cirrhosis, kidney disease, mood disorders (Graffeo et al., 2017) and myocardial infarction (Gallerani et al., 1990). It's worth noting that most of the studies on clinical outcomes associated with β-thalassaemia have been carried out in non-malaria endemic regions and may not be directly translatable to malaria endemic countries. In malaria endemic regions, the coexistence of factors such as presence of iron deficiency anaemia, malaria, hookworm and schistosomiasis may further exacerbate effects of anemia.

Table 1. Classification of the β-thalassaemias.

Common genotype	Clinical syndrome	Phenotype
β/β	Normal	Normal
β/β^{+} β/β^{0}	β-thalassaemia trait	Thalassaemia minor: asymptomatic, mild microcytic anaemia
β^{+}/β^{+} β^{+}/β^{0} β^{+}/β^{0}	β-thalassaemia intermedia	Variable severity Mild to moderate anaemia Possible extramedullary haematopoiesis Iron overload
β^{0}/β^{0}	β-thalassaemia major (Cooley's Anaemia)	Severe anaemia Transfusion dependence Extramedullary haematopoiesis Iron overload
β^{s}/β^{+} β^{s}/β^{0}	HbS/β-thalassaemia	Sickle cell disease

Because of the association between haemoglobin disorders and malaria (discussed in subsequent sections), the structural mutations and the thalassaemias often co-exist in the same populations and can be co-inherited to result in a range of clinical outcomes. For example, when β-thalassaemia is inherited with the sickle mutation (β^{s}) on the contralateral chromosome, it results in HbS/β-thalassaemia, a specific cause of sickle cell disease (SCD). Although in general, the clinical manifestations of HbS/β-thalassaemia are similar to those of HbSS (whereby the two

chromosomes have a sickle mutation), they can vary depending on the type of β-thalassaemia mutation that is co-inherited, which in turn varies from region to region (M. H. Steinberg, Forget, B. G., Higgs, D. R., & Weatherall, D. J., 2009). In particular, depending on the amount of β-globin produced, individuals with HbS/β^{+}-thalassaemia often have a milder form of SCD than those with HbS/β^{0}-thalassaemia or HbSS (Bennani et al., 1993; Jha, Mishra, Verma, Pandey, & Lakkakula, 2018; Serjeant, Sommereux, Stevenson, Mason, & Serjeant, 1979; Yadav et al., 2016). The few studies that have been conducted have found that patients with HbS/β-thalassaemia are more likely to have splenomegaly, elevated HbA_2 values and reduced mean cell volume (MCV) and mean cell haemoglobin (MCH) values in comparison to individuals with HbSS disease. Lower MCV and MCH values suggest that HbS/β-thalassaemia patients could have reduced haemolysis and therefore less anaemia that not only results in fewer haemolytic complications but also in increased blood viscosity and a consequent increase in the frequency of vaso-occlusive events. Most studies on this condition have been conducted in Indian populations where, coincidentally, clinical sequelae of HbSS are also more mild; moreover, none have accounted for the effect of α-thalassaemia genotype status (Jha et al., 2018; Yadav et al., 2016) making their interpretation a little more difficult. The type of HbSS found in most sub-Saharan countries has been observed to be a more severe phenotype that originated on a Bantu/CAR haplotype (Pagnier et al., 1984; Powars, 1991). This together with the high prevalence of $\alpha^{-3.7}$-thalassaemia within the region presents the opportunity for interaction with β-thalassaemia that could potentially alter the clinical picture. This has been seen in other regions where, for example, the coinheritance of β-thalassaemia intermedia with two α-globin gene deletions (-α/-α) results in a milder disease that requires fewer transfusions, whereas inheritance of β-thalassaemia trait with the triplicated α-globin may result in a severe form anaemia that resembles that of thalassaemia intermedia (M. H. Steinberg, Forget, B. G., Higgs, D. R., & Weatherall, D. J., 2009; Traeger-Synodinos et al., 1996).

1.5 β-thalassaemia and HbA_2

The association between elevated levels of HbA_2 and β-thalassaemia heterozygosity was first identified by Kunkel and colleagues in 1957 (Kunkel, Ceppellini, Muller-Eberhard, & Wolf, 1957). They found a mean range for HbA_2 of 5.11% (SD 1.6%) in 34 β-thalassaemia carriers and of 2.54% (0.35%) in 300 normal adults. In the same study twelve children with homozygous β-thalassaemia had variable HbA_2 levels, most having levels within the normal range (Kunkel et al., 1957). These observations have subsequently been confirmed in numerous other studies (Gerald & Diamond, 1958; Mosca, Paleari, Ivaldi, Galanello, & Giordano, 2009; Stephens et al., 2012; Van Delft et al., 2009; Weatherall, 1964).

Although the mechanism resulting in high levels of HbA_2 among β-thalassaemia heterozygotes remains unknown, it has been suggested that it might be linked to post transcriptional and post-translational effects. For post translation effects, reduced β-chain synthesis in β-thalassaemia carriers results in excess unpaired α-chains that likely pair up with δ-chains to form α/δ dimers and consequently result in an increase in HbA_2. Transcription effects occur when there is reduced competition for transcription factors due to inefficient binding of transcription factors to the altered β-globin promoters. The free transcription factors are more likely to bind to the inefficient δ-globin promoter, resulting to increase in transcription of δ-globin gene. Evidence supporting this mechanism was observed by Codrington and colleagues (Codrington et al., 1990), who found that δ-globin gene transcription is likely to be increased only when in *cis* with the β-thalassaemia mutation.

It is estimated that heterozygous β-thalassaemia is the cause in 95% of cases of elevated HbA_2 levels (Weatherall & Clegg, 2001). Alternative but less common causes include SCD, megaloblastic anaemia, Zidovudine HIV treated patients and Hyperthyroidism (Bain, 2020; Kendall & Bastomsky, 1981; Routy et al., 1993). HbA_2 is reduced in iron deficiency (Kattamis, Lagos, Metaxotou-Mavromati, & Matsaniotis, 1972; Mohanty et al., 2014) and by α- and δ-thalassaemia (Mosca et al., 2009; M. H. Steinberg, Forget, B. G., Higgs, D. R., & Weatherall, D. J., 2009; Stephens et al., 2012; Van Delft et al., 2009). The reduction of HbA_2 in the presence

of α-thalassaemia is related to increased competition for dimerization between β- and δ-globin chains for the limited number of β-globin chains present in α^+-thalassaemia that favors α/β as compared to α/δ dimers. This is supported by studies conducted in vitro that show evidence that preferential dimerization occurs in the order of $\alpha/\beta > \alpha/\delta > \alpha/\beta^s$ (Martinez & Menendez, 1983; Shaeffer, 1980). The same principles may also explain why elevated HbA_2 is observed among HbSS participants with co-inherited α^+-thalassaemia, whereby preferential dimerization would favor α/δ as compared to α/β^s dimers (Macharia et al., 2019; M. H. Steinberg & Embury, 1986).

In a study by Van Delft and colleagues, the normal range of HbA_2 in healthy adults after excluding iron deficient and α-thalassaemic individuals and accounting for differences in equipment precision was found to be between 2.5% and 3.5% (Van Delft et al., 2009). Similarly, in 766 individuals from other cohorts, as summarized by Stephens and colleagues (Stephens et al., 2012) , the normal range was found to lie between 2.2% and 3.3% and rarely outside the range of 2.0% to 3.5% unless in the presence of iron deficiency where the range went below 2.0%. In individuals with common types of β-thalassaemia trait from the same study, HbA_2 concentrations were found to lie within 4.0% to 6.0% of total haemoglobin and rarely outside the range of 3.5% to 7.0% unless when inherited with α/δ-thalassaemias or HbS (Colaco & Nadkarni, 2021; Stephens et al., 2012). Similar observations were made by Mosca and colleagues (Mosca et al., 2009) and by Weatherall and Clegg (Weatherall & Clegg, 2001). Based on these findings, the International Council for the Standardization of Haematology (ICSH) recommends that in the diagnosis of β-thalassaemia carriers, all haematological parameters should be considered, and experience used when interpreting chromatograms and electropherograms that include other haemoglobin variants (Stephens et al., 2012).

1.6 The molecular basis of β-thalassaemia

To date, more than 300 different β-thalassaemia mutations have been identified, most being single nucleotide substitutions, small insertions or deletions involving a few nucleotides and

with large deletions being rare (M. H. Steinberg, Forget, B. G., Higgs, D. R., & Weatherall, D. J., 2009; Thein, 2018). The non-deletional mutations are generally grouped based on the stage of gene expression that is affected. Some examples are listed below:

- Transcriptional mutations: These mutations affect the 5´UTR and promoter sequences (either CACCC or TATA box) of *HBB* to result in mild or minimal reduction in β-globin synthesis. These mutations largely result in β^+-thalassaemias. One example of such mutations is rs34598529, an A>G single nucleotide polymorphism at position −29 within the TATA box. This mutation is mainly found in black people and is associated with a β^+-thalassaemia phenotype among black population but a more severe phenotype among people of Chinese ancestry (Gonzalez-Redondo et al., 1988; Huang et al., 1986).
- RNA processing mutations: These include mutations that affect the exon-intron splice sites and cryptic splice sites. Mutations at the exon-intron splice site prevent normal splicing of pre-mRNA to result in β^0-thalassaemias. Mutations affecting the cryptic splice sites result in the creation of a new splice site that causes premature splicing and subsequent generation of defective mRNA that includes introns. Cryptic site mutations can result in both β^+ or β^0-thalassaemia mutations depending on the ratio of normal to abnormal mRNA generated. An example is rs35004220, a β^+ pathogenic variant resulting from a G>A change at position IVS-I-110 that causes creation of a new splice that results in synthesis of 80% abnormal RNA (Macharia et al., 2020; Thein, 2013).
- Translational mutations: These mutations cause premature termination of RNA synthesis at the beginning (initiation codon mutation) or at any subsequent stage of RNA translation. The mutations include nonsense, frameshift, and initiation codon mutations. Translational mutations account for the most common types of β-thalassaemia and often result in β^0-thalassaemia. An example of this mutation is rs33941849 an ATG to ACG initiation codon mutation that abrogates the transfer RNA binding site (Thein, 2018; Wildmann et al., 1993).

- Deletional mutations: Although rare, β-thalassaemia can be caused by deletions affecting the β-globin gene, most ranging from 25bp to 6kb. The type of β-thalassaemia resulting from these deletions is often the β^0-thalassaemia. An example is the rs193922563 β^0-thalassaemia mutation that results from a 25bp deletion and causes inactivation of an acceptor splice site (Macharia et al., 2020; McGann et al., 2018; Thein, 2018).

1.7 Global distribution of β-thalassaemia

The prevalence of β-thalassaemia is highest in the Mediterranean region, Middle East, Southeast Asia, India, Melanesia and parts of Africa, specifically within countries in the North. It is estimated that out of the 300 mutations identified worldwide, more than 80% of cases are caused by approximately 20. Each region has a unique cluster of β-thalassaemia mutations out of which only a handful (usually <5) are dominant (Thein, 2018; Weatherall & Clegg, 2001).

1.7.1 β-thalassaemia in Africa

Most reports on the prevalence of β-thalassaemia in Africa are from the North. For example, in Morocco, the carrier frequency of β-thalassaemia has been estimated at between 1.5% to 3% (WHO, 1989) while in Algeria a prevalence of 3% has been reported by Labie and colleagues (Labie, Bennani, & Beldjord, 1990). Similar frequencies of between 3 and 4% have been reported in Egypt (Habib & Book, 1982; Novelletto et al., 1990). Although the mutations causing β-thalassaemia in Tunisia have been documented in a number of studies (Fattoum, Messaoud, & Bibi, 2004; Sahli et al., 2016), none have reported their carrier frequencies. In other parts of Africa, carrier frequencies of between 0.8% to 1.7% have been reported in limited parts of Nigeria (Esan, 1970) and Ghana (Weatherall & Clegg, 2001), whereas in East and central Africa, reports have largely been limited to case reports (Frischknecht, Troxler, Greiner, Hengartner, & Dutly, 2008; McGann et al., 2018; Waye et al., 2015). In west Africa, frequencies

as high as 9% have been recorded in specific ethnolinguistic groups within Liberia (M. C. Willcox, 1975).

Reports on the molecular characterisation of β-thalassaemia mutations have also been largely limited to North Africa (Table 2). Like other regions, more than 50% of β-thalassaemia in each of these countries has been attributed to only three mutations. For example, in Morocco, CD39 (C>T), FSC6 (-A) and FSC8 (-AA) are the most common mutations, accounting for >50% of β-thalassaemia chromosomes reported. Similar observations were seen for the rest of the region: Algeria (CD39 (C>T), IVS-I-110(G>A), FSC6 (-A)); Tunisia (CD39 (C>T), IVS-I-110(G>A), IVS-I-I (G>A)); and Egypt (IVS-I-110(G>A), IVS-I-I (G>A), IVS-I-6 (C>T)) (Table 2).

Even at a regional level, there are still differences in the prevalence of β-thalassaemia chromosomes. One example is IVS-I-6 (C>T), which is found at a high frequency in Egypt but has not been described at all in the other North African countries. Another mutation that shows a difference in prevalence is the IVS-I-110(G>A) which is found at a high prevalence in Algeria, Tunisia, and Egypt but not in Morocco. Several studies suggest the mutation might have originated from the Eastern Mediterranean, having been associated with ~50% of the β-thalassaemia chromosomes identified in Turkey (Guzelgul 2020 and Keser 2004). Tadmouri and colleagues (Tadmouri, Garguier, Demont, Perrin, & Basak, 2001) hypothesized that the mutation then spread through gene flow during the period of the Ottoman/Turkish empire in the 16th century, which controlled Southeast Europe, Western Asia, and Northern Africa (specifically Egypt, Algeria, and Tunisia). Morocco was not part of the Ottoman empire which could explain the low prevalence of IVS-I-110(G>A) in Morocco (Cherry, Calo, Talmaci, Perrin, & Gavrila, 2016; Perrin et al., 1998)).

Table 2. The distribution of β-thalassaemia in North African countries.

Mutation	MOR[1]	MOR[2]	MOR[3]	ALG[4]	ALG[5]	TUN[6]	TUN[7]	EGY[8]	EGY[9]	EGY[10]
CD39 (C>T)	**71.2**	**26.2**	**28.0**	**26.0**	**43.0**	**49.0**	**37.5**	4.0	1.5	
IVS-I-110(G>A)	3.9	3.2	0.0	**26.4**	**26.0**	**21.0**	**22.9**	**48.0**	**32.9**	**22.0**
FSC6 (-A)	3.9	13.4	9.8	13.0	6.7	2.6	2.0			
-29 (A>G)	3.9	4.3	8.5	1.4	3.3	0.0	0.0			
IVS-I-I (G>A)	1.9	8.6	7.3	9.1	10.0	4.5	8.3	24.0	11.3	7.0
IVS-II-I (G>A)	1.9	3.2	2.4	1.0	0.0	0.6	0.0	7.0	3.0	
FSC5 (-CT)	5.8	0.0	0.0	0.0	0.0	0.4	0.0			
IVS-I-6 (C>T)	0.0	13.9	2.4	6.2	0.8	0.6	4.2	**40.0**	**13.6**	**28.0**
FSC8 (-AA)	0.0	9.6	22.0	1.0	1.7	0.2	0.0		1.8	
IVS-II-745 (C>T)	0.0	0.5	11.0	0.0	0.0	2.6	4.2			
IVS-II-848 (C>A)				3.8		0.4		9.0	8.6	9.0
IVS-I-5 (G>C)				1.9		1.0	4.2	10.0	0.6	
IVS-I-2 (T>C)		2.1	1.2	0.9					1.8	
IVS-II-745 (C>G)		0.5				2.6		8.0	5.6	6.0
-87(C>G)						1.7		3.0	1.2	6.0
CD5(-CT)									2.4	3.0
Chromosomes (n)	52	90	82	208	120	475	48	400	337	94

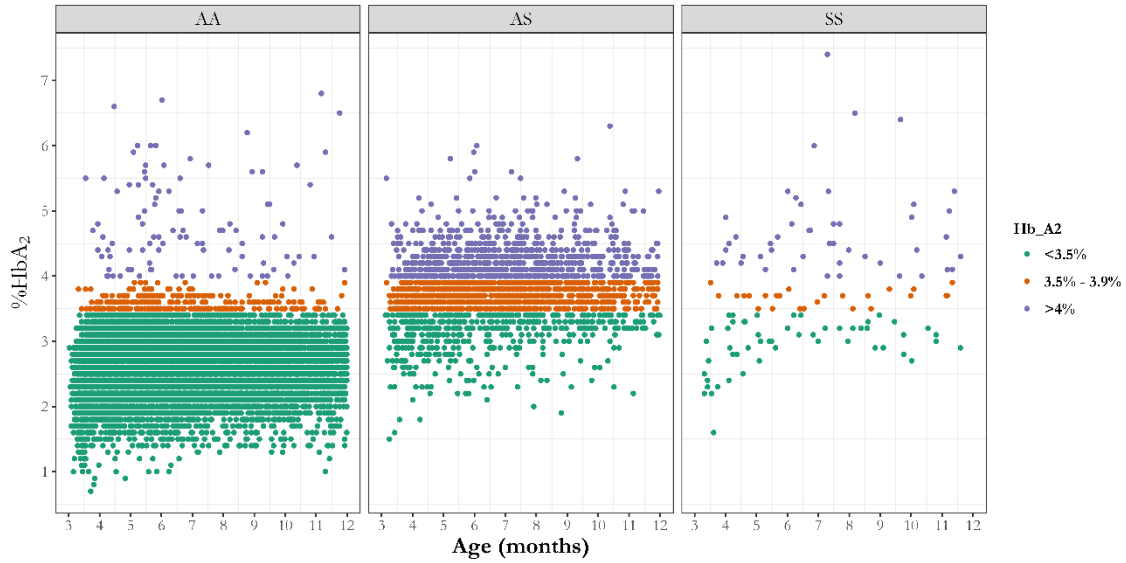
The rows of the table indicate types of mutations, while the columns represent the distribution of these mutations as identified in various studies. The studies referenced on the table are include: **(Belmokhtar et al., 2022)** (1),**(Lemsaddek et al., 2003)** (2),(Agouti et al., 2007) (3),(Boudrahem-Addour et al., 2009) (4),(Abdaoui et al., 2019) (5),(Fattoum et al., 2004) (6),(Sahli et al., 2016) (7),(El-Shanshory et al., 2014) (8),(Waye et al., 1999) (9),**(Elmezayen, Kotb, Sadek, & Abdalla, 2015)** (10). Countries: MOR (Morocco), ALG (Algeria), TUN (Tunisia), EGY (Egypt). The most common mutations in each country are highlighted in bold text.

The prevalence and mutations associated with β-thalassaemia in sub-Saharan Africa remains poorly described. The few case reports published to date suggest that β-thalassaemia might exist in these populations, albeit at a low prevalence (Frischknecht et al., 2008; Veten, Ghaber, Habti, & Houmeida, 2015; Waye et al., 2015).

1.8 β-thalassaemia in Kilifi

Similar to much of sub-Saharan Africa, until very recently, β-thalassaemia has not been reported in Kenya. Nevertheless, in a recent study conducted in Kilifi, we found elevated HbA_2 levels of ≥4% in 0.8% of HbAA infants identified by their HPLC-defined sickle cell are suggestive of β-thalassaemia (Figure 4) (Macharia et al., 2019).

Figure 4. HbA_2 in levels by age and by HbS status in children from Kilifi Birth Cohort.



The graph shows distribution of HbA_2 in among 15,737 children from Kilifi genetic birth cohort, quantified using high performance liquid chromatography (HPLC). HbA_2 values are colour coded according to the ranges normally helpful for the presumptive diagnosis of β-thalassaemia: green: <3.5%; orange: 3.5-3.99%; purple: >3.99%.

Simultaneously, in a separated study in which we were investigating the use of hydroxyurea in children with sickle cell anaemia, we observed that 10/151 children (10%) had HbS/β-thalassaemia (McGann et al., 2018). Eight of these children were affected by a mutation at rs33959855, a nonsense variant that results in premature termination of β-mRNA resulting in β^0-thalassaemia whereas the other two were affected by a mutation at rs193922563, a deletion that causes inactivation of an acceptor splice site resulting in β^0-thalassaemia. These are the first reports of β-thalassaemia within East Africa where, because the prevalence is very low, its greatest clinical significance is likely to relate to its coinheritance with HbS, which results in HbS/β-thalassaemia. More studies are needed to better characterise β-thalassaemia in this region.

1.9 β-thalassaemia and malaria

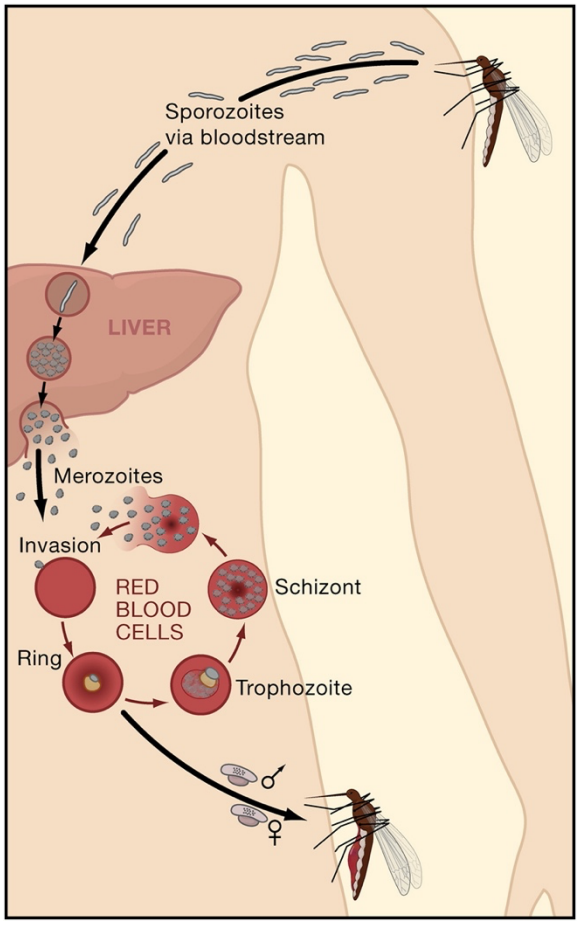
The literature review on pathophysiology of β-thalassaemia (section 1.4) suggests it is a deleterious condition resulting in a spectrum of clinical outcomes ranging from mild anaemia in carriers to a state of transfusion dependence in homozygotes. This prompts the question of why

a condition with such adverse effects would persist within a population. One plausible explanation is that β-thalassaemia like other RBC polymorphisms is protective against malaria and that this has led to balancing selection driven through survival advantage in heterozygotes at the expense of high mortality in homozygotes. To better understand this relationship, in this section I survey existing literature connecting β-thalassaemia and malaria.

1.9.1 Life cycle of *Plasmodium falciparum*

Infection with malaria is initiated by the injection of sporozoites into dermis by a female anopheline mosquito. The sporozoites enter the vasculature and are transported to the liver where they migrate through Kupffer cells and invade hepatocytes. Here, they undergo schizogony where they multiple to produce up to 30,000 merozoites (Sturm et al., 2006). This stage is asymptomatic and takes between 5 and 15 days. The merozoites are then released into the blood stream, where they invade red blood cells (RBCs), marking the beginning of the erythrocytic life cycle. The parasites develop through various stages, beginning with early stages referred to as the ring-stage, then trophozoite and schizont stages. At the schizont stage the parasite replicates producing between 16 and 32 daughter merozoites. The schizont then ruptures to release merozoites that then invade fresh RBCs. Each erythrocytic cycle lasts approximately 48hrs. Clinical symptoms including fever manifest at this stage. As a result of stress, environmental or chemical cues some parasites commit to develop into asexual stage parasites, or gametocytes, that are taken up by mosquitoes when feeding on the infected host. Male and female gametes undergo gametogenesis in the mosquito midgut, the male gamete fertilizing the female gamete to form a zygote. The zygote transforms and develops into an ookinete, crosses the midgut epithelium, and encysts to become an oocyst, which then bursts to release sporozoites. The sporozoites then migrate to the mosquito salivary gland ready to infect the next human host during a blood meal (Figure 5) (Cowman & Crabb, 2006).

Figure 5. Asexual life cycle of *Plasmodium falciparum* in the human host.



The life cycle of *P. falciparum* in the human host begins with a bite from an infectious Anopheles mosquito which injects sporozoites into the dermis. The parasites migrate to the liver where they invade hepatocytes marking the beginning of the erythrocytic cycle. After 5 to 15 days the parasites are released into the blood stream where they invade RBCs marking the beginning of the erythrocytic cycle. The parasite spends most of its lifetime in the erythrocytic stage and commits to gametocyte formation under stress or when there is a change in the environment (see text for more detailed explanation). Reproduced with permission from (Cowman & Crabb, 2006).

Molecular basis of *P. falciparum* invasion

The invasion of RBCs by merozoites involves several receptor ligand interactions. A number of these RBC receptors have been identified experimentally by blocking receptor ligand interactions using either RBCs lacking certain receptors, antibody binding, peptides and small molecules or enzymatic receptor cleavage (Table 3) (Weiss et al., 2015).

Upon egress from a schizont-infected RBC, the released merozoites bind to the surface of fresh RBC's using one of the many parasite ligands on the parasite surface (Das et al., 2015). After binding to the RBC, the parasites reorient itself so that the apical tip is in contact with the RBCs. This is followed by several interactions that primarily involve ligands from the erythrocyte

binding antigen (EBA) and *P. falciparum* reticulocyte binding homolog 5 (PfRH5) families, followed by the interaction of Apical membrane protein 1 (AMA1) and host receptor rhoptry neck (RON) protein complex resulting in the formation of a moving tight junction. The parasite then propels into the RBC by forming a parasitophorous vacuole that surrounds the merozoite which subsequently develops into a trophozoite and then to a schizont (Cowman & Crabb, 2006; Weiss et al., 2015).

Table 3. *P. falciparum* ligands and their corresponding receptors on the surface of the RBC.

***P. falciparum* ligands on merozoite surface**	**Host erythrocyte receptor**
Erythrocyte binding antigen 140 (EBA-140)	Glycophorin C
Erythrocyte binding antigen 175 (EBA-175)	Glycophorin A
Erythrocyte binding antigen 181 (EBA-181)	Receptor W
Erythrocyte binding ligand 1 (EBL-1)	Glycophorin B
P. falciparum reticulocyte binding homolog 4 (PfRh4)	Complement receptor 1 (CR1)
P. falciparum reticulocyte binding homolog 5 (PfRh5)	Basigin
P. falciparum reticulocyte binding homolog 2a (PfRh2a)	Unknown
P. falciparum reticulocyte binding homolog 2b (PfRh2b)	Receptor Z
Apical membrane protein 1 (AMA-1)	RON2
Merozoite surface protein 1 (MSP1)	Proteoglycan with heparin-like chains

The table shows some of the best characterised *P. falciparum* ligands and receptors. Some of the host erythrocyte receptors such as that of PfRh2a are unknown (Baum, Maier, Good, Simpson, & Cowman, 2005; Koch & Baum, 2016; Weiss et al., 2015).

1.9.2 β-thalassaemia and resistance to malaria

In 1949, when considering the suggestions made by Neil and Valentine that β-thalassaemia heterozygotes might be less fit than normal, Haldane proposed that it might well be the opposite. He suggested that red cells from β-thalassaemia heterozygotes might be more resistant to invasion by malaria parasites, and that this might explain why the condition was found at such a high frequency in Greece, Italy and Sicily, which were malaria endemic regions at that time (Haldane, 1949). Now known famously as the "malaria hypothesis", in the years that followed several compelling lines of evidence have emerged that support Haldane's assertion. Although the hypothesis was initially based on β-thalassaemia, the strongest evidence so far has

related to HbS and α-thalassaemia, with only very few studies having been conducted on β-thalassaemia (Weatherall & Clegg, 2001; Williams & Weatherall, 2012).

The primary evidence that β-thalassaemia protects against malaria is the geographic overlap between the two conditions in regions such as Asia and Southern Europe where malaria has historically been endemic. Similar observations have also been made at a micro-epidemiologic scale. In Sardinia for example, the incidence of malaria was historically known to correlate with altitude. When Siniscalco and colleagues investigated the prevalence of β-thalassaemia in this region, they observed that it was inversely correlated with increase in altitude, a feature that itself is correlated with a decreasing incidence of malaria (Siniscalco, Bernini, Latte, & Motulsky, 1961). Similar observations were subsequently made by Hill and colleagues in Melanesia (Hill, Bowden, O'Shaughnessy, Weatherall, & Clegg, 1988). In the only case control study published to date that has specifically addressed this question, β-thalassaemia was associated with an 50% reduction in clinical *Plasmodium falciparum* malaria in Northern Liberia (M. Willcox et al., 1983). The second line of evidence that β-thalassaemia protects against malaria comes from laboratory-based studies, a topic discussed in further detail below.

1.10 Mechanisms of protection against malaria

The reduced synthesis of β-globin in the β-thalassaemias results in various changes in RBCs that could consequently affect the survival of malaria parasites. One of the observations made by Haldane was that the RBCs from β-thalassaemia heterozygotes are smaller than normal and more resistant to hypotonic solutions (Haldane, 1949). In later studies it was observed that free unbound α-chains play a central role of these changes. In β-thalassaemia major, excess unbound α-chains precipitate in RBCs and their precursors in the bone marrow leading to premature destruction of the RBCs through apoptosis and ineffective erythropoiesis (Weatherall, 1998). In RBCs that are in circulation, the free α-chains undergo an irreversible process involving oxidization of the free α-chains to methaemoglobin with subsequently formation of hemichromes that precipitate in RBCs to form inclusion bodies. Hemichromes promote

clustering of band 3 (Aljurf et al., 1996) and also likely result in the expression of neo-antigens on the surface of the RBC (Yuan, Kannan, Shinar, Rachmilewitz, & Low, 1992). These neo-antigens are recognised by autologous IgG and as a result these RBCs are cleared by the reticuloendothelial system (RES) akin to senescent RBCs (Low, Waugh, Zinke, & Drenckhahn, 1985; Weatherall & Clegg, 2001). Evidence to support this mechanism is based on the observation that among β-thalassaemic individuals, RBCs containing inclusion bodies have only been observed in subjects who have undergone splenectomy, suggesting either that these cells are trapped in the spleen or that inclusion bodies are removed by the spleen (Jacob, 1970). Similar though milder changes have been observed in β-thalassaemia carriers (Knox-Macaulay & Weatherall, 1974; Vettore, Falezza, Cetto, & De Matteis, 1974). These changes create an environment that may be less hospitable for the survival of malaria parasites and, as a result, protect the human host.

Investigations into the mechanisms of protection against malaria by β-thalassaemia have mainly been laboratory-based, specifically using parasite culture techniques. These mechanisms can be classified into four broad groups based on the stages of interaction of the RBC with malaria parasites, as discussed below.

1.10.1 Invasion of β-thalassaemia RBCs by *P. falciparum* merozoites

Studies on the more severe forms of β-thalassaemia, such as HbE/β-thalassaemia, have consistently shown reduced rates of malaria parasite invasion and subsequent growth (Brockelman, Wongsattayanont, Tan-ariya, & Fucharoen, 1987). In one of the studies, a reduction in parasite multiplication rate of up to 30% in RBCs from HbE/β-thalassaemia was observed when compared to normal controls (Udomsangpetch et al., 1993). However, this cannot be directly translated to the less severe forms of β-thalassaemia, including carrier forms, as one of the main differences between the two conditions is HbF, which is found in levels of >20% in HbE/β-thalassaemia carriers (Weatherall & Clegg, 2001). In a study by Udomsangpetch and colleagues, a negative correlation was found between HbF levels and the

parasite multiplication rate, indicating that HbF might also play a role in protection (Udomsangpetch et al., 1993).

To date, studies on malaria parasite invasion of RBCs from β-thalassaemia carriers have produced mixed results. Whereas Senok and colleagues observed a reduction in invasion, Ayi and colleagues and Introini and colleagues found no difference (Ayi, Turrini, Piga, & Arese, 2004; Introini et al., 2022; Senok et al., 1997). The discordance seen between these studies might relate to methodological considerations. For example, whereas in the study by Ayi, the β-thalassaemia carriers were normal for G6PD, α-thalassaemia and HbAS, those in the Senok study were only normal for G6PD and in the Introini study α-thalassaemia and G6PD were not considered. Nonetheless, important observations were made in these studies, including that by Introini that the success of *P. falciparum* invasion involves a combination of factors that reduce invasion rate, including higher tension, and those that increase them, including bending modulus and a higher expression of the basigin receptor.

1.10.2 Growth of *P. falciparum* in β-thalassaemic RBCs

As for invasion, observations on the growth of *P. falciparum* parasites in β-thalassaemic RBCs are inconsistent. Whereas observations by Kaminsky and colleagues, Senok and colleagues, Brockelman and colleagues, and Udomsangpetch and colleagues all support inhibited growth in β-thalassaemic when compared to normal RBCs (Brockelman et al., 1987; Kaminsky et al., 1986; Udomsangpetch et al., 1993), this was not supported in studies by Ayi and colleagues and Roth and colleagues (Ayi et al., 2004; Roth, Raventos-Suarez, Rinaldi, & Nagel, 1983). This disagreement might also be explained by methodological issues that might include the confounding effects of other RBC polymorphisms. While Brockelman and Roth accounted for confounding by G6PD deficiency in their studies, they observed a reduction in parasite growth although the difference was not statistically significant in the Roth study. Conversely, in the Ayi study, which also accounted for G6PD, α-thalassaemia and HbAS, there was no difference in parasite growth after 3 growth cycles (Ayi et al., 2004).

1.10.3 Growth inhibition, recognition, and clearance of *P. falciparum* infected β-thalassaemic RBCs by the immune system

A third potential mechanism involves serum factors that either enhance clearance of *P. falciparum*-infected β-thalassaemic RBCs by the immune system or inhibit the multiplication of malaria parasites. In the Ayi study the authors observed that markers of membrane damage, namely hemichromes, membrane bound autologous IgG and c3c, and aggregated band 3, were already increased at baseline in β-thalassaemic compared to normal RBCs. When invaded by malaria parasites, expression of these markers was enhanced resulting in early recognition by phagocytic cells (Ayi et al., 2004). The recognition and clearance of early-stage parasites could potentially result in reduced parasite burdens in the infected individuals and result in reduced disease severity.

Several studies have looked at the inhibitory effects of serum factors on parasite growth and multiplication. This was done by supplementing the culture media with serum from a β-thalassaemia carriers in place of normal serum. Kaminsky and colleagues (Kaminsky et al., 1986) observed that growth suppression only occurred when RBCs from the β-thalassaemia carriers were incubated with serum from β-thalassaemia carriers and that no growth inhibition was observed when β-thalassaemic RBCs were incubated in culture media supplemented with normal serum. Similarly, Thanomsub and colleagues (Thanomsub, Fucharoen, Brockelman, & Bhisutthibhan, 1989) and Vattanaviboon and colleagues (Vattanaviboon, Sriklup, Polkaew, Prawatmurng, & Ittarat, 1999) observed that culture media supplemented with serum from HbE/β-thalassaemia was also able suppress growth of *P. falciparum* when compared to culture media supplemented with normal serum.

1.10.4 Reduced rosetting and cytoadhesion

A final potential mechanism, that has not been intensively studied, is that β-thalassaemia modulates the severity of infections to result in a reduction of the more severe forms of malaria

such as severe malaria anaemia and cerebral malaria. *P. falciparum* erythrocyte membrane protein 1 (PfEMP-1) is a protein expressed on the surface of trophozoite-stage falciparum-infected RBCs, which mediates their attachment to microvasculature endothelial cells, enabling the parasite to avoid splenic clearance. The sequestration of these parasites in microvasculature of various organs such as the brain, results in vaso-occlusion and various clinical complications. Infected RBCs expressing PfEMP-1 also bind to uninfected RBCs to form rosettes, which are also associated with cerebral malaria (Rowe, Obeiro, Newbold, & Marsh, 1995). Both cytoadherence and rosetting are thus considered to be correlates of disease severity. In the studies conducted on rosetting and cytoadhesion using RBCs from β-thalassaemia carriers, Carlson and colleagues and Undosangpetch and colleagues, observed reduced rosetting and cytoadhesion, respectively (Carlson, Nash, Gabutti, al-Yaman, & Wahlgren, 1994; Udomsangpetch et al., 1993) which suggests β-thalassaemia might confer protection by stifling the pathophysiological process that lead to severe disease.

1.11 Research questions raised during this literature review

In summary based on studies conducted to date, β-thalassaemia is most prevalent in regions where malaria has been historically present. However, the prevalence of β-thalassaemia has not been studied extensively in sub-Saharan Africa where malaria is also endemic, only a handful of reports having been published to date. In addition, the relationship between β-thalassaemia and protection against malaria remains poorly studied with only one case control study conducted to date, and the few laboratory-based studies have produced mixed and conflicting results.

The literature review raises the following questions related to β-thalassaemia in Kilifi that I will address through this thesis:

What is the prevalence of β-thalassaemia in Kilifi and what mutations are involved?

Although we found that 10% of children in the REACH study were suffering from HbS/β-thalassaemia and that 2 β-thalassaemia mutations were responsible for β-thalassaemia in this group (McGann et al., 2018), the true prevalence of β-thalassaemia and the spectrum of

mutations in this population remains unknown. This is because the participants in the REACH study were from a selected group of children with sickle cell anaemia and, as such, were not completely representative of the entire Kilifi population.

To the best of my knowledge, the evaluation of HbA_2 in the diagnosis of β-thalassaemia in sub-Saharan countries has not been performed before which could explain the few reports on β-thalassaemia from this region. It is however important to evaluate the precision of these cutoffs in this population mainly due to the presence of other haemoglobinopathies such as HbS and factors such as iron deficiency anaemia and HIV that are also known to cause variations HbA_2 levels.

In Chapter 3 of this thesis therefore, I investigate the prevalence and spectrum of β-thalassaemia mutations in a cohort of children representative of the whole Kilifi population in whom I also investigate the specificity and sensitivity of HbA_2 in the diagnosis of β-thalassaemia.

What are the origins of these mutations?

The identification of these mutations in Kilifi also raises questions about where else these mutations have been reported and whether the mutations first occurred on the chromosomal background of the population in Kilifi or were introduced from elsewhere through geneflow.

To answer this question, I perform population genetic analysis in Chapter 4, where I explore the global distribution of these mutations and their associated haplotypes, providing information about their origin and evolutionary history.

Does β-thalassaemia confer protection against malaria and what are the mechanisms involved?

This question remains unanswered, mainly because so few epidemiological studies have been conducted to date, and the laboratory-based investigations show conflicting results. One plausible explanation for the conflicting outcomes in the laboratory-based studies relates to

confounding from the co-inheritance of other protective polymorphisms in the RBCs under investigation and differences in sensitivity in laboratory methods, as discussed in the literature review.

In Chapter 5, I describe a recall by genotype study to investigate the impact of β-thalassaemia mutations on malaria parasite invasion and on the mechanism of enhanced clearance of ring-parasitised RBCs through opsonic phagocytosis. I perform these investigations through laboratory based functional assays on fresh RBCs from β-thalassaemia carriers and matched controls. The participants were matched for, among other things, known malaria protective polymorphisms.

In chapter 6, I address this question further through a new analysis of a birth cohort study. In this analysis, I investigate the impact of β-thalassaemia on under 5 mortality and incidence of hospital admission with malaria and other illnesses.

1.12 Study aims and objectives

The overall aim of this thesis is to describe the prevalence, genetic spectrum and origins of β-thalassaemia in Kilifi, Kenya, and to further investigate the malaria-protective mechanisms and health consequences of these mutations. The outcome of this work can be viewed in two broad themes. In Chapters 3 and 4, I investigate the prevalence, type, and origin of β-thalassaemia mutations while in Chapters 5 and 6, I address the question of protection against malaria and the wider health consequences of β-thalassaemia in Kilifi through laboratory-based investigations and by the analysis of data from a birth cohort study. The specific objectives of my thesis are thus:

1. To determine the prevalence and spectrum of β-thalassaemia mutations in Kilifi.
2. To investigate the origin of β-thalassaemia in Kilifi.

3. To investigate the impact of β-thalassaemia mutations on malaria parasite invasion and on the mechanism of enhanced clearance of ring-parasitised RBCs through opsonic phagocytosis.
4. To determine the impact of heterozygous β-thalassaemia on mortality and admission to hospital with malaria and other childhood illnesses.

Chapter 2: General materials and methods

In this chapter, I describe the study populations included in the individual studies conducted within my thesis, my community engagement activities, general laboratory procedures, statistical analytical procedures and ethical approvals for my studies.

2.1 Study Site

All of the studies included in my thesis were conducted among the population of Kilifi County on the Coast of Kenya. The county covers an area of 12,540 km^2 and had a total population of 1,453,787 in 2019, of whom 42% were below 14 years of age (census 2019, https://www.knbs.or.ke/). The county is largely populated by rural communities who make their living through subsistence farming. The majority of the population comes from the Mijikenda ethnolinguistic group, a broad group consisting of nine language communities that speak the Bantu dialect. The main sub-tribes are the Giriama, Chonyi, Kauma, Rabai, Kambe, Ribe, Digo, Jibana and Duruma.

Kilifi has historically been a malaria endemic region with seasonal transmission occurring during the long rains between April and June and short rains between October and December. Malaria is mainly caused by *P. falciparum* which is transmitted by *Anopheles gambiae* sensu lato with minor contributions by *A. funestus* vector species (Mbogo et al., 1995). The *P. falciparum* parasite prevalence declined between 1998 to 2011 after which it began to rise again through to 2014. Emerging resistance to chloroquine and the introduction of treatment with sulfadoxine/pyrimethamine has been suggested as a plausible explanation for this decline (Snow et al., 2015).

The studies were conducted through the KEMRI-Wellcome Trust Research Programme (KWTRP), which is located within the Kilifi County Hospital (KHC) compound. KCH is a government hospital located in Kilifi and serves as a primary point of care and as a first referral

centre for the majority of the population in the county. The hospital admits between 4000 and 5000 paediatric patients and 3000 to 4000 adults per year (Etyang et al., 2014; O'Meara et al., 2008).

A system of epidemiological and demographic surveillance, known as the Kilifi Health and Demographic Surveillance System (KHDSS), was established by the KWTRP in 2002, for collecting data on morbidity and mortality from common childhood diseases, migration events, and host genetic associations with infectious diseases. The KHDSS was established within a defined area of 891km^2 surrounding KCH. The total population at the time of initiation was approximately 260,000, of which approximately 100,000 were children <14 years. Approximately 80% of the patients admitted to KCH reside within the KHDSS study area (Scott et al., 2012). Every 3 months, enumerators visit households and update records on births, deaths, pregnancies, and migration events. These details are also linked to hospitalisation events at KCH using a system of computer based unique personal identification numbers (PIDs). Data such as the number of admissions, laboratory data, age, sex and migration status can be retrieved from this database. Over the years the KHDSS platform has been used to describe the epidemiology of SCD (Uyoga, Macharia, Ndila, et al., 2019), malaria (Njuguna et al., 2019), pneumonia (Hammitt et al., 2019) and rotavirus (Nokes et al., 2008) among other diseases. It has also been used to evaluate the impact of insecticide-treated bed-nets and vaccines on morbidity and mortality (Hammitt et al., 2019; Kamau et al., 2017; Otieno et al., 2020).

2.2 Study Populations

The samples used in answering the aims of my study were drawn from two studies conducted at the KWTRP, namely, the Kilifi Genetic Birth Cohort (KGBC) and the Kilifi Sickle Cell Disease Cohort (KSCDC) studies, each of which are described in further detail below.

2.2.1 The Kilifi Genetic Birth Cohort (KGBC) Study

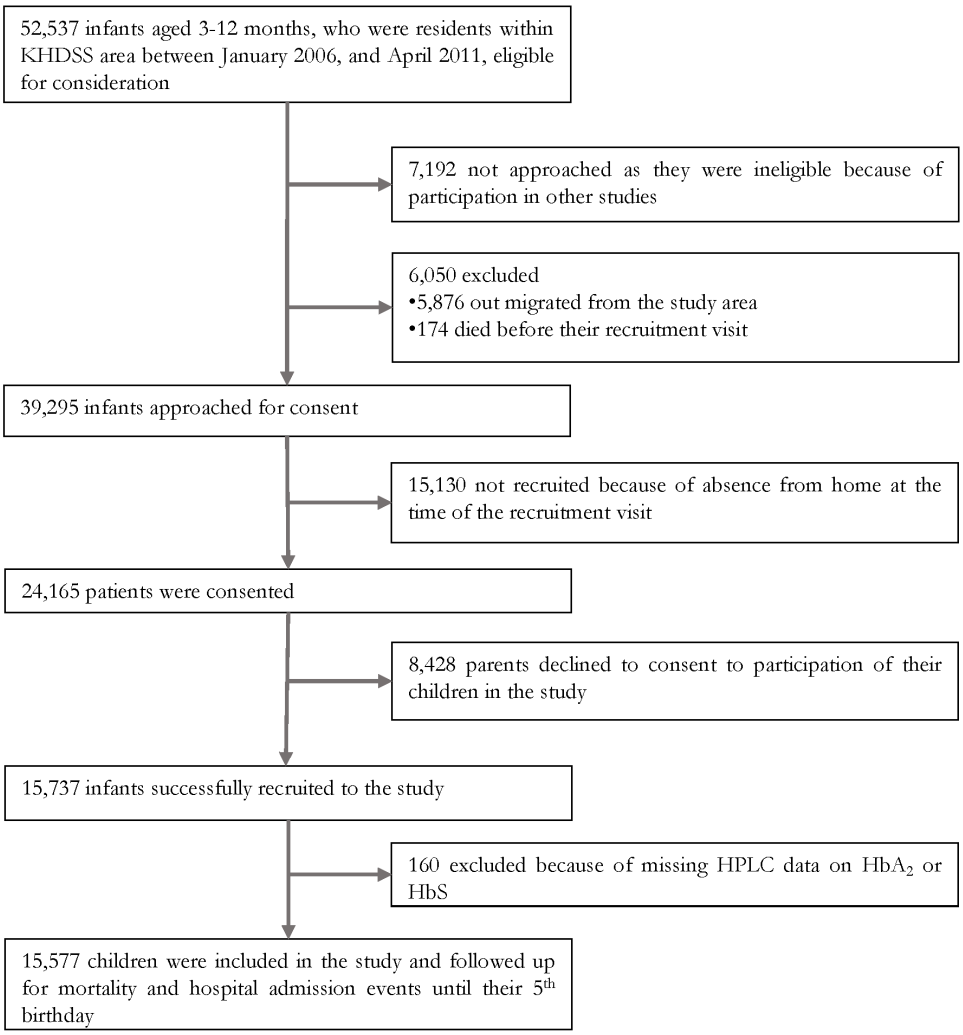
This study was designed to investigate the impact of a range of host genetic factors on a range of common child health outcomes including malaria. The study aimed to recruit 15,000 children between 3- and 12-months of age, and to follow them up for various outcomes of interest. Selection and recruitment of participants to the KGBC study was facilitated through the KHDSS. The homes of eligible participants were visited by the KGBC study team who sought consent for recruitment from their parents.

Upon consent, demographic data and capillary blood samples were collected into EDTA and transported to our laboratory at the KWTRP for further processing. There, within five days of sample collection, we measured the proportions of specific types of haemoglobins including HbS and HbA_2 using a Variant Classic^{TM} HPLC analyser and the β-thalassaemia Short Program (BioRad, Hercules, CA, USA). DNA was extracted and genotyped for the common African 3.7kb α-thalassaemia deletion by PCR. Suspected cases of β-thalassaemia were identified using HbA_2 cut-off values after which the β-globin region was sequenced to identify β-thalassaemia inherited in the form HbA/β-thalassaemia or HbS/β-thalassaemia. The remaining DNA was then stored at -80^{0}C for further analysis (see Section 2.3 for a more detailed description of the sample processing procedures).

Test results on sickle cell status reported as "sickle cell disease" or "no sickle cell disease" were fed back to the participants. Parents whose children were positive for SCD were offered basic counselling and were invited for confirmatory testing at the KCH SCD clinic, which is the only clinic that offers specialised services for children with SCD within the region. Once confirmed, the participants were invited to join the KSCDC.

Between 1st January 2006 and 30th April 2011, a total of 52,537 children aged 3-12 months were born in the KHDSS study area, out of which 15,737 were recruited to KGBC study. 160 of 15,577 successfully recruited patients were excluded from the analysis because they were missing HPLC data on HbA_2 or HbS (Figure 6).

Figure 6. Flowchart showing recruitment of study participants to the KGBC study.



Those who joined the KGBC study were followed passively for various outcomes of interest. During the 3-monthly visits by the KHDSS study team, the vital status of participants (alive, dead, or out-migrated) was recorded. If admitted to KCH, the children were attended by trained personnel who recorded data on both their admission and discharge diagnoses. At admission, data from routine laboratory measurements, including full blood counts, malaria microscopy and blood cultures, were also recorded in a database linked to the KHDSS database.

2.2.2 Recall-by-genotype study of *P. falciparum* parasite invasion in β-thalassaemia

This study was designed to investigate the impact of β-thalassaemia mutations on malaria parasite invasion and on the proposed mechanism of enhanced clearance of ring-parasitised

RBCs. This was done through laboratory based functional assays on fresh RBCs from β-thalassaemia carriers and matched controls. Before the study, I undertook a number of community engagement activities whereby I presented the benefits and risks of the study to various community representatives and received useful feedback on how best to conduct the study. The study aimed at recruiting fifteen β-thalassaemia heterozygotes and fifteen matched controls with normal haemoglobin genotype (HbAA). The determination of sample size in this study was pragmatic, primarily being guided by the sample sizes employed in previous studies with similar objectives and design (Introini et al., 2022; Kariuki et al., 2020). The cases and controls were selected from among those identified as having HbA/β-thalassaemia and HbAA in KGBC, respectively. To eliminate possible confounding, each case was matched as closely as possible to their control for age, sex, area of residence, and the following co-inherited malaria protective polymorphisms: α-thalassaemia, Dantu and ABO blood group. The study team visited the homes of the selected case and control children to seek parental consent for participation in the study. If granted, the matched case and control children were scheduled for a clinic visit on the same day as one another. On the day of the clinic visit, finger prick samples were taken, and malaria tests were carried out by microscopy. If negative, 10mls of blood was then drawn into EDTA for use in the subsequent experiments. Children found to be malaria parasite positive were treated using standard of care medications and given a repeat appointment to re-attend at a later date.

2.2.3 The Kilifi Sickle Cell Disease Cohort (KSCDC) Study

A specialist clinic for children with SCD has been running at KCH for the past 30 years. The clinic was set up and is managed by staff at the KWTRP, and offers clinical, laboratory and pharmacological support to patients with the condition. The clinic currently has a registered population over 900 children <18 years, and between 80 and 120 additional patients are diagnosed and enrolled annually. Members of the cohort are mainly identified and recruited to the clinic from KCH, or through cohort studies conducted in KHDSS such as the KGBC, while

some cases are also referred from other hospitals. About 60% of the patients are residents within the KHDSS area and can be linked to the KHDSS database using their unique PID.

The KSCDC study is a descriptive study embedded in the SCD clinic that was established with the aim of describing the clinical characteristics of children with SCD in Kilifi both at steady state and during inter-current episodes and hospital admissions. The study aims to recruit all participants of the SCD clinic. Upon primary diagnosis, written informed consent is obtained from parents and guardians and blood samples are taken for baseline laboratory indices and for diagnostic confirmation by HPLC. A separate EDTA blood sample is also collected for investigation of potential genetic modifiers by use of DNA-based methods. Upon diagnostic confirmation, consenting participants are scheduled for 3-monthly clinic visits during which they were offered routine clinical surveillance and chemoprophylaxis for malaria and pneumococcal disease. All clinical and laboratory data are entered and stored in a centralised digital database. The children included in the current study were participants in the KSCDC study between the years 2014 and 2019 who had a blood sample available for DNA analysis. During this period, a total of 700 children were members of the KSCDC of which 644 had an EDTA blood sample available for DNA extraction.

The 1000 genomes population

The 1000 Genomes dataset is a public dataset that contains genetic sequence data from 2504 healthy individuals from 26 superpopulations sampled in Africa (AFR), East Asia (EA), Europe (EUR), South Asia (SAS) and Americas (AMR) (Genomes Project et al., 2015). The dataset contains 88 million variants 84.7 million single nucleotide polymorphisms (SNPs), 3.6 million short insertions/deletions (indels), and 60,000 structural variants), all phased onto high-quality haplotypes.

I downloaded the VCF for chromosome 11 from

http://ftp.ensembl.org/pub/data_files/homo_sapiens/GRCh37/variation_genotype/.

2.3 Laboratory methods

2.3.1 Quantification of haemoglobins using high performance liquid chromatography (HPLC)

Haemoglobin variants were quantified using the Variant Classic™ HPLC analyser and the β-thalassaemia Short Program (BioRad, Hercules, CA, USA) that utilizes principles of ion-exchange high performance liquid chromatography (HPLC). The procedure involves the following steps. RBCs are first lysed and diluted by adding 5µls of whole blood to 1.25mLs of deionised water. Haemolysed samples are then placed in a sample rack where an injector needle aspirates the sample and introduces it to the analytical flow path. A programmed buffer gradient of different ionic strength is then mixed with the samples and delivered to the analytical cartridge. Different forms of haemoglobin bind to the matrix in the analytical cartridge. Because of differences in amino acid substitutions, they bind at different ionic strengths to the cartridge matrix. The haemoglobins are then eluted from the cartridge using buffers of increasing ionic strength. Because of the differences in binding strength, the haemoglobins elute from the cartridge at different "retention times", a characteristic which is reproducible for each haemoglobin type. The eluted haemoglobins then pass a detector that measures absorbance at 415nm. The concentrations of the different haemoglobin variants (including HbA, HbF, HbA_2 and HbS) are then determined by calculating the area under the curve which are then assigned to the respective retention times (Figure 7) (Stephens et al., 2012).

In both the KGBC and KSCDC studies the HPLC chromatograms were reviewed by at least two experienced personnel and classified as HbAA, HbAS, HbSS or as other variants using the following criteria.

- HbAA
    - A combination of HbF and HbA_0 of greater than 70%
    - HbA_2 ≤4%

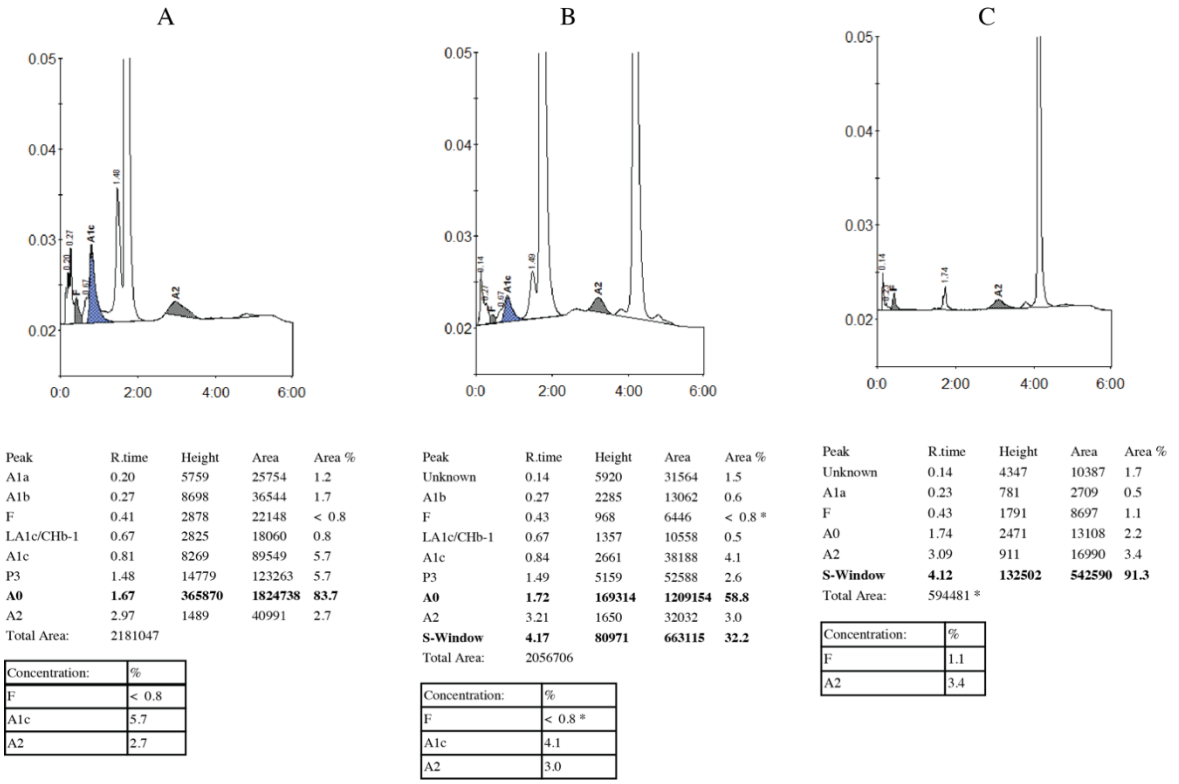
- Absence of HbS
- Absence of other pathogenic variants

- HbAS
  - Presence of both HbA_0 and HbS whereby HbA_0 is greater than HbS at a ratio of ~60:40 (HbA_0: HbS).
  - Absence of other pathogenic variants.

- HbSS
  - Presence of HbS at a concentration that is greater than 50%
  - Absence of HbA_0
  - Absence of other pathogenic variants.

- Other
  - Presence of other haemoglobin variants apart from those identified as HbA, HbS, HbF, HbA_2 and minor Hb peaks found in normal chromatogram.

Figure 7. Example HPLC chromatograms from participants diagnosed as HbAA, HbAS and HbSS.



Peak	R.time	Height	Area	Area %
A1a	0.20	5759	25754	1.2
A1b	0.27	8698	36544	1.7
F	0.41	2878	22148	< 0.8
LA1c/CHb-1	0.67	2825	18060	0.8
A1c	0.81	8269	89549	5.7
P3	1.48	14779	123263	5.7
A0	**1.67**	**365870**	**1824738**	**83.7**
A2	2.97	1489	40991	2.7
Total Area:	2181047			

Concentration:	%
F	< 0.8
A1c	5.7
A2	2.7

Peak	R.time	Height	Area	Area %
Unknown	0.14	5920	31564	1.5
A1b	0.27	2285	13062	0.6
F	0.43	968	6446	< 0.8 *
LA1c/CHb-1	0.67	1357	10558	0.5
A1c	0.84	2661	38188	4.1
P3	1.49	5159	52588	2.6
A0	**1.72**	**169314**	**1209154**	**58.8**
A2	3.21	1650	32032	3.0
S-Window	**4.17**	**80971**	**663115**	**32.2**
Total Area:	2056706			

Concentration:	%
F	< 0.8 *
A1c	4.1
A2	3.0

Peak	R.time	Height	Area	Area %
Unknown	0.14	4347	10387	1.7
A1a	0.23	781	2709	0.5
F	0.43	1791	8697	1.1
A0	1.74	2471	13108	2.2
A2	3.09	911	16990	3.4
S-Window	**4.12**	**132502**	**542590**	**91.3**
Total Area:	594481 *			

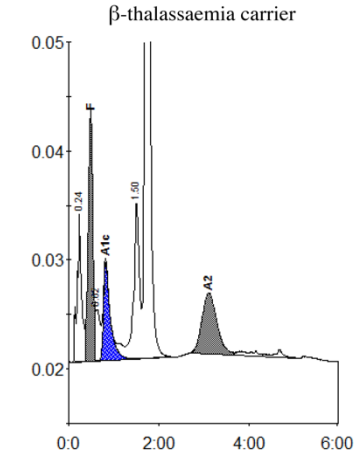
Concentration:	%
F	1.1
A2	3.4

Chromatogram from a normal individual (A) showing a single dominant peak at the A_0 region with the rest being minor haemoglobins. The second chromatogram (B) shows an individual identified as HbAS based on 2 dominant peaks at A_0 and S-window, with the peak at A_0 region being greater than that in the S-window (58.8 versus 32.2). The last chromatogram (C)is from a person categorised as HbS showing one major peak at the S-window and absence of a peak at A_0 region.

HbA/β-thalassaemia carriers (Figure 8) were identified from among the population with the phenotype HbAA using the following criteria:

- A combination of HbF and HbA_0 of >70%
- HbA_2 ≥4%
- Absence of other pathogenic variants

Figure 8. HPLC from a suspected HbA/β-thalassaemia heterozygous participant.



Peak	R.time	Height	Area	Area %
A1a	0.24	13732	93084	3.9
F	0.49	22768	162035	7.0
LA1c/CHb-1	0.62	4706	30401	1.3
A1c	0.82	8921	93413	5.8
P3	1.50	14336	122454	5.1
A0	1.70	370923	1765903	73.6
A2	**3.10**	**5622**	**131419**	**6.4**
Total Area:	2398710			

Concentration:	%
F	7.0
A1c	5.8
A2	6.4

The chromatogram shows an individual identified as HbAA but with an HbA_2 value of ≥4%, a characteristic of for β-thalassaemia heterozygosity.

2.3.2 DNA Extraction

For the KGBC study, DNA was extracted from EDTA blood samples with volumes of between 100 and 500µl using an ABI PRISM 6100 Nucleic Acid PrepStation™ (Applied Biosystems, Waltham, Massachusetts, USA). In the KSCDC study, DNA was extracted from EDTA samples of ~500µl using the selective filter membrane method (Qiagen DNA Blood Mini KitTM, Qiagen Ltd, West Sussex, and UK). In both studies, the DNA was quantified using a NanoDrop ND1000 device (Thermo Fisher Scientific, Inc., Wilmington, USA) and samples diluted to a final concentration of 15 to 20ng/µl with DNase free water (Invitrogen; Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA) where necessary. A volume of between 1µl and 1.5µl of diluted sample was used in the PCR and sequencing reactions.

2.3.3 Genotyping for the β^s globin mutation

As discussed previously, HbS results from a single nucleotide substitution of the adenine for thymine at codon 6 of the β-globin gene which results in substitution of the amino acid glutamic acid for valine in the β-globin chain (Ingram, 1957). Genotyping for this SNP was performed using an allele-specific amplification (ASA) PCR method that I adapted from that described by Waterfall and colleagues (Waterfall & Cobb, 2001). In ASA PCR, primers targeting both wildtype and mutant alleles, that are similar in sequence to the complementary strand apart from the last base on the 3´ end (either Adenine for the wildtype or Thymine for the mutant primer), are included in the same reaction. When PCR is performed, amplification only happens for primers that perfectly match the complementary strand. Using this method, it is possible to genotype both wildtype and mutant alleles in a single tube reaction. The primers and conditions used in this reaction are listed below:

WT-AS: 5´ ATGGTGCACCTGACTCCTGA 3´

WT-CP: 5´ CCCCTTCCTATGACATGAACT 3´

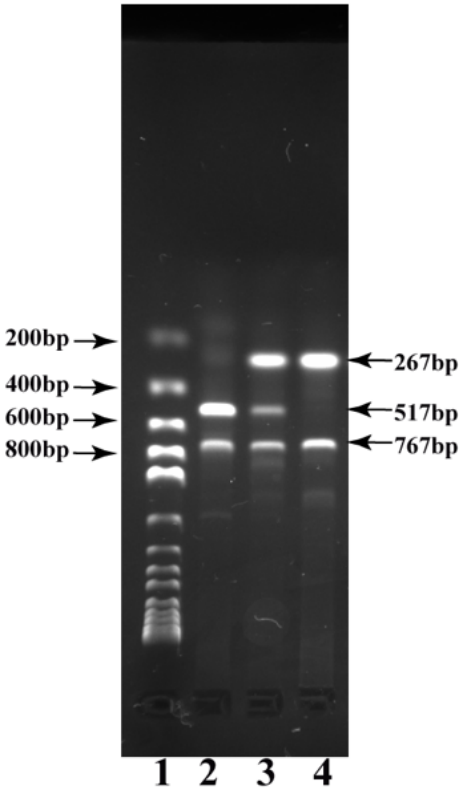
MUT-AS: 5´ CAGTAACGGCAGACTTCTCCA 3´

MUT-CP: 5´ GGGTTTGAAGTCCAACTCCTA 3´

A typical reaction consisted of 1.5µls of diluted DNA at a concentration of 15 to 20ng/µl, 6.45µl of deionised water (Invitrogen, Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA), 1.25µls PCR reaction buffer (Invitrogen; Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA), 1.0µls of 25mM $MgCl_2$, 0.75µls of 8 mM dNTP mix (Invitrogen; Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA), 0.05µls of Platinum Taq polymerase (Invitrogen; Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA), 1µl of 10µM primers WT-AS and WT-CP, 0.25µls of 10µM primers MUT-AS and MUT-CP. Amplification was performed using a GeneAmp 9700 PCR system (Applied

Biosystems, Foster City, CA, USA) under an initial denaturation of 95^0C for 5 minutes followed by 30 cycles of denaturation at 95^0C for 30 seconds, primer annealing at 60^0C for 30 seconds, primer extension at 72^0C for 30 seconds and a final extension of 72^0C for 5 minutes. Amplicons were then separated on a 1% agarose gel (Invitrogen; Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA) containing 0.5µg/ml of ethidium bromide (Sigma-Aldrich, Dorset, UK) in 1x Tris-borate-EDTA buffer (TBE) (Sigma-Aldrich, Dorset, UK) at a voltage of 95V for 2.5h. Gels were visualised in GEL DOC 2000^{TM} UV transilluminator (Bio-Rad, USA) (Figure 9).

Figure 9. An agarose gel showing separation of PCR amplicons from ASA PCR targeting the HbS allele.



Lane 1 shows the 1000bp molecular ladder. Lanes 2-4, show PCR amplicons from individuals with HbAA, HbAS and HbSS genotypes. A common fragment of 767bp amplified by outer primers (WT-CP and MUT-CP) is amplified in all samples and acts as a positive control for amplification success. HbAA is diagnosed by the presence of the 767bp fragment and a HbA specific fragment of 517bp. HbSS is diagnosed by the presence of the 767bp fragment and a HbS specific fragment of 267bp. HbAS is diagnosed by the presence of the 3 fragments.

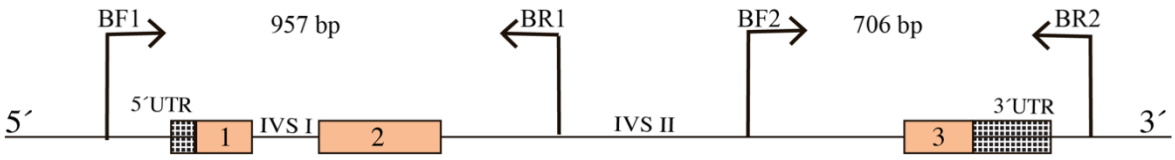
2.3.4 Sequencing

β-globin gene sequencing was performed on DNA extracted as described in Section 2.3.2 using the dye primer chemistry method developed by Clark and colleagues (Clark & Thein, 2004). The procedure involves four main steps, namely, PCR amplification, PCR product purification, BigDye™ terminator sequencing reaction and capillary electrophoresis.

PCR Amplification

Two separate PCR reactions were performed to amplify products from the regions covered by two sets of primers: BF1/BR1 and BF2/BR2 (Figure 10). The PCR conditions for the two reactions were similar and are discussed in detail below.

Figure 10. β-globin gene showing location of sequencing primers.



The lower panel shows the β-globin gene with the respective exons labelled 1, 2 and 3. Sequencing was performed on two regions. The first region was sequenced using primers BF1 and BR1 that covers 5´ promoter, 5´ untranslated region (UTR), exon 1 and 2, and the intronic sequences flanking exon 2. The second region was sequenced using primers BF2 and BR2 that covers a 706bp consisting of exon 3 and 3´ UTR and the intronic regions flanking these sequences.

The primer sequences used in sequencing were as follows:

BF1 5´ CGATCTTCAATATGCTTACCAA 3´

BR1 5´ CATTCGTCTGTTTCCCATTCTA 3´

BF2 5´ CAATGTATCATGCCTCTTTGCA 3´

BR2 5´ TGCAGCCTCACCTTCTTTCAT 3´

The four primers were first diluted to 10μM and added into separate 20μls reaction consisting of 12μls Qiagen fast-cycling PCR master mix, 2μls of RNAse-Free water, 1.2μls of each primer: BF1 and BR1 for reaction 1 and BF2 and BR2 for reaction 2. The reaction mix was then

amplified using a GeneAmp 9700 PCR system (Applied Biosystems, US) with the following temperature settings: initial denaturation of 950C for 5m followed by 30 cycles of denaturation at 950C for 30s, primer annealing at 650C for 30s and primer extension at 680C for 40s and a final extension of 680C for 5m. The amplicons were then separated on a 1% UltraPure™ Agarose (Invitrogen; Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA) containing 0.5µg/ml of ethidium bromide (Sigma-Aldrich, Dorset, UK) in 1xTBE at a voltage of 95V for 1.5h. The gels were then visualised in a UV transilluminator GEL DOC 2000 (Bio-Rad, USA).

PCR amplicon purification

Purification of PCR amplicons was performed to prevent carryover of unused primers and nucleotides, which could result in synthesis of non-specific products during the cycle sequencing reaction stage. I used the EXOSAP-IT (Affymetrix, Santa Clara, California, United States) commercial kit to perform the PCR amplicon purification. This contains a single mix with the active ingredients being Exonuclease 1, that breaks down single stranded primers into dNTPs, and Shrimp Alkaline Phosphatase (SAP), that removes the phosphate group from dNTPs. A single reaction mix consists of the following: 5µls PCR product and 3.4µls of the EXOSAP-IT reagent. This is then incubated at 37^0C for 15m to degrade primers and nucleotides followed by 80^0C for 15m to denature the enzymes and a final incubation at 15^0C for 5m to cool the samples.

BigDye™ terminator sequencing reaction

BigDye™ terminator reaction is based on the Sanger sequencing method that involves incorporation of modified fluorescent labelled nucleotides (di-deoxynucleotides; ddNTPs) that once incorporated terminate DNA elongation. ddNTPs lack a hydroxyl group at the 3´ end, that is required for phosphodiester bond formation and, as a result, synthesis is terminated when DNA polymerase adds a ddNTP. This results in synthesis of PCR products of different lengths.

In the current study, BigDye™ sequencing reaction was set up separately for each of the four primers mixed together with the respective purified PCR product as follows: 0.5µls BigDye™ terminator Ready Reaction Mix, 2µls sequencing buffer, 1µl of primer (BF1/BF2/BR1/BR2), 5.5µls of deionised water and 1µl of purified PCR product. Cycle sequencing was then performed using the following temperatures: initial denaturation of 95^0C for 60s followed by 25 cycles of denaturation at 96^0C for 10s, primer annealing at 50^0C for 15s and primer extension at 60^0C for 4m.

Purification of sequenced amplicons

The Ethanol/EDTA precipitation method was used to remove unincorporated ddNTPs. Briefly to each sequenced sample 2µls of EDTA (125mM), 2µls sodium acetate, 30µls of absolute ethanol were added and mixed using a pipette. The plates were sealed and incubated at room temperature for 20m to allow DNA to precipitate. The plates were then spun at 4000 revolutions per minute (rpm) for 20m using a 5810R benchtop centrifuge (Eppendorf, Hamburg, Germany) to precipitate the DNA. The seal was then removed, and the plates inverted on top of clean paper towels to drain the solution mix. To drain remnant solution, the plates were inverted on paper towels and spun at 400 rpm for 1m. The amplicons were then washed by adding 90µls of ice cold 70% ethanol to each well, then sealed and spun at 4000rpm for 5m. After centrifugation, the plates were inverted on paper towels to drain ethanol. Excess ethanol was drained by inverting the plates on paper towels and spinning at 400 rpm for 1m. All centrifugation steps were done at 4^0C. The plates were then placed unsealed on a heating block at 65^0C for 5m to vaporise remnant ethanol. 10µls of Hi-Di formamide (Invitrogen; Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA) was added to each sample and the plates were sealed. Denaturation was performed by incubating the plates at 96^0C for 3m

on a GeneAmp 9700 PCR system (Applied Biosystems, US). The plates were then stored at 4^0C before shipment to the sequencing laboratory.

Size separation using capillary electrophoresis

The plates containing the sequenced products were shipped for capillary electrophoresis to the sequencing laboratory at the International Livestock Research Institute (ILRI) using an ABI 3730xl (Applied Biosystems, US). Briefly, capillary electrophoresis separates the amplicons by size and records the fluorescence signal per strand, the signal is then fed into a base calling software that assigns a base depending on the light spectrum emitted. The data is then stored as fluorescent peak trace chromatograms which contains the full sequence of the DNA sample.

Sequence analysis for detection of SNPs

I performed the analysis using CLC Main Workbench v7.9.1 (Qiagen, UK). I first imported the sequence chromatograms for each primer (BF1/BF2/BR1/BR2) and separately aligned them to chromosome 11 β-globin gene region from the reference genome (GRCh37). I then looked at the quality of the chromatograms by looking at the height and distinctness of the peak and the quality scores (QS), which is the average quality value (QV) of all the bases. The QV score is a metric generated by the sequencing software and measures the probability that the base is called incorrectly. A QV of 20 represents a 1% chance that a base has been incorrectly called. I re-sequenced all the samples without distinct peaks and with QS scores of <20.

SNPs and insertions/deletions (INDELS) were identified by looking at variations in nucleotides between the aligned sequences and reference genome (GRCh37) on both forward and reverse strands. SNPs were identified by looking for variants at a single nucleotide position and the genotype entry was made as either homozygous mutant if one variant nucleotide was present or heterozygous if both the reference and mutant allele were present at the same potion. INDELS

were identified as differences in sequence in more than one variant following each other consecutively and the breakpoints for the INDELS were determined by looking at both forward and reverse strands. The mutations were then annotated using the online databases Ensembl (https://www.ensembl.org/index.html), Varsome (https://varsome.com/) and NCBI (https://www.ncbi.nlm.nih.gov/).

2.3.5 Parasite culture

The 3D7 strain of *P. falciparum* parasites malaria parasite isolate was used in performing functional assays towards understanding the mechanisms through which β-thalassaemia protects against malaria. This was done through culturing 3D7 in RBCs from the β-thalassaemia carriers and non-carriers and looking at differences in invasion, growth, and on the proposed mechanism of enhanced clearance of ring-parasitised RBCs using opsonic-phagocytosis. 3D7 is a clone of NF54 parasite isolate, which was derived from a patient living near Schiphol Airport, Amsterdam (Walliker et al., 1987). 3D7 is well characterised, and its genome has been fully sequenced.

2.3.6 Culture of 3D7 parasite isolates

3D7 parasites were cultured in blood group O^+ human RBCs at 3% haematocrit (Hct) in RPMI 1640 medium (Sigma-Aldrich, Dorset, UK) supplemented with 25mM HEPES (Gibco; Thermo Fisher Scientific, Waltham, MA, USA), 20mM glucose (Sigma-Aldrich, Dorset, UK) and 25µg/ml gentamicin (Gibco; Thermo Fisher Scientific, Waltham, MA, USA) (incomplete medium) containing 10% Albumax (Gibco; Thermo Fisher Scientific, Waltham, MA, USA) (compete medium). Culture volumes of between 10 to 40mls were maintained in tissue culture flasks (Falcon; Corning, NY, USA) depending on number of study participants. The cultures were gassed to maintain an atmosphere of 3% O_2, 5% CO_2 and 92% N_2 and placed in an incubator at a temperature of 37°C. Culture media were changed daily to replenish nutrients for

the parasites. The parasites were maintained at a parasitaemia of between 3 and 5% by diluting the cultures with fresh O^+ RBCs.

Monitoring parasite growth using microscopy

Parasitaemia was monitored daily using standard light microscopy. For each sample, 100µl of culture was aspirated and transferred to a 1.5ml Eppendorf tube. The tube was spun at 700xg for 1m using a Eppendorf 5424R microfuge (Epperdorf, Hamburg, Germany). The supernatant was removed, and the pellet resuspended in the remnant volume then smeared on a microscope slide, air dried, then fixed for 1m using absolute methanol. The slide was then dried and stained with filtered Giemsa stain (10% Giemsa (Sigma-Aldrich, Dorset, UK) in phosphate buffered saline (Sigma-Aldrich, Dorset, UK) for 10m. Slides were then washed using tap water and air dried. Parasitaemia was determined using a light microscope (Nikon E400, Tokyo, Japan) with a x100 objective and oil emulsion. A total of 1000 RBCs were counted in at least 5 random fields with ~200 RBCs each. The parasitaemia was calculated from the percentage of parasitised RBCs per 1000 RBCs.

Synchronisation of parasites using 5% D-sorbitol

Parasite cultures become asynchronous in long term culture. To maintain them in synchronous cycles, sorbitol synchronisation was performed. D-sorbitol selectively causes lysis of RBCs containing late-stage parasites (trophozoites and schizonts), retaining only the ring stage parasites. Sorbitol synchronisation was performed on a weekly basis when parasites were at majority ring stages and at ~5% parasitaemia. The cultures were transferred to sterile 50 ml tubes and spun at 2000 rpm for 5m, after which the supernatant was discarded, and the culture pellet resuspended in 5% (w/v) of D-sorbitol (Sigma-Aldrich, Dorset, UK) in distilled water. The D-sorbitol/culture mix was incubated at 37°C for 5m in a water bath and then washed twice using incomplete medium. The culture pellet was then resuspended in complete medium,

re-gassed, and returned to the incubator. A smear was collected before and after synchronisation to determine parasite stages using microscopy as described above.

2.3.7 THP1 monocyte cell line culture

The THP1 human monocyte cell line is commonly used to study phagocytosis. THP1 cells were cultured in RPMI-1640 medium containing 10% heat inactivated fetal calf serum (FCS) and penicillin-streptomycin (Sigma-Aldrich, Dorset, UK) at 37°C and 5% CO_2. The cells were maintained at a concentration of 3-7 x 10^5 cells per ml in a 200ml vented culture flask (Corning, NY, USA). Twenty hours before phagocytosis assay was performed the monocytes were stimulated with TNF-α (250U/ml) and IFN-γ (50U/ml).

2.3.8 Pre-processing samples

Because RBC samples from β-thalassaemic individuals degrade rapidly, I aimed to perform all experiments within a period of seven days of collection. As described under Section 2.2.2, the participants were invited as case (HbA/β-thalassaemia) and control (HbAA) pairs. A 10ml EDTA blood sample was then drawn in tandem from both case and control children and the downstream procedures performed simultaneously.

After sample collection the paired EDTA samples were spun at 440xg for 10m using an Eppendorf 5810R microfuge (Eppendorf, Hamburg, Germany) and plasma was removed. This was followed by removal of WBCs by density centrifugation using Lymphoprep (Axis Shield PoC AS, Oslo, Norway). The RBCs pellet was then washed twice using incomplete media and resuspended to 50% Hct using incomplete medium. Remnant WBCs were then removed by passing the blood through a Plasmodipur filters (EuroProxima, Arnhem, Netherlands) using a 20ml syringe. The filtrate containing purified RBCs was washed twice with incomplete media and resuspended to 50% Hct in complete media. The RBCs were then counted using a haemocytometer and a dilution ratio determined by dividing the RBC counts for the paired

samples. The sample with higher RBC count was diluted by adding complete media according to the dilution ratio. The RBCs samples were then stored at 4°C.

2.3.9 Preference invasion assay

This assay was adapted from Theron and colleagues (Theron, Cross, Cawkill, Bustamante, & Rayner, 2018), with slight modification. The design of the preference assay allows RBCs from case and control children to be co-incubated with parasites in the same well. Quantifying the parasitaemia for each erythrocyte population provides a measure of malaria parasite RBC preference.

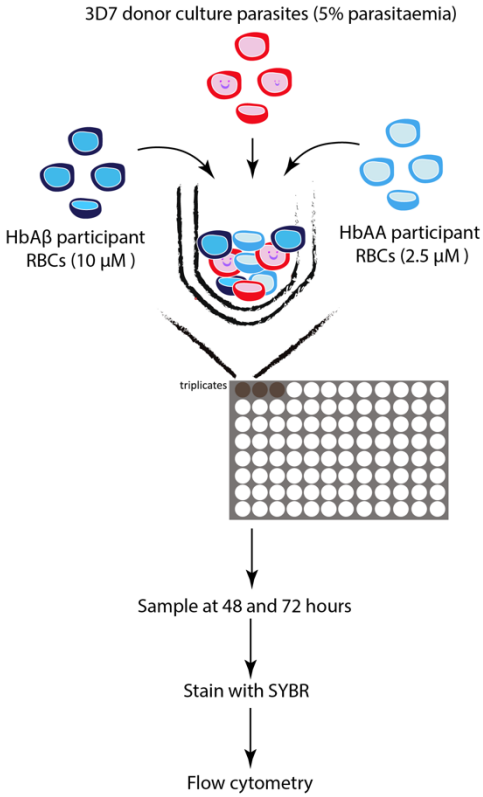
A 500µl aliquot of the resuspended RBCs (see Section 2.3.8) was transferred to a labelled Eppendorf tube. The RBCs were spun at 700xg, and the supernatant discarded. The RBCs for the paired cases and controls were then resuspended in different concentrations (either 2.5µM or 10µM) of the cell membrane dye CellTrace Far Red™ cell proliferation kit (Invitrogen; Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA). Allocation to either concentration was performed at random. The cells were then incubated at 37°C for 2h for staining to take place, then washed three times using incomplete RPMI. After the final wash the cells were resuspended to 2% Hct in complete medium and stored in the dark at 4°C for a maximum of 24h until use.

To set up the invasion assay, synchronised ring-stage 3D7 parasites at 4 to 5% parasitaemia (see Section 2.3.5) were diluted to 2% Hct in complete medium. In a 96 well round bottom plate (Falcon; Corning, NY, USA), donor parasites together with the stained RBCs samples were added in equal volumes (33µl of 3D7 donor culture, 33 µl of β-thalassaemia case and 33 µl of control) to the same well in triplicates (Figure 11). The plates were then incubated at 37°C in a sealed humified chamber and gassed to maintain an environment of 3% O_2, 5% CO_2 and 92% N_2. After 48h (the first invasion cycle) 50µl of the culture was sampled for determination of differences in parasitaemia post invasion using flow cytometry. The remaining culture was replenished with complete media, gassed, and returned to the incubator for a further 24h. Plates

were then removed and stained to determine differences in parasite growth using flow cytometry as described below.

The cultures sampled at 48h or 72h post invasion were first washed once using PBS. To remove residual RNA, the cultures were treated with 0.5µg/ml Ribonuclease A (Sigma-Aldrich, Dorset, UK) in phosphate buffer saline with incubation at 37°C for 1h. The cultures were then washed twice using PBS and then incubated at 37°C for 1h with 2x SYBR green I DNA dye (Invitrogen; Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA) in 1x PBS which stains malaria parasites. This was followed by 2x washes with 1x PBS after which the cultures were resuspended in PBS ready for reading on a flow cytometer.

Figure 11. Design of preference invasion assay for β-thalassaemia and non-β-thalassaemia RBCs.



The paired RBCs were stained with different concentrations of CellTrace Far Red™ membrane dye and then co-cultured with the donor culture at 5% parasitaemia. The cultures were then sampled at 48h and 72h to determine differences in parasite invasion and growth through staining parasites and quantifying parasitaemia using flow cytometry.

The cultures were examined with a 488-nm blue laser and a 633-nm red laser on a BD FACS Canto flow cytometer (Beckman Coulter, High Wycombe, UK). CellTrace Far Red™ was

excited by a red laser and detected by a 660/20 filter whereas SYBR green I was excited by a blue laser and detected by a 530/30 filter. A total of 60,000 events were acquired for each sample using BD FACS Diva software (Beckman Coulter, High Wycombe, UK). The data was then analysed using FlowJo™ software (TreeStar, Ashland, OR, USA) whereby the difference in parasitaemia between β-thalassaemia carriers and non-carriers was assessed at 48h for invasion and the difference in mean florescence intensity (MFI) between 48h and 72h as a measure of parasite growth (Figure 12).

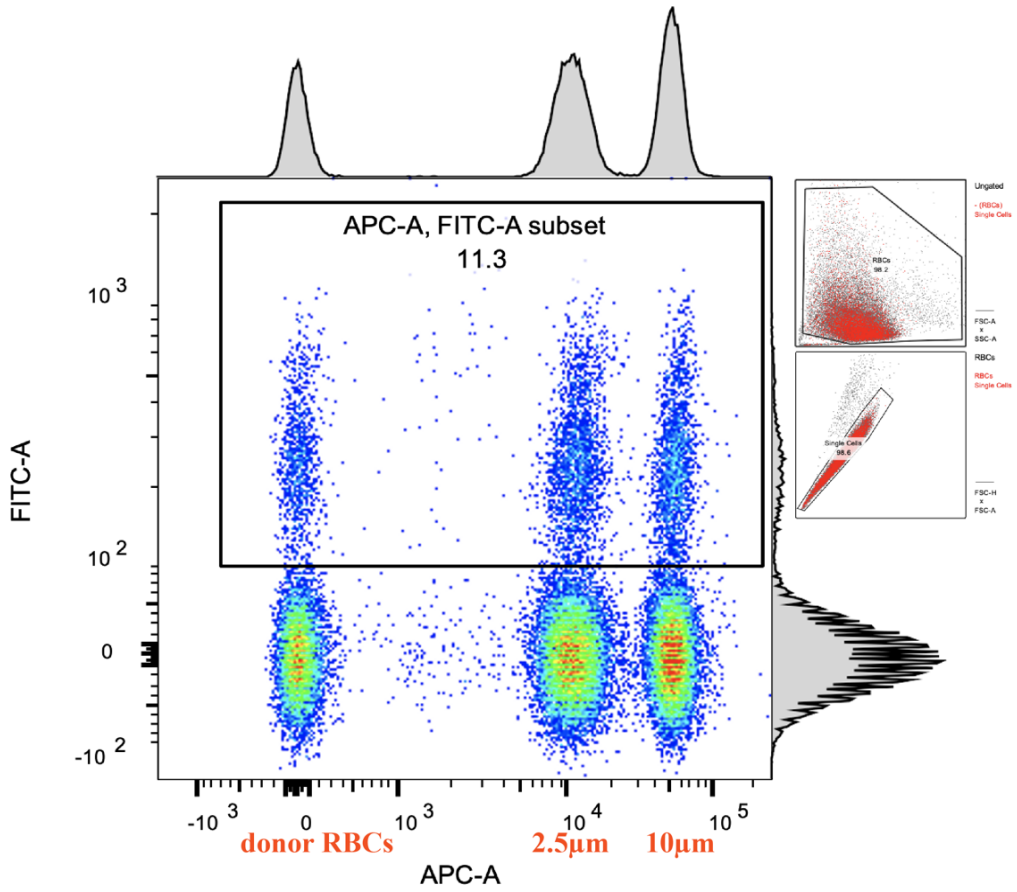
To determine the invasion at 48h, I corrected for variation of number of RBCs added to each well by dividing the number of parasitised RBCs from each genotype by the total of number of labelled RBCs that belong to the same genotype as determined using flow cytometry. The percentage of invasion was then calculated as the mean of the triplicate wells.

Because of variation in invasion parasitaemia, I also analysed invasion results as relative parasitaemia for better visualization of minor changes in parasitaemia between β-thalassaemia carriers and HbAA controls. This was calculated as follows. For each pair, I calculated a divisor which when used in dividing the mean parasitaemia of normal control would result in a value of 1. This divisor was calculated by dividing the value of 1 by the mean invasion parasitaemia of the triplicate wells of the normal control. Parasitaemia's from the triplicate wells of both normal control and β-thalassaemia carriers were then normalised by diving using the divisor. Finally, the mean of the normalised triplicates was calculated, of which the resulting mean for the normal control was always 1. For example, if the calculated invasion parasitaemia for each triplicate well for normal control was 3.2%, 3.3%, 3.1% (mean=3.2%) and that of paired β-thalassaemia carrier was 4.2%, 4.1%, 4.0% (mean=4.2%), the divisor was calculated as 1 divided by 3.2% (31.25). When all the triplicates are divided using this value (31.25), the normalised values are 1.00, 1.03, 0.97 (mean=1) for normal control and 1.31, 1.28, 1.25 (mean=1.28) for β-thalassaemia heterozygote.

To assess parasite growth, I employed the following approach: I calculated the ratio of average MFI at 72hr to that of 48h for each sample. To obtain the average MFI readings, I first

computed the mean MFI of the triplicate samples at both 48h and 72h for each participant. Next, I divided average MFI at 72h by average MFI at 48h then conducted a statistical comparison between the mean ratios of β-thalassaemia heterozygotes to those of the corresponding paired HbAA controls.

Figure 12. Analysis of flowcytometry data on preference assay.



RBCs labelled with 2.5µm and 10µm CellTrace Far Red™ (labelled APC-A) were mixed and incubated with 3D7 *P. falciparum* donor culture parasites in the same well. The cultures were sampled at 48 and 72h. The samples were then stained using 2x SYBR green I DNA dye which stains parasite DNA. Gating was first done for all RBCs followed by singlets (small graphs on the right) and then by SYBR green stain (labelled FITC-A) and CellTrace Far Red™ (labelled APC-A) stains (middle graph). The X-axis shows the 3 populations of cells separated based on CellTrace Far Red™ staining, and the Y-axis shows the parasitised RBCs stained using SYBR green I DNA stain. The parasitised RBCs are gated for the whole population: for this experiment the parasitaemia was 11.3%. The parasitaemia for each RBC population was further determined through gating the parasitised RBCs for 2.5 µm and 10µm dyes separately.

2.3.10 Opsonic-phagocytosis assay

The procedure was adapted from Gallo and colleagues (Gallo, Skorokhod, Schwarzer, & Arese, 2012) with slight modifications. The procedure involves three main steps: invasion of

participants' RBCs, opsonisation, and phagocytosis. A comparison is then made between the paired β-thalassaemia cases and controls to determine the differences in the number of monocytes having phagocytosed RBCs.

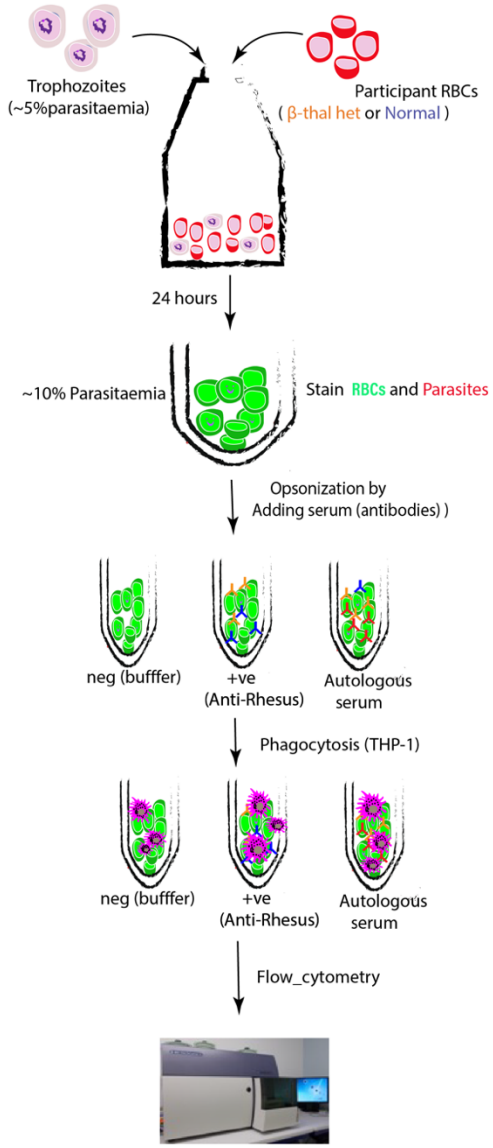
To set up invasion of participants' samples, trophozoites were first harvested from 3D7 culture using Percoll density gradient (Sigma-Aldrich, Dorset, UK). A 3D7 culture at 5% parasitaemia (see Section 2.3.6) was first layered on top of a Percoll density gradient consisting of a bottom layer of 70% and top layer of 40% Percoll diluted in RPMI. The layered Percoll-culture gradients were spun at a 3900 rpm for 30m without brakes. The trophozoites were harvested from the interface between 70% and 40% Percoll and washed three times using incomplete medium.

Blood samples from the paired cases and controls were diluted to 3% haematocrit in complete RPMI in separate 50ml culture flasks (Falcon; Corning, NY, USA). Trophozoites were then added to the culture to achieve a parasitaemia of 5%. The culture was then gassed (3% O_2, 5% CO_2 and 92% N_2) and incubated at a temperature of 37°C for 24h then transferred to centrifugation tubes, spun at 2000 rpm for 5m and washed once using incomplete medium. To distinguish parasitised from non-parasitised RBCs, the RBCs were stained for 30m with 10μg/ml ethidium bromide (Sigma-Aldrich, Dorset, UK), which stains the parasite DNA. The RBCs were then washed three times with RPMI after which the RBC membrane was stained by incubating with 0.1μm carboxyfluorescein diacetate succinimidyl ester (CFDA-SE) (Invitrogen; Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA) at 37°C for 10m. Staining was stopped by adding FCS and washing three times with RPMI.

Opsonisation is important as it aids in recognition and phagocytosis by monocytes. The RBCs were opsonised using anti-D IgG (Rhophylac, CLS Berling, USA) as positive control and autologous serum at a ratio of 1:1:25 (RBC:RPMI:anti-D IgG) and 1:9:10 (RBC:RPMI:serum) respectively. A negative control consisting of non-opsonised RBCs was also included. The cells were then incubated at 37°C for 30m. After washing three times with RPMI, phagocytosis assays were set up in duplicate wells, each containing 1μl of RBCs to 0.1 million monocytes suspended in 300μl of RPMI-1640 medium containing 10% heat inactivated FCS and incubated at 37°C

for 3h in 5% CO_2. Phagocytosis was stopped through centrifugation at 350xg for 3m at 4°C, and un-phagocytosed RBCs lysed by re-suspending monocytes in 150µl Puregene RBC lysis solution (Qiagen Ltd, West Sussex, and UK). The monocytes were washed twice with cold FACS buffers (2% FCS in PBS) with centrifugation at 350xg for 3m at 4°C and then resuspended in 200µl of cold FACS buffer. The samples were examined on a BD FACS Canto flow cytometer (Beckman Coulter, High Wycombe, UK) to determine the proportion of monocytes that had taken up RBCs. This was done by determination of mean fluorescent intensity (MFI) of Etbr and CFDA in 60,000 monocytes per sample (Figure 13).

Figure 13. Design of the opsonic phagocytosis experiments.



Trophozoites were harvested from 3D7 culture using a Percoll density gradient. Experiments were conducted in three main stages: parasite invasion of participant RBCs, opsonisation, and phagocytosis.

Analysis to determine the proportion of monocytes with phagocytosed RBCs was performed using FlowJo™ software (TreeStar, Ashland, OR, USA). First, a cut-off was set for monocytes that were positive for CFDA-SE or Etbr using the negative control (Figure 14). Monocytes above this threshold were then counted and expressed as a proportion of the total monocytes after subtraction of background from the negative controls. The results were then presented as:

- CFDA-SE positive: Representing the proportion of monocytes with phagocytosed parasitised and non-parasitised RBCs.
- CFDA-SE and Etbr positive: Representing the proportion of monocytes with phagocytosed parasitised RBCs.

Figure 14. Analysis of flow cytometry data on opsonic phagocytosis.

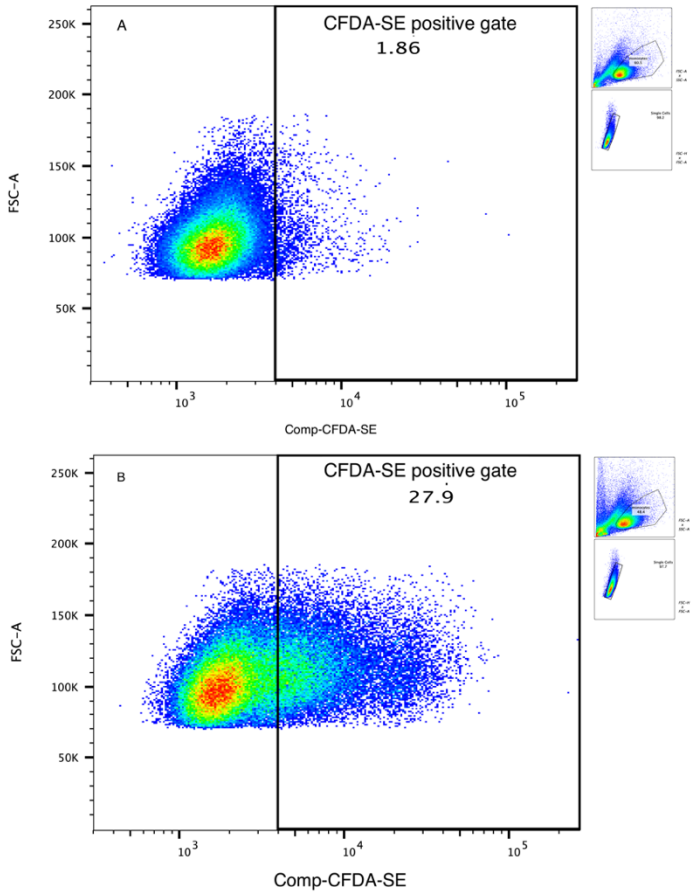


Figure A and B show the gating of monocytes that had taken up RBCs (CFDA-SE positive) for negative control (A) and anti-D IgG opsonised positive control (B). After gating for monocytes and singlets (small graphs on the right), the monocytes were interrogated for CFDA-SE fluorescence (X-axis). Using the negative and positive control a gate was set to select the monocytes with phagocytosed RBCs.

2.3.11 Characterising RBC membrane protein expression with fluorescent monoclonal antibodies

To characterise the RBC membrane, I selected a panel of 11 monoclonal antibodies targeting various RBC antigens either known or suspected to be involved in *P. falciparum* cell adhesion and invasion (Table 4).

Table 4. Table showing 11 monoclonal antibodies targeting different membrane proteins on the surface of RBCs suspected to be involved in *P. falciparum* adhesion and invasion.

Fluorescent-conjugated Antibody	Membrane protein	Dilution	Manufacturer
CD35-APC	CR1	1:50	Invitrogen, Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA
CD44-FITC	P-glycoprotein 1	1:400	IBGRL, Bristol, UK
CD49d-APC	Integrin	1:25	Invitrogen, Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA
CD55-FITC	DAF	1:200	IBGRL, Bristol, UK
CD71-PE	Ferritin receptor	1:25	Invitrogen, Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA
CD147-FITC	Basigin	1:100	Invitrogen, Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA
CD233-FITC	Band 3	1:2000	IBGRL, Bristol, UK
CD234-APC	Duffy antigen	1:100	Milteny Biotec
CD235a-FITC	Glycophorin A	1:1000	IBGRL, Bristol, UK
CD235b-primary	Glycophorin B	1:100	Abcam, Cambridge, UK
CD235b-FITC	Glycophorin B	1:1000	Abcam, Cambridge, UK
CD236-FITC	Glycophorin C	1:2000	IBGRL, Bristol, UK

The RBCs from the paired cases and controls were analysed concurrently and in duplicate to reduce variation. Apart from Glycophorin B, each sample was diluted to 0.5% Hct using PBS and then incubated with primary antibodies conjugated to fluorophores for 1 hour at 4°C. For Glycophorin B the RBCs were first diluted to 0.5% Hct, then incubated with anti-Glycophorin B primary antibody for 1h at 4°C, and finally washed twice and incubated with a secondary antibody conjugated to a fluorophore for 1h at 4°C. For each sample, control RBCs without

antibodies were included to measure background fluorescence. After washing three times with PBS the MFI for 60,000 cells was measured on a BD FACS Canto flow cytometer (Beckman Coulter, High Wycombe, UK). These data were then analysed using FlowJo™ software (TreeStar, Ashland, OR, USA). The MFI for the whole population after gating for singlets was determined after which a comparison was made between the paired case and control.

2.3.12 Enzyme-Linked Immunosorbent Assay (ELISA) for determination of serum transferrin receptor (sTfR)

Quantification of sTfR was done using solid-phase sandwich ELISA commercial kit (Invitrogen; Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA). Diluted serum and standards were first added to a plate pre-coated with the capture antibody. After incubation the plate was washed 4 times to remove unbound antigen, after which biotin-conjugated antibody specific to sTfR was added and incubated. The plate was then incubated with Horseradish peroxide (HRP) conjugated streptavidin (streptavidin-HRP), then washed and incubated with substrate for colour development. Colour intensity was measured spectrophotometically at 450nm and the sTfR quantities extrapolated from the standard curve.

2.4 Ethical considerations

Written informed consent was provided by the parents of the study participant in both KGBC and KSCDC.

Ethical approval for KGBC study was granted by the Kenya Medical Research Institute Scientific and Ethics Review Unit in Nairobi, Kenya (references SCC1192, SERU3420 and SERU3500).

Ethical approval for KSCDC study was granted by the Kenya Medical Research Institute Ethical Review Committee in Nairobi, Kenya (Number: SCC3891).

Chapter 3: Prevalence of β-thalassaemia in Kilifi

3.1 Abstract

Background

β-thalassaemia is rare in sub-Saharan Africa. In two separate studies we conducted in Kilifi, we reported that a small proportion of children with HbA2 values of ⩾4%, indicating presence of β-thalassaemia carriers and two β^0-thalassaemia mutations (rs33959855 and rs193922563) were identified(Macharia et al., 2019; McGann et al., 2018). The objective of this chapter was to investigate the true burden of β-thalassaemia in Kilifi by assessing the prevalence, causal mutations, and clinical significance. Additionally, I evaluated the accuracy of HPLC-derived HbA_2 cut-off values in the diagnosis of β-thalassaemia carriers.

Methods

I screened participant samples to identify β-thalassaemia mutations in two cohorts: the KGBC, representative of the entire KHDSS population, and KSCDC consisting of children with sickle cell disease. I performed HBB sequencing on individuals with raised HbA_2 values of ⩾ 4.0%, <3.5% to 3.9% and <3.5% in KGBC. Genotyping of the rs334 allele in KSCDC helped distinguish HbS/β-thalassaemia from HbSS. Additionally, I sequenced a sample from a child suspected to have β-thalassaemia. To determine the diagnostic accuracy of an HbA_2 cut-off of ⩾4%, I calculated the sensitivity, specificity, positive- (PPV) and negative predictive values (NPV) in the presence or absence of HbS using sequencing as a gold standard.

Results

In the two cohorts I identified 73 subjects with HbA/β-thalassaemia, 40 subjects with HbS/β-thalassaemia and one β-thalassaemia homozygote (β^0/β^0). Four β-thalassaemia pathogenic variants were identified: three β_0-thalassaemia (rs33959855, rs33941849, and rs193922563) and one β+-thalassaemia (rs35004220). I estimated the minimum allele frequency of all variants combined within the study population at 0.3%. Among HbAA participants an HbA2 value of

⩾4% was associated with a sensitivity of 91% and a PPV of 75% for β-thalassaemia heterozygosity. However, when those with HPLC phenotypes consistent with either HbAS or HbSS were considered, the use of an HbA_2 level of ⩾4% resulted in a considerably higher false positivity rate, reflected in a PPV of only 9%.

Conclusions

β-thalassaemia is present in Kilifi, Kenya, an observation that has implications for the diagnosis and clinical care of children from the East Africa region. In addition, when using HPLC-derived HbA_2 cut-off values to diagnose β-thalassaemia, genetic modifiers such as HbS should be considered.

3.2 Introduction

Before this work was conducted, β-thalassaemia had rarely been described in sub-Saharan Africa and, to the best of my knowledge, no specific mutations or confirmed cases of homozygous β-thalassaemia had been reported from within the region. The aim of this chapter, therefore, was to determine the prevalence of β-thalassaemia and the spectrum of mutations associated with β-thalassaemia in Kilifi. The work described in this chapter has resulted in two publications (Macharia et al., 2020) and (Macharia et al., 2022).

I first suspected the presence of β-thalassaemia in Kilifi when I was conducting analyses on the determinants of variation in haemoglobins A, A_2, F and S among infants aged 3 to 12 months who were recruited to the KGBC study in Kilifi, Kenya (Macharia et al., 2019). To my surprise, I observed that a group of children from this study had elevated haemoglobin A_2 (HbA_2) levels. HbA_2 is a minor haemoglobin synthesised from birth at levels of between 2.3% and 3.5% in normal children and is retained at these levels thereafter. HbA_2 levels are the most important parameter in the diagnosis of β-thalassaemia carriers (Van Delft et al., 2009). In most diagnostic facilities, an HbA_2 value of ≥4% has been used as a cut-off for diagnosis of heterozygous β-thalassaemia. However, the use of this cut-off has been queried since, in a few studies,

heterozygous β-thalassaemia cases have been identified among subjects with borderline values of between 3.5% and 4% (Colaco & Nadkarni, 2021; Mosca et al., 2009), an observation that requires further investigation. In the KGBC, study I observed that HbA_2 levels of ≥4% were detected in 97/12,853 (0.8%) HbAA infants (Macharia et al., 2019). This intrigued me because, to the best of my knowledge, β-thalassaemia had not previously been reported from the East African coast. Concurrently, in a study in which we were investigating the use of hydroxyurea among children with SCD (REACH trial), we found that despite manifesting both the clinical and phenotypic HPLC characteristics of SCD, 10/151 (10%) participants had genotypes that were suggestive of HbAS. On further investigation through sequencing of the β-globin gene region we found that, rather than being homozygotes for the sickle mutation (HbSS), all these individuals were affected by an alternative form of SCD caused by the coinheritance of the sickle mutation and one of two different β-thalassaemia mutations (HbS/β-thalassaemia): either rs33959855, a nonsense variant that results in premature termination of β-mRNA or rs193922563, a deletion that results in inactivation of an acceptor splice site (McGann et al., 2018). We concluded, therefore, that 10% of these children were suffering from HbS/β-thalassaemia out of which 8/10 (80%) was caused by coinheritance of HbS allele in combination with rs33959855 and 2/10 (20%) from coinheritance of HbS allele with rs193922563 mutation. It is from these observations that I became interested in documenting the true burden of β-thalassaemia within the Kilifi population, and whether these were the only mutations present.

In this chapter, therefore, I investigate the prevalence of β-thalassaemia and the spectrum of β-thalassaemia mutations in Kilifi and highlight the clinical significance through a case report of an individual born with homozygous β-thalassaemia. I investigated these questions using two cohorts of children, the KGBC and KSCDC studies, described in detail in Chapter 2. In the KGBC study, which was a sample of the entire KHDSS population, I screened participant samples to identify β-thalassaemia mutations inherited in the form of either homozygous β-thalassaemia (where both chromosomes are affected), HbA/β-thalassaemia or HbS/β-thalassaemia where one *HBB* allele carries a β-thalassaemia mutation while the contralateral

allele carries the HbS mutation. Because of inconsistencies in the literature regarding the cut-off values for HbA_2 that are considered to be diagnostic of β-thalassaemia carriage, I also performed a sensitivity and specificity analysis of HbA_2 in the diagnosis of β-thalassaemia. In the KSCDC study, which consists of children diagnosed with SCD, I screened for the presence of HbS/β-thalassaemia and the causative mutations. Within the KSCDC study I also report one case of homozygous β-thalassaemia. In the subsequent sections, I discuss the observations made in each cohort.

3.3 KGBC

3.3.1 Methods

3.3.1.1 Study Population

In this cohort, I selected and sequenced participant samples based on HbA_2 levels measured at the time of recruitment. Because HbA_2 levels measured using Bio-Rad Variant Classic™ HPLC analyser are artefactually elevated in the presence of HbS (Suh, Krauss, & Bures, 1996), I first grouped the participants into the categories HbAA, HbAS and HbSS (see Section 2.3.1 for further details on this classification).

In HbA/β-thalassaemia cases one chromosome is affected by a β-thalassaemia mutation, but the individual does not carry the HbS mutation and therefore appears as HbAA by HPLC. For this reason, to identify HbA/β-thalassaemia carriers, I first sequenced all children with HbA_2 values of $\geqslant$4.0% in those with the phenotype HbAA. Next, I sequenced a random sample of 110 participants whose values were between 3.5 and 3.9%, a cut-off associated with a high sensitivity but low specificity for β-thalassaemia. Finally, I sequenced a random sample of 114 participants whose values were <3.5%, in whom β-thalassaemia should be rare.

To identify HbS/β-thalassaemia carriers, I sequenced samples from participants with HPLC patterns suggestive of HbSS and HbAS. Because HbA_2 values as measured by Bio-Rad Variant Classic™ are falsely elevated in the presence of HbS (Suh et al., 1996), I sequenced all

participants with an HPLC pattern suggestive of HbSS. In participants with HPLC pattern suggestive of HbAS, β-thalassaemia can be found in the form of HbS/β^+-thalassaemia whereby it results in reduced synthesis of HbA and as a result increased HbS (Weatherall & Clegg, 2001). HbS/β^+-thalassaemia can thus be diagnosed through sequencing HbAS participants in whom the concentration of HbS is higher than of HbA. Based on this observation, among participants with HPLC pattern suggestive of HbAS, I sequenced samples from all participants in whom concentration of HbS exceeded that of HbA. Additionally, to have a representative sample of all HbA_2 categories from the HbAS group, I sequenced a random sample of ~100 participants with each of the HbA_2 cut-offs as outlined for HbAA subjects above, in whom β-thalassaemia mutations would not be expected.

3.3.1.2 Human β-globin sequencing

I sequenced the β-globin region using the Sanger sequencing method as described in Section 2.3.4.

3.3.1.3 Statistical Analysis

Differences in mean HbA_2 percentages between different β-thalassaemia pathogenic variants were analysed using one-way ANOVA. P-values of <0.05 were considered statistically significant. All statistical analyses were performed using the R Foundation for Statistical Computing Platform Version 3.1.1. (R Core Team (2020)).

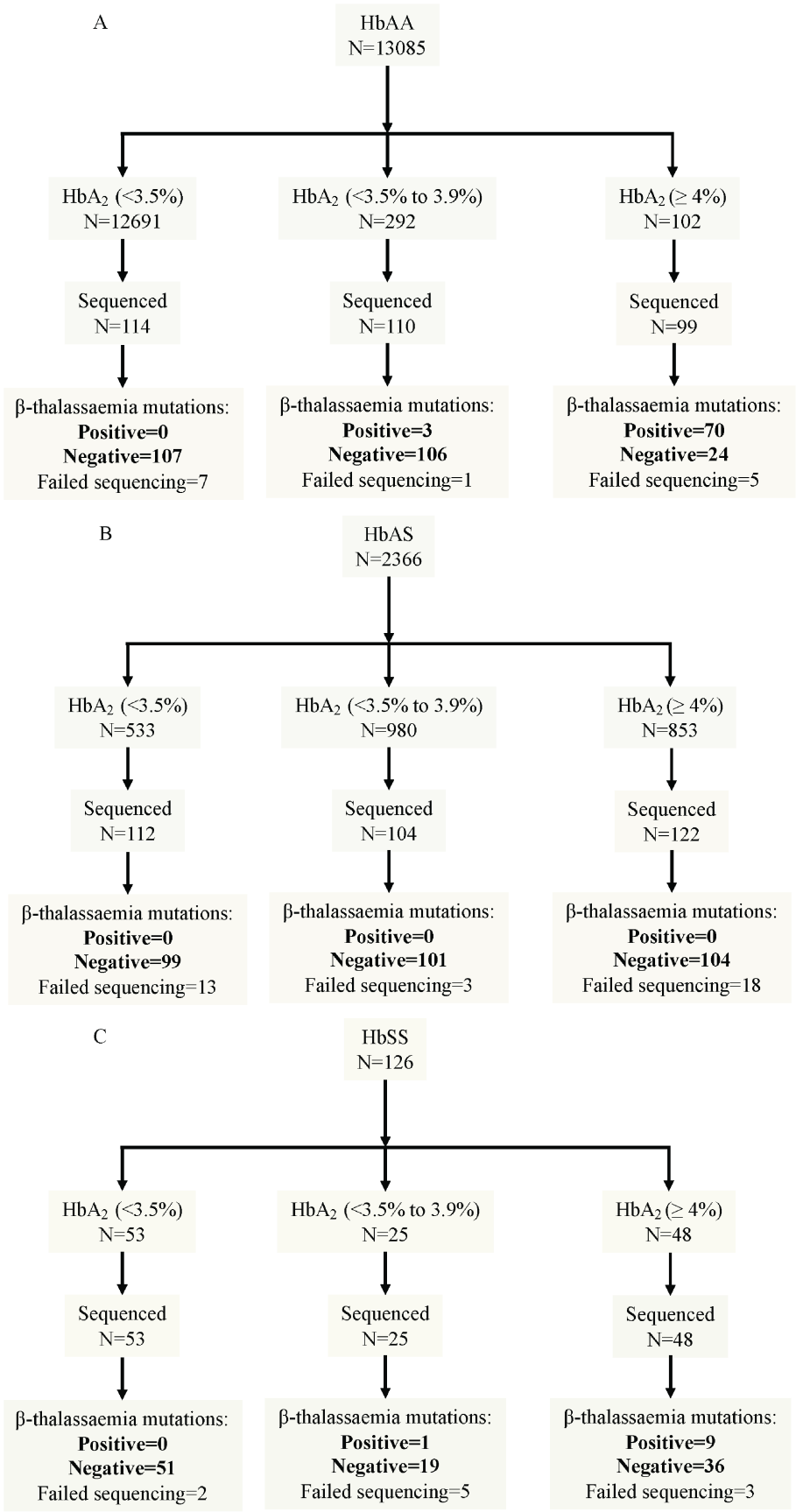
3.3.2 Results

3.3.2.1 Prevalence of β-thalassaemia in KGBC

Overall, 15,577 participants from KGBC were included in this analysis (see chapter 2, section 2.2.1). The median age was 6.5 (IQR 5.0-8.4) months. In total, 13,085 (84%) showed an HPLC pattern suggestive of HbAA, 2,366 (15.2%) of HbAS and 126 (0.8%) of HbSS.

Based on my sampling strategy, I selected 787 participants distributed for sequencing as follows: HbAA (n=323), HbAS (n=338) and HbSS (n=126). I successfully sequenced a total of 730 samples out of which 83 participants were found to have β-thalassaemia pathogenic variants, all as heterozygous carriers. Overall, 73/310 (24%) and 10/116 (9%) of the β-thalassaemia alleles were found among the presumed HbAA and HbSS groups respectively. I found no β-thalassaemia alleles among any of the participants with an HPLC pattern suggestive of HbAS (Figure 15). I only identified two participants within this group in whom the concentration of HbS exceeded that of HbA. The proportions of HbA, HbS, HbF and HbA_2 for the two individuals were (i) 21%, 26%, 51%, 2% and (ii) 34%, 36%, 27%, 3% respectively. No β-globin mutations were identified in either of these participants.

Figure 15. Sample selection and sequencing results for identification of β-thalassaemia pathogenic variants in participants with different HbS phenotypes.



Note: Total number of study participants was 15,577 distributed as (a) HbAA = 13,085 (84.0%); (b) HbAS = 2,366 (15.2%), and (c) HbSS = 126 (0.8%).

Estimation of the number of β-thalassaemia alleles in the population

Since the samples selected for sequencing were representative of the different HbA_2 categories for the various subgroups of HbAA, HbAS and HbSS in KGBC, I extrapolated the allele frequencies for the whole population in KGBC using the allele frequencies generated in each of the HbA_2 categories as shown in Table 5. An example of how I did this is in the HbAA category whereby in the group with HbA_2 of <3.5% to 3.9% the total number individuals successfully genotyped was 109 and total number of β-thalassaemia carriers was 3 making the allele frequency 3/218 (1.4%). To calculate the total number of β-thalassaemia alleles for the whole population from the same category of HbA_2, I multiplied the β-thalassaemia allele frequency of 1.4% with the total number of alleles for the whole population in KGBC with HbA_2 of <3.5% to 3.9% which was 584 (292 participants). In this case the total number of β-thalassaemia alleles in HbA_2 category of between <3.5% to 3.9% within the HbAA group was extrapolated to be 8. The overall minimum allele frequency was 0.3% which was calculated by summing the total number of extrapolated β-thalassaemia alleles in all categories and dividing by the total number of alleles in KGBC population (95/31,154).

Estimating the prevalence of β-thalassaemia homozygotes and heterozygotes in the population

Assuming Hardy-Weinberg equilibrium, the estimated prevalence of β-thalassaemia carriers and homozygotes was calculated using the Hardy-Weinberg equation as follows.

Hardy-Weinberg equation: $p + q = 1$ and $p^2 + 2qp + q^2=1$

q = proportion of β-thalassaemia alleles in the population.

p= proportion of normal alleles in the population.

$q = 95/31,154=0.003$

$p=1- 0.003 =0.997$

2pq = expected genotype frequency for heterozygous β-thalassaemia

$2pq = 2*(0.003*0.997) = 0.006$

q^2= expected genotype frequency for homozygous β-thalassaemia

$q^2 = (0.003*0.003) = 0.000009$

When rounded off to the nearest whole number, the estimated prevalence's for β-thalassaemia heterozygotes and homozygotes within the study population is 6/1000 and 1/100,000, respectively.

Estimation of prevalence of sickle cell homozygotes

Assuming Hardy-Weinberg equilibrium, the estimated prevalence of HbSS was calculated as follows.

q = proportion of HbS alleles in the population.

p= proportion of normal alleles in the population.

q= (HbSS alleles (126*2) +HbAS alleles (2366)) ÷ total alleles in the population (31,154)

q=0.084

p=1-0.084=0.916

q^2= expected genotype frequency for homozygous sickle cell anaemia cases

$q^2 = (0.084*0.084) = 0.007$

Based on these results the estimated prevalence of HbSS within this population is 1/100. The sequencing results showed that 10/117 (~10%) participants identified as HbSS using HPLC were HbS/β^0-thalassaemia carriers. Based on this observation, I estimate the prevalence of HbS/β-thalassaemia to be 1/1000, meaning that approximately 10% of all cases of SCD within the Kilifi population are due to HbS/β^0-thalassaemia.

Table 5. The prevalence of β-thalassaemia by HbA_2 categories.

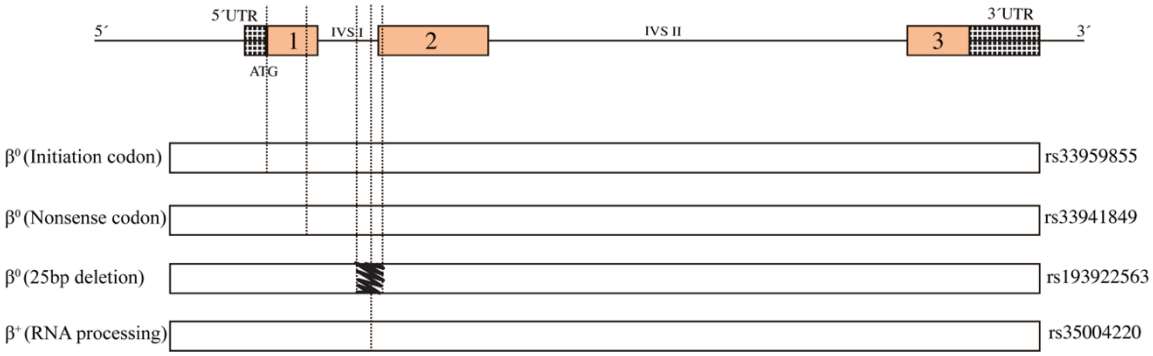
HbA_2 category	HbAA			HbSS			HbAS*		
	Observed β–thalassaemia alleles (n/N %)	**Alleles in total population (P)**	**Estimated β–thalassaemia alleles within the total population**	**Observed β–thalassaemia alleles (n/N) (%)**	**Alleles in total population (P)**	**Estimated β–thalassaemia alleles within the total population**	**Observed β–thalassaemia alleles (n/N) (%)**	**Alleles in total population (P)**	**Estimated β–thalassaemia alleles within the total population**
HbA_2 (<3.5%)	0/214 (0)	25,382	0	0/102 (0)	106	0	0/198 (0)	1066	0
HbA_2 (3.5-3.9%)	3/218 (1.4%)	584	8	1/40 (2.5%)	50	1	0/202 (0)	1960	0
HbA_2 (≥4.0%)	70/188 (37.2%)	204	76	9/90 (10%)	96	10	0/208 (0)	1706	0
Total			**84**			**11**			

Note: The estimated numbers of β-thalassaemia alleles within each subgroup were calculated from the fractions within each Hb type and HbA_2 category, multiplied by the number of alleles within the same subgroups in the whole population. * No β^+-thalassaemia alleles were identified in any members of this group, including the two infants in whom the proportion of HbS exceeded that of HbA.

3.3.2.2 Spectrum of mutations associated with β-thalassaemia in the KGBC

I observed four different β-thalassaemia pathogenic variants in total (Figure 16 and Table 6). Two of these mutations rs33959855 and rs193922563 of the β^0-thalassaemia type had been identified in one of our previous studies (McGann et al., 2018), the newly identified mutations were rs33941849, an initiation codon mutation that abrogates the transfer RNA binding site (Wildmann et al., 1993) and rs35004220, a β^+ pathogenic variant that results in formation of a new splice site resulting in 80% abnormal RNA (Giardine et al., 2014). β^0-thalassaemia pathogenic variants were the most common, representing 97.6% of all pathogenic variants identified, while rs33959855 was the most common overall (66.3%). Although the group sizes were too small to allow definitive conclusions, I saw no differences in HbA_2 values between these pathogenic variants (Table 6).

Figure 16. Pathogenic variants causing β-thalassaemia in Kilifi.



The upper panel shows the β-globin gene with the respective exons labelled 1, 2 and 3. The dashed lines represent the positions where the mutations are located. The boxes below show the type of mutations identified.

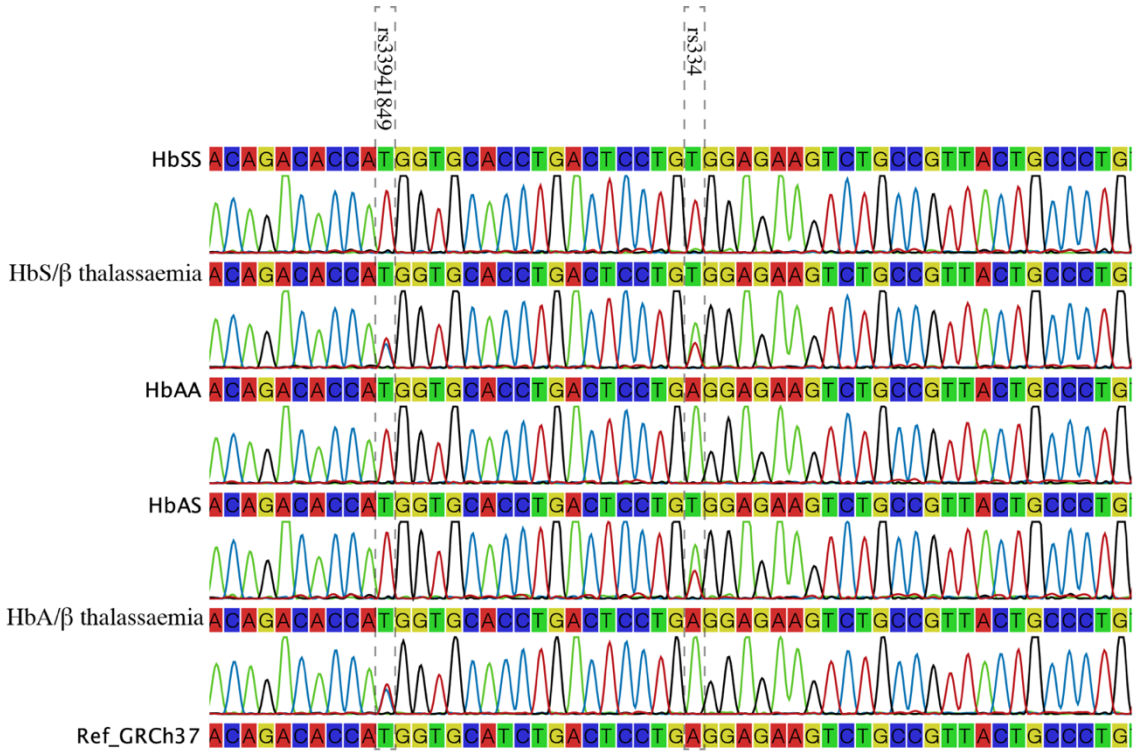
Table 6. The distribution of β-thalassaemia pathogenic variants within the study population.

SNP	HGVS coding	Location	Type	Alleles n (%)	Mean HbA_2 (%; 95% CI)[§]
rs33959855	NM_000518.5:c.67G>T	CD22 (GAA→TAA)	β^0	55 (66.3%)	4.95 (4.76-5.14)
rs33941849	NM_000518.5:c.2T>C	Initiation codon (ATG→ACG)	β^0	20 (24.1%)	5.37 (4.67-6.06)
rs193922563	NM_000518.5:c.93-22_95del	IVS1-3′ end del 25 bp	β^0	6 (7.2%)	4.88 (4.18-5.58)
rs35004220	NM_000518.5:c.93-21G>A	IVS-I-110 (G→A)	β^+	2 (2.4%)	4.30 (NA)

[§]No significant differences were seen in % HbA_2 means when comparing β-thalassaemia pathogenic variants to each other. P=0.25

As shown in Figure 17 these mutations were inherited in combination with the rs334 (HbS locus) resulting in either HbA/β-thalassaemia or HbS/β-thalassaemia. In individuals identified as HbSS using HPLC, sequencing results for the 10 individuals with β-thalassaemia mutations revealed that the genotype at the HbS loci was heterozygous (HbAS). Thus, in individuals identified as HbSS using HPLC, identification of those with β-thalassaemia mutations can be done through genotyping for the rs334 mutation (Figure 17).

Figure 17. Sequencing chromatograms showing rs33941849 and rs334 mutations in individuals identified as HbAA, HbAS, HbSS, HbA/β-thalassaemia and HbS/β-thalassaemia.



The chromatogram shows sequencing alignment for five individuals identified as HbAA, HbAS, HbSS, HbA/β-thalassaemia and HbS/β-thalassaemia. The x-axis shows the various nucleotides present at those positions upon alignment with the reference genome (GRCh37). Individuals who were homozygous for an allele had only one peak at the loci whereas those who were heterozygous had two peaks. The nucleotide position for rs33941849 and rs334 is labelled at the top of the image. The difference between HbSS and HbS/β-thalassaemia is seen at these positions. The HbSS individual is homozygous for the mutant allele at rs334 loci and wildtype allele at rs33941849 loci whereas HbS/β-thalassaemia individual is heterozygous at both sites.

3.3.2.3 HbA_2 levels and the diagnosis of β-thalassaemia in KGBC.

To assess the accuracy of using an HbA_2 level of ≥4% as the diagnostic cut-off for β-thalassaemia carriage, I calculated the sensitivity, specificity and positive- (PPV) and negative predictive values (NPV) with sequencing as the gold standard. Because HbS has been suggested to result in incorrectly elevated HbA_2 (Suh et al., 1996), I calculated the diagnostic accuracy of the HbA_2 cut-offs in the presence or absence of HbS.

In the absence of HbS, the sensitivity, specificity and positive- (PPV) and negative predictive values (NPV) for a threshold of ⩾3.5% was 100%, 98%, 21% and 100% respectively while one of ⩾4% gave values of 91%, 100%, 75% and 100% respectively (Table 7). When all participants were included in this analysis, including those with presumed HbAS and presumed HbSS,

specificity and PPV estimates for an HbA_2 threshold of ⩾3.5% dropped substantially to 86% and 4% respectively and those for a threshold of ⩾4% fell to 94% and 9% respectively.

Table 7. The diagnostic accuracy of HbA_2 values in the prediction of β-thalassaemia within the sub-group of participants with HPLC patterns consistent with HbAA.

β-thalassaemia	**HbA_2 <3.5%**	**HbA_2 ≥3.5%**	**HbA_2 ≥4%**
Present[§]	0	84	76
Absent[§]	12694	310	26
Sensitivity (%; 95% CI)	N/A	100.0 (94.5-100.0)	90.5 (81.6-95.5)
Specificity (%; 95% CI)	N/A	97.6 (97.3-97.9)	99.8 (99.7-99.9)
PPV (%; 95% CI)	N/A	21.3 (17.4-25.8)	74.5 (64.7-82.4)
NPV (%; 95% CI)	N/A	100.0 (99.9- 100)	100.0 (99.9-100.0)

N/A not applicable – no β-thalassaemia alleles were detected within this group. [§]The estimated numbers of β-thalassaemia carriers within each sub-group were calculated from the fractions within each Hb type and HbA_2 category, multiplied by the number of participants within the equivalent sub-group within the whole population.

3.4 The Kilifi Sickle Cell Disease Cohort (KSCDC)

3.4.1 Methods

3.4.1.1 Study population

The KSCDC consists of children clinically diagnosed with SCD and confirmed in the laboratory using HPLC (see Section 2.2.3 for more details on the cohort). All children in the KSCDC with an EDTA blood sample available were included in this analysis.

3.4.1.2 Human β-globin sequencing

In the analysis of the β-thalassaemia prevalence within the KGBC study group, I observed that 10% of the children had inherited SCD in the form of HbS/β-thalassaemia. The sequencing results showed that in individuals identified as HbSS using HPLC, HbS/β-thalassaemia cases could be identified through genotyping the rs334 mutation (Figure 17). Based on this observation I genotyped for HbS using ASA PCR (see Section 2.3.3) to discriminate between participants who were homozygous for rs334 (HbSS) and those who were heterozygotes

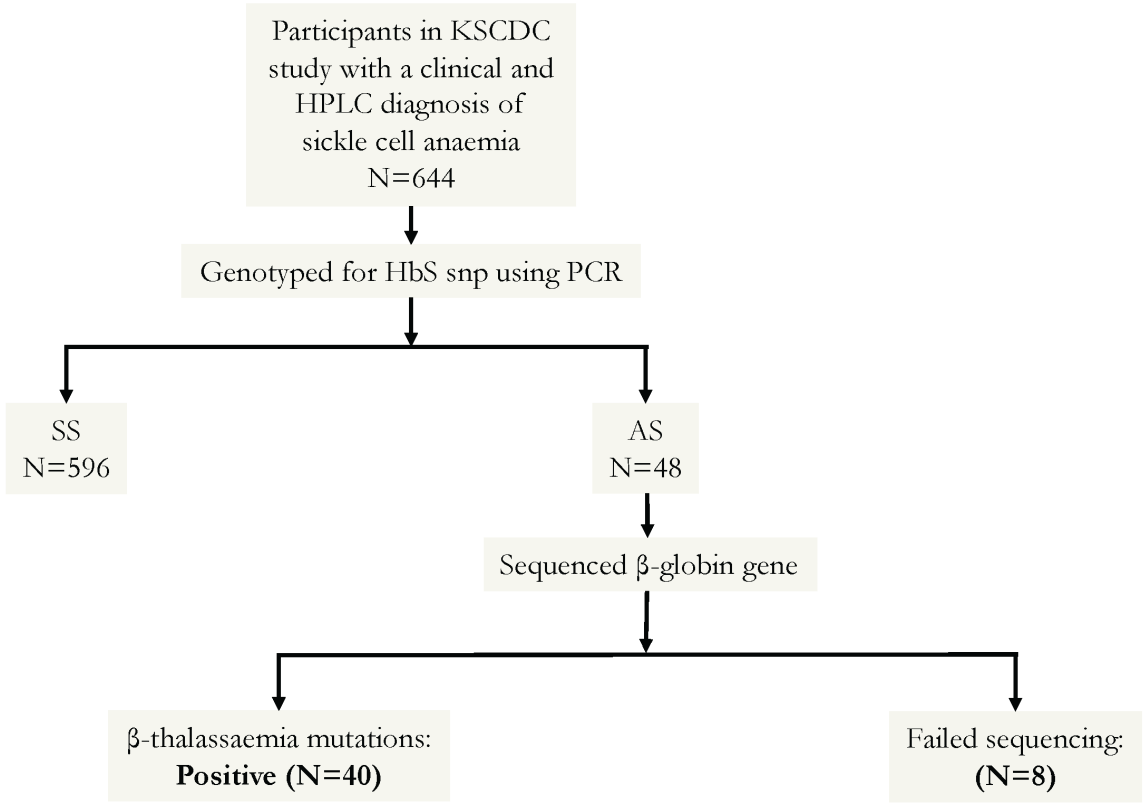
(HbAS). I then sequenced the β-globin gene region by use of the Sanger sequencing method (Section 2.3.4) in those whose PCR genotype revealed a diagnosis of HbAS.

3.4.2 Results

3.4.2.1 Prevalence of β-thalassaemia in the KSCDC

Genotyping revealed that 48/644 (~8%) members of the KSCDC study showed PCR results that were compatible with a diagnosis of HbAS. On sequencing, I found β-thalassaemia mutations in all but 8 of these participants in whom the sequencing failed after several attempts. The prevalence of HbS/β-thalassaemia is this group was 6% (40/636) (Figure 18).

Figure 18. Flow chart showing identification of β-cases in the KSCDC study.



DNA from all the participants was genotyped for the rs334 mutation. β-globin gene was then sequenced in those whose PCR genotype was consistent with HbAS.

3.4.2.2 Spectrum of mutations associated with β-thalassaemia in the KSCDC

Three mutations were identified among the 40 individuals with HbS/β-thalassaemia. The three mutations identified were rs33959855, rs33941849 and rs193922563. Apart from rs35004220 which was not observed in this cohort, the rest of the mutations had already been identified in the KGBC, with the only difference being their frequencies. The most common mutation in the current cohort was rs33941849, which accounted for 40% of all the mutations found (Table 8), whereas in KGBC it was rs33959855 which accounted for 66% of the mutations identified (Table 6). No new mutations were identified in this cohort.

Table 8. The prevalence and spectrum of β-thalassaemia pathogenic variants in patients attending the SCD clinic in Kilifi County Hospital.

SNP	Location	Type	Alleles n* (%)
rs33959855	CD22 (GAA→TAA)	β^0	14 (35%)
rs33941849	Initiation codon (ATG→ACG)	β^0	16 (40%)
rs193922563	IVS1-3′ end del 25 bp	β^0	10 (25%)

The table shows the mutations identified among 40 individuals whose samples were successfully sequenced.
*n represents the number of alleles identified per mutation.

3.5 A case of homozygous β-thalassaemia

While conducting my PhD investigations a clinician who was aware of my PhD work brought to my attention the case of a child who had been clinically diagnosed with SCD but whose laboratory diagnosis by HPLC, though abnormal, did not show the presence of HbS. I discuss this case in further detail below.

The child, a two-and-a-half-year-old female, presented to Kilifi County Hospital, with a one-week history of left sided abdominal swelling. No previous hospital admissions were reported. Clinical history suggested delayed developmental milestones; specifically, she was unable to walk without support. The child was the fourth born of five siblings, all of whom were alive and well as were both of her parents. Both her parents were of Mijikenda ethnolinguistic ancestry and no recent genetic admixture was apparent from the clinical history. On physical examination,

the child was pale but had no signs of clinical jaundice. Her vital signs were essentially normal with the exception of a fever measured at 38.8°C per axilla. Frontal-maxillary skull bossing was apparent. Her abdomen was distended, soft and non-tender, massive splenomegaly being detected at 8cm below the costal margin. She was severely malnourished with a weight of 8.8kg, a height of 78.5cm, a height for age z-score (HAZ) of -3.79, a weight for age z-score (WAZ) of -3.20 and a weight for height z-score (WHZ) of -1.23. Further examination was essentially normal. A full hemogram revealed marked anaemia (Hb 6.6g/dL), a low mean corpuscular volume (MCV) of 64fL, a low mean corpuscular haemoglobin (MCH) of 19.4pg, and a raised total white blood cell (WBC) count of 49.6 × 10^9/μl which were predominantly lymphocytes. Her platelet count was normal at 321 × 10^6/L while her plasma creatinine level was mildly elevated at 32μmol/L. Blood cultures and tests for malaria were negative. A peripheral blood film revealed nucleated red blood cells (RBCs), microcytes, dacrocytes, acanthocytes, giant platelets and a marked lymphocytosis (Table 9).

Table 9. Complete blood count and peripheral blood film from the child with β-thalassaemia.

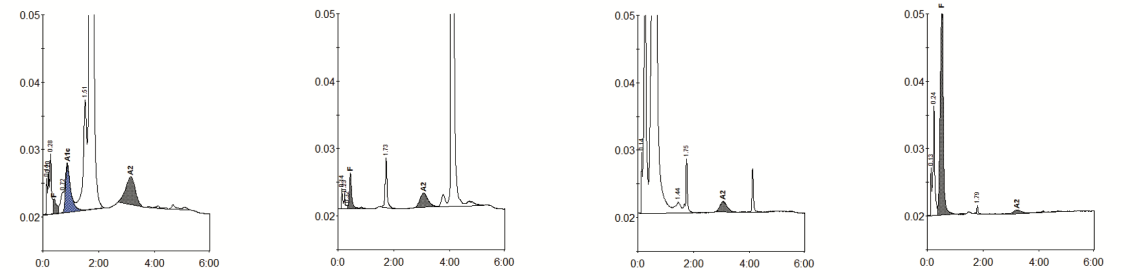
Parameter	First admission [¢]	Second Admission [‡]
HbA_2 (%)	2.5	5.0
HbF (%)	>80%	>80%
WBC count (x10^3/ml)	49.6	20.5
RBC count (x10^6/ml)	3.38	1.10
Hb (g/dl)	6.6	2.2
HCT (%)	21.5	6.2
MCV (fL)	64	56
MCH (pg)	19.4	20.1
Platelets (x10^6/L)	321	330
Peripheral blood film	Leucocytosis, lymphocytosis, nucleated red blood cells, dacrocytes, anisocytosis and giant platelets	

Abbreviations: WBC, white blood cells; RBC, red blood cells; Hb, haemoglobin; HCT, haematocrit; MCV, mean cell volume; MCH, mean cell haemoglobin; PBF, peripheral blood film. [¢]Age=2.5 years, [‡]Age=3.5 years.

The child was admitted to the general paediatric ward with a working diagnosis of iron deficiency anaemia, potentially complicated by bacterial sepsis, and with a differential diagnosis of SCD. She was treated empirically with iron and folic acid supplementation for her anaemia and with intravenous penicillin and gentamicin to cover sepsis. She was also prescribed malaria prophylaxis with proguanil pending analysis for SCD by high-performance liquid chromatography (HPLC). Her fever subsided within two days of admission, at which point she was discharged home on oral amoxicillin, with follow-up planned for the following week.

The results of her HPLC analysis, received after discharge from hospital, revealed the absence of normal adult haemoglobin (HbA), normal levels of HbA_2 at 2.5% and elevated levels of foetal haemoglobin (HbF) (>80% of total Hb) that eluted in adjacent peaks A1b (16%) and LA1C/cHb1 (76.5%) (Figure 19). The complete absence of HbA suggested a diagnosis of β^0-thalassaemia major. On suspecting the child as a possible case of homozygous β-thalassaemia, I sequenced the β-globin gene and confirmed that she was homozygous for the rs33941849 initiation codon mutation.

Figure 19. HPLC chromatograms from study participants with normal HbA individual (HbAA), homozygous haemoglobin S (HbSS) and homozygous β-thalassaemia patient at first admission (age 2.5 years) and at second admission (age 3.5 years).



Peak	R.time	Height	Area	Area %
Unknown	0.14	5516	13495	0.4
A1a	0.20	6243	17484	0.6
A1b	0.28	9074	40453	1.3
F	0.41	2173	23712	< 0.8 *
LA1c/CHb-1	0.72	3026	26485	0.9
A1c	0.88	6906	89786	4.5
P3	1.51	16370	148900	4.9
A0	1.72	401677	2584872	84.9
A2	3.14	4110	98538	3.1
Total Area:			3043724	

Concentration:	%
F	< 0.8 *
A1c	4.5
A2	3.1

Peak	R.time	Height	Area	Area %
Unknown	0.14	3072	7957	0.6
A1a	0.23	2256	6611	0.5
A1b	0.32	579	2271	0.2
F	0.45	5375	29479	1.8
A0	1.73	7383	34515	2.5
A2	3.08	2118	43521	3.6
S-Window	4.09	259324	1281753	91.2
Total Area:	1406107			

Concentration:	%
F	1.8
A2	3.6

Peak	R.time	Height	Area	Area %
Unknown	0.14	9534	19501	1.4
A1b	0.26	41288	224095	16.0
LA1c/CHb-1	0.63	93362	1070118	76.5
P3	1.44	1601	23465	1.7
A0	1.75	8139	33049	2.4
A2	3.06	1575	28938	2.5
Total Area:	1399165			

Concentration:	%
A2	2.5

Peak	R.time	Height	Area	Area %
Unknown	0.13	7112	16242	4.9
A1a	0.24	16276	70033	21.2
F	0.52	31333	230159	80.1 *
A0	1.79	1188	4412	1.3
A2	3.22	527	9657	5.0
Total Area:	330503 *			

Concentration:	%
F	80.1 *
A2	5.0

On each chromatogram the x-axis represents the retention time for the haemoglobin and the y-axis the concentration. The area under the peak (Area %) represents the concentration of each haemoglobin.

Initially lost to follow-up, the child re-presented at the age of three years 11 months with a one-week history of a cough and fever. On examination at that time, her spleen remained grossly enlarged at 10 cm, and she remained malnourished with a HAZ of -4.98, a WAZ of -4.01 and a WHZ of -0.99. Although hemodynamically stable, she was profoundly anaemic (Hb 2.2 g/dL) and was therefore transfused and treated with folic acid supplementation and nutritional support. Repeat HPLC analysis revealed the continued absence of HbA together with elevated levels of HbF (>80%) and HbA_2 (at 5%) (Figure 19). PCR for the $\alpha^{-3.7}$ deletional form of α -thalassaemia was negative. The timeline of events is given in Table 10.

Table 10. Timeline of events.

Age at presentation	Symptoms	Diagnostic Testing	Interventions
2 years 6 months	• Delayed developmental milestones • Left sided abdominal swelling • 8 prior transfusions (other hospital) • Low grade fever • Bossing of the skull • Massive splenomegaly • Malnutrition	• HPLC analysis; absence of HbA and elevated HbF (>80% of total Hb) • Sequenced her *HBB* gene region, which revealed she was homozygous for the initiation codon (ATG→ACG) mutation (rs33941849) • Full haemogram; hb 6.6gm/dl	• Admission • Treated for suspected sepsis
3 years 11 months	• Cough • Fever • Splenomegaly • Malnutrition • Transfusion dependent anaemia	• PCR for the $\alpha^{-3.7}$ deletional form of α-thalassaemia was negative • Repeat HPLC analysis revealed the continued absence of HbA, HbF (>80% of total Hb) and HbA_2 at 5%	• Began regular monthly blood transfusions • Referred for surgical splenectomy

3.6 Discussion/Conclusion

In this chapter, I set out to determine the prevalence, causal mutations and clinical significance of β-thalassaemia on the Coast of Kenya. To answer this objective, I screened members of the KGBC and KSCDC for β-thalassaemia mutations and made several important observations.

The first observation was on the utility of HbA_2 cut off values used in the diagnosis of β-thalassaemia. Whereas in most laboratories HbA_2 value of ⩾4% is used as a cut-off for diagnosis of β-thalassaemia carriers there has been several studies showing β-thalassaemia carriers with HbA_2 values of between 3.5% to 4% (Colaco & Nadkarni, 2021; Mosca et al., 2009) indicating that this cut-off might not be sensitive enough to identify all cases of β-thalassaemia carriers. In the Colaco meta-analysis of the causes of borderline HbA_2 values, the authors observed that 20% were β-thalassaemia carriers, 15% had some other genetic defect in either their α-, β- or δ-globins with the cause in the remaining mainly unknown. For these reasons, I decided to evaluate the diagnostic utility of HbA_2 cut-off values in the diagnosis of β-thalassaemia in our own study population. To do this I looked at the sensitivity and specificity of HbA_2 cut off values of ⩽3.5%, 3.5% to 4% and ⩾4% in diagnosis of β-thalassaemia carriers selected according to sickle cell status (HbAA, HbAS and HbSS) as determined using HPLC. I found that among HbAA participants, an HbA_2 value of ⩾4% was associated with a sensitivity of 91% and a positive predictive value of 75% for β-thalassaemia heterozygosity whereas that of ≥3.5% was associated with a sensitivity of 100% and a positive predictive value of 21%. Importantly was the diagnosis of β-thalassaemia carriers in individuals with borderline HbA_2 levels of 3.5% to 4%, who would have otherwise been missed had I only used the commonly used cut-off of HbA_2 ⩾4%. Nonetheless and in keeping with previous reports (Colaco & Nadkarni, 2021; Gasperini et al., 1993), I found no β-thalassaemia alleles in 26% of individuals with HbA_2 of ≥4% among HbAA participants. This could be explained by the presence of rare pathogenic variants in the δ-globin gene, α-globin gene defects, mutation in KLF1 gene or by other causes of elevated HbA_2, which include megaloblastic anaemia, iron deficiency anaemia, HIV infections and hypothyroidism (Weatherall & Clegg, 2001), none of which I investigated in the current study. When considering the entire population, including those with HPLC phenotypes consistent with either HbAS or HbSS, the use of an HbA_2 level of ⩾4% resulted in a considerably higher false positivity rate, reflected in a PPV of only 9%. This is almost certainly explained by the presence of glycated HbS and adducts associated with HbS, which

have been shown to co-elute with HbA_2 and to result in artefactually raised HbA_2 values when using the BioRad Variant system (Macharia et al., 2019; Suh et al., 1996). These observations underscore the importance of including borderline HbA_2 range of 3.5% to 4% and accounting for the effect of genetic modifiers such as HbS in diagnosis and in population screening for β-thalassaemia pathogenic variants.

The second observation was on the spectrum of mutations causing β-thalassaemia in our population. To date, more than 300 β-thalassaemia pathogenic variants have been identified worldwide (Kountouris et al., 2014), although approximately 90% of cases are caused by only 40 (Thein, 2018). The majority are point mutations or small deletions or insertions that affect *HBB* gene function at either the transcriptional, post transcriptional or translational stages (De Sanctis et al., 2017; Thein, 2018; Weatherall & Clegg, 2001). Pathogenic variants differ by geographical region with only one or two accounting for >50% of cases within any given region (Weatherall & Clegg, 2001). In North Africa for example, rs35004220 is the most common pathogenic variant, which together with CD39 C→T, IVS1-1 G→A and IVS1-6 T→C is responsible for over 60% of cases (Douzi et al., 2015; Elmezayen et al., 2015; Weatherall & Clegg, 2001). The cluster of pathogenic variants defining β-thalassaemia in Kilifi, observed through this study, is dominated by two β^0-thalassaemia variants, rs33959855 and rs33941849, accounting for 90% of the cases while the remaining cases were explained by one additional β^0-thalassaemia variant (rs193922563) and a single β^+-thalassaemia variant (rs35004220). Most are rare in other populations: for example, with the exception of our own previous study (McGann et al., 2018), to the best of our knowledge the main pathogenic variant we found (rs33959855) has only previously been reported in one individual from the Reunion Republic (Ghanem et al., 1992) in whom the clinical outcome was not described. Similarly, to date, rs33941849 has only been reported in three members of one family from Yugoslavia (Wildmann et al., 1993), two family members of Swiss origin (Beris, Darbellay, Speiser, Kirchner, & Miescher, 1993), and two family members of Russian origin (Molchanova, Postnikov Yu, Gu, & Huisman, 1998). Notably, in all three studies it was reported that HbA_2 levels were higher than those commonly

seen in other β-thalassaemia pathogenic variants. In KGBC I identified 18 participants who were carriers of this mutation and in whom HbA_2 levels were also higher than in carriers of other variants, although this observation did not reach statistically significance. Finally, the third most common pathogenic variant observed in this study was rs193922563. First identified by Orkin and colleagues (Orkin, Sexton, Goff, & Kazazian, 1983) in a patient of Indian origin, this pathogenic variant has since been found to be common in a number of Middle Eastern populations. The highest frequencies have been reported in Bahrain, where in one study it accounted for 36% of all β-thalassaemia alleles (Jassim & Krishnamoorthy, 2002). However, in other Middle Eastern countries the frequencies of this pathogenic variant are not as high, being 7.3% in Kuwait (Adekile et al., 1994), 9.5% in United Arab Emirates (el-Kalla & Mathews, 1997) and 12.9% in Saudi Arabia (el-Hazmi, al-Swailem, & Warsy, 1995). In the current study rs193922563 explained 6.8% of the β-thalassaemia alleles in our population. Two occurrences of the β^+-thalassaemia pathogenic variant rs35004220, a common variant in North African, European and the Middle Eastern countries (Agouti, Badens, Abouyoub, Levy, & Bennani, 2008; Bennani et al., 1993; Douzi et al., 2015; Elmezayen et al., 2015; Weatherall & Clegg, 2001), were identified in our study. The mutation results in formation of a new splice site resulting in synthesis of 80% abnormal mRNA and 20% normal mRNA (Spritz et al., 1981; Weatherall & Clegg, 2001).

The third observation was the contribution of these mutations towards the disease burden affecting this community. The three β^0-thalassaemia variants observed in this study are of clinical importance as they contribute to clinical disease by causing SCD in the form of HbS/β-thalassaemia and β-thalassaemia major. HbS/β-thalassaemia results from coinheritance of both a β-thalassaemia and a β^s pathogenic variant on contralateral chromosomes. Although the clinical manifestations of HbS/β-thalassaemia are thought to be generally similar to those of HbSS, they can vary depending on the type of β-thalassaemia mutation that is co-inherited, which in turn varies from one region to another (M. H. Steinberg, Forget, B. G., Higgs, D. R., & Weatherall, D. J., 2009). In particular, depending on the amount of β-globin produced,

individuals of the HbS/β^+-thalassaemia can have a milder form of SCD compared to HbS/β^0-thalassaemia and HbSS (Jha et al., 2018; Serjeant et al., 1979; Yadav et al., 2016). Describing the disease phenotypes that are associated with these mutations is important because it can inform guidelines on better management of individuals suffering from these conditions. In the current study, HbS/β^0-thalassaemia was the main syndrome identified in the two cohorts. The syndrome accounts for ~10% of SCD cases in this community and are part of the patients receiving clinical care under the KSCDC study.

The clinical significance of these mutations is further seen in the diagnosis of a case of β-thalassaemia major associated with rs33941849 mutation. To the best of my knowledge there has been no case of β-thalassaemia major reported from East African region. The child in this study had visited the hospital severally and because the clinicians were unaware of existence of β-thalassaemia in the region, the child had been misdiagnosed and managed as a case of SCD. The management of these two conditions is however different since individuals who are homozygous for β^0-thalassaemia do not produce any adult haemoglobin and as a result they are transfusion dependent for survival (Weatherall & Clegg, 2001). The only other case of β-thalassaemia major associated with rs33941849 mutation ever reported was a case of a male child of Pakistani origin who presented at 10 months of age with a palpable liver and spleen at 7cm and 3cm below costal margin, respectively. His Hb was 9.2g/dL, MCV of 73fL and MCH of 33pg. He was also found to be homozygous for the $\alpha^{-3.7}$-thalassaemia deletion and to have a Bantu β-globin gene cluster haplotype. He was managed with regular blood transfusions (Khan, Riazuddin, & Galanello, 2000). On comparing the current and previously described cases, all had anaemia, a low MCV and massive splenomegaly. In our current patient, I also observed elevated levels of HbF and varying levels of HbA_2 at the two points of testing, an observation which is common in β-thalassaemia major (M. H. Steinberg & Rodgers, 2015). I have estimated the prevalence of β-thalassaemia major in this region to be 1 in 100,000, which means cases of homozygotes should be seen. Nevertheless, low awareness of this condition among clinicians and the low availability of diagnostic facilities within the region mean that historically,

individuals with β-thalassaemia major have probably been misdiagnosed with other conditions such as SCD or iron deficiency anaemia as was the case with this child. As such, I hope that this study will raise awareness about the existence and clinical importance of β-thalassaemia major as a public health problem within the East Africa region and lead to the development of locally appropriate diagnostic and treatment guidelines.

This work had two important limitations: the age range of the children studied and the method by which blood was collected within the KGBC study. Switching to adult patterns of haemoglobin production may well have been incomplete, particularly in the youngest subgroup, making the classification of HbA_2 values potentially misleading. In addition, the method of blood collection onto filter paper did not allow us to refine our screening strategy to take account of data from complete blood counts or to investigate children for alternative causes of raised HbA_2 values, which include megaloblastic anaemia, HIV infections and hypothyroidism (Weatherall & Clegg, 2001). Although deficiencies in the study design might have led me to miss some cases in the KGBC, the identification of the same mutations in a second cohort, consisting entirely of children with SCD, indicates that these are probably the most clinically relevant mutations. Additionally, through the study I was able to provide a minimum estimate of the true prevalence of β-thalassaemia within my study population.

In conclusion, in this chapter, I set out to determine the prevalence, spectrum of mutations and clinical significance of β-thalassaemia on the Coast of Kenya. I found that four mutations were associated with β-thalassaemia overall that together exist at a minimum allele frequency of 0.3%. These mutations are of clinical importance as they contribute to 10% of SCD cases with one of the mutations rs33941849 also being identified as the cause of β-thalassaemia major in a misdiagnosed patient visiting the SCD clinic.

Chapter 4: The origin of β-thalassaemia mutations in Kilifi

4.1 Abstract

Introduction

In Chapter 3, I identified four mutations defining β-thalassaemia in Kilifi, with rs33959855 being the most prevalent. This chapter explores the global distribution of these mutations and their associated haplotypes, providing information about their origin and evolutionary history.

Methods

I first conducted a literature and genomic database search for these mutations to determine whether and where else these mutations have been reported in the global datasets. Secondly, I looked for evidence of ethnic clustering of the mutations in members of the KGBC from the following ethnolinguistic groups: Chonyi, Giriama, Kauma, Kambe, Jibana and Others (non-Mijikenda). Lastly, I selected SNPs that were common in both KGBC and 1000 genomes, a public genome database for estimating continent specific allele frequencies. Using these SNPs, I generated haplotypes and compared them between β-thalassaemia participants, non-β-thalassaemia individuals and the 1000 Genomes superpopulations.

Results

With the exception of rs35004220, the remaining β-thalassaemia mutations identified were not reported in any of the global genomic datasets studied (1000 genomes, GGVP, gnomAD, NCBI ALFA, TOPMED, UK10K, GEM-J, NHLBI Exome sequencing project) and were also very rarely reported in studies identified through my literature search. For example, to the best of my knowledge, the rs33959855 (G>T) has only been reported once previously, in a case from Reunion Island.

Among groups in Kilifi County, ethnolinguistic analysis revealed that rs33959855 was at higher frequency among the Chonyi [0.382%; 95% confidence interval (CI) 0.288-0.508] than other groups (e.g. 0.029% (95% CI 0.012-0.076, P=0.001) in the Giriama), whereas rs33941849 was

at higher frequency in the Kambe (1.47%; 0.45-5.17) and Kauma (0.27%; 0.14-0.51) ethnolinguistic groups with $P<0.05$ observed when comparing both to Chonyi (0.008%; 0.002-0.045) and Giriama (0.0597%; 0.031-0.012).

Analysis of haplotypes in non-β-thalassaemia individuals, using an overlapping set of 8 SNPs, revealed 4 common haplotype clusters shared in both KGBC and 1000 Genomes. Of these, one cluster ("Hap 4") was associated with the β^s mutation and at highest frequency in African populations (~10% of haplotypes) while a second haplotype ("Hap 3") was at highest frequency in South Asian populations (~20% of haplotypes). Meanwhile, chromosomes from individuals with β-thalassaemia mutations all clustered with Hap 1 haplotype except for chromosomes carrying rs33959855 for which Hap 3 was most common, accounting for ~90% of the observed haplotypes.

Conclusion

Based on my studies, I hypothesise that rs33959855 is a recent mutation introduced to the Chonyi ethnolinguistic through interactions with a population carrying the Hap 3 haplotype. One potential route would have been during the historical interaction with the South Asian population that occurred during the Asia-Swahili trade that took place along the East African Coast in 1200CE. I further hypothesise that the remaining three mutations were either introduced through gene flow or emerged locally on the common Hap 1 haplotype.

4.2 Introduction

In the previous chapter, I set out to investigate the burden of β-thalassaemia in Kilifi, a condition that was not previously described in this population. I found four mutations that were associated with β-thalassaemia which, together, occur on approximately 0.3% of all haplotypes within this population. The discovery of these four mutations was interesting and made me curious about their origins and global distribution. There are several hypotheses that could explain the origin of these mutations in Kilifi. One is that they were introduced to these communities from elsewhere, either through gene flow or population migration. A second hypothesis is that the

mutations arose *de novo* on the chromosomal background of this population. Whatever the explanation, the mutations continue to exist in these populations, possibly due to balancing selection having been driven through a survival advantage against *P. falciparum* at the expense of mortality of homozygotes, which I investigate further in Chapters 5 and 6.

In this chapter, I have conducted population genetic analyses to determine the distribution of these β-thalassaemia mutations and to investigate their possible origins. I did this by first performing a literature and database search to determine the frequencies of these mutations on a global scale. I then investigated their origins by looking for clustering by ethnicity, which if present might indicate that a mutation had arisen recently enough in Kilifi that it had not fully spread by admixture. Lastly, I looked for evidence of haplotype sharing within the mutations and compared these haplotypes to those of participants without β-thalassaemia in Kilifi and in other populations outside of Kenya from the 1000 Genomes Project. I present the investigation and outcome of this work in three sections as follows.

1. A literature review and database search of the β-thalassaemia mutations identified in Kilifi cohorts.
2. Evidence for ethnic clustering of β-thalassaemia mutations identified in Kilifi cohorts.
3. Haplotype sharing between β-thalassaemia chromosomes, non-β-thalassaemia chromosomes, and chromosomes from 1000 genomes populations.

4.3 A literature review and database search of β-thalassaemia mutations identified in Kilifi cohorts.

4.3.1 Methods

4.3.1.1 Literature review

I first undertook a broad literature search in PubMed and African Journals Online (AJOL) to determine the number of studies on β-thalassaemia by different continents. I did this using the following search terms in PubMed; Africa AND " beta-Thalassemia "[Mesh]. "Africa" was then

replaced with either "Asia", "Americas", or "Europe" in the search term. To determine the number of studies conducted on β-thalassaemia in sub-Saharan Africa I performed a search using the search terms "Africa South of the Sahara"[Mesh] AND " beta-Thalassemia "[Mesh]. A similar search was done in AJOL using the term "β-thalassaemia" with either North African countries by name (Algeria, Egypt, Tunisia, Libya, Morocco, and Sudan) or the sub-Saharan countries (African countries not in the list of North African countries). As a comparator I did the same search for α-thalassaemia.

To determine the prevalence of each of the four mutations in other populations, I performed a literature search for each mutation using the rs ID number, HGVS and the haematological coding in PubMed and AJOL. An example of this is with the β-thalassaemia mutation rs35004220 whereby I searched for rs35004220, NM_000518.5:c.93−21G>A and IVS-I-110 (G>A) separately. I then retrieved the publications for each allele and recorded the origin of the population and the prevalence of the mutation.

4.3.1.2 Genomic database search

To further establish the distribution of these mutations, I performed an online search of the four β-thalassaemia mutations in various genomic databases. I did this using the Ensembl genome browser which performs a search across several genome databases including:

- 1000 Genomes: dataset that contains 88 million variants from 2504 healthy individuals of different ancestries.
- Genome aggregation database (gnomAD): A resource developed by a coalition of international investigators with the aim of aggregating and harmonizing exome and genome sequencing data (https://gnomad.broadinstitute.org/). The v2.1.1 dataset contains 125,748 exome seuences and 15,508 genome sequences from individuals of diverse ancestries.

- National Library of Medicine Allele Frequency Aggregator dataset (ALFA): Contains frequencies of more than 447 million variants from individuals of different ancestries (Table 11) https://ncbiinsights.ncbi.nlm.nih.gov/2020/03/26/alfa/ .
- Trans-Omics for Precision Medicine (TOPMed): Contains more than 700 million variants from individuals suffering from various diseases and from different ancestries (Table 11) https://topmed.nhlbi.nih.gov/ .
- UK10K: Variants from whole genome and Exome sequence individuals (4000 healthy and 6000 with health problems) of different ancestries based in the United Kingdom.
- National Heart Lung and Blood Institute (NHLBI) Exome sequencing project: consists of more than 1.8 millions of variants identified in target populations with problems related to heart, lungs and blood https://esp.gs.washington.edu/drupal/ (Auer et al., 2016).
- Gambian Genome Variation Project: Whole genome sequences from healthy individuals from Gambia https://www.internationalgenome.org/gambian-genome-variation-project/ .

Details of the populations present in these datasets and their ancestries are highlighted in Table. 11.

Table 11. List of projects and populations that are queried when conducting a search using Ensembl genome browser.

Project	Populations	Ancestry (N)
1000 Genomes	2504	African (661) American (347) East Asian (504) European (503) South Asian (489)
GEM-J	7609	Japanese (7609)
NCBI ALFA (release 2 Version: 20201027095038)	192,710	European (163,190) African American (5989) African Others (211) Africans (6200) South Asian (2619) East Asian (2515) Other Asian (1000) Asian (3515) Latin American 1(817) Latin American 2 (4703)
NHLBI Exome Sequencing Project (ESP)‡	200,000	United states populations.
TOPMed‡	132,345	United states populations.
UK10K‡	10,000	Population in United Kingdom
gnomAD genomes v3.1.2	125,748	African/African American (20744) Amish (456) Latino/Admixed Americans (7647) Ashkenazi Jewish (1736) East Asian (2604) European Finnish (5316) European non-Finnish (34029) Middle Eastern (158) South Asian (2419) Other‡ (1047)
Gambian Genome Variation Project	580	Gambians (580)

‡ Some of the population datasets did not have a classification available on the ancestry of the populations.

4.3.2 Results

4.3.2.1 Results from literature search

A literature search of the word β-thalassaemia or α-thalassaemia and the continent in which the study was conducted revealed that most of the studies were conducted in Asia (Table 12). Only 35 studies on β-thalassaemia were conducted in sub-Saharan Africa while four times that number were conducted in North Africa. In contrast there were four times as many articles mentioning the term α-thalassaemia and any of the sub-Saharan African countries when compared to North African countries.

Table 12. Number of β-thalassaemia studies classified by the continent mentioned in the article.

	β-thalassaemia		**α-thalassaemia**	
Continent	**PubMed**	**AJOL**	**PubMed**	**AJOL**
Asia	1778	NA	856	NA
Europe	668	NA	201	NA
Americas	224	NA	153	NA
Africa	221	60	134	59
North Africa	139	29	18	6
Sub-Saharan Africa	35	13	88	24

The data show the total number of research articles containing the term β-thalassaemia or α-thalassaemia classified by the region mentioned in the article. The search was done in PubMed and AJOL.

A literature search for the four mutations identified in the studies described in Chapter 3 revealed rs35004220 as the most common mutation in regions outside of Kenya. I identified a total of 146 publications that mentioned this mutation, reducing to 88 after dropping review papers, and articles that did not mention the origin of the population. I downloaded these publications and extracted data on the ancestry of the study participants and the prevalence of rs35004220 allele calculated as a proportion of total β-thalassaemia mutations identified in the population.

The highest prevalences of rs35004220 were observed in Western Asia, Southern Europe and North Africa. The country with the highest prevalence was Turkey in which 10 studies observed

that rs35004220 accounted for greater than 30% of β-thalassaemia mutations identified among study participants. The frequency of β-thalassaemia in Turkey has been estimated at 2% overall, with the highest frequency at 10% reported in Thrace region of Northwest Turkey (Altay, 2002). The geographic distribution of β-thalassaemia in Turkey differs by region, with the highest frequencies being reported in western and central Turkey and lowest in southern and eastern Turkey (Altay, 2002; Ozkinay et al., 2015; Tadmouri et al., 2001) (Table 12). In the haplotype analysis done by Tadmouri and colleagues (Tadmouri et al., 2001), the authors observed that 87% of chromosomes harbouring the rs35004220 mutation occurred on a sequence of a single haplotype of Anatolian origin (present day Turkey) that dated back to the era when agriculture was first introduced in this region (6500 to 2000 B.C) (Tadmouri et al., 2001). It has been reported that during this era, there were malaria outbreaks that could have caused positive selection of β-thalassaemia mutations including the rs35004220 mutation (Angel, 1966; Tadmouri et al., 2001). In subsequent years, the region witnessed a number of civilisations including the Hatti and Hittite, Urartu, Phyria, Persian, Byzantine, Seljuk and the Ottomans that led to the spread of this mutation to other regions outside of present day Turkey (Tadmouri et al., 2001). High prevalence rates of this mutation were further sustained by the high percentage of consanguineous marriages in this region (Altay, 2002).

In Africa, reports on the prevalence of β-thalassaemia are mainly from North African countries where frequencies of β-thalassaemia trait range from 1.5 to 4% in Morocco, Algeria, Tunisia and Egypt (see Chapter 1 Table 2). Carrier frequencies as high as 9% have also been reported in some ethnic groups within Liberia, (M. C. Willcox, 1975). Apart from Morocco, rs35004220 is the second most common mutation in North African countries where it accounts for between 20% to 30% of β-thalassaemia mutations (Table 13). In Morocco the prevalence is much lower ranging from 3 to 4% (Belmokhtar et al., 2022; Lemsaddek et al., 2003), although a prevalence of 26% was reported in one study (Agouti et al., 2008) (Table 13).

Table 13. Prevalence of rs35004220 by region in studies that have identified the mutation.

Mutation	Ancestral origin	Country	Total sampled	β-thalassaemia positive	Proportion of β-thalassaemia chromosomes positive for rs35004220 (%)‡	References
rs35004220	Americas	Argentina	71	71	16	(Lazarte et al., 2014)
rs35004220	Americas	Brazil	27	24	17	(Carrocini, Venancio, Pessoa, Lobo, & Bonini-Domingos, 2017)
rs35004220	Americas	Brazil	35	35	7	(da Silveira et al., 2011)
rs35004220	Americas	Brazil	33	33	3	(Silva et al., 2016)
rs35004220	Americas	Mexico	13	1	5	(Economou et al., 1991)
rs35004220	Americas	Venezuela	127	127	6.6	(Bravo-Urquiola et al., 2012)
rs35004220	Central Europe	Czecholovakia	93	93	6	(Indrak et al., 1992)
rs35004220	South Europe	Brazil	35	35	19	(Fonseca et al., 1998)
rs35004220	North Africa	Algeria	105	105	26	(Boudrahem-Addour et al., 2009)
rs35004220	North Africa	Algeria	31	31	19	(Bouhass, Perrin, & Trabuchet, 1994)
rs35004220	North Africa	Algeria	32	32	19	(Perrin et al., 1998)
rs35004220	North Africa	Algeria	89	89	28	(Bennani et al., 1993)
rs35004220	North Africa	Angola	359	2	0.6	(Borges et al., 2019)
rs35004220	North Africa	Egypt	37	37	41	(Hussein et al., 1993)
rs35004220	North Africa	Egypt	94	94	57	(Jiffri, Bogari, Zidan, Teama, & Elhawary, 2010)
rs35004220	North Africa	Egypt	47	47	22	(Elmezayen et al., 2015)
rs35004220	North Africa	Morocco	80	80	26	(Agouti et al., 2008)
rs35004220	North Africa	Morocco	39	39	4	(Belmokhtar et al., 2022)
rs35004220	North Africa	Morocco	90	90	3	(Lemsaddek et al., 2003)
rs35004220	North Africa	Tunisia	2	2	NA	(Khelil et al., 2003)
rs35004220	North Africa	Tunisia	118	118	11	(Chouk et al., 2004)
rs35004220	Northern Europe	Ireland	23	23	15	(Knott et al., 2006)
rs35004220	South Asia	Pakistan	8716	67	1.4	(Hussain et al., 2017)
rs35004220	South Europe	Albanian	68	12	23	(Babameto-Laku, Mitre, Berisha, Mokini, & Roko, 2011)

Mutation	Ancestral origin	Country	Total sampled	β-thalassaemia positive	Proportion of β-thalassaemia chromosomes positive for rs35004220 (%)‡	References
rs35004220	South Europe	Albanian	201	4	1	(Baghernajad-Salehi et al., 2009)
rs35004220	South Europe	Bulgaria	64	64	24	(Petkov et al., 1990)
rs35004220	South Europe	Bulgaria	723	723	23	(Petkov & Efremov, 2007)
rs35004220	South Europe	Greece	1	1	NA	(Spritz et al., 1981)
rs35004220	South Europe	Greece	1179	1179	13	(Georgiou et al., 2003)
rs35004220	South Europe	Greece	3931	85	22	(Kalleas et al., 2012)
rs35004220	South Europe	Greece	174	174	43	(Kattamis et al., 1990)
rs35004220	South Europe	Greece	705	705	80	(Baysal et al., 1992)
rs35004220	South Europe	Italy	2	2	NA	(Ragusa et al., 2003)
rs35004220	South Europe	Italy	1	1	NA	(Cannata et al., 2019)
rs35004220	South Europe	Romania	100	100	49	(Cherry, Calo, Talmaci, Perrin, & Gavrila, 2016)
rs35004220	South Europe	Serbia and Montenegro	73	73	17	(Pavlovic et al., 2005)
rs35004220	South Europe	Sicily	547	547	24	(Schiliro et al., 1995)
rs35004220	South Europe	Spain	22713	175	12	(Lopez-Escribano et al., 2013)
rs35004220	South Europe	Turkey	106	106	35	(Fettah et al., 2013)
rs35004220	South Europe	Yugoslavia	180	180	37	(Dimovski et al., 1990)
rs35004220	Western Asia	Azerbaijan	1757	143	10	(Aliyeva et al., 2020)
rs35004220	Western Asia	Azerbaijan	49	49	20	(Curuk et al., 1992)
rs35004220	Western Asia	Iran	41	41	15	(Jalilian et al., 2017)
rs35004220	Western Asia	Iran	176	176	7	(Moradi et al., 2020)
rs35004220	Western Asia	Iran	995	995	14	(Nezhad, Nezhad, Choghakabodi, & Keikhaei, 2018)
rs35004220	Western Asia	Iran	50	50	2	(Sajadpour, Amini-Farsani, Motovali-Bashi, Yadollahi, & Khosravi-Farsani, 2020)
rs35004220	Western Asia	Iran	1635	24	1	(Akhavan-Niaki et al., 2011)
rs35004220	Western Asia	Iran	392	392	3	(Derakhshandeh-Peykar et al., 2007)

Mutation	Ancestral origin	Country	Total sampled	β-thalassaemia positive	Proportion of β-thalassaemia chromosomes positive for rs35004220 (%)‡	References
rs35004220	Western Asia	Iran	1241	1241	14	(Galehdari et al., 2010)
rs35004220	Western Asia	Iran	60	60	6	(Haghi, Khorshidi, Hosseinpour Feizi, Pouladi, & Hosseinpour Feizi, 2009)
rs35004220	Western Asia	Iran	164	164	18	(Hosseinpour Feizi, Hosseinpour Feizi, Pouladi, Haghi, & Azarfam, 2008)
rs35004220	Western Asia	Iran	65	65	12	(Kiani, Mortazavi, Zeinali, & Shirkhani, 2007)
rs35004220	Western Asia	Iran	2632	1346	7	(Moghadam et al., 2015)
rs35004220	Western Asia	Iran	1053	1053	5	(Najmabadi et al., 2001)
rs35004220	Western Asia	Iran	8135	416	21	(Yuzbasioglu Ariyurek, Yildiz, Yalin, Guzelgul, & Aksoy, 2016)
rs35004220	Western Asia	Iraq	159	159	5	(Amin et al., 2020)
rs35004220	Western Asia	Iraq	94	94	34	(Eissa, Kashmoola, Atroshi, & Al-Allawi, 2015)
rs35004220	Western Asia	Iraq	103	31	30	(Al-Allawi, Al-Mousawi, Badi, & Jalal, 2013)
rs35004220	Western Asia	Iraq	104	2	2	(Al-Allawi, Jubrael, & Hughson, 2006)
rs35004220	Western Asia	Israel	412	2	0.5	(Zlotogora, Hujerat, Barges, Shalev, & Chakravarti, 2007)
rs35004220	Western Asia	Jordan	49	49	21	(Sadiq & Huisman, 1994)
rs35004220	Western Asia	Lebanon	126	60	29	(Farra et al., 2021)
rs35004220	Western Asia	Lebanon	255	255	34	(Makhoul et al., 2005)
rs35004220	Western Asia	Palestine	51	51	7.8	(Faraon, Daraghmah, Samarah, & Srour, 2019)
rs35004220	Western Asia	Palestine	16	16	NA	(Samha, Sirdah, Reading, Karmi, & Agarwal, 2020)
rs35004220	Western Asia	Palestine	95	95	9.5	(El-Latif, Filon, Rund, Oppenheim, & Kanaan, 2002)
rs35004220	Western Asia	Palestine	225	225	34	(Sirdah et al., 2013)
rs35004220	Western Asia	Saudi Arabia	172	172	27	(Abuzenadah et al., 2011)

Mutation	Ancestral origin	Country	Total sampled	β-thalassaemia positive	Proportion of β-thalassaemia chromosomes positive for rs35004220 (%)‡	References
rs35004220	Western Asia	Saudi Arabia	93	93	27	(el-Hazmi et al., 1995)
rs35004220	Western Asia	Syria	19	19	10.4	(Gunes & Gozden, 2021)
rs35004220	Western Asia	Syria	49	49	NA	(Murad, Moassas, Ghoury, & Mukhalalaty, 2018)
rs35004220	Western Asia	Syria	189	189	17	(Jarjour, Murad, Moasses, & Al-Achkar, 2014)
rs35004220	Western Asia	Syria	82	82	24	(Kyriacou et al., 2000)
rs35004220	Western Asia	Syria	55	55	6	(Murad, Moassas, Jarjour, Mukhalalaty, & Al-Achkar, 2014)
rs35004220	Western Asia	Syria	730	730	22	(Murad et al., 2018)
rs35004220	Western Asia	Turkey	1	1	NA	(Westaway & Williamson, 1981)
rs35004220	Western Asia	Turkey	380	86	39	(Oner et al., 1990)
rs35004220	Western Asia	Turkey	89	89	31	(Ulasli et al., 2015)
rs35004220	Western Asia	Turkey	517	71	20	(Arpaci et al., 2021)
rs35004220	Western Asia	Turkey	45	45	29	(Gunes & Gozden, 2021)
rs35004220	Western Asia	Turkey	52	52	58	(Guzelgul, Seydel, & Aksoy, 2020)
rs35004220	Western Asia	Turkey	34	34	39	(Yilmaz, 2019)
rs35004220	Western Asia	Turkey	268	268	36	(Basak et al., 1992)
rs35004220	Western Asia	Turkey	197	197	39	(Bilgen, Arikan, Canatan, Yesilipek, & Keser, 2011)
rs35004220	Western Asia	Turkey	202	202	47	(Keser et al., 2004)
rs35004220	Western Asia	Turkey	420	201	42	(Mendilcioglu et al., 2011)
rs35004220	Western Asia	Turkey	1171	1171	42	(Ozkinay et al., 2015)
rs35004220	Western Europe	Iraq	123	123	20	(Jalal et al., 2010)

‡The proportion is calculated as percentage of the total β-thalassaemia mutations identified in the study.

Because few cases of rs33941849, rs193922563 and rs33959855 have been identified previously, I present the data from my literature search for these mutations as the absolute number of cases positive for the mutation (Table 14).

rs33941849 was the second most common mutation after rs35004220; however, unlike rs35004220 which was most frequently reported in Western Asia, Southern Europe and North Africa, the reported positive cases for rs33941849 were spread across seven regions including broadly across Europe. The ancestry of five African individuals identified with rs33941849 mutation included one African Cuban identified in the study by Muniz 2000 (Muniz, Martinez, Lavinha, & Pacheco, 2000), the other four being from Mayotte Island, which is mainly composed of individuals of East African origin (Muszlak et al., 2015). The six cases of rs193922563 were of Asian origin. A single case of rs33959855 was identified by Ghanem *et al* 1992 (Ghanem et al., 1992) in the Reunion Island but their ancestry was not defined (Table 14).

Table 14. Summary of studies that have identified rs33941849, rs193922563 and rs33959855 and the ancestry of the population identified in each study.

Mutation	Population region of origin	Country sampled	Total sampled	β-thalassaemia positive	Positive for mutation	References
rs33941849	African	Cuba	232	1	1	(Muniz et al., 2000)
rs33941849	African	Mayotte island	82	4	4	(Muszlak et al., 2015)
rs33941849	Central Europe	Switzerland	8	2	2	(Beris et al., 1993)
rs33941849	Central Europe	Germany	256	2	2	(Vetter, Schwarz, Kohne, & Kulozik, 1997)
rs33941849	East Asia	China	1	1	1	(Lam et al., 1990)
rs33941849	East Asia	Korea	195	20	1	(Park et al., 2013)
rs33941849	East Asia	China	1441	175	1	(Zhang et al., 2015)
rs33941849	Eastern Europe	Russia	1	1	1	(Molchanova et al., 1998)
rs33941849	South Asia	India	1545	2	2	(Edison et al., 2008)
rs33941849	South Asia	Pakistan	21	1	1	(Khan et al., 2000)
rs33941849	South Asia	India	3	3	3	(Gupta, Hattori, & Agarwal, 2002)
rs33941849	South Asia	India	257	10	8	(Gorakshakar et al., 2018)
rs33941849	South Europe	Yugoslavia	100	3	3	(Jankovic et al., 1990)
rs33941849	South Europe	Yugoslavia	70	1	1	(Najmabadi et al., 2002)
rs33941849	Western Europe	Belgium	9	3	3	(Wildmann et al., 1993)
rs193922563	South Asia	India	1	1	1	(Orkin et al., 1983)
rs193922563	Western Asia	Kuwait	41	17	3	(Adekile et al., 2015)
rs193922563	Western Asia	Iran	45	2	2	(Miri-Moghaddam, Bahrami, Naderi, Bazi, & Karimipoor, 2016)
rs33959855	Unknown	Reunion Island	36	1	1	(Ghanem et al., 1992)

4.3.2.2 Results from genome database search

I next inspected data regarding the four mutations from global panels of genetic variation accessed through the Ensembl Genome Browser. The search includes datasets from the following prominent projects; 1000 Genomes Project, gnomAD, NCBI ALFA, TOPMed, UK10K, NHLBI Exome sequencing project and the Gambian Genome Variation Project. The

mutations, rs33959855 and rs35004220, were the only mutations in which positive cases were identified (Table 15).

The mutations identified at the SNP position for rs33959855 were, however, different from those observed among Kilifi β-thalassaemia carriers. The G>C mutation is a missense variant that results in the substitution of glutamic acid for glutamine (NP_000509.1: p. Glu23Gln). The change does not affect the size of the polypeptide chain, but it does change the charge of the protein from negative to positive. The haemoglobin formed, referred to as Hb D-Iran, is benign (Patrinos et al., 2004; Thornburg, Zimmerman, Schultz, & Ware, 2001). Similarly, the second mutation, a change from G>A is also another missense variant that results from substitution of glutamic acid for lysine (NP_000509.1:p.Glu23Lys). The resulting haemoglobin referred to as HbE-Saskatoon is also considered benign (Birben et al 2001). However, these two mutations are different from the G>T mutation identified among our Kilifi β-thalassaemia carriers, that results in the creation of a stop codon (NP_000509.1:p.Glu23Ter). The individuals identified with the G>C mutation were of South Asian origin and were all identified in the gnomAD dataset (Table 15).

The majority of cases identified through my search with rs35004220 (~80%) were of European origin and were identified in the following datasets: gnomAD (n=39), NCBI ALFA (n=26), TOPMED (n=1), UK10K (n=1) and NHLBI Exome sequencing project (n=2).

No cases of rs33941849 and rs193922563 were reported, possibly due to limited number of participants that were genotyped for these mutations. For example, only 426 individuals of African American origin were genotyped for rs33941849 (NCBI ALFA dataset) and none were sequenced for rs193922563 deletion (Table 15).

Taken together my literature review and database search for these mutations suggests an underrepresentation of research into β-thalassaemia in African countries, with sub-Saharan Africa having the least number of studies. Among the β-thalassaemia mutations, only rs35004220 was found in the public genome databases, which posed a significant limitation in directly evaluating their haplotypes in section 4.5 below.

Table 15. Prevalence of mutations causing β-thalassaemia in Kilifi in other regions.

rsID	Pathogenicity	Mutant (Alleles)	AFR	CEU	SAS	EAS	AMR	Other
rs33959855								
	Total genotyped		18634	152278	30708	18590	37552	6598
	Normal	G	18634	152278	30693	18590	37552	6598
	β–thalassaemia	C	0	0	15	0	0	0
	β–thalassaemia	A	0	1	0	0	0	0
	β–thalassaemia	T	0	0	0	0	0	0
rs33941849								
	Total genotyped		426	0	0	0	0	52
	Normal	T	426	0	0	0	0	52
	β–thalassaemia	C	0	0	0	0	0	0
rs35004220								
	Total genotyped		23020	148693	30801	19734	56726	13763
	Normal	G	23020	148633	30801	19734	56722	13763
	β–thalassaemia	A		60	1		4	4
rs193922563								
	Total genotyped	All	0	0	0	0	0	0
	Normal	No_deletion	0	0	0	0	0	0
	β–thalassaemia	AGCCTAAGGGTGGGAAAATAGACCAA/A	0	0	0	0	0	0

The table shows the total alleles identified per SNP tested and grouped by superpopulations, as reported by the Ensembl Genome Browser. Underlying datasets include 1000 Genomes Project (N=2504), gnomAD (N=125,748), NCBI ALFA (N=192,710), TOPMed (N=132,345), UK10K (10,000), NHLBI Exome sequencing project (200,000), GEM-J (7609) and the Gambian Genome Variation Project (N=580). The superpopulations used in grouping the SNPs include Africans (AFR), Europeans (CEU), South Asians (SAS), East Asians (EAS), Americans (AMR) and unknown populations (Other). The rows shaded in grey represent the β-thalassaemia alleles identified in Kilifi. The data was extracted from Ensembl, accessed November 2022. Data on the specific studies and populations from which the data was collected is highlighted in Table 11.

4.4 Evidence of ethnic clustering of β-thalassaemia mutations in the KGBC

In this section, I look for evidence for a recent origin of these mutations by looking for clustering of the mutations within ethnolinguistic groups. This is based on the hypothesis that if a mutation was introduced or originated in a community recently, it is likely to be at its highest frequency within the same community as it would not have had enough time to spread to other communities through admixture. Alternatively, the mutations might have drifted to high frequencies within one ethnic group but not the others due to a selective force present in that group. In this section, therefore, I perform an analysis to determine if there exist any differences in allele frequencies by ethnicity.

4.4.1 Methods

4.4.1.1 Study populations

All 15,577 participants in the KGBC were included in this analysis. As outlined in Chapter 3, I sequenced all the individuals without the sickle mutation (HbAA) who had HbA_2 levels of >4%, a commonly used diagnostic cut-off for the diagnosis of β-thalassaemia carriers. Overall, I identified 83 β-thalassaemia heterozygotes who I included in the current analysis with an assumption that these were all the cases that existed within the cohort. Based on this, I make an important assumption that the rest of the population are free of β-thalassaemia mutations. The study population and the identification of β-thalassaemia carriers is described in detail in Chapters 2 (Section 2.2.1) and 3 (Section 3.2.1.1).

4.4.1.2 Statistical analysis

I calculated the allele frequencies for each SNP as the total number of β-thalassaemia alleles divided by the total number of chromosomes in the sampled population. To visually compare frequencies between groups and compute credible intervals, I used a binomial likelihood formulation with uniform prior, and computed credible intervals from the resulting posterior distribution, which is mathematically described by a β-distribution. I used Fisher's exact test to compare allele frequencies between ethnolinguistic groups. All statistical analyses and plotting of graphs were undertaken using R statistical software (version 4.0.2) using the following packages: ggplot2 v3.4.0, tidyverse v1.3.1 and ggsci v0.6.0.

4.4.2 Results

The four β-thalassaemia mutations showed different patterns of association with ethnic groups. For rs33959855 mutation, 47 out of the total of 55 individuals identified with this mutation were from the Chonyi ethnic group, with the remaining distributed as follows: Giriama (4), Duruma

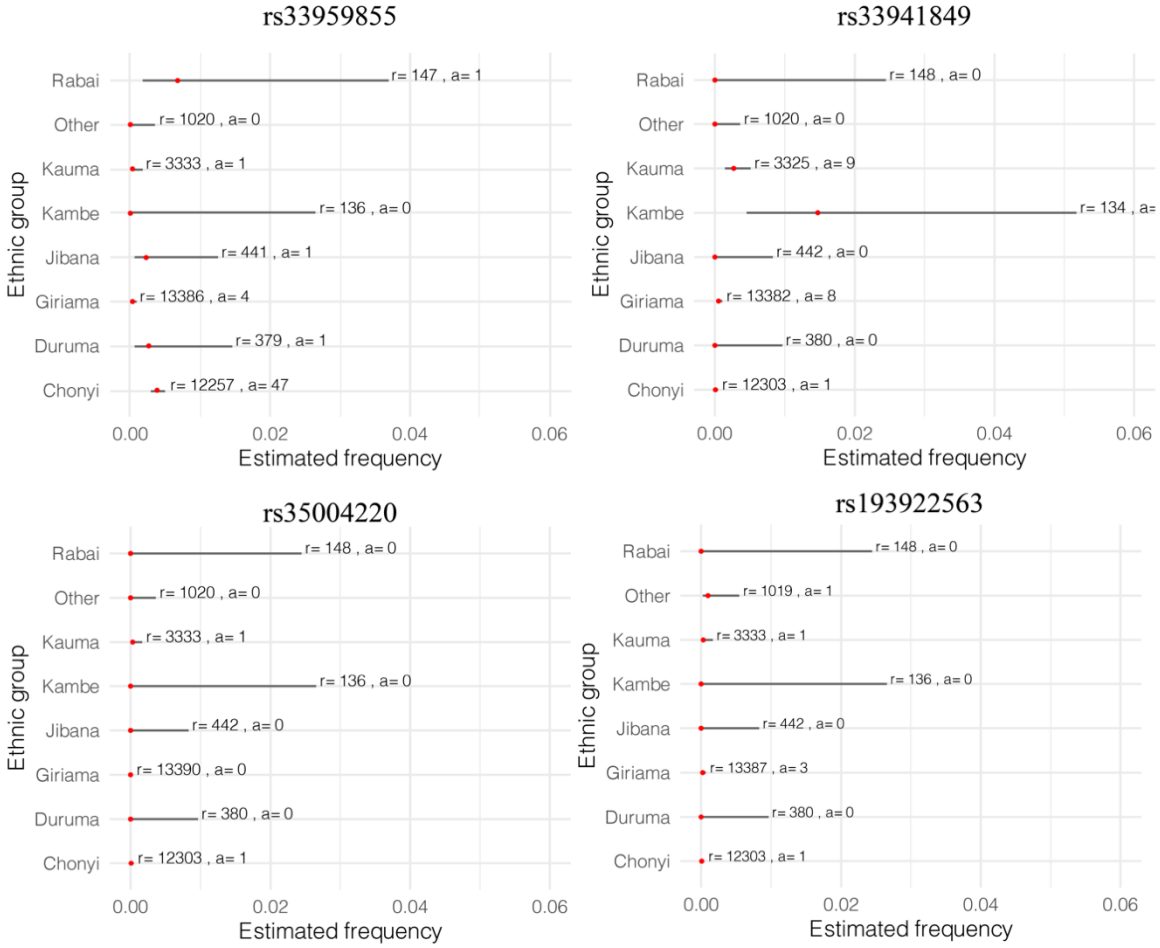
(1), Jibana (1), Kauma (1), Rabai (1). The allele frequency for rs33959855 among the Chonyi [0.382%; 95% CI 0.288-0.508] was higher compared to Giriama, Kauma and non-Mijikenda ethnicities, where the frequencies were especially low (Figure 20) (e.g. 0.029%; 0.012-0.076, P=0.001) in the Giriama). Similar differences in observed frequencies were observed between the Chonyi and the Kambe, Duruma, and Jibana, but not the Rabai, but the sample sizes of these populations were considerably smaller, such that *P*-values were greater than 0.05. Additionally, a difference in observed frequency was observed when comparing the prevalence of rs33959855 in the Giriama 0.029% (0.012-0.076) versus the Rabai 0.678% (0.162-3.68, P=0.053) (Figure 20).

The 20 cases of rs33941849 were distributed across 4 ethnicities as follows: Kauma (9/3,325; 0.27%), Giriama (8/13,382; 0.06%), Kambe (2/134; 1.5%) and Chonyi (1/12,303; 0.008%). When I performed a comparison of the allele frequencies by ethnicity, I found that rs33941849 was highest among Kambe (1.470%; 0.454-5.170) and Kauma (0.270%; 0.144-0.512) compared to other ethnicities with *P*-values of <0.05 observed when comparing frequencies between the Chonyi (0.008%; 0.002-0.045) and Giriama (0.0597%; 0.031-0.012) ethnolinguistic groups (Figure 20).

The 6 individuals identified with the rs193922563 mutation were spread between different communities as follows: Giriama (3), Kauma (1), Chonyi (1) and 1 from non-Mijikenda community. No association was observed between the mutation and any specific ethnic group (Figure 20).

The two cases of rs35004220 mutation identified in Kauma and Chonyi did not show any association with either of the ethnicities (Figure 20).

Figure 20. Estimated allele frequencies of β-thalassaemia mutations across ethnic groups.



Each figure shows the distributions of estimated allele frequencies (red dots) of the four mutations across the indigenous ethnic groups found in Kilifi referred to as Mijikenda. The ethnic group referred to as other are individuals from other tribes that are not Mijikenda (see Chapter 2 for details). The straight line striking through the red dots represents the 95% credible intervals calculated using posterior distribution assuming a flat prior probability. In each ethnic group, r represents the participants without β-thalassaemia mutations whereas a represents participants identified with β-thalassaemia mutations.

4.5 Haplotype sharing between β-thalassaemia chromosomes, non-β-thalassaemia chromosomes, and chromosomes from 1000 genome superpopulations.

A haplotype refers to a combination of alleles that occur close to each other on the same chromosome and tend to be inherited together. Similarity in haplotypes between individuals or populations indicates shared ancestry and can be used to show historical interactions between the two groups due to admixture.

In this section, I use statistical phasing approach to generate haplotypes from β-thalassaemia chromosomes and compare them to haplotypes from non-β-thalassaemic chromosomes from the KGBC and 1000 Genomes superpopulations. The objective of this analysis was to

determine whether there exists a shared haplotype among β-thalassaemia chromosomes which could inform on the origin of the mutations.

4.5.1 Methods

4.5.1.1 Study populations

The study populations included in this chapter were drawn from KGBC and 1000 Genomes. The study populations have been described in detail in Chapter 2, here I give a brief description on the specific populations selected for this chapter.

Kilifi Genetic Birth Cohort (KGBC)

My analysis included all the 730 participants from the KGBC whose samples were sequenced, out of which 83 participants were identified as heterozygous for β-thalassaemia pathogenic mutations (Chapter 3). In brief, the sequencing method covered a 1.8kb region that included the 5′ promoter, 5′ and 3′ untranslated regions, exons 1–3, and the intervening sequence I and II regions flanking exons 2 and 3. Further details of the sequencing approach can be found in Chapter 2 Section 2.3.4.

1000 Genomes

The 1000 Genomes dataset contains genetic variation data from 2504 healthy individuals from 26 superpopulations sampled in Africa (AFR), East Asia (EA), Europe (EUR), South Asia (SAS) and Americas (AMR). The dataset contains 88 million variants all phased into haplotypes. I downloaded the VCF for chromosome 11 from

http://ftp.ensembl.org/pub/data_files/homo_sapiens/GRCh37/variation_genotype/. All participants in this study were included in the analysis.

4.5.1.2 Phasing of the KGBC genotypes

Phasing is the process of assigning alleles to either of the parental chromosomes and can be done either using laboratory based experimental methods or computational methods. The KGBC data was in the form of unphased genotypes and required haplotype phasing to be performed which is a prerequisite for haplotype analysis. A total of 15 SNPs identified in the KGBC were phased using BEAGLE, a statistical phasing software. To investigate reliability of the phased genotypes, I generated 10 phased datasets by repeating the software algorithm with the following conditions: 25 iterations, different random seeds, and no imputation. I then examined the quality of the phased data through comparison of switch error rate between the first dataset and the rest. Switch errors occur when a heterozygous site is incorrectly phased when compared to predecessor SNPs. If phasing is accurate, then differences in heterozygous alleles (switch errors) between the 10 phased datasets should be minimal since the phased datasets were generated from the same data. QCTOOL (https://www.well.ox.ac.uk/~gav/qctool_v2/) was used to compare the phased datasets and determine the switch errors.

I calculated the average switch error rate per individual for each of the nine phasings compared to the first phased dataset. This was done by dividing the total number of switches by the total number of heterozygous sites per participant to obtain the switch error rate. The average switch error rate was then calculated by averaging over samples and informative sites.

4.5.1.3 Haplotype analysis

Selection of SNPs for haplotype analysis

To investigate haplotype structures in phased genotypes, I first selected SNPs from the sequenced region in KGBC samples. I included all SNPs that were common in both the KGBC and 1000 Genomes. Having a similar set of SNPs defining haplotypes in both datasets allowed for comparison to be done between the populations.

Haplotype construction

I first used phased chromosomes from all the participants in 1000 Genomes (N=2,504) to examine how well the 1.8kb sequenced region in the KGBC samples captures haplotypes structures present in the β-globin region. I did this by selecting SNPs within a MAF of >5% and within a region of 8kb (selected so as to capture a large number of SNPs that can clearly define haplotypes in this region), a region flanking the 1.8kb region sequenced in the KGBC samples. I then grouped the haplotypes based on similarity by calculating the distance metrics using Manhattan distance followed by clustering using hierarchical clustering. I visualised the resulting patterns using a heat map in which the ancestral alleles were represented by the black spots and the derived alleles by the white spots. I identified the haplotype patterns visually by looking for patterns of homogeneity on the heatmap and labelled them accordingly. I repeated the same analysis on the same dataset for the 1.8 kb region using the SNPs that were common in both the 1000 Genomes and the KGBC datasets. I then compared haplotype structures defined using the 8kb region and the 1.8kb region to determine how well my sequenced region captured the major haplotypes.

Next, I investigated haplotypes in the KGBC participants without β-thalassaemia mutations (normal chromosomes) by looking at haplotype structures within the 1.8kb region defined using SNPs that were common in both the 1000 Genomes and the KGBC datasets. The haplotypes were clustered, sorted and visualised using Manhattan distance and hierarchical clustering as described previously in 1000 Genome analysis. The haplotype structures were then compared with those from the 1.8kb region in 1000 Genomes after which common haplotypes were identified and labelled accordingly.

Finally, I performed haplotype analysis on phased genotypes harbouring the β-thalassaemia mutations (β-thalassaemia chromosomes). This was performed on the 1.8kb region using the same approach as that of normal chromosomes described previously. I then looked for evidence of haplotype sharing between chromosomes from 1000 Genomes, normal chromosomes, and β-thalassaemia chromosomes.

4.5.1.4 Statistical analysis

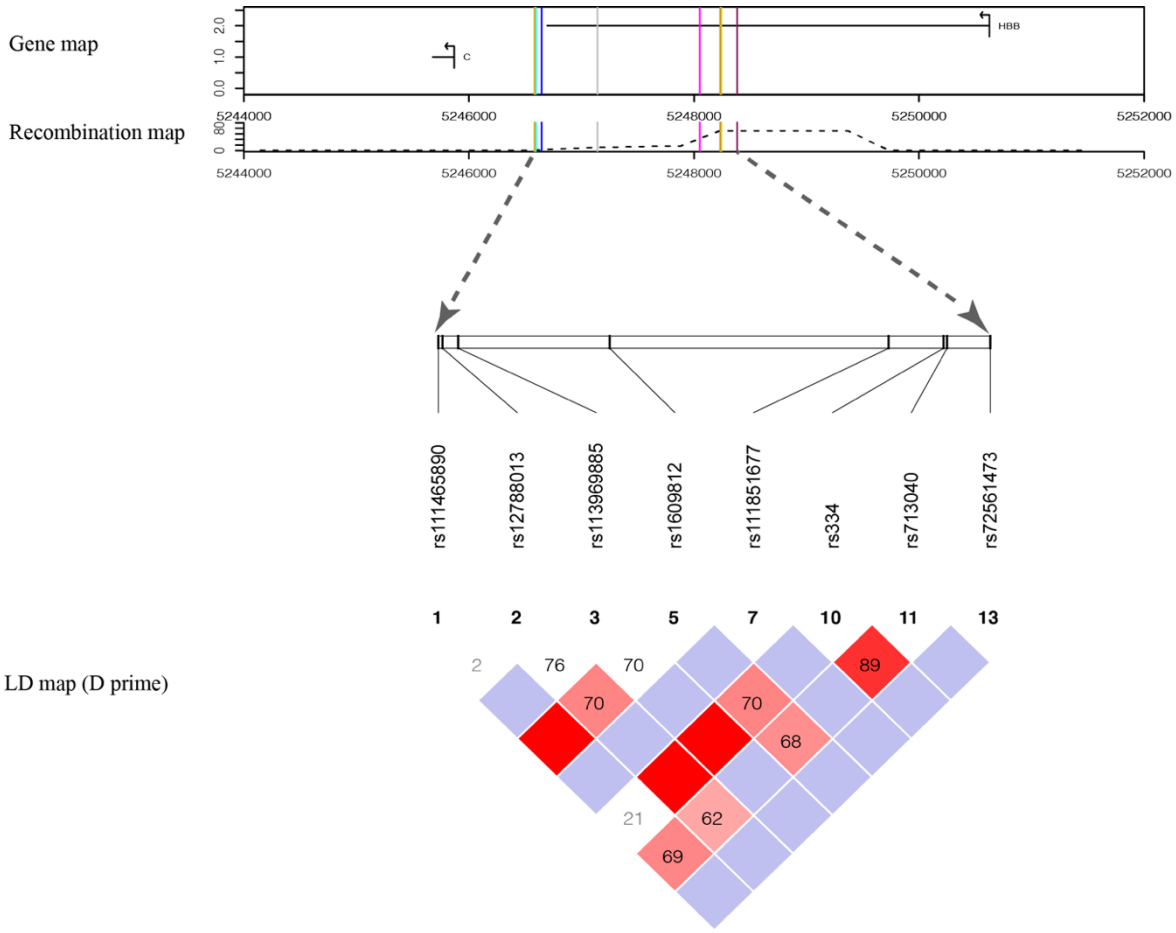
R statistical software (version 4.0.2) was used in plotting graphs and tables on haplotype distributions. Haplotype analysis was done using hierarchal clustering using hclust() function in R statistical software (version 4.0.2). hclust analysis requires phasing of chromosomes to be performed to generate two haplotypes per person before analysis. In the KGBC, phasing was performed using BEAGLE as described in Section 4.5.1.2. The 1000 Genomes dataset was available in a pre-phased format.

4.5.2 Results

4.5.3 Selection of SNPs for haplotype analysis

A total of fifteen SNPs including the β-thalassaemia mutations were identified in the sequenced region of 1.8kb in KGBC samples. Out of the fifteen SNPs, eight were common in both KGBC and 1000 Genomes. The location of the eight SNPs and recombination map in normal chromosomes from KGBC are shown in Figure 21. Out of the fifteen SNPs, three were dropped because they were at very low minor allele frequency and absent from 1000 Genomes. The β-thalassaemia mutations identified in KGBC were not in 1000 Genomes, making the haplotype information mainly reliant on these eight common SNPs.

Figure 21. Gene, recombination, and LD maps showing the chromosomal location, recombination, and linkage disequilibrium of eight SNPs spanning a region of 1.8kb, that were used in constructing haplotypes.



Each coloured vertical line on the gene and recombination map represents location of one of the 8 SNPs that were used in constructing the haplotypes. On the recombination map there is a large recombination hotspot of 70cM/Mb shown by the dotted line. Two SNPs rs713040 and rs72561743 are located within this hotspot. The LD map shows the linkage disequilibrium as D′ in normal chromosomes from the KGBC. The standard colour scheme of Haploview was used to display LD (logarithm of likelihood odds ratio [LOD] (a measure of confidence of D') score ≥2 and D'=1 is shown in bright red; LOD score <2 and D'=1 is shown in blue, LOD score ≥2 and D' <1 is shown in pink and LOD <2 and D' <1 is shown in white.

4.5.3.1 Quality of phased genotypes

Because the true haplotype phase of genetic variants in KGBC is unknown, I used computational method to infer the most likely haplotypes in each participant. To assess the accuracy of the phasing logarithm, I created 10 phased datasets with random starting points and evaluated the differences between them by looking at the switch error rates. Switch error occurs when two adjacent genetic markers on the same chromosome are incorrectly assigned to the opposite haplotypes. I calculated the switch error rates for KGBC as the average number of

differences in haplotype phase observed between each dataset and the first phased dataset, averaged over samples and informative sites.

The switch error rates comparing the first dataset and the remaining nine phasings were consistently low and exhibited little variation. In order of the comparator datasets 2 to 9, the average switch error rates were as follows: 2.6%, 2.2%, 1.7%, 1.9%, 2.3%, 2.7%, 2.0%, 2.3%, 2.2%. To illustrate these values, a switch error rate of 2.6% here means that approximately 2.6% of heterozygous sites were phased in the opposite sense in the two datasets, averaged across samples. Since we do not know the true haplotype phase, this number is based on two runs of the phasing algorithm both of which may have erroneous phase and is thus likely to overestimate true switch error rates.

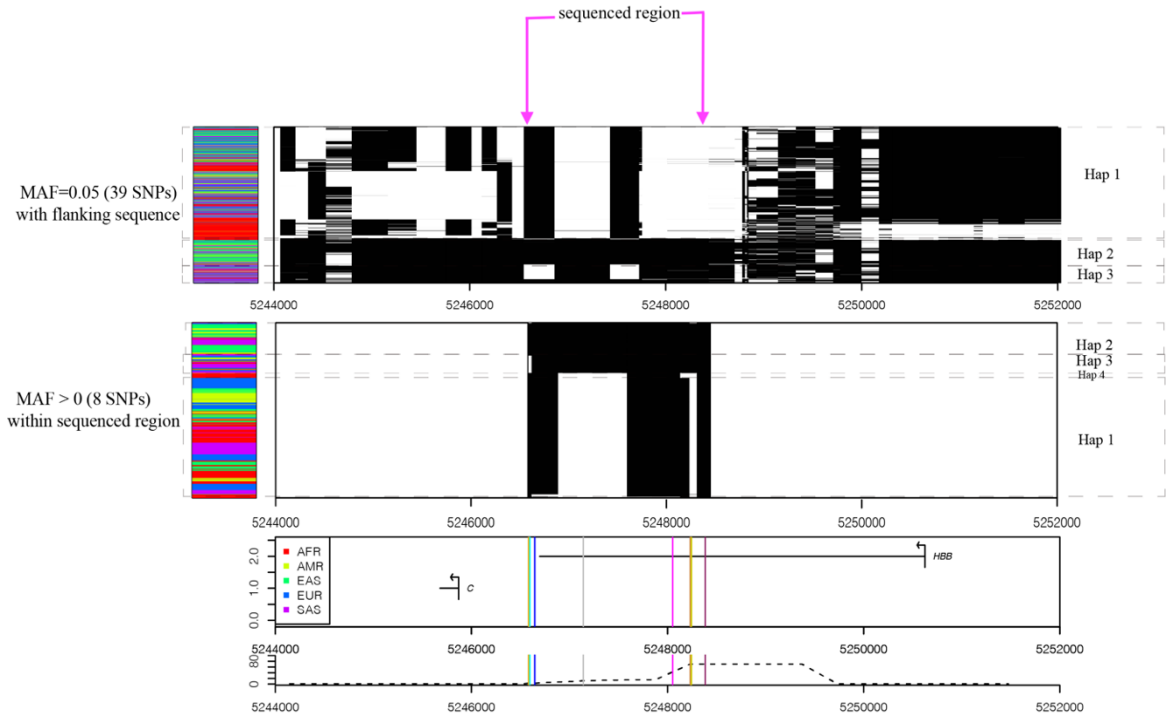
To provide a point of comparison, a study conducted by Choi and colleagues et al (Choi, Chan, Kirkness, Telenti, & Schork, 2018) utilised BEAGLE software and 1000 genomes dataset as a reference for phasing the publicly available NA12878 genome. By comparing to a high-quality phased version of the NA12878 genome previously constructed, they estimated a switch error rate of 1.53% for a haplotype block of 0.4 to 0.5 Mb.

4.5.3.2 Haplotypes defined using SNPs from the 1.8kb sequenced region capture same major haplotypes as those from a large region of 8kb in the 1000 Genome superpopulations.

I next performed an analysis to determine whether the 1.8kb region that I sequenced in my study was able to capture information on major haplotype structures found in this region. I did this by comparing haplotypes defined using a larger region of 8kb consisting of 39 SNPs at a MAF of >5% versus the 1.8kb region defined by eight selected SNPs in the same set of samples from 1000 Genomes.

In Figure 23, I show haplotypes in the β-globin locus defined using an 8kb region consisting of 39 SNPs and a 1.8kb region consisting of eight SNPs which does not include the β-thalassaemia mutations as discussed in section 4.5.3. I observed four important features on the haplotype structures. Firstly, in both 8kb and 1.8kb analyses there were three main haplotype clusters, which I have labelled as Hap 1, Hap 2 and Hap 3 in Figure 22. Hap 1 was the most common followed by Hap 2 and Hap 3, respectively. The second observation was that Hap 2 and Hap 3 were more common in EAS and SAS respectively and less common in AFR, EUR or AMR superpopulations as seen on the coloured line graphs representing the different populations. The third observation was that minor haplotype Hap 4 (labelled on the eight SNP haplotype map) was more common in African populations. The fourth observation was that the 1.8kb region was able to identify the same major haplotypes as the 8kb region. In the subsequent analysis I focus on the 1.8kb region analysed using the selected eight SNPs mainly because this was the region that was sequenced in my study.

Figure 22. Haplotypes structures within the β-globin region defined using 8kb region consisting of 39 SNPs and 1.8kb region consisting of eight SNPs.



The chromosomes are represented on the Y axis with the upper panel showing the haplotypes constructed using 39 SNPs from the 8kb region and the middle panel showing haplotypes constructed using eight SNPs from the 1.8kb region. The haplotypes are represented by the black and white lines; the black lines representing the reference allele and the white lines the minor allele. The major haplotypes are shown using the dotted triangles and labelled Hap 1, Hap 2 and Hap 3. A minor haplotype Hap 4 is also labelled in the middle panel. The coloured lines to the left side of the haplotypes represents the origin of the population from which the samples were drawn. The graphs at the bottom show the recombination map and the gene maps for this region with the horizontal-coloured lines representing the positions of the 8 SNPs that were used in defining haplotypes for the 1.8kb region. The effect of the recombination hotspot between positions 5248000 to 525000 on the haplotypes structure is seen on the right side of the 8kb region.

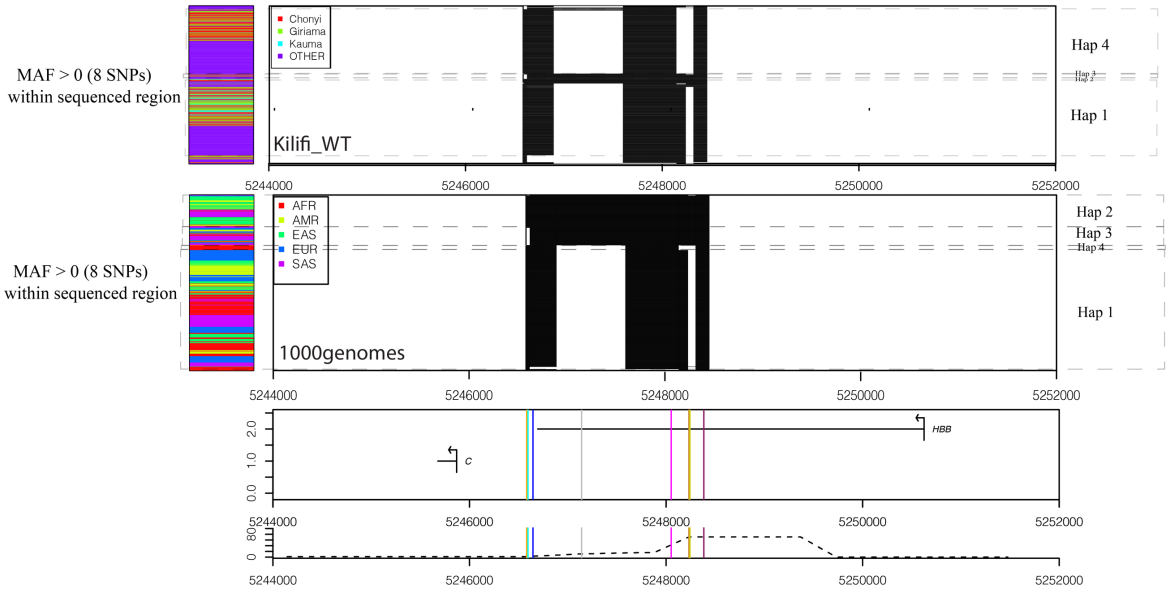
4.5.3.3 Haplotype sharing in normal chromosomes from the KGBC compared to 1000 Genomes superpopulations.

This analysis was performed to examine haplotype sharing between normal chromosomes from KGBC and 1000 Genomes. Two main haplotypes (Hap 1 and Hap 4) and two minor haplotypes Hap 2 and Hap 3 were observed in chromosomes from the KGBC (Figure 23). The most common haplotype in both the KGBC and 1000 Genomes datasets was Hap 1. Unlike 1000 Genomes, Hap 3 and 2 represented a smaller fraction of the total haplotypes observed in the KGBC.

Interestingly, Hap 4, a minor haplotype in 1000 genomes, was the second most dominant haplotype in KGBC and was observed in ~30% of the chromosomes. Further investigation of

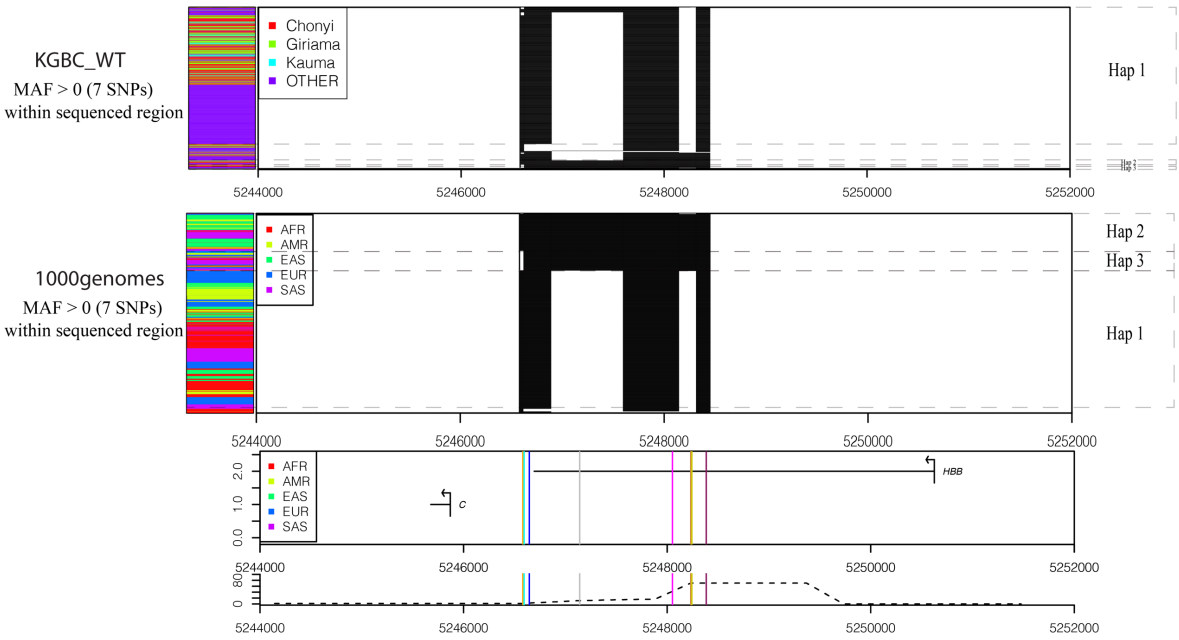
Hap 4 revealed that it was a haplotype associated rs334 which encodes for sickle haemoglobin. This explained its exclusive presence in individuals of African origin in 1000 genomes dataset. Figure 24 shows the haplotypes generated after exclusion of rs334. Absence of Hap 4 in both 1000 genomes and KGBC after exclusion of r334 confirms that it was a sickle cell-associated haplotype.

Figure 23. Haplotype structures in normal chromosomes from the KGBC compared to chromosomes from the 1000 Genomes superpopulations.



The top panel represents haplotypes from normal chromosomes from the KGBC, labelled Kilifi_WT, and the middle panel shows chromosomes from the 1000 Genomes superpopulations. The coloured line graphs on the left of the haplotype image represents the population/superpopulation from which the chromosomes were obtained. The major haplotypes are labelled on the left side of the haplotype image (Hap 1-4). The graphs at the bottom show the recombination and the gene map for this region with the horizontal-coloured lines representing the positions of the eight SNPs that were used in defining haplotypes.

Figure 24. Haplotype structures in normal chromosomes from the KGBC compared to chromosomes from the 1000 Genomes superpopulations with exclusion of the sickle SNP (rs334).



Haplotypes generated using seven SNPs after exclusion of rs334. The top panel represents haplotypes from normal chromosomes from KGBC labelled Kilifi_WT and the middle panel shows chromosomes from 1000 Genomes superpopulations. The coloured line graphs on the left of the haplotype image represents the population/superpopulation from which the chromosomes were obtained. The haplotypes are labelled on the left side of the haplotype image (Hap 1-4).

4.5.3.4 Haplotype structures in β-thalassaemia chromosomes from KGBC compared to normal chromosomes from KGBC and 1000 Genomes superpopulations.

In this analysis I examine the haplotypes present in β-thalassaemia chromosomes and look for evidence of haplotype sharing with non β-thalassaemia encoding chromosomes from the KGBC and the 1000 Genomes. Whereas the ideal analysis would have been to compare haplotypes containing β-thalassaemia mutations in the KGBC to those containing the same mutations in the 1000 Genomes dataset, this comparison was not feasible because β-thalassaemia mutations identified in KGBC study were not genotyped in 1000 genomes dataset (see section 4.3.5 for SNPs included in this analysis).

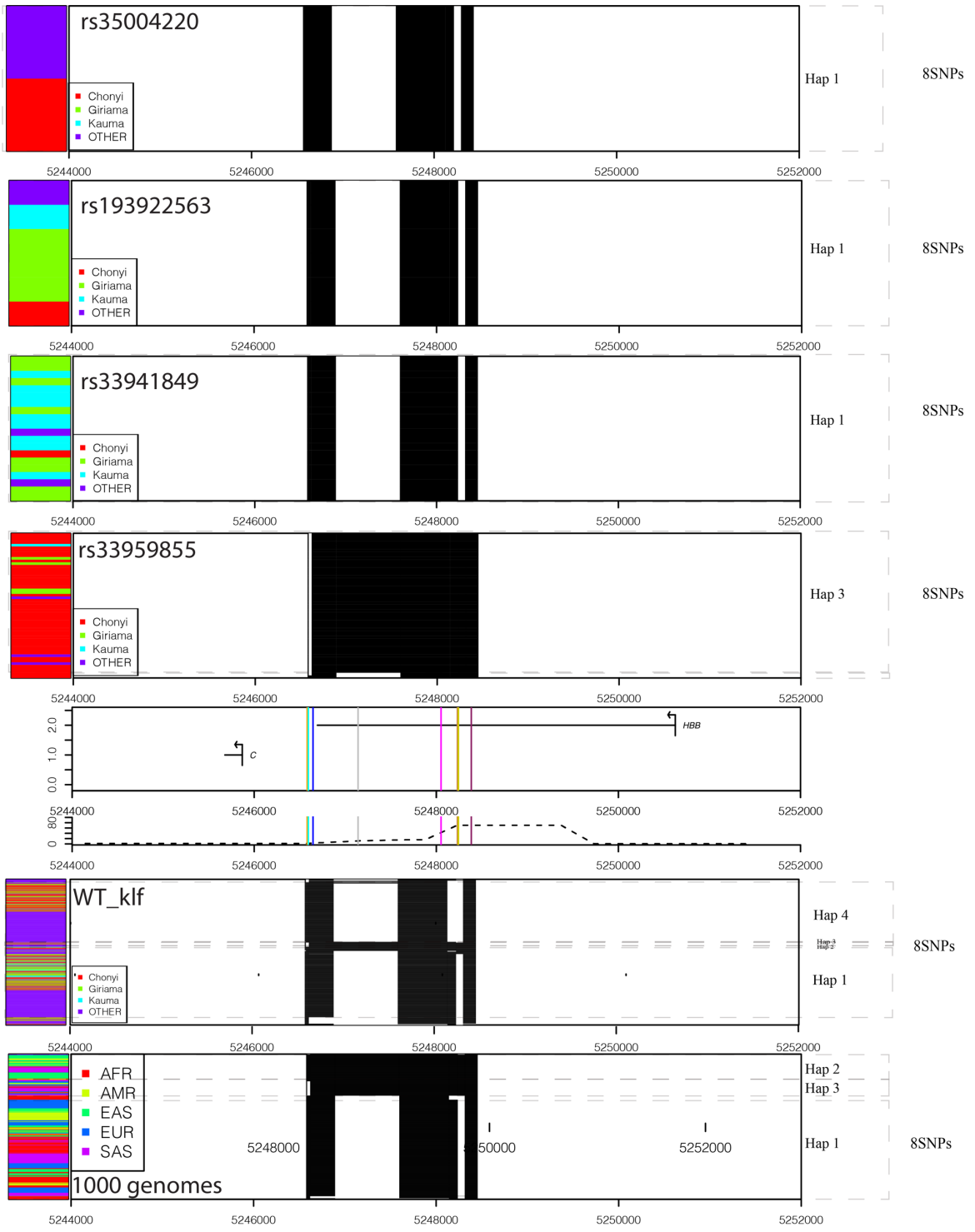
The analysis revealed that apart from chromosomes harbouring the rs33959855 mutations, Hap 1 was the most common haplotype in β-thalassaemia chromosomes accounting for 100% of the haplotypes in chromosomes containing rs33941849, rs35004220 and rs193922563 mutations (Figure 25). Conversely, Hap 3 was the most dominant haplotype in chromosomes containing

rs33959855 mutation accounting for ~90% of the haplotypes observed, with the remaining proportion being mainly from minor haplotypes. This may be noteworthy, because Hap 3 accounts for less than 5% of non-β-thalassaemia haplotypes in KGBC and suggests an origin of rs33959855 on the rare haplotype background Hap 3.

In the 1000 Genomes dataset, the prevalence of Hap 3 was highest in the SAS superpopulation, where it accounts for ~25% of the haplotypes observed, followed by the EUR population at ~15%. It was lowest in EAS and AFR populations where it accounted for ~5% of the haplotypes. This observation was clearer when the 1000 Genomes superpopulations were analysed separately (Figure 26). However, despite Hap 3 haplotypes being dominant in chromosomes harbouring the rs33959855 mutation in the KGBC, the mutation was not observed in any of the Hap 3 haplotypes from the 1000 Genomes superpopulations.

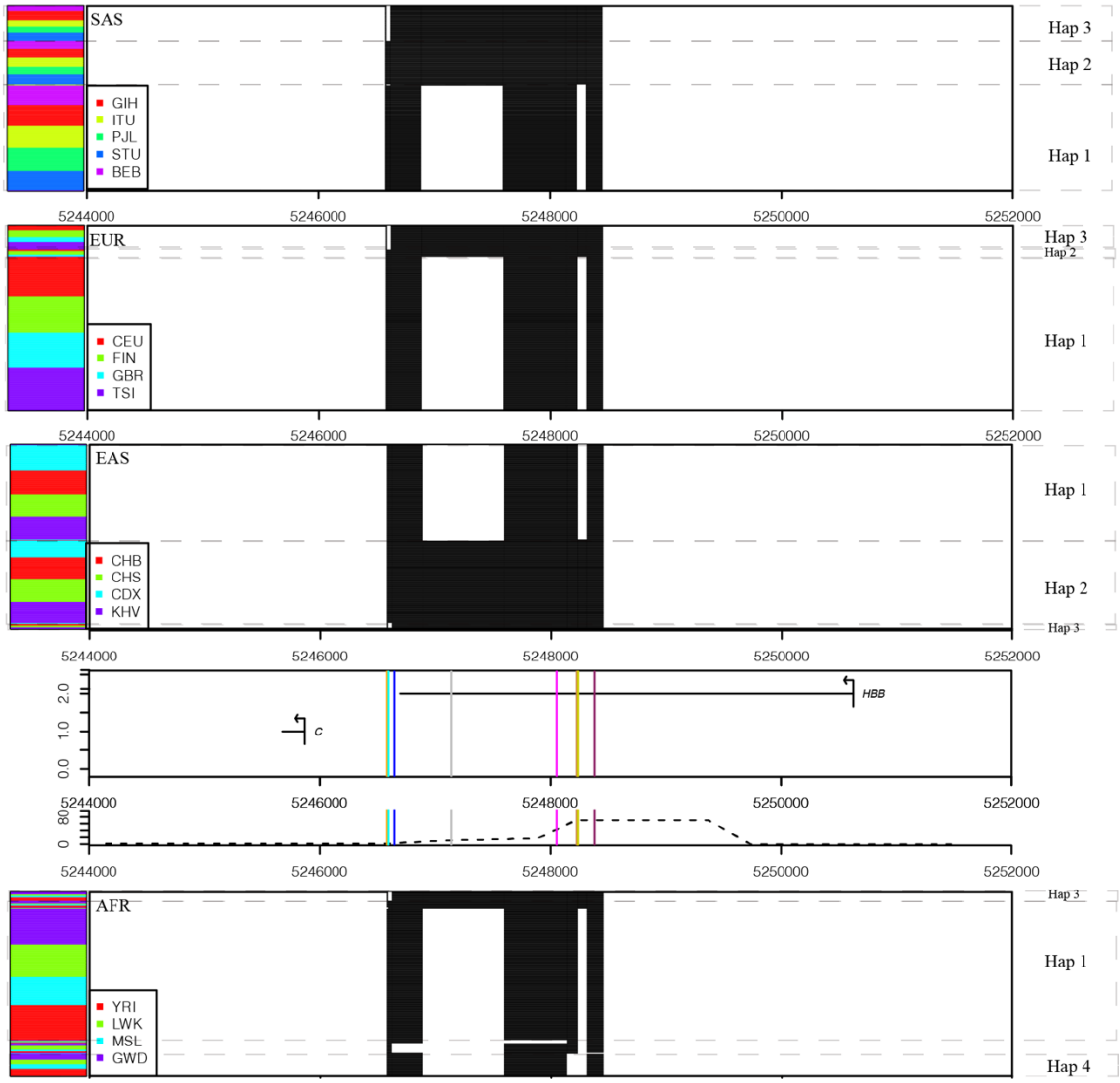
The observation regarding clustering of β-thalassaemia mutations by ethnolinguistic group was also seen in the haplotype analysis, the fact that more than 90% of the chromosomes containing the rs33959855 mutations were from Chonyi ethnolinguistic group being one example.

Figure 25. Haplotypes structures in β-thalassaemia chromosomes compared to non-β-thalassaemia chromosomes from the KGBC and the 1000 Genomes.



The top four panels show chromosomes containing β-thalassaemia mutations with each panel labelled according to the type of mutation present in the chromosomes. The fifth and sixth graphs from the top show the recombination and gene map for this region with the horizontal-coloured lines representing the positions of the eight SNPs that were used in defining haplotypes. The bottom two image graphs show the haplotype structures in normal chromosomes from the KGBC (WT_klf) and from the 1000 Genomes superpopulation for comparison. On the left side of the image graph is a line graph with different colours representing different superpopulations from which the samples were obtained. Within the KGBC datasets ethnolinguistic groups with the highest case numbers have been highlighted: Chonyi(n=50), Giriama (n=15), Kauma(n=12). The remaining ethnolinguistic groups with a single case each were classified under others (n=4). On right side are the major haplotypes labelled Hap 1-4.

Figure 26. Detail of haplotype structures in the 1000 Genomes sub-populations.



The four graphs labelled SAS, EUR, EAS, AFR represent haplotype structures for the following 1000 Genomes super populations: South Asia, European, East Asia and Africa, respectively. The colour codes represent the sub-populations namely: GIH(Gujarati Indian from Houston), ITU(Indian Telegu from UK), PJL(Punjabi),STU(Sri Lankan Tamil), BEB(Bengali), CEU(Utah residents with Northern and Western European ancestry), FIN(Finnish),GRB (), TSI(Toscani), CHB(Han Chinese), CHS (Southern Hans Chinese), CDX (Chinese Dai), KHV (Kihn in Vietnam), YRI(Yoruba), LWK (Luhya), MSL(Mende), GWD (Gambian). The fifth and sixth graphs from the top show the recombination and gene map for this region with the horizontal-coloured lines representing the positions of the eight SNPs that were used in defining haplotypes.

4.5.4 Discussion/Conclusion

Having identified four mutations associated with β-thalassaemia in Kilifi, in this chapter I aimed to determine the frequencies of these variants and their haplotype structures across diverse global populations, thereby providing valuable insights into the history and origins of these

mutations. This investigation involved conducting a literature review on the prevalence of these mutations, examining evidence of ethnic clustering, and exploring haplotype sharing patterns.

After performing a literature search on β-thalassaemia and the countries mentioned in the articles, I found there were between three- and five-times more articles mentioning β-thalassaemia and any Asian or European countries than there were on β-thalassaemia and any African country. Of the studies that mentioned African countries, the majority mentioned North African countries with at most 35 studies mentioning any of the sub-Saharan countries. In comparison, there were eight-times more studies mentioning α-thalassaemia and any of the sub-Saharan countries than north African countries. The few studies reporting on β-thalassaemia in sub-Saharan Africa might reflect the low prevalence of the condition resulting in fewer clinical cases and as a result it rarely comes to the attention of public health officials unlike α-thalassaemia and sickle cell disease which have more studies done in sub-Saharan countries.

The most common β-thalassaemia mutation identified in the current study was the rs33959855 G>T mutation. Analysis of ethnicity in the study revealed that rs33959855 was significantly associated with the Chonyi ethnolinguistic group as compared to other ethnicities. The association of rs33959855 with the Chonyi ethnolinguistic group suggests that, whatever its origin, the mutation may not have been present in Kilifi for long enough to spread to all sympatric ethnic groups. If this observation is true, then we would expect to see evidence of haplotype sharing between chromosomes carrying the rs33959855 mutation, similar to what we observed for chromosomes harbouring the rs334 mutations, which are mainly found on a single haplotype referred to as BANTU haplotype (McGann et al., 2018). Haplotype analysis on rs33959855 chromosomes revealed that more than 90% of the chromosomes shared the same haplotype (Hap 3), as defined using 8 SNPs found in both KGBC and 1000 Genomes datasets that do not include the β-thalassaemia mutations. This haplotype was less common in normal chromosomes from both the KGBC and the 1000 Genomes. Minor haplotypes were also observed in rs33959855 chromosomes, which suggest there might be recombination with other

haplotypes, an observation which is supported by the presence of a large recombination hotspot in the adjacent region. I note that this analysis relies on statistical phasing of heterozygous SNPs along the chromosome, and another explanation for presence of these minor haplotypes might therefore be phasing errors; however, these appear to be limited based on my analysis of repeat phasings. Altogether the dominance of Hap 3 in rs33959855 mutation and not in other β-thalassaemia or normal chromosomes was an important observation.

If the conclusion that rs33959855 mutation was introduced on the background of Hap 3 is true, then it is possible that the mutation might have been introduced through gene flow since the chromosomes of the rest of the population (>90%) were mainly of the Hap 1 haplotype. The population in which Hap 3 was highest was the SAS superpopulation in the 1000 Genomes database (although we remind the reader that these do not carry the β-thalassaemia mutations, which were not observed in the 1000 Genomes). Interestingly in the genomic database search, 15/16 participants with a mutation at the same site were of the SAS population, though this was a G>C mutation and not the G>T observed in this study and observed exclusively in the large gnomAD dataset.

Furthermore, in the KGBC data, Hap 3 was mainly observed in the Chonyi, raising the question of whether there is evidence of contact between the Chonyi and south Asian populations. In a study we conducted on admixture into and within sub-Saharan Africa, which included a subset of samples from this study, we inferred admixture events contributing Eurasian haplotypes into the Chonyi (approximately dated as 1138 CE: 1080-1182CE) and Kauma (1225CE: 1167-1254CE) (Busby et al., 2016). The most closely matched present-day surrogate population for the admixture were the Gujarati (GIH) and Kinh Vietnamese (KHV). A plausible explanation is that admixture could have occurred during the era of Asia - Swahili trade that took place along the East African coast in 1200CE (Roberts, 2007). The only other case ever reported in the literature that carried the rs33959855 G>T mutation was an individual from Reunion Island whose ancestry was not reported. Reunion Island is in the Indian Ocean close to the East

African Coast. The main inhabitants of the Island are of African, South Asian, and European mixed ancestry (Berniell-Lee et al., 2008).

The second most common β-thalassaemia mutation in my study was the rs33941849 (T>C), which is also rare with only 34 cases reported in literature. The cases are spread in seven regions and in eleven countries with no specific country dominating. Among the five African individuals identified with this mutation, four were from Mayotte island which is located along the East African coast. The individuals were of Comoran origin, a population which is thought to have originated from Eastern Africa (Muszlak et al., 2015). In the current study the mutation was significantly associated with Kambe and Kauma ethnolinguistic groups. All the chromosomes with this mutation were of Hap 1 haplotype which was the dominant haplotype in the general population. Taken together this would suggest the mutations might have first occurred in the Kambe or Kauma groups either through gene flow or locally within the genetic background of the current population.

The third most common β-thalassaemia mutation, rs193922563 is also a rare mutation that has so far been described in six individuals of Asian origin. The six carriers identified in Kilifi represent the highest number of cases with this mutation reported from a single site. The chromosomes carrying this mutation did not show evidence of ethnic clustering or any unique haplotype that was different from that of normal chromosomes.

Based on my literature review, rs35004220 is one of the most common β-thalassaemia mutations with majority of cases having been reported in Western Asia, Southern Europe, and North Africa. Several studies suggest this mutation might have originated from East-Mediterranean countries especially Turkey where it is found at high prevalence. The mutation then spread to north Africa during the Ottoman empire rule that controlled Southern Europe, Western Asia and Northern Africa between the 16th and 19th century (Cherry et al., 2016; Perrin et al., 1998). The two participants identified in the current study were from Chonyi and Kauma ethnicities. The chromosomes carrying this mutation belonged to the common Hap 1 haplotype group.

Based on this evidence, I hypothesize that rs33959855 might be a recent mutation introduced to the Chonyi from a population carrying the Hap 3 haplotype and might be undergoing admixture with the local chromosomes dominated by the common Hap 1 haplotype. The other three mutations were probably either introduced through gene flow or emerged locally on the common Hap 1 haplotype.

The results presented in this chapter should, however, be interpreted with caution as there are several important limitations associated with the study. First the length of the sequence was short and only captured a limited number of SNPs, limiting my ability to distinguish haplotypes. Extending the sequence region would have captured more sequence diversity and would result in more precise analysis of haplotype sharing. A second limitation is that β-thalassaemia haplotypes could not be inspected directly in public genome datasets such as the 1000 genomes project, mainly because the mutations were not observed in these datasets. Lastly, having sequenced all the individuals with HbA_2 of >4% among HbAA participants, and identified 83 participants in total, I made assumptions that the remaining participants were free of β-thalassaemia mutations. Although a HbA_2 cut-off of >4% is used in clinical diagnosis of β-thalassaemia carriers, in Chapter 3 I observed that some individuals have HbA_2 values of between 3.5 and 3.9%. Using the allele frequency, I estimated the number of β-thalassaemia carriers as 95 in the whole cohort meaning that approximately twelve cases might have been mis-classified as normal. Whereas the misclassification of these cases might change the credible values, it is more likely to affect rs35004220 which is the mutation in which 2 out of 3 cases had levels of HbA_2 that were within this range as observed in Chapter 3 (Table 3).

Despite these limitations the observations made in this chapter may be important, as results such as clustering of these mutations by various ethnicities could be used in future new-born screening programmes in identification of β-thalassaemia homozygotes. Evidence for possible haplotype sharing between these communities and populations from outside this region is important genetic epidemiology information that could guide us in answering questions relating to the origin and age of these mutations. To answer these questions, a larger section of the

chromosome would need to be sequenced. In addition, there is also a need to have more African genomic datasets represented as part of the global genomic datasets to enable a broad comparison of haplotype structures.

Chapter 5: The impact of β-thalassaemia on malaria parasite invasion and on the mechanism of enhanced clearance of ring-parasitised RBCs through opsonic phagocytosis.

5.1 Abstract

Introduction

Malaria has been suspected to be the selective force behind high frequencies of β-thalassaemia in malaria endemic regions. The evidence is however limited to one epidemiological study and a few laboratory-based studies with conflicting outcomes. Having observed β-thalassaemia is present in Kilifi at a minimum allele frequency of 0.3%, in this chapter I investigated the impact of β-thalassaemia on invasion and the proposed mechanism of enhanced clearance of ring-parasitised RBCs.

Methods

A total of 17 pairs β-thalassaemia heterozygotes and HbAA controls, matched for various factors including RBC protective polymorphisms were recruited into the study. RBCs from these individuals were examined using flow-cytometer based assays for differences in malaria parasite invasion and growth, expression of RBC membrane proteins and the proposed mechanism of enhanced ring-parasitised RBCs.

Results

Compared to HbAA controls, β-thalassaemia heterozygotes exhibited an overall 20% reduction in invasion efficiency when compared to HbAA controls. However, co-inheritance with α-thalassaemia was associated with a loss of this effect. Several differences were noted in β-thalassaemia RBCs compared to normal RBCs that could potentially affect invasion mechanism including increased levels of basigin, reduced levels CD71 and CD49d, along with changes in RBC indices. However, no significant difference in parasite growth during the first cycle was observed, and the results from opsonic phagocytosis were inconclusive.

Conclusion

β-thalassaemia is associated with a modest reduction in invasion. Resistance to invasion could be driven by several factors including changes in receptors on the RBC surface and coinheritance of other malaria protective polymorphisms with some having an additive and others an antagonistic effect.

5.2 Introduction

In Chapters 3 and 4 I showed that β-thalassaemia is found in the Kilifi region at an allele frequency of approximately 0.3%. I identified four causative mutations: (1) the β^0-thalassaemia type rs33959855 G>T, a nonsense variant that results in premature termination of β-mRNA; (2) rs33941849 T>C, an initiation codon mutation that abrogates the β-globin gene transfer RNA binding site; (3) rs193922563 a 25 bp deletion that causes inactivation of an acceptor splice site, and (4) rs35004220 G>A, a β^+ pathogenic variant that results in the formation of a new splice site resulting in 80% reduction in synthesis of normal RNA. Among the 83 members of the Kilifi Genetic Birth Cohort (KGBC) identified as β-thalassaemia heterozygotes, the rs33959855 mutation was the most prevalent, accounting for 66.3% of the cases identified, followed by rs33941849 at 24.1%, then rs193922563 at 7.2% and rs35004220 at 2.4%. In Chapter 4, I observed differences in the ethnolinguistic distribution and haplotype structures of these mutations which suggest that some might have been introduced from other regions through gene flow (rs33959855) or have occurred *de novo* in the Kilifi population. Whatever the explanation, the persistence of these mutations in this community is a subject of interest. One potential explanation is that β-thalassaemia, like HbS (Uyoga, Macharia, Ndila, et al., 2019) and the Dantu blood group (Kariuki et al., 2020), is protective against malaria, and that this has led to its Mendelian selection within our population.

Support for this protective effect comes from epidemiological observations and laboratory-based investigations. Epidemiological evidence is based notable correlation between the population frequency of β-thalassaemia and the historic incidence of malaria (Hill et al., 1988;

Siniscalco et al., 1961) and the observation made by Willcox and colleagues that β-thalassaemia was 50% protective against uncomplicated *P. falciparum* malaria infections (M. Willcox et al., 1983). Apart from these two epidemiological observations, most of the further evidence has been based on *in vitro* studies, where results are more mixed.

The oxidative damage caused by free α-chains and other changes such as reduced MCV and MCH (Ebel, Kuypers, Lin, Petrov, & Egan, 2021), and changes in biophysical properties (Introini et al., 2022), are likely to create an environment that makes it difficult for malaria parasites to invade and thrive. Laboratory investigations have been performed to determine how these changes impact malaria parasite invasion, growth, and survival, with the aim of understanding the mechanism of protection. Investigations into these mechanisms can be grouped into the following categories:

1. Reduced *P. falciparum* invasion.
2. Restriction of intraerythrocytic growth.
3. Recognition and clearance of *P. falciparum* RBCs by the immune system.
4. Protection against severe malaria through interference with processes such as rosetting and cytoadhesion.

The mechanisms of impaired invasion and growth of *P. falciparum* in β-thalassaemia RBCs are subject to conflicting outcomes with some finding impaired invasion (Brockelman et al., 1987; Udomsangpetch et al., 1993) and growth (Brockelman et al., 1987; Kaminsky et al., 1986; Senok et al., 1997; Udomsangpetch et al., 1993) while others observed no difference in invasion (Ayi et al., 2004; Introini et al., 2022; Senok et al., 1997) or growth (Ayi et al., 2004; Roth et al., 1983). The discrepancies in outcomes could be caused by differences in experimental design including variation in background genetics and sensitivity of the assays used as discussed previously (Chapter 1 section 1.9). The third mechanism proposed by Ayi et al (Ayi et al., 2004) whereby β-thalassaemic RBCs were recognised by the immune system and phagocytosed more readily than normal RBCs, has not been replicated in other studies.

While acknowledging the potential significance of all four mechanisms, in this chapter I primarily concentrate on the first three. A detailed review on past work on these mechanisms is discussed in Chapter 1 section 1.9.

In this chapter therefore, I investigate the impact of β-thalassaemia on malaria parasite invasion and phagocytosis of infected RBCs. I have addressed this objective through a recall by genotype study in which known β-thalassaemia heterozygotes were matched to controls and then invited to donate blood samples for functional assays. To overcome confounding due to other malaria protective polymorphisms, I matched the cases and controls as closely as possible based on the other major malaria protective polymorphisms found in this region. I then used a preference invasion assay to investigate the effect of β-thalassaemia on malaria parasite invasion and growth. I then look at differences in haematological indices and expression of RBC surface proteins. Lastly, I investigate the proposed mechanism of enhanced clearance of ring-parasitized RBCs using an opsonic phagocytosis assay.

5.3 Methods

5.3.1 Study populations

To answer this objective, I conducted a recall by genotype study in which I aimed to recruit 15 β-thalassaemia heterozygotes (HbA/β-thalassaemia) and 15 matched normal HbAA controls from the KGBC. With the aim of minimizing confound, each case-control pair were matched on the basis age, gender, area of residence, and other malaria protective polymorphisms namely α-thalassaemia and the ABO and Dantu blood groups. Participants with HbS mutation were excluded from the study (more details in sections 2.2.2).

5.3.2 Sample processing

Samples for this study were drawn in tandem (one β-thalassaemia heterozygote and one HbAA control) and downstream sample processing was thereafter performed simultaneously. Because RBCs from β-thalassaemia individuals are potentially more prone to deterioration in vitro, I

aimed to perform all assays within a maximum period of 7 days. The sample processing procedures are discussed in detail in sections 2.3.5 to 2.3.11.

Pre-processing samples

Upon sample collection, complete blood counts were performed after which leucocytes were removed using Lymphoprep and Plasmodipur filtre methods. The paired RBC samples were then brought to 50% Hct and diluted to same RBC count.

Invasion assay

As described in the Methods, I assessed invasion using erythrocyte preference assays (Theron et al., 2018). Briefly, RBCs from different donors were stained with different concentrations of dye before being mixed together and co-incubated with ring stage 3D7 strain *P. falciparum* parasites at 5% parasitaemia. All assays were performed in triplicate, and after incubating for 48h for assessment of invasion and 72h for growth, parasites were labelled using a fluorescent DNA dye. Two-colour flow cytometry was then used to count parasites present in each stained RBC cell type. Invasion was calculated both as % infected cells relative to the total number of labelled RBCs for that genotype (averaged across triplicates) and as relative parasitaemia, by dividing the mean parasitaemia of the β-thalassaemia heterozygotes with that of the HbAA controls. Growth was calculated by dividing the average MFI at 72h (averaged across triplicates) by average MFI at 48h for the same sample. Statistical comparisons were then conducted to compare the paired samples and evaluate any significant differences in invasion and growth rates. More details of the calculations involved are found in the Methods, section 2.3.9.

Analysis of RBC membrane protein expression by flow cytometry

I used a panel of monoclonal antibodies targeting 11 antigens on the surface of the RBC that have been confirmed or could potentially be involved in cell adhesion and parasite invasion.

The samples were diluted to 0.5% Hct before incubation with monoclonal antibodies, after which the cells were washed and analysed on a two-colour flow cytometer. The average MFI for the duplicate wells was determined and comparison done for the sample sets. More details on the assay be found under methods section 2.3.11.

Opsonic Phagocytosis assay

As described in Methods, opsonic phagocytosis was assessed using a previously described assay (Gallo et al., 2012). In brief, I first incubated heterozygous and normal control RBCs with 3D7 parasites at trophozoite stage for 24h to achieve 10% parasitaemia. The RBCs were then stained with an intracellular membrane dye CFDA-SE and ethidium bromide for identification of all RBCs and parasitised RBCs, respectively. The stained RBCs were then opsonised with autologous serum and incubated with THP1 monocytes for phagocytosis to take place. Positive controls were a replica of the same experiment with opsonisation done using anti-D IgG in place of autologous serum. Flow cytometry was used to count the number of monocytes with phagocytosed parasitised and non-parasitised RBCs. The average proportion of monocytes with phagocytosed parasitised and non-parasitised RBCs was then calculated for the duplicate wells and comparison done between β-thalassaemia heterozygotes and HbAA controls.

5.3.3 Statistical analysis

Differences between group means were evaluated using non-parametric tests (Wilcoxon paired test, Kruskal-Wallis test) or parametric test (paired t-test) as discussed under each result. P-values of <0.05 were considered statistically significant. All statistical analyses were performed using the R Foundation for Statistical Computing Platform Version 3.1.1 (R Core Team (2020)).

5.4 Results

5.4.1 Baseline characteristics of the study population

A total of 17 β-thalassaemia heterozygotes paired with 17 HbAA controls were recruited to the study. The mean age for both groups was 13 years with more boys than girls being enrolled (at ratio of 12/5; Table 16). The β-thalassaemia mutations identified in this group included rs33959855 (n=13), rs33941849 (n=3) and rs193922563 (n=1). The distribution of α-thalassaemia genotypes among the matched pairs was as follows; wild type normal (n=5), heterozygous (n=9) and homozygous mutant (n=3), while blood group O was predominant, comprising of 76% of the total. All participants were wild type homozygotes at the Dantu marker SNP rs186873296 A>G.

5.4.2 The effect of β-thalassaemia on haematological indices

Full blood counts were performed on all samples to determine the haematogical indices of all study participants. This was performed both to characterise the blood samples, but also to explore potential confounding variables for malaria parasite invasion and growth, as reduced MCV and MCH have both been associated with significant reduction in parasite growth in a previous study (Ebel et al., 2021), and reticulocytes are known to be preferentially invaded by *P. falciparum* parasites (Pasvol, Weatherall, & Wilson, 1980).

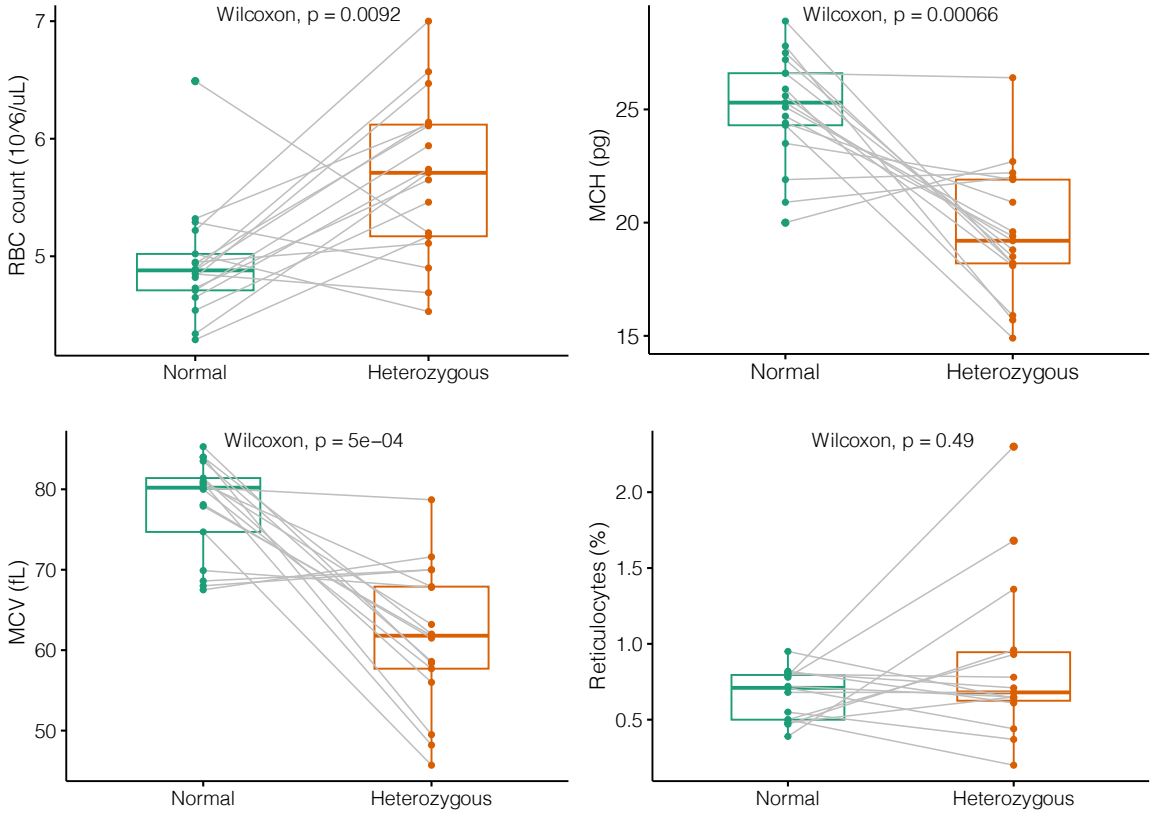
Consistent with previous studies (Kanavakis et al., 1982; Weatherall & Clegg, 2001), I observed that all indices involving RBCs apart from reticulocyte counts and MCHC varied significantly when comparing β-thalassaemia heterozygotes to HbAA controls. β-thalassaemia heterozygotes had significantly elevated RBC counts (5.7 vs. 4.9 $x10^6/\mu l$; P=0.009) and lower MCVs (61.7 vs. 78.0fL; P=0.001), Hb concentrations (11.0 vs. 12.3g/dL; P=0.006), MCH values (19.6 vs. 25.1pg; P=0.001) and Hct values (34.6 vs. 38.3%; P=0.006). No differences were seen in reticulocyte counts (mean 0.7% (SD=0.2) in HbAA vs 0.9% (SD=0.5) in β-thalassaemia heterozygotes (P=0.56), n=15 pairs) (Table 16 and Figure 27). Also notable, three β-thalassaemia heterozygotes had extremely low MCVs of less than 50fL: approximately 40% lower than the average MCV of HbAA controls (Figure 27).

Table 16. Baseline data showing the haematological and demographic characteristics of study participants.

Characteristic	HbAA controls	β-thalassaemia heterozygotes	*P*-value
Sex (Male/Female)	12/5	12/5	
Mean age (years)	12.9 (1.4)	13.40 (1.7)	0.178
β-thalassaemia			
Normal	17	0	
rs33959855	0	13	
rs33941849	0	3	
rs193922563	0	1	
Haematological indices			
Hct (%)	38.3 (2.6)	34.6 (3.1)	**0.006**
Hb (g/dL)	12.3 (0.9)	11.0 (1.0)	**0.006**
MCH (pg)	25.1 (2.4)	19.6 (2.9)	**0.001**
MCHC (g/dL)	32.16 (1.4)	31.7 (0.8)	0.276
MCV (fL)	78.0 (6.0)	61.7 (8.9)	**0.001**
Platelet count (10^3/μl)	293.0 (59.0)	345.6 (106.0)	0.068
Red blood cells (10^6/μl)	4.9 (0.5)	5.7 (0.7)	**0.009**
Reticulocytes (%) ‡	0.7 (0.2)	0.9 (0.5)	0.49
WBC (10^3/μl) (s.d.)	4.8 (1.3)	6.3 (2.3)	0.055
α-thalassaemia			
Normal (αα/αα)	5	5	
Heterozygous (-α/αα)	9	9	
Homozygous (-α/-α)	3	3	
Blood group ABO			
A/A	1	1	
A/AB	1	1	
AB/AB	1	1	
B/B	1	1	
O/O	13	13	
Dantu			
Normal (A/A)	17	17	

Figures show means (standard deviations) for continuous variables and absolute numbers for discrete variables. Haematocrit (Hct), Haemoglobin (Hb), mean cell haemoglobin (MCH), mean cell haemoglobin concentration (MCHC), mean cell volume (MCV), white blood cells (WBC). Statistical comparison of haematological data between HbAA controls and β-thalassaemia heterozygote was performed using the Wilcoxon matched pairs test. Values with *P* values of <0.05 are shown in bold. ‡ only 15 samples were analysed for reticulocytes.

Figure 27. Distribution of MCVs, RBC counts, MCH values and reticulocyte counts between paired samples.

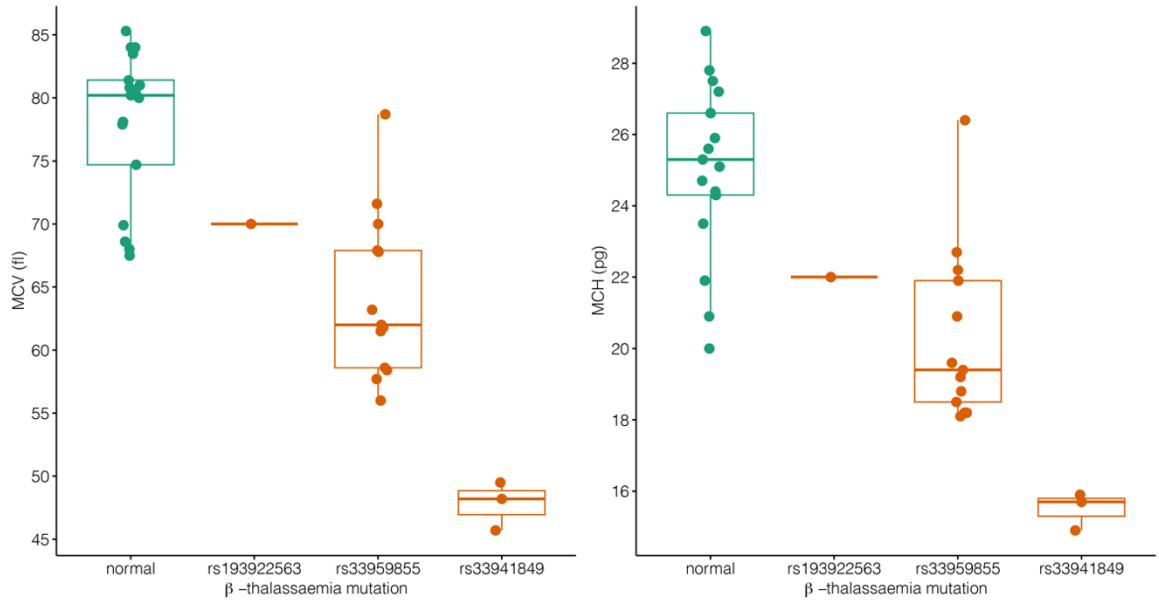


The panels show the distributions of four RBC indices namely, RBCs, MCH, MCV and reticulocytes. The y-axis shows the RBC indices and the x-axis the β-thalassaemia status. The boxes represent the mean, the whiskers show the 95% confidence intervals, and the dots represent the percentage of parasitised RBCs per participant. The grey lines join the sample pairs. Comparison between paired samples was performed using the pairwise two sample Wilcoxon test.

5.4.2.1 Effect of β-thalassaemia mutation type and α-thalassaemia genotype on MCV and MCH values

Having enrolled participants with three different β-thalassaemia mutations (rs33959855, rs33941849 and rs193922563), I investigated whether MCV and MCH values might vary by mutation. I observed that 3 participants had MCV values of <50 fL and MCH of <16 pg. Interestingly, all three were heterozygous for the same mutation: rs33941849 (Figure 28).

Figure 28. Effect of β-thalassaemia mutation type on MCV and MCH levels.



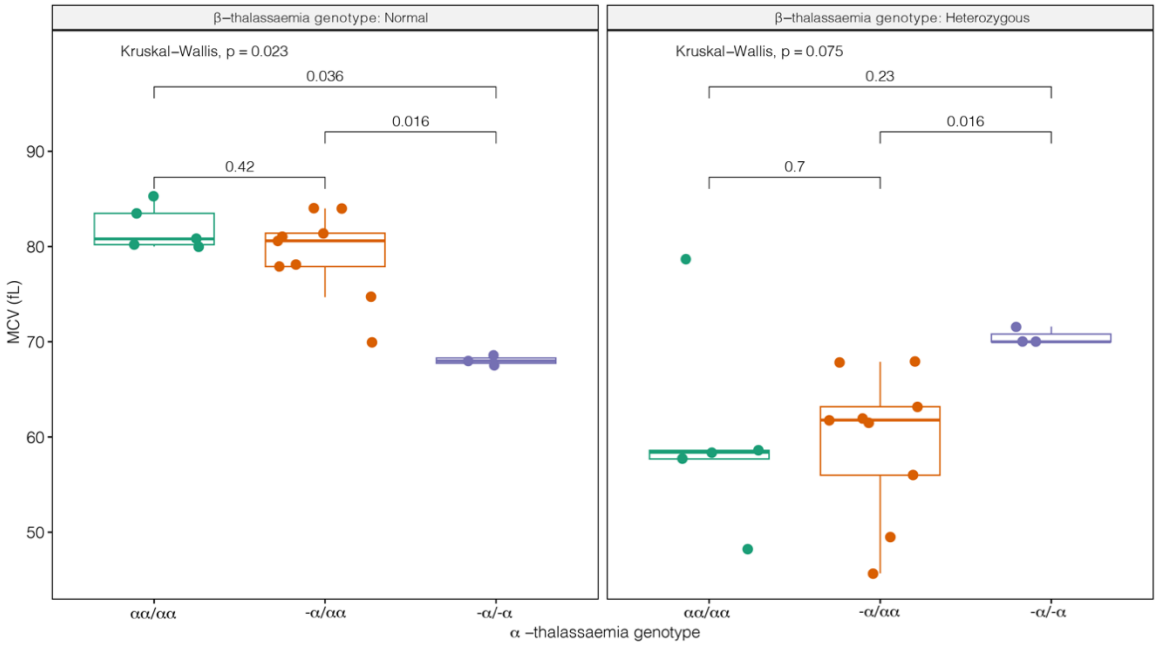
The two panels of graphs show the distribution of MCV and MCH on the y-axis. The status and type of β-thalassaemia mutations is shown on the x-axis with normal representing the HbAA controls. Within each panel the boxes represent the mean, the whiskers show the 95% confidence intervals, and the dots represent the MCV and MCH value per participant.

Previous studies have consistently shown that coinheritance of both β-thalassaemia and α-thalassaemia has a significant effect on haematological indices. It has been observed that the co-inheritance of α-thalassaemia results in increased MCV and MCH values in a dose dependent manner. This effect is opposite to what has been observed when α-thalassaemia is inherited in the absence of β-thalassaemia, whereby MCV and MCH values are reduced in a dose dependent manner (Kanavakis et al., 1982; Saleh-Gohari, Khademi Bami, Nikbakht, & Karimi-Maleh, 2015). Because these changes might interfere with the protective mechanism of β-thalassaemia, I analysed whether MCV and MCH values changed by α-thalassaemia genotype, and whether the interaction between β-thalassaemia and α-thalassaemia led to different RBC index outcomes.

In normal HbAA controls, as expected, I found that α-thalassaemia was associated with a dose-dependent reduction in MCV (Figure 29) and MCH (Figure 30) values, the lowest values being seen in homozygotes. However, the differences in MCV were only significant when comparing means between α-thalassaemia homozygotes (68.0; 95% CI 66.6-69.4 fL)] to normal (82.0; 79.1-84.9 fL) and heterozygotes (79.1; 75.6-82.6 fL) genotypes (Figure 29).

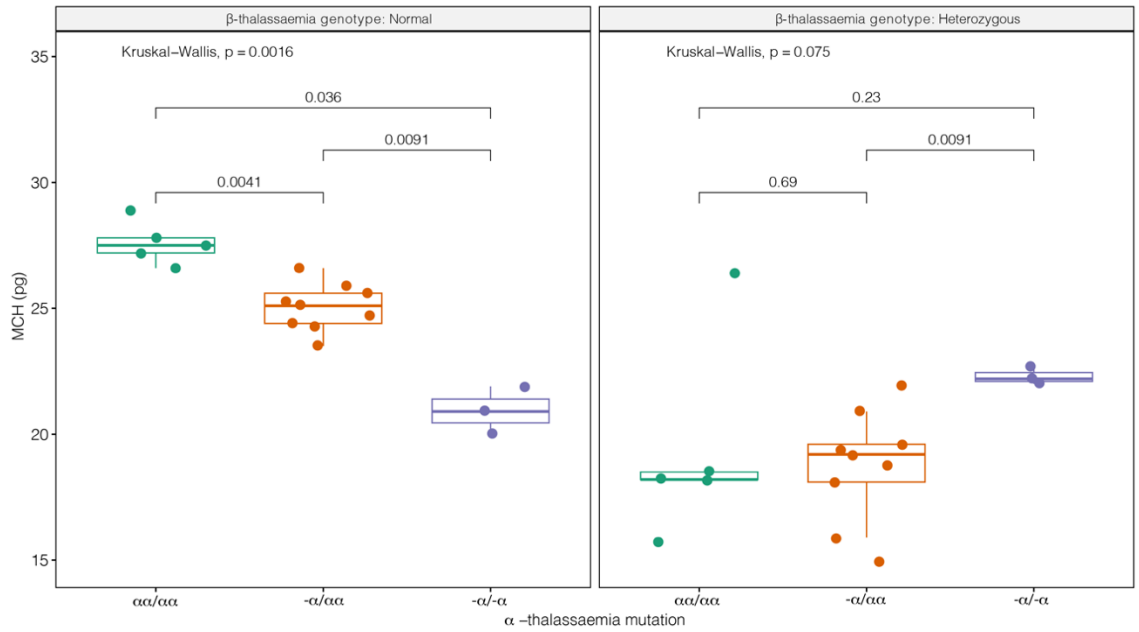
A reversal of this trend was observed among β-thalassaemia heterozygotes, whereby α-thalassaemia was associated with a dose-dependent increase in MCV and MCH values, the highest values being seen in homozygotes (Figure 29 and Figure 30). However, the difference in MCV was only significant when comparing means between α-thalassaemia homozygotes (70.5; 68.2-72.8 fL) vs heterozygotes (59.5; 53.6-65.4 fL) (Figure 29). A similar trend was observed in MCH (Figure 30), although not all comparisons were significant within β-thalassaemia heterozygotes, probably due to the outlier value in the group with a normal genotype for α-thalassaemia.

Figure 29. The effect of α-thalassaemia genotype on MCV levels among β-thalassaemia heterozygotes and HbAA controls.



The two panels show the distribution of MCVs (y-axis) in HbAA controls (β-thalassaemia genotype: Normal) and β-thalassaemia heterozygotes (β-thalassaemia genotype: Heterozygous), stratified by the α-thalassaemia genotype: αα/αα=normal; -α/αα =heterozygous; -α/-α=homozygous). Within each panel the boxes represent the mean, the whiskers show the 95% confidence intervals, and the dots represent the MCV values for each participant. Statistical comparison across genotypes was done using the Kruskal-Wallis test, while pairwise comparison was done using Dunn's test.

Figure 30. Effect of α-thalassaemia genotype on MCH values in β-thalassaemia heterozygotes and HbAA controls.



The two panels of graphs show the distribution of MCH (y-axis) in β-thalassaemia controls (β-thalassaemia status: Normal - HbAA) and β-thalassaemia heterozygotes (β-thalassaemia status: Heterozygous) stratified by the α-thalassaemia genotype. Within each panel the boxes represent the mean, the whiskers show the 95% confidence intervals, and the dots represent the MCH values for each participant. Statistical comparisons across genotypes were done using the Kruskal-Wallis test, while pairwise comparison was done using Dunn's test.

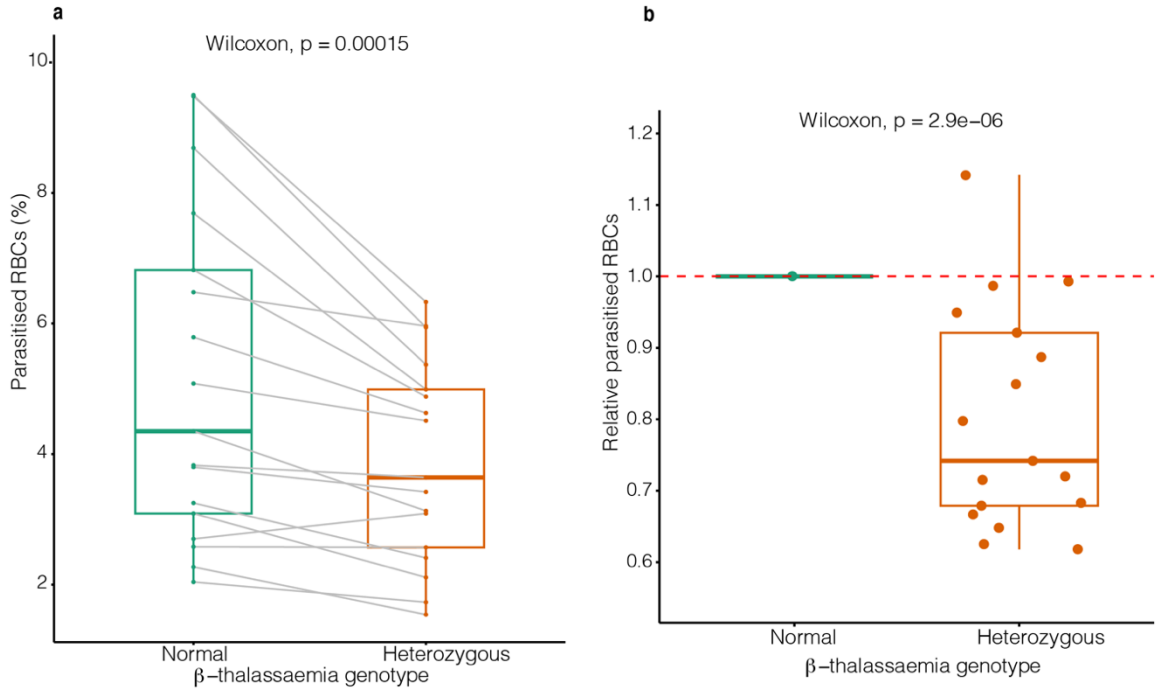
It is worth noting that this analysis is based on only a limited number of participants per category, meaning that these results should be interpreted with some caution. However, it does show a trend in haematological indices by α-thalassaemia genotype that has also been observed in previous studies (Kanavakis et al., 1982; Saleh-Gohari et al., 2015).

5.4.3 Invasion by 3D7 *P. falciparum* parasites is reduced into β-thalassaemia RBCs

Next, I investigated the impact of β-thalassaemia on *P. falciparum* invasion and growth using FACS-based preference invasion assays. The paired samples were first differentially labelled and co-incubated with 3D7 strain *P. falciparum* parasites at 5% parasitaemia. Invasion and growth events were then measured using a fluorescent DNA dye (SYBR-Green) at 48 hrs and 72hrs respectively. Invasion was calculated as % infected RBCs relative to the total number of labelled RBCs for that genotype and as relative parasitaemia (see section Methods section 2.3.9 for more details).

At the 48h timepoint (Figure 31a), I observed a significant difference in parasitaemia between β-thalassaemia heterozygotes (mean 4.1%; 95% CI 3.1-5.1%) compared to that of the paired HbAA controls (5.5%; 3.9-7.1%). When parasitaemia in β-thalassaemia heterozygotes was analysed relative to the paired HbAA controls, β-thalassaemia heterozygosity was associated with a mean invasion efficiency of 0.80 (0.72-0.88) compared to controls which is equivalent to an average of 20% reduction in invasion efficiency (Figure 31b).

Figure 31. Invasion in β-thalassaemia heterozygotes compared to paired controls at 48h post incubation with 3D7 parasites.

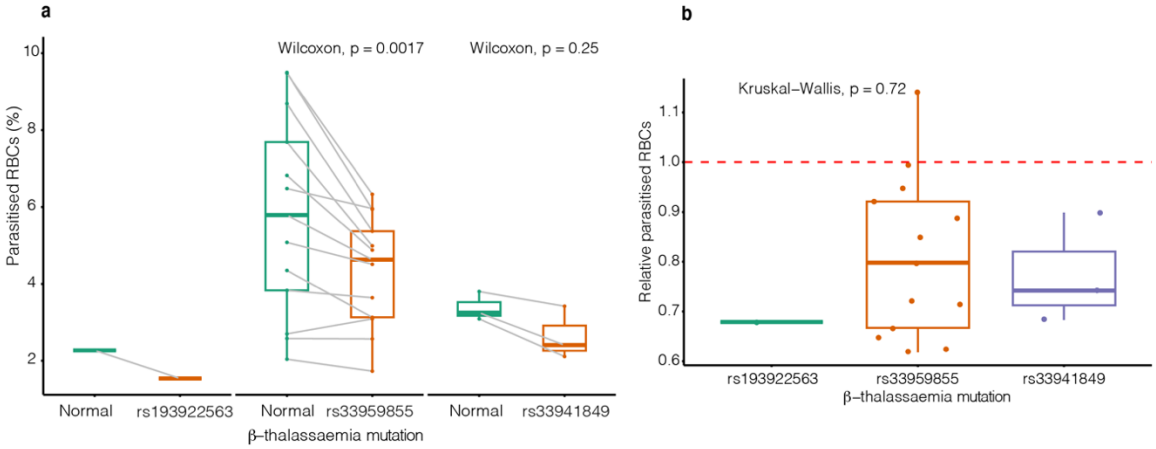


Using flowcytometry I measured the relative ability of 3D7 parasites to invade RBCs from β-thalassaemia heterozygotes (Heterozygotes) compared to HbAA controls (Normal). **a)** The graph shows proportion of parasitised RBCs at 48h post co-incubation of paired RBCs with 3D7 parasites. The y-axis shows the proportion of parasitised RBCs and the x-axis the β-thalassaemia genotype. The boxes represent the mean, the whiskers show the 95% confidence intervals, and the dots represent the percentage of parasitised RBCs per participant. The grey lines show the sample pairs. Comparison between paired samples was performed using pairwise two sample Wilcoxon test**. b**) The graph shows relative parasitaemia at 48h post-incubation in β-thalassaemia heterozygotes versus HbAA controls. For each pair of participants, the proportion of parasitised RBCs in β-thalassaemia heterozygotes was normalized relative to the paired control. The y-axis represents the relative parasitised RBCs after normalisation and the x-axis the β-thalassaemia status. The boxes represent the mean, the whiskers show the 95% confidence intervals, and the dots represent the normalised values per participant. After normalisation the values for HbAA controls were all equals to 1. β-thalassaemia heterozygotes with a relative parasitaemia of below 1 indicates those whose parasitaemia was lower compared to controls and above 1 indicates those whose parasitaemia was higher compared to controls. The mean for the β-thalassaemia heterozygotes thus represents the average invasion efficiency of β-thalassaemia heterozygotes compared to controls. Comparison between paired samples was performed using pairwise two sample Wilcoxon test.

5.4.4 Effect of different types of β-thalassaemia mutations on malaria parasite invasion.

As noted above, there was a trend in changes in haematological indices with β-thalassaemia genotype, particularly for rs33941849 where all the three heterozygotes had the lowest MCVs of less than 50 fL. As MCV has been associated with invasion previously (Senok et al., 1997), I compared relative parasitaemia in β-thalassaemia participants with and without rs33941849 mutation. No difference was observed: those without rs33941849 averaged a relative parasitaemia of 0.77 (95% CI 1.05-0.49) and those with rs33959855 mutation also averaged 0.81 (0.71-0.91) (Figure 32).

Figure 32. Effect on invasion of β-thalassaemia mutation type.



Differences in relative ability of 3D7 to invade RBCs of different β-thalassaemia mutations. **a**) Proportion of parasitised RBCs at 48h post co-incubation (y-axis) grouped by the different β-thalassaemia mutations (x-axis). The grey lines show the sample pairs. Comparison between paired samples was performed using pairwise two sample Wilcoxon test**. b)** Relative ability of 3D7 to invade RBCs of different β-thalassaemia mutations relative to the HbAA controls. For each pair of participants, the proportion of parasitised RBCs in β-thalassaemia heterozygotes was normalized relative to the paired control. The y-axis represents the relative parasitised RBCs after normalisation and the x-axis the β-thalassaemia status. The boxes represent the mean, the whiskers show the 95% confidence intervals, and the dots represent the normalised values per participant. After normalisation the values for HbAA controls were all equals to 1. β-thalassaemia heterozygotes with a relative parasitaemia of below 1 indicates those whose parasitaemia was lower compared to controls and above 1 indicates those whose parasitaemia was higher compared to controls. Therefore, the mean relative parasitised RBCs for each β-thalassaemia mutation represents the average invasion efficiency in RBCs of different β-thalassaemia mutations compared to controls.

5.4.5 Effect of coinheritance of β-thalassaemia and α-thalassaemia mutations on malaria parasite invasion.

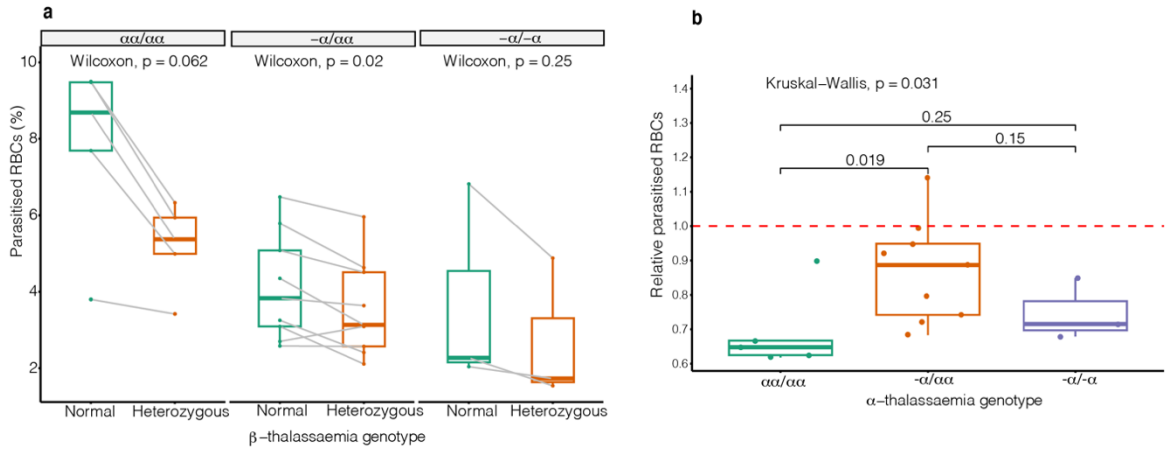
The co-inheritance of some specific malaria protective polymorphisms has been shown to cancel out their individual effects (Williams, Mwangi, et al., 2005). Having observed that α-

thalassaemia affects haematological indices in β-thalassaemia heterozygotes, specifically MCH and MCV, I hypothesised that these changes might also have indirect effects on the relative parasitaemias among β-thalassaemia heterozygotes.

When parasitaemia was stratified by α-thalassaemia genotypes, there was a trend towards reduction in parasitaemia with increased loss of α-genes in both normal and β-thalassaemia heterozygotes (Figure 33a). These differences could partially be explained by minor variations in the proportion of donor parasites added or by the physical properties of the RBCs such as differences in MCV and MCH.

Among β-thalassaemia heterozygotes, when comparing relative parasitaemia across different α-thalassaemia genotypes, a trend was observed towards a greater reduction in invasion efficiency among individuals without α-thalassaemia (31%) compared to either heterozygotes (13%) or homozygotes (25%) (Figure 33b). However, these differences were only significant when comparing relative parasitaemia in α-thalassaemia heterozygotes 0.87 (95% CI 0.75-0.98) to normals 0.69 (0.51-0.90; P=0.042). This difference in parasitaemia did not appear to be dose dependent, as there was no trend towards lower parasitaemia when comparing α-thalassaemia homozygotes to heterozygotes. An observed outlier in the α-thalassaemia normal group, with a relative parasitaemia of 0.99, might potentially explain the non-significance observed between this group and α-thalassaemia homozygotes (Figure 33b). However, the limited number of participants per group precludes definitive conclusions.

Figure 33. Effect of α-thalassaemia on invasion in β-thalassaemia heterozygotes



The graphs show invasion in β-thalassaemia heterozygotes RBCs compared to controls, stratified by α-thalassaemia genotype. **a)** Proportion of parasitised RBCs at 48h post co-incubation (y-axis) grouped by the different α-thalassaemia genotypes (Normal= αα/αα, Heterozygotes = -α/αα, and Homozygotes = -α/-α). The grey lines show the sample pairs. Comparison between paired samples was performed using pairwise two sample Wilcoxon test**. b)** Invasion of β-thalassaemia heterozygotes relative to HbAA controls. For each pair of participants, the proportion of parasitised RBCs in β-thalassaemia heterozygotes was normalized relative to the paired control (represented by the red dotted line). The y-axis represents the relative parasitised RBCs after normalisation and the x-axis the α-thalassaemia genotype. The boxes represent the mean, the whiskers show the 95% confidence intervals, and the dots represent the normalised values per participant. Individuals with a relative parasitaemia below 1 (indicated by the dotted red line) had lower parasitaemia compared to controls, whereas those with values above 1 had higher parasitaemia in β-thalassaemia heterozygotes compared to controls. The mean for each box thus represents the average invasion efficiency of α-thalassaemia genotype compared to controls. Statistical comparison across genotypes was done using Kruskal-Wallis test, while pairwise comparison between α-thalassaemia genotypes was performed using Dunn's test.

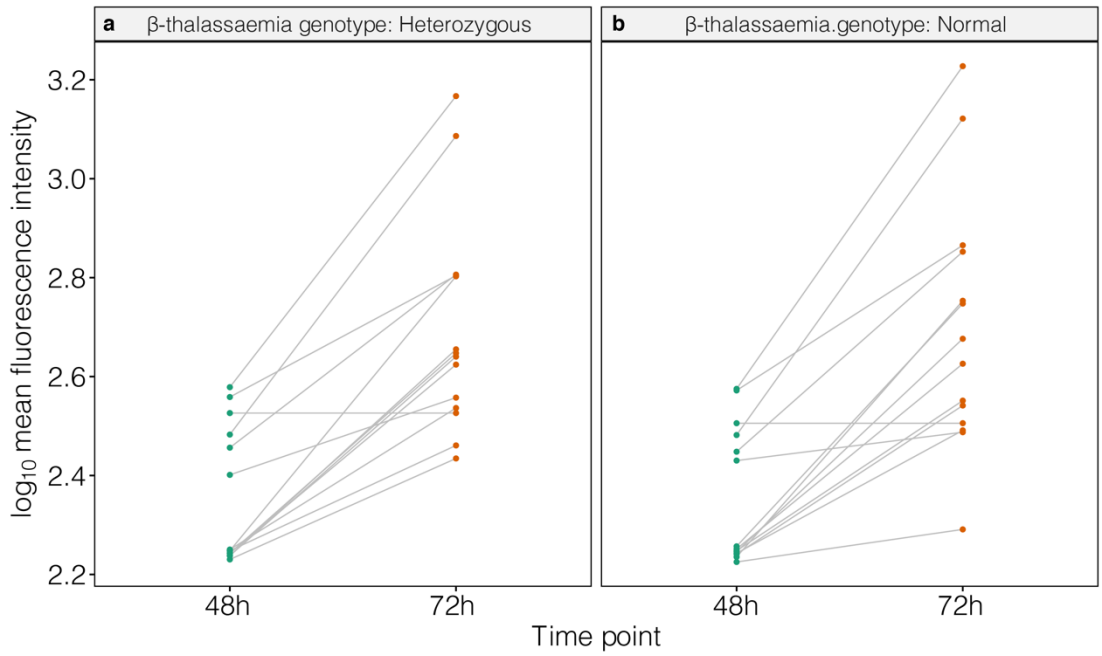
5.4.6 Malaria parasites grow at a similar rate between 48 and 72hr post invasion in β-thalassaemic and control RBCs

To assess differences in growth between β-thalassaemia heterozygotes and HbAA controls, I divided the SYBR-Green MFI at 72h (when parasites are transitioning from trophozoites to early schizogony) by that at 48h (ring stage parasites) and then performed a statistical comparison for the paired samples. As SYBR-Green labels parasite DNA, it is an indication of parasites starting to multiply their DNA, and hence of growth.

RBCs from 3 of the paired sample sets haemolysed during sample processing, so this analysis was performed on only 14 out of the total 17 pairs. In all samples MFI of SYBR-stained parasitised RBCs increased as parasites matured from rings to early schizonts (Figure 34). However, this increase was not significantly different when comparing β-thalassaemia heterozygotes (2.3; 1.7-2.9) vs paired HbAA controls (2.4; 1.8-3.0; *P*=0.78) as determined using

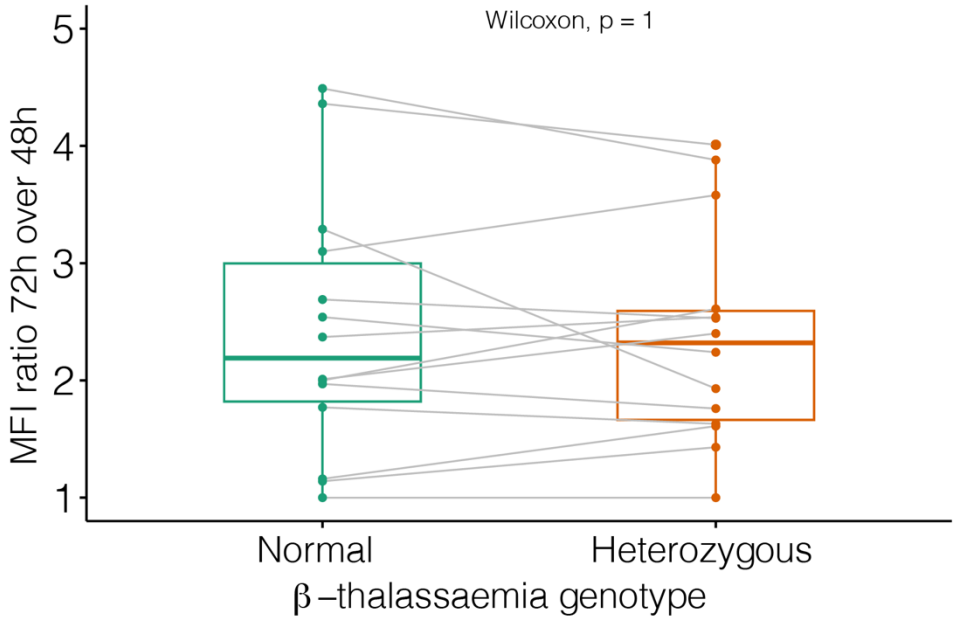
MFI ratios (Figure 35). The presence of early schizonts in the 72-hour samples was validated by microscopy.

Figure 34. Change in SYBR-Green MFI of parasitised RBCs between 48h and 72h post incubation.



The two panels of graphs show the change in SYBR-Green MFI of parasitised RBCs in β-thalassaemia heterozygotes (A) and HbAA controls (B) at 48h and 72h post co-incubation with 3D7 parasites. In both panels the y-axis shows the log transformed MFI and the x-axis the time points. Each dot represents the MFI reading for each sample at 48hrs and the grey lines connect to the corresponding reading at 72hrs post-incubation.

Figure 35. Comparison of 72h/48h SYBR-Green MFI ratios between β-thalassaemia heterozygotes and HbAA controls



Parasite growth was determined comparing SYBR-Green MFI ratios at 48 and 72h between paired normal and β-thalassaemia heterozygotes. The y-axis represents the ratio of MFI at 72h divided by MFI at 48h for each sample. The x-axis shows the β-thalassaemia genotype (Norm= normal and Het= heterozygous). The boxes represent the mean, the whiskers show the 95% confidence intervals, and the dots represent the 72h/48h MFI ratio for each participant. The grey lines show the sample pairs. Comparison between paired samples was performed using pairwise two sample Wilcoxon test.

5.4.7 β-thalassaemia is associated with variation in RBC surface expression of Integrin, basigin and transferrin receptor (CD71)

The experiments above suggest that there is a small, but consistent effect of β-thalassaemia on *P. falciparum* invasion, but not growth. Invasion of RBCs involves interaction between multiple receptors on the surface of the RBC membrane and parasite ligands. Because multiple receptors are involved, I used flow cytometry to determine whether expression of specific RBC proteins, including some known to act as invasion receptors, varied between β-thalassaemia heterozygotes and paired controls, in order to explore whether the abundance of a given receptor correlated with invasion efficiency.

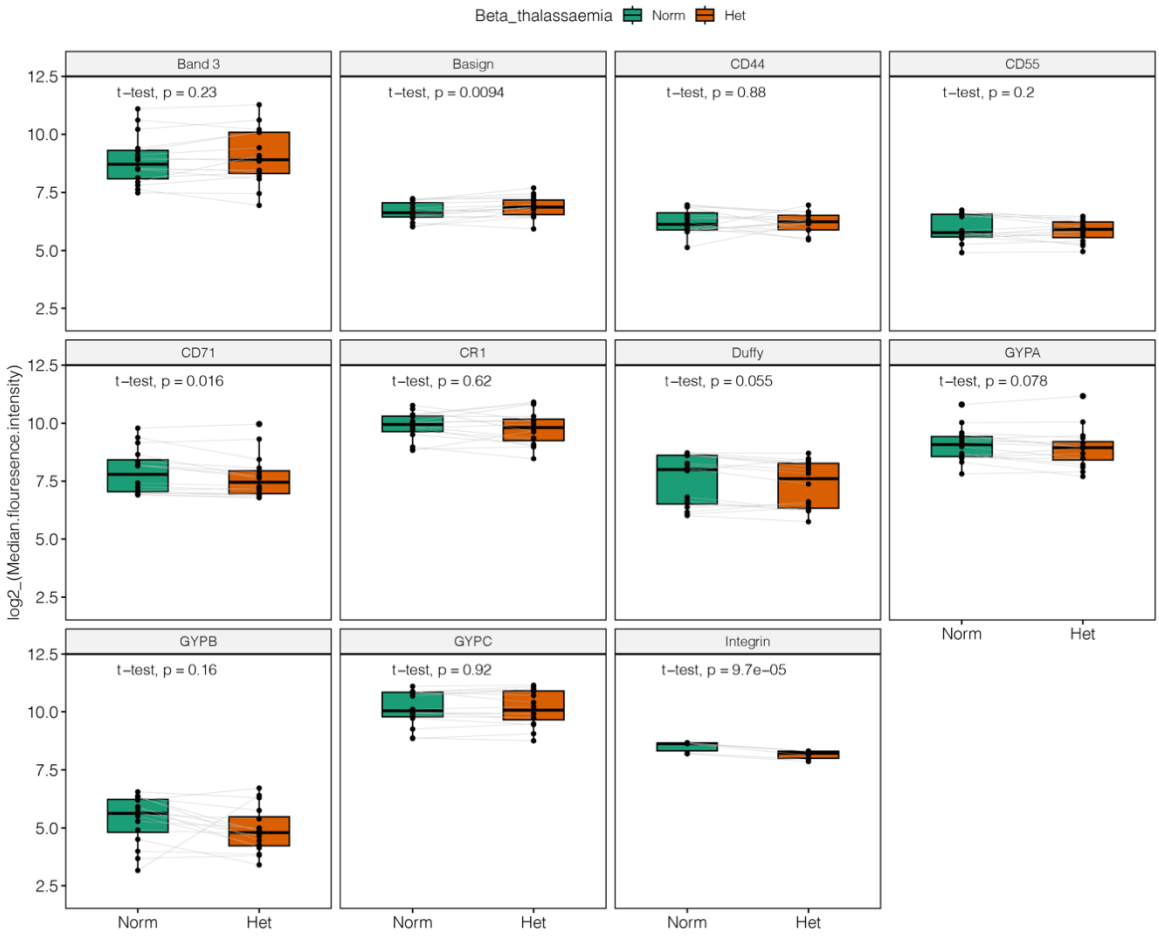
The RBCs from paired β-thalassaemia heterozygotes and HbAA controls were analysed concurrently and in duplicate to reduce variation. Each sample was first diluted to 0.5% Hct and incubated with the monoclonal antibodies in different wells. The cells were then washed

after which the MFI readings were measured on a flow cytometer. The average MFI for the duplicate sample wells was calculated and a comparison in means between the paired β-thalassaemic heterozygotes and HbAA controls performed using the Student's *t*-test.

In total 17 pairs of samples were analysed for all antigens apart from integrin in which 14 pairs were analysed. Three pairs could not be analysed for this receptor due to a shortage of reagents from the manufacturers during the period of the COVID19 epidemic.

I observed a small but significant increase in expression level of basigin in β-thalassaemic RBCs. In addition, I observed reduced expression of Intergrin-alpha-4 (CD49d), an important receptor for adhesion and contact of bone marrow stroma during RBC development, and reduced expression of transferrin receptor (CD71), a marker for younger RBCs that is lost with maturation (Figure 36).

Figure 36. Expression of RBC membrane proteins in β-thalassaemia heterozygotes and HbAA controls.



Box plots showing relative expression of 11 RBC membrane proteins in β-thalassaemia heterozygotes compared with HbAA controls. Participant RBCs were incubated with monoclonal antibodies targeting the specific proteins, washed and the MFI assessed using flow-cytometry. The paired samples were analysed co-currently. In each box plot the y-axis represents the log transformed MFI and the x-axis the β-thalassaemia status. The boxes represent the mean, the whiskers show the 95% confidence intervals, and the dots represent the log transformed MFI reading of each participant. The grey lines show the sample pairs. All the 17 pairs were analysed for the rest of the proteins apart from Integrin in which 14 pairs were analysed. Statistical comparison across paired samples was performed using paired t-test.

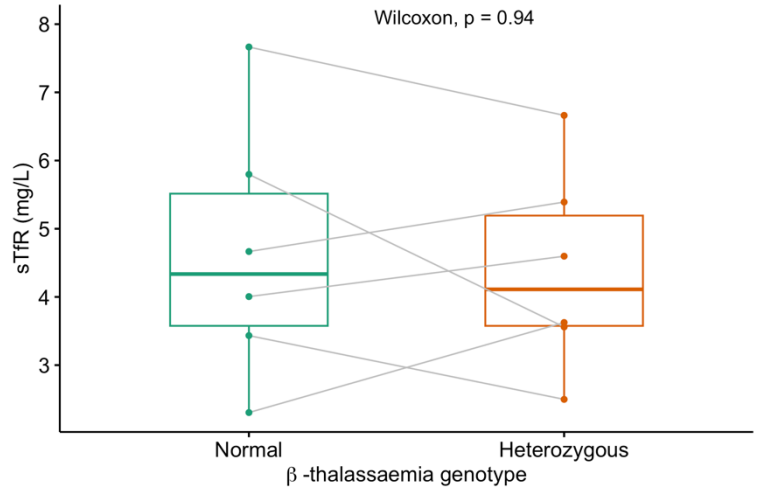
5.4.7.1 Differences in serum transferrin receptor levels (sTfR) levels

The observation that expression of CD71, a marker of reticulocytes, was significantly higher in the control group compared to cases, yet no difference in the absolute number of reticulocytes between groups was identified using a Sysmex XT-2000i Haematological Analyser, prompted further investigation to confirm this finding. sTFR is a soluble form of CD71 generated through cleavage of membrane bound CD71 the bulk of which is generated from erythroblast and to some extent reticulocytes (Beguin et al., 1993). By measuring sTFR, we can indirectly determine the presence and abundance of reticulocytes.

sTFR was measured on a subset of samples (n=6) using a solid-phase sandwich Enzyme-Linked Immunosorbent Assay (ELISA) commercial kit ((Invitrogen; Life Technologies, Thermo Fisher Scientific, Inc., Waltham, MA, USA) (see Methods section 2.3.12).

The reagents available were sufficient for analysis of 6 pairs and the results for the 6 pairs are presented on Figure.38. Upon examination there was no statistical difference in sTfR levels comparing the mean of the HbAA controls 4.6 (95% CI 2.7 – 6.6) verses that of β-thalassaemia heterozygotes 4.4 (2.8 – 5.9; P=0.94) (Figure 37).

Figure 37. Difference in serum transferrin levels between β-thalassaemia heterozygotes and HbAA controls



Concentration of serum transferrin in HbAA controls and β-thalassaemia heterozygotes measured using ELISA. Y-axis represents concentration of sTfR and the x-axis the β-thalassaemia genotype. Quantification was done on six random pairs. Comparison between paired samples was performed using pairwise two sample Wilcoxon test. The boxes represent the mean, the whiskers show the 95% confidence intervals, and the dots represent the concentration of sTfR per participant. The grey lines show the sample pairs.

5.4.8 Differences in immune recognition and phagocytosis of parasitised and non-parasitised normal and β-thalassaemic RBCs

Finally, I also performed opsonic phagocytosis assays as enhanced clearance of ring stage parasitised RBCs has been proposed as one of the mechanisms through which β-thalassaemia protects against malaria (Ayi et al., 2004).

Due to the COVID-19 outbreak, there was a scarcity of reagents for these assays, which limited my ability to conduct thorough testing. Consequently, I was only able to test a total of 3 pairs for this study.

Table 17. Proportion of monocytes phagocytosing opsonised RBCs from β-thalassaemia heterozygotes and HbAA controls.

	Opsonin	β-thalassaemia heterozygotes			HbAA controls		
Sample pairs‡		**P1.2**	**P3.4**	**P7.8**	**P1.2**	**P3.4**	**P7. 8**
% Monocytes with RBCs	Anti-D IgG	10.8%	25.4%	37.7%	25.1%	8.3%	45.0%
% Monocytes with pRBCs	Anti-D IgG	6.0%	17.1%	20.9%	10.8%	10.0%	25.0%
% Monocytes with RBCs	autologous	**0.1%**	**0.2%**	0.0%	**1.1%**	**1.0%**	**1.0%**
% Monocytes with pRBCs	autologous	**0.6%**	0.0%	0.0%	**0.6%**	**0.0%**	0%

‡ The samples were analysed in pairs, comprising both β-thalassaemia heterozygotes and matched non-β-thalassaemia controls, the pairs are denoted by similar sample identifiers, such as P1.2 for β-thalassaemia heterozygotes and P1.2 for non-β-thalassaemic controls, indicating that they belong to the same pair.

For the positive controls (anti-d IgG opsonised), uptake of both parasitised and non-parasitised RBCs by monocytes occurred on all samples as expected. In the experiment (opsonised with autologous serum), the mean percentage of monocytes that had phagocytosed RBCs was 1.0% for HbAA controls and 0.1% for β-thalassaemia heterozygotes. The majority of the phagocytosed RBCs for both genotypes were non-parasitised (Table 17).

Although the data suggest a trend towards increased phagocytosis of non-parasitised RBCs in HbAA controls, it remains inconclusive since it was based on only 3 pairs of samples. Despite my inability to analyse all the samples, the assays that were set up, along with the preliminary observations, could be helpful for planning future investigations of this mechanism within the same cohort.

5.5 Discussion

The objective of this chapter was to investigate the impact of β-thalassaemia on both malaria parasite invasion and on the clearance of ring-parasitised RBCs through opsonic phagocytosis. The results revealed a significant reduction in invasion among β-thalassaemia heterozygotes compared to HbAA controls, but no difference in parasite growth during the first cycle. The findings on opsonic phagocytosis by THP1 monocytes were intriguing but inconclusive, due to limited number of assays run.

The observation that β-thalassaemia is associated with reduced invasion agrees with that of Senok and colleagues (Senok et al., 1997), but contradicts those of both Ayi and colleagues and Introini and colleagues (Ayi et al., 2004; Introini et al., 2022). These discrepancies could potentially be explained by differences in methods used to match study participants or to the sensitivity of the invasion assays employed. In the current study, I observed a modest 20% reduction in invasion efficiency between β-thalassaemia heterozygotes and HbAA controls. This small difference could easily go unnoticed in smaller studies or if confounded by the presence of other genetic polymorphisms. As observed in the current study, in individuals with the normal α-globin genotype, β-thalassaemia was associated with a 31% reduction in invasion when compared to HbAA controls. However, the coinheritance of β-thalassaemia and α-thalassaemia appeared to reverse this trend with heterozygotes and homozygotes showing a 13% and 25% reduction in invasion efficiency, respectively. This observation could be true for other malaria protective polymorphisms and might partly explain the discrepancies observed in these studies. For instance, in the Introini study (Introini et al., 2022), the authors ruled out the presence of HbS, HbC, HbE but did not rule out the presence of α-thalassaemia and G6PD; there were also fewer samples included in that study, and they were not drawn in the paired manner that I used here. While they did not observe overall significant differences in invasion using four different parasite strains (3D7, 7G8, Dd2 and GB4), one parasite strain (Dd2) was associated with significantly reduced preference invasion in heterozygous β-thalassaemic RBCs when compared to HbAA controls. Furthermore, when they performed invasion of 3D7 parasites under live

video imaging, they observed increased invasion events in HbAA controls compared to β-thalassaemia heterozygotes, although the difference was again not significant. Considering that the difference in invasion observed in the current study was fairly modest, it is plausible that methodological differences, specifically the processing of samples, could explain the discordant result in the Introini study. In the study by Ayi and colleagues (Ayi et al., 2004), various polymorphisms including, G6PD, HbS, α-thalassaemia were controlled for, yet no differences in invasion and growth were observed. However, in this work I utilised the more sensitive erythrocyte preference assay in which both erythrocyte types are exposed to the same culture conditions, specifically by culturing them in the same well rather than separately, which may provide an explanation for why we observed a difference in invasion and Ayi and colleagues did not. My experimental setup could have enhanced detection of subtle variations in invasion efficiency that were not previously detected; indeed this was the logic behind the development of the preference assay approach, which identified a preference across ABO blood groups that has not been previously identified (Theron et al., 2018). Additionally, in the current study, participants were matched for location, increasing the likelihood of genetic similarity for other minor RBC polymorphisms.

Similar to invasion, previous studies of the impact of β-thalassaemia on growth and development of malaria parasites have produced mixed results. Whereas some authors (Brockelman et al., 1987; Senok et al., 1997; Udomsangpetch et al., 1993) have observed impaired growth in heterozygous β-thalassaemic RBCs, others have found no difference (Ayi et al., 2004; Roth et al., 1983). In the current study I investigated parasite growth during the first cycle by comparing SYBR-Green MFI readings at 72h (early to late schizonts) to 48h (ring stage). This timeframe is important as it is during this period that the parasite replicates its DNA content. If there was any growth retardation, then it would result in reduced DNA synthesis between 48h and 72h. The DNA content, measured as MFI using flow cytometry, increased at the same rate in both HbAA controls and β-thalassaemia heterozygotes during this period. The results suggest that once *P. falciparum* parasites invade RBCs from β-thalassaemia heterozygotes,

they develop at a normal rate during the first cycle. However, this does not rule out the possibility of reduced parasite growth in subsequent cycles. In a study by Senok et al, for example, the authors observed reduced growth in RBCs from β-thalassaemia heterozygotes with the difference being more apparent in later cycles. However, the study by Ayi and colleagues did not yield similar results in the same growth period. The discrepancies observed in these studies might also result from methodological differences including sensitivity of the methods used and selection criteria of cases and controls as discussed earlier under invasion mechanisms.

So far, I have shown that erythrocytes from β-thalassaemia heterozygotes are more resistant to invasion by *P. falciparum* parasites of the 3D7 strain, but that after invasion the parasites grow normally during the first 72hrs. However, an important question arises on the mechanisms through which the reduction in invasion is mediated. In an attempt to answer this question, I quantified 11 proteins on the surface of the RBC and found differential expression in three membrane proteins. RBCs from β-thalassaemia heterozygotes were associated with increased expression of basigin receptor and reduced expression of Intergrin-alpha-4 (CD49d) and transferrin receptor (CD71).

Basigin is a member of the immunoglobin superfamily and is widely expressed in many tissues including RBCs, leucocytes and endothelial cells. It has multiple functions including, induction of metalloproteases, association with monocarboxylate transporters to catalyse transport of lactate and other monocarbohydrates across plasma membrane, immune regulation as a receptor for cyclophorins and also through interaction with integrins. However, the physiological role of basigin in RBCs is poorly understood (Muramatsu, 2016). It has however been identified as a crucial receptor for PfRH5 ligand, which is essential for the invasion of RBCs (Crosnier et al., 2011). Studies conducted by Crosnier and colleagues and Zenono and colleagues demonstrated complete blockade of invasion using anti-basigin antibodies in various parasites including *P. falciparum* (Crosnier et al., 2011; Zenonos et al., 2015).

The observation of increased basigin expression in β-thalassaemic RBCs in the current study was also recently reported by the Introini study (Introini et al., 2022). It is possible that increased expression of basigin in β-thalassaemia RBCs might result in increased invasion as the parasite would have more receptors available, but that is only true if the quantity of basigin on the surface is limiting for invasion, and there have been no studies correlating basigin expression levels and invasion efficiency. The mechanism behind increased basigin expression in β-thalassaemic RBCs is not known. One possible explanation is that the increased expression of basigin could be as a result of oxidative damage inflicted on RBC membranes by free α-chains, similar to the damage in band 3 caused by these free α-chains (Yuan et al., 1992).

CD71 (transferrin receptor) is a transmembrane glycoprotein that imports iron into the cell. In the blood stream a soluble form of CD71 known as serum transferrin receptor (sTFR) binds and transports iron to the cells where it attaches to membrane bound CD71. This complex is then internalised into the cell through endocytosis (Beguin et al., 1993; Yiannikourides & Latunde-Dada, 2019). The expression of CD71 decreases with maturation of RBCs. Early erythroblasts which require iron for heme synthesis express the highest levels of CD71 and as erythroblasts mature expression of CD71 diminishes. Reticulocytes which serve as transitional stages of RBCs partially express CD71 while mature RBCs do not express CD71 at all (Thomson-Luque et al., 2018). Because of this difference in expression, CD71 is used as a marker for reticulocytes. Another marker that is also lost with maturity of RBCs is CD49d. CD49d is involved in adhesion and migration of erythroblasts to the bone marrow through its ligand VCAM-1 (Telen, 2000; Thomson-Luque et al., 2018). It is rapidly lost at the very early steps of reticulocyte development right after enucleation and as such it is a marker for very immature reticulocytes (Thomson-Luque et al., 2018). The increased expression of both CD71 and CD49d in control samples observed in this study would therefore suggest an increased proportion of reticulocytes among controls. *P. falciparum* has a known preference for reticulocytes, so the preferential invasion of controls over β-thalassaemia could be largely driven by the presence of more reticulocytes in control samples. However, when reticulocytes were

measured immediately after blood draw using Sysmex XT-2000i Haematological Analyser, no significant difference in reticulocyte counts was observed between the groups. The discrepancies in reticulocyte count between flow cytometry-determined CD71 and that measured using the Sysmex XT-2000i has been investigated previously by Kono and colleagues (Kono, Kondo, Takagi, Wada, & Fujimoto, 2009). The Sysmex XT-2000i measures reticulocytes by staining and counting RBCs that are positive for reticulum (remnant RNA) which is also lost with RBC maturity. By comparing the two methods Kono observed that reticulocytes measured using Sysmex Analyser were not equal to those measured using CD71. Whereas Sysmex Analyser quantifies all stages of reticulocytes, CD71 measures highly immature reticulocytes. Considering in the current study, CD49d, another marker for highly immature reticulocytes was also elevated in controls, this would suggest presence of very immature reticulocytes in controls and not in β-thalassaemia heterozygotes. The presence of these immature reticulocytes in controls would suggest increase erythropoietic drive in controls possibly due to anaemia. However, haemoglobin levels in controls were within the normal range for local population and were not suggestive of anaemia whereas those of β-thalassaemia heterozygotes were in the lower range. Based on later observation, it is expected that increased erythropoietic drive is more likely to be present in the β-thalassaemia heterozygotes as compared to controls. To elucidate this conundrum, I measured sTFR levels in both heterozygotes and controls, which is informative of both increase in erythropoietic drive and CD71 levels released from immature erythrocytes. I found no difference in sTFR levels between β-thalassaemia heterozygotes and controls, suggesting there was no increase in erythropoietic drive in both groups which agreed with the evidence on reticulocyte counts measured using the Sysmex Analyser. Taken together, this evidence suggests that reticulocyte counts in both β-thalassaemia heterozygotes and controls are within the normal range, however the reticulocytes from β-thalassaemia heterozygotes express less CD71 and CD49d when compared to controls.

This observation has been reported elsewhere by Kossiva and colleagues 2003 (Kossiva et al., 2003) although their study focused on a more severe form of β-thalassaemia known as β-

thalassaemia intermedia. Unlike heterozygous β-thalassaemia, β-thalassaemia intermedia may result in a severe form of anaemia that could result in transfusion or splenectomy. In this study they compared expression of CD49d and measured sTFR in 3 groups of participants: healthy controls, β-thalassaemia intermedia, Iron deficiency anaemia (IDA) and various diseases (VD). Apart from CD71 and CD49d, the other markers associated with anaemia namely sTFR concentrations, Ferritin and reticulocyte counts were elevated in β-thalassaemia intermedia. Interestingly despite the raised reticulocyte counts, they found CD71 and CD49d was significantly reduced in both bone marrow and peripheral blood in β-thalassaemia intermedia participants compared to IDA, VD and normal controls. The authors hypothesize that reduced CD71 and CD49d was likely the result of oxidative damage caused by free α-chains. The evidence from this study suggests these markers are lost early in the bone marrow before the release of reticulocytes to peripheral blood. It is possible the same mechanism could explain reduced expression of these markers in β-thalassaemia heterozygotes from the current study. Considering reticulocytes are preferentially invaded by *P. falciparum* and *P. vivax*, the reduction of these receptors on the surface of RBCs, before leaving the bone marrow, could potentially reduce invasion of reticulocyte population in β-thalassaemia heterozygotes and as a result protect the host. Further investigations are required to validate this hypothesis.

Consolidating this evidence, we have seen that RBCs from β-thalassaemia heterozygotes are more resistant to invasion by 3D7 parasites, but upon invasion the parasites grow at a normal rate for the first 72hrs. This reduction invasion could be attributed to the differential expression of three membrane proteins, namely basigin, CD71 and CD49d, on the surface of β-thalassaemic RBCs. Whereas increased expression of basigin might support invasion, reduced expression of CD71 and CD49d would counteract this effect. In addition to this, the co-inheritance of β-thalassaemia with other RBC polymorphisms may influence the overall outcome of invasion. For instance, the negative effect observed when β-thalassaemia is coinherited with α-thalassaemia suggests the presence of additional genetic factors might modulate further the effect of β-thalassaemia on invasion efficiency. To add to this complexity

are other RBC properties of β-thalassaemia RBCs that are likely to affect invasion and were not investigated in the current study such as the biophysical properties of the RBCs as described in the Kariuki (Kariuki et al., 2020) and Introini (Introini et al., 2022). It is the sum of this complex interactions that determines the success of malaria parasite invasion into β-thalassaemia RBCs.

In addition to the mechanism of invasion, there may be other mechanisms of protection associated with β-thalassaemia. One such mechanism proposed by Ayi and colleagues is enhanced recognition and clearance of ring-parasitised RBCs (Ayi et al., 2004). However, the attempt to replicate these findings in the current study were unsuccessful mainly due to technical challenges faced in setting up the study during the COVID19 pandemic period as discussed in the results section. Nevertheless, I managed to set up the necessary methods to investigate this mechanism in the future. Another mechanism not investigated in this study is protection from severe forms of malaria such as severe malaria anaemia and cerebral malaria.

In conclusion the findings of this study suggest that β-thalassaemia is associated with a subtle decline in invasion susceptibility mediated through RBC receptors and polymorphisms with some having a synergistic effect and others an antagonistic effect. However, the more important question is how these changes impact the hosts susceptibility to malaria. This is a question that I will address in the next chapter.

Chapter 6: Effect of β-thalassaemia on health and survival during childhood in Kilifi

6.1 Abstract

Introduction

β-thalassaemia is thought to be under Mendelian selection for the survival advantage it confers to heterozygotes against death from malaria. Nevertheless, epidemiological data in support of this hypothesis from regions where malaria is prevalent are extremely limited. In this chapter, I aimed to describe the impact of β-thalassaemia on both under 5 mortality and the incidence of hospital admission with malaria and other childhood illnesses.

Methods

The data presented in this chapter represent a secondary analysis of data from the Kilifi Genetic Birth Cohort (KGBC) study, a prospective study conducted among children living in Kilifi, Kenya (see Chapter 2 section 2.2.1). The children were recruited at ages 3-12 months and followed up for survival and admission to KCH until their 5th birthday. Recruitment blood samples were analysed by HPLC. β-thalassaemia mutations were determined through sequencing DNA of participants with elevated HbA_2 and α-thalassaemia genotyped by PCR.

Results

A total of 72 infants with HbA/β-thalassaemia and 15,080 non-β-thalassaemia carriers were included in this analysis. The rates of both death and hospital admission were similar in both groups, with adjusted hazard ratios (aHR) of 1.3 (95% CI 0.2-9.1) and adjusted incidence rate ratios (aIRR) of 1.4 (0.8-2.4) respectively. No differences were found in the incidence of admission to hospital with specific conditions with the exception of severe anaemia, where the aIRR was 5.1 (1.2-21.0; p=0.02). The risk of admission to hospital with anaemia of <10g/dL among β-thalassaemia heterozygotes decreased in a stepwise manner with the increasing loss of α-globin genes due to α-thalassemia, with α-thalassemia homozygotes having the lowest

incidence. The aIRRs were 3.4 (1.4-8.1; p=0.007), 2.2 (1.0-4.8; p=0.06), and 1.1 (0.2-8.0; p=0.918) for αα/αα, -α/αα and -α/-α groups respectively. The aIRR for admission to hospital with a diagnosis of malaria was 1.4 (0.3-5.5; p=0.67).

Conclusion

In a birth cohort of children in Kilifi, Kenya, I found few effects of heterozygous β-thalassaemia on the incidence of death or specific childhood illnesses. However, I did find an increased incidence of severe anaemia among β-thalassaemia heterozygotes. I hypothesise that this might relate to the lower haemoglobin levels seen in such children at steady state. Consistent with this hypothesis, the coinheritance of β-thalassaemia with α-thalassaemia mitigates this susceptibility in a dose-dependent manner with individuals homozygous for α-thalassaemia showing least susceptibility to severe anaemia.

6.2 Introduction

In Chapter 5, I showed that β-thalassaemic RBCs are modestly resistant to invasion by *P. falciparum* malaria parasites in vitro. However, to date, there remains a dearth of studies that have investigated this question in vivo. Here, I utilise data from a prospective cohort study to investigate the impact of heterozygous β-thalassaemia on both under-5 mortality and the incidence of hospital admission with malaria and other childhood diseases.

To the best of my knowledge, the only other study that has attempted to answer this question in a malaria endemic region was a case control study conducted by Willcox and colleagues in Liberia (M. Willcox et al., 1983). In that study the authors found that β-thalassaemia was associated with 50% protection against clinical malaria; however, they did not report on other conditions such as anaemia, a common sequel of malaria infection. Most studies conducted in non-malaria endemic regions have shown that β-thalassaemia heterozygotes lead a normal life without complications, with the possible exception of mild anaemia during pregnancy (Premawardhena, Arambepola, Katugaha, & Weatherall, 2008; Tsatalas et al., 2009; Weatherall & Clegg, 2001). Nevertheless, it is important to consider that the applicability of these findings

to malaria-endemic regions is not straight forward, and that further studies are needed to better understand the relationship between β-thalassaemia and various health conditions present in these regions.

6.3 Methods

6.3.1 Study populations

The study was conducted on 15,152 out of the 15,577 children recruited to the KGBC study, which was described in detail in Chapter 2. Given that we have previously shown that sickle cell disease was associated with a high mortality rate in this cohort (Uyoga, Macharia, Mochamah, et al., 2019), participants with either HbSS (N=116) or HbS/β-thalassaemia (N=10) were excluded from the current analysis. In brief, upon recruitment, an EDTA sample was collected and measurements of HbS and HbA_2 performed at KWTRP using HPLC. Suspected cases of β-thalassaemia were identified using HbA_2 cut-off values and their β-globin gene sequenced as described in Chapter 3. The β-thalassaemia heterozygotes included in this analysis were those identified as having HbA/β-thalassaemia (N=72/73), as outlined in Chapter 3. One HbA/β-thalassaemia participant and 298 HbAA participants were excluded from this analysis because no linked epidemiological data were available.

The median age at recruitment was 6.5 months (IQR 5.0-8.4). The participants were followed up passively for mortality and admission events until their fifth birthday. The vital status of participants (dead, alive or out-migrated) was recorded through 3-monthly home visits by the KHDSS study team and ward admission events were captured through the system of ward surveillance described in Chapter 3. For each admission, routine laboratory tests, including full blood counts, malaria blood films and blood cultures, were conducted and the discharge diagnosis recorded. To better understand the impact of β-thalassaemia on the incidence of admission with anaemia, I further categorised admission haemoglobin levels as severe anaemia (<5.0 g/dL), moderate anaemia (5.0-6.9 g/dL), mild anaemia (7.0-9.9 g/dL) and any level of anaemia (<10.0 g/dL).

6.3.2 Statistical analysis

Survival analysis was conducted on 72 children identified as HbA/β-thalassaemia heterozygotes and 15,080 non-carriers. Person-years of observation (PYO) for death and hospital admission were calculated based on follow-up times from the date of recruitment and the date of birth, respectively. Age served as the basis for survival time, with the date of recruitment indicating the entry point for the risk of death and the date of birth indicating the entry date for the risk of hospital admission. Participants were censored on death, outmigration, or on their 5th birthday. Mortality rates and the incidence of syndrome specific admission to Kilifi County Hospital were calculated by dividing the number of deaths and admission events by the corresponding PYOs, respectively. Incidence rate ratios (IRRs) were derived using Poisson regression models, both with and without adjustment for the confounders ethnic group, HbS phenotype (HbAA, HbAS), age at recruitment and α-thalassaemia genotype. A P value of <0.05 was considered significant for indicating differences between the two groups. All analyses were undertaken using STATA version 15.1 (StataCorp, Timberlake, USA).

6.4 Results

During the 5-year follow-up period, a total of 146 deaths were recorded during a cumulative 59,132 PYO. The crude mortality was 2.5 per 1000 PYO (95% CI 2.0-2.9) overall. Only 1 death occurred among the group with heterozygous β-thalassaemia in 285 PYO, giving a crude mortality rate of 3.5 per 1000 PYO. In contrast 145 deaths occurred in non-β-thalassaemia carriers during 58,847 PYO, a rate of 2.5 per 1000 PYO. The difference between these rates was not significant (IRR 1.26; 0.17-9.08; P=0.814) (Table 18).

Table 18. Mortality rates among children in KGBC stratified by β-thalassaemia

	N	Deaths (n)	PYO	Deaths /1000 PYO (95% CI)	Hazard ratio (95% CI)	*P*	Adjusted hazard ratio (95% CI) *	*P*
Non-β-thalassaemia	15,080 (99.52%)	145	58,847	2.46 (2.09-2.89)	1	n/a	1	n/a
HbA/β-thalassaemia	72 (0.48%)	1	285	3.50 (0.49-24.8)	1.42 (0.19-10.2)	0.723	1.26 (0.17-9.08)	0.814

PYO=person years of observation, *Adjusted for ethnic group, HbS, age at recruitment and α-thalassaemia genotype.

During the follow-up period, the total number of admissions was 2,930 during a cumulative 67,678 PYO. The overall Incidence Rate was 43.3/1000 PYO (41.7-44.8) PYO. Among these, 21 admissions occurred in β-thalassaemia heterozygotes, with 328 PYO, resulting an incidence of 63.9/1000 PYO. In comparison, 2,909 admissions occurred in non-β-thalassaemia carriers, during 67,350 PYO, an incidence rate of 43/1000 PYO. There were no differences in the incidence rates of admission to hospital among β-thalassaemia carriers and non-carriers (aIRR 1.4; 0.8-2.4; P=0.201) (Table 19).

Table 19. Incidence of admission to Kilifi County Hospital in β-thalassaemia carriers and non-carriers.

	N	ADM (n)	PYO	ADM/1000 PYO (95% CI)	Incidence ratio (95% CI)	*P*	Adjusted incidence ratio (95% CI) *	*P*
Non-β-thalassaemia	15,080 (99.52%)	2909	67,350	43.1 (41.6-44.7)	1	n/a	1	n/a
HbA/β-thalassaemia	72 (0.48%)	21	328	63.9 (41.6-98.0)	1.34 (0.80-2.25)	0.263	1.40 (0.83-2.35)	0.201

ADM= admission, PYO=person years of observation. *Adjusted for ethnic group, HbS, age at recruitment and α-thalassaemia genotype.

However, when compared to non-β-thalassaemia carriers, β-thalassaemia carriers did show a significant difference in the incidence of admission to hospital with WHO defined severe anaemia (Hb<5 g/dl). β-thalassaemia carriers were almost 5 times more likely than non-carriers to be admitted with severe anaemia (aIRR 5.1; 95% CI 1.2-21.0) (Table 20 and 21). Moreover, they were more than twice as likely (aIRR 2.2; 1.3-4.0) to be admitted with an Hb <10g/dL, which captures both mild and moderate, as well as severe anaemia. A similar trend was observed when anaemia was classified as moderate Hb 7-9.9g/dL with an aIRR of 2.0 (1.0-3.8) and Hb

5-6.9g/dl (aIRR 3.0; 0.7-12.3), although the *P*-value in the latter group was non-significant (Table 21).

Table 20. The incidence of admission to hospital with various clinical diseases in β-thalassaemia carriers and non-carriers.

	Non-β-thalassaemia		**HbA/β-thalassaemia**			
	Episodes	**Incidence (95%CI)**	**Episodes**	**Incidence (95%CI)**	**aIRR* (95%CI)**	***P***
Clinical syndromes						
All cause hospital admission	2909	43.1 (41.6-44.7)	21	63.9 (41.6, 98.0)	1.40 (0.83-2.35)	0.201
Neonatal conditions	471	407 (372, 446)	3	541 (174, 1677)	1.48 (0.47, 4.62)	0.498
Malaria	228	3.4 (2.9, 3.8)	2	6.0 (1.5, 24.3)	1.35 (0.33-5.48)	0.666
Severe pneumonia	953	14.2 (13.3, 15.1)	8	24.4 (12.1, 48.7)	1.47 (0.61-3.56)	0.383
Very severe pneumonia	491	7.3 (6.7, 7.9)	5	15.2 (6.3, 36.5)	1.53 (0.49-4.69)	0.459
Meningitis /Encephalitis	1061	15.1 (14.2, 16.0)	9	27.4 (14.2, 52.6)	1.71 (0.82-3.58)	0.149
Severe malnutrition	243	3.6 (3.1, 4.1)	2	6.0 (1.5, 24.3)	0.99 (0.13-7.10)	0.995
Gastroenteritis	577	8.6 (7.9, 9.3)	3	9.1 (2.9, 28.3)	1.33 (0.43-4.15)	0.621
Jaundice	149	2.2 (1.9, 2.6)	0			
Other	561	8.3 (7.6, 9.0)	3	9.1 (2.9, 28.3)	1.27 (0.40-3.97)	0.675
Laboratory features and outcome						
Bacteraemia	103	1.5 (1.2, 1.9)	2	6.0 (1.5, 24.3)	1.34 (0.81-2.22)	0.257
Malaria blood film positive	255	3.8 (3.4, 4.3)	3	9.1 (2.9, 28.3)	1.34 (0.33-5.32)	0.677
Severe anaemia£	72	1.1 (0.8, 1.3)	2	6.1 (1.5, 24.3)	**5.09 (1.23-21.0)**	**0.024**
Transfused	92	1.4 (1.1, 1.7)	0			
Hospital deaths	45	0.6 (0.4, 0.9)	0			

Incidence= incidence per 1000 years of follow-up, £ Severe anaemia defined as haemoglobin concentration of less than 5g/dL, *Adjusted for ethnic group, HbS, age at recruitment and α-thalassaemia genotype. Children with SCD were excluded from the current analysis.

In Chapter 5, I showed that β-thalassaemia was associated with changes in RBC indices, specifically with reduced MCV and MCH, and that coinheritance of α-thalassaemia reversed this trend in a dose dependent manner resulting in an increase in MCH levels similar to those observed in individuals without β-thalassaemia. As a result of these observations, I decided to

explore the effect of the coinheritance of both β-thalassaemia and α-thalassaemia on the incidence of anaemia.

Among β-thalassaemia carriers, the risk of admission to hospital with anaemia reduced with the loss of α-genes in a dose dependent manner. The difference was statistically significant for αα/αα with an aIRR of 3.4 (1.4-8.1) but did not reach significance for -α/αα and -α/-α where the aIRRs were 2.2 (1.0-4.8) and 1.1 (0.2-8.0) respectively (Table 21).

Table 21. Incidence of different levels of anaemia stratified by α-thalassaemia status.

	Non-β-thalassaemia		**HbA/β-thalassaemia**			
	Episodes	**Incidence (95%CI)**	**Episodes**	**Incidence (95%CI)**	**aIRR*** **(95%CI)**	**P**
Anaemia defined using different cut-offs						
Hb <5g/dL	72	1.1 (0.8, 1.3)	2	6.1 (1.5, 24.3)	**5.09 (1.23-21.0)**	**0.024**
Hb 5-6.9g/dL	126	1.9 (1.5, 2.2)	2	6.0 (1.5, 24.3)	3.03 (0.74-12.3)	0.121
Hb 7-9.9g/dL	1108	16.4 (15.5, 17.4)	11	33.4 (18.5, 60.4)	**1.97 (1.02-3.80)**	**0.042**
Anaemia defined as Hb <10g/dL						
Hb <10g/dL	1307	19.4 (18.3, 20.5)	15	45.6 (27.5, 75.7)	**2.25 (1.27-3.98)**	**0.005**
Anaemia defined as Hb <10g/dL stratified by a-thalassaemia genotypes						
αα/αα	434	19.2 (17.5, 21.1)	5	51.9 (21.6, 124)	**3.35 (1.38, 8.12)**	**0.007**
-α/αα	609	18.9 (17.4, 20.5)	9	49.0 (25.5, 94.2)	2.15 (0.96, 4.79)	0.061
-α/-α	238	20.8 (18.3, 23.6)	1	20.5 (2.8, 145)	1.11 (0.15, 7.96)	0.918

*Adjusted for ethnic group, HbS, age at recruitment. Apart from α-thalassaemia genotypes, the rest of the groups were also adjusted for α-thalassaemia status.

Clinical malaria was defined as fever in the presence of *P. falciparum* parasitaemia at any density in <1 year old children or at a density of >2500 parasites/μl in older children. The total number of malaria admissions recorded among β-thalassaemia carriers were just 2 episodes in 328 PYO, resulting in an incidence rate of 6.0 (95% CI, 1.5-24.3) / 1000 PYO. This was in comparison to non-carriers, with 228 admissions in 67,350 PYO resulting in an incidence rate of 3.4 (2.9-3.8)/ 1000 PYO (aIRR 1.4, 95% CI 0.3-5.5; p=0.67). Similarly, no significant difference was seen in incidence of malaria defined as blood film positivity (aIRR 1.3; 0.3-5.3) (Table 20). Finally, heterozygous β-thalassaemia was not associated with either an increased or reduced risk of a

range of other specific diseases, including pneumonia, meningitis, severe malnutrition, jaundice, and neonatal conditions (Table 20).

6.5 Discussion

In this Chapter I have explored the impact of β-thalassaemia trait on both under-five mortality and admission to hospital with malaria and other childhood diseases during the first five years of life. I found no significant differences in the risk of death, or admission to hospital with malaria or any other illnesses with the exception of a higher incidence of severe anaemia among β-thalassaemia carriers.

To the best of my knowledge, this is the first cohort study that has been undertaken to date that has investigated mortality among children with heterozygous β-thalassaemia in a malaria endemic region. After adjusting for HbS, α-thalassaemia, ethnicity, and age at recruitment, I found no significant difference in mortality between β-thalassaemia carriers and non-carriers. These results align with the observations made in two large studies conducted in Europe (Gallerani et al., 1990; Graffeo et al., 2017). In the prospective study by Gallerani and colleagues, conducted on Italian subjects, no difference in average life expectancy was found between β-thalassaemia carriers and non-carriers during the 6-year follow-up period. Similarly, Graffeo and colleagues reported similar observations in a 10-year retrospective study involving 4,943 β-thalassaemia carriers and 21,063 controls which was also conducted on Italian subjects.

Most previous studies have found that β-thalassaemia carriers live a relatively normal life without any complications, with the potential exception of an exacerbation of anaemia during pregnancy in females, which in most cases is also usually mild (Premawardhena et al., 2008; Tsatalas et al., 2009). Nevertheless, associations between heterozygous β-thalassaemia and selected clinical conditions including an increased risk of cholelithiasis, cirrhosis, mood disorders and kidney disease (Gardikas, 1968; Graffeo et al., 2017) and reduced myocardial infraction (Gallerani et al., 1990) have been documented in a number of individual studies. Finally, a population-based study undertaken in Sri Lanka, which was declared malaria-free in

2016 (Dharmawardena et al., 2022), found that β-thalassaemia carriers were more likely than controls to experience symptoms associated with mild anaemia and mild pyrexia (Premawardhena et al., 2008). However, as highlighted by Weatherall and Clegg (Weatherall & Clegg, 2001), most of these studies probably included subjects with a more severe phenotype or did not include normal controls. Importantly, all of these studies were conducted in non-malaria-endemic regions and may not necessarily be directly translatable to malaria-endemic countries where a different spectrum of tropical diseases exists.

In the current study I found the incidence of admission with anaemia of <10g/dl in β-thalassaemia carriers was twice that seen among controls and almost 5 times higher for the most severe category (<5g/dL). In the recall by genotype study involving 17 children from this cohort (see chapter 5), β-thalassaemia carriers exhibited significantly lower haemoglobin levels, with a mean of 11g/dL (SD=1.0) compared to HbAA controls with a mean of 12.3g/dL (SD=0.9). Due to their lower steady state haemoglobin levels, I hypothesise that the impact of any underlying trigger for anaemia might be more pronounced in β-thalassaemia carriers than in non-β-thalassaemic individuals. However, the resulting severe anaemia might not be severe enough to result in increased transfusion or mortality rates during the first 5 years of life. More investigations are required to identify these triggers for anaemia, which may include previously identified illnesses in this region such as malaria, hookworm and schistosomiasis (Kassebaum et al., 2014) and their potential role in causing severe anaemia in β-thalassaemia carriers.

Whereas previous reports have examined the effect of coinheritance of β-thalassaemia with α-thalassaemia on the more severe forms of β-thalassaemia, there are no studies investigating this effect in β-thalassaemia carriers. In the more severe forms of β-thalassaemia, including β-thalassaemia major (Furbetta et al., 1983; Galanello et al., 2009), β-thalassaemia intermedia (Galanello et al., 2009) and HbE/β-thalassaemia (Sripichai et al., 2008), the loss of one or more α-genes was reported to result in a less severe form of disease. This protective effect has been attributed with the reduction of free α-chains resulting in restoration of a balance in α/β-globin (Mettananda, Gibbons, & Higgs, 2015; Thein, 2018). In the current study I found that co-

inheritance of both β-thalassaemia and α-thalassaemia resulted in a reduced incidence of admission to hospital with anaemia when compared to β-thalassaemia alone. This reduction may be attributed to the effect of coinheritance of α-thalassaemia on haemoglobin levels in β-thalassaemia carriers. In Chapter 5 I observed that an increasing loss of α-chains was associated with an increase in MCH in a dose dependent manner, reaching levels that were similar to those observed in individuals without β-thalassaemia. Interestingly, the aIRR for admission with anaemia (<10g/dL) in the current study also reduced in a dose dependent manner with increasing loss of α-globin chains from 3, 2 and 1 for -α/-α, -α/αα and αα/αα, genotypes respectively. The increase in baseline haemoglobin levels in carriers of both α- and β-thalassaemia, may contribute to the attenuation of anaemia severity when triggered by underlying condition. Further investigation is required to confirm the observed findings and their potential impact on severe malaria anaemia.

Despite strong evidence that β-thalassaemia might protect against malaria based on its geospatial distribution, most of the direct evidence in support of this hypothesis has come from laboratory based in vitro studies, with only one case control study having been conducted to date (M. Willcox et al., 1983). In this study, the investigators found that β-thalassaemia was associated with 50% protection against clinical malaria. In our cohort study, we did not find any evidence for a protective effect of β-thalassaemia against hospital admission with clinical malaria or malaria defined as blood slide positive. There are a number of differences that could potentially explain the discrepancies between the two studies. While Willcox and colleagues stratified their analysis based on age, tribe and residence to control for confounding, we used a multivariate analysis to not only control for sickle cell anaemia, age, and sex but also α-thalassaemia, which, as observed in Chapter 5 could affect malaria parasite invasion. Moreover, our study used a birth cohort approach, resulting to better matched controls than Willcox et al study. However, one of the limitations of our study was that the follow-up was done at a time when there was decline in malaria, which explains the few malaria cases observed. Furthermore, the small number of β-thalassaemia carriers was a major weakness, resulting in a very limited number of significant

events. In particular, considering that in Chapter 5 we observed that β-thalassaemia has a subtle effect on malaria parasite invasion, it is possible that such an effect might be difficult to detect given the limited number of malaria events observed in the current study. As a result, further research is needed to validate these observations, preferably in a region with higher malaria transmission.

In conclusion, in the current study, I found no evidence to suggest that heterozygous β-thalassaemia confers either a significant health advantage or disadvantage in our setting on the coast of Kenya. While I did see an increased risk of admission to hospital with severe anaemia among β-thalassaemia carriers, this observation was leveraged on only 2 admissions within the β-thalassaemia group and may simply be related to their lower haemoglobin levels at steady state. This conclusion is supported by the fact that the increased risk of severe anaemia declined with the coinheritance of α-thalassaemia, a condition that reverses the negative effect of β-thalassaemia on steady state haemoglobin concentrations.

Chapter 7: Concluding remarks and recommendations

Before initiating my investigations in 2018, there was little known about β-thalassaemia in the Kilifi region or in much of sub-Saharan Africa. Our initial investigations confirmed elevated HbA_2 levels in members of a birth cohort study that suggested the presence of β-thalassaemia within the region (Macharia et al., 2019), which was followed by the discovery of two mutations (rs33959855 G>T and the rs193922563 25bp deletion) among children involved in a separate study of sickle cell disease (McGann et al., 2018). Through this thesis, I investigated these observations further by studying the prevalence, genetic types, and origins of β-thalassaemia in Kilifi. Additionally, I explored the malaria protective mechanisms associated with these mutations and examined their wider health consequences in an epidemiological study. In this final chapter, I summarise my findings and their significance before outlining their implications and considering future studies.

7.1 Four mutations defining β-thalassaemia in Kilifi: prevalence, incidence, and implications in diagnosis

Firstly, my research revealed that β-thalassaemia is present in Kilifi at a minimum allele frequency of 0.3% with an incidence of homozygosity of approximately 1 in 100,000. Secondly, I identified two new mutations (rs33941849 T>C and rs35004220 G>A), bringing the total number β-thalassaemia mutations in Kilifi to four. Thirdly, I found that the presence of HbS is associated with a high false positive rate when using HPLC determined HbA_2 cut-off values for the diagnosis of β-thalassaemia. These findings have implications for the diagnosis of β-thalassaemia in regions such as sub-Saharan Africa, where HbS is common. In such regions, when using HPLC determined HbA_2 values in the diagnosis of β-thalassaemia carriers, the participants should first be grouped according to the presence or absence of HbS into those

with HbAA, HbAS or HbSS. The diagnosis of HbA/β-thalassaemia, HbS/β^+-thalassaemia and HbS/β^0-thalassaemia can then be approached as follows.

1. HbA/β-thalassaemia carriers can then be diagnosed through sequencing participants with an HbA_2 value of >4% among HbAA participants, although it should be noted that a few cases occur within the range of 3.5% to 3.9% range.
2. HbS/β^+-thalassaemia carriers can be diagnosed through sequencing of samples from HbAS participants in whom the concentration of HbA exceeds that of HbS.
3. HbS/β^0-thalassaemia can be diagnosed by first genotyping the HbS allele (rs334) in HPLC-defined HbSS individuals, followed by sequencing for β-thalassaemia in participants whose PCR genotype at the rs334 allele is heterozygous.

Where the causative mutations have already been established, such as in Kilifi, β-thalassaemia sequencing can be replaced with more cost-effective PCR-based genotyping methods that target the four mutations defined in this thesis and sequencing can then be considered only for samples that test negative by these methods.

7.2 Possible diverse origins of β-thalassaemia mutations in Kilifi

After identifying these mutations, my next inquiry was focused on investigating the origin of these mutations. Through literature review, ethnolinguistic and haplotype analysis I found that these mutations might either have been introduced to this populations through geneflow or have emerged de novo. Notably, evidence on rs33959855 suggests the mutation might have been introduced on a haplotype that is more common in South Asian populations than the local populations. The mutation was also dominantly found in the Chonyi ethnolinguistic group which has previously been identified as the earliest route of Asian ancestry into Africa dating back to approximately at 1138CE (Busby et al., 2016). Based on the evidence from this analysis, I hypothesised that rs33959855 might have been introduced to Kilifi around 1200CE during the

Asia-Swahili trade took place along the East African Coast (Allen, 1993; Brielle et al., 2023) and is currently undergoing admixture with haplotypes from the indigenous populations. However, I was unable to determine the origin of the other 3 mutations as they were found on the most common haplotype present in both the indigenous and other populations outside of Kenya.

To study the origins of these mutations more definitively, it would be necessary to sequence a larger region capturing more sequence diversity. Nevertheless, the outcome of this analysis indicates potential regions where β-thalassaemia might be found. These regions include countries that were in contact with the Asian populations during the Asian-Swahili trade around 1200CE. These regions encompasses all the countries and islands along the coast between Kiwaiyu near Lamu to Mozambique (Allen, 1993). In these regions therefore, investigations of β-thalassaemia should be considered in children presenting with anaemia, after ruling out other causes.

7.3 β-thalassaemia has a modest effect on invasion with no impact on clinical malaria and mortality

After identifying the mutations present in this population and their possible origins, I next wanted to investigate whether malaria might be responsible for maintaining these deleterious mutations in the population through balancing selection. I used laboratory-based RBCs functional-assays to investigate the mechanisms of invasion, growth, and enhanced phagocytosis of ring-parasitised RBCs. Additional I conducted a survival analysis to investigate the impact of β-thalassaemia on under 5 mortality and admission to hospital with malaria and other childhood illnesses. I found that β-thalassaemia carriers have a modest resistance to invasion by 3D7 parasites and that this might be mediated through a balance between receptors that were differentially expressed in β-thalassaemia carriers, some that likely result in increased invasion namely increased basign expression and others that could result in reduce invasion namely reduced expression of CD71 and CD49d. Additionally, success of invasion might also be influenced by other factors such as coinheritance of β-thalassaemia with α-thalassaemia that

seems to reverse the protective effect β-thalassaemia on invasion. Furthermore, other factors not investigated in this study such as biophysical properties of the RBCs might also play a role (Introini et al., 2022). Whatever the mechanism, in our setting β-thalassaemia does not seem to offer any added advantage against mortality or admission to hospital with clinical malaria or malaria defined as blood slide positive during the first 5 years of life. The lack of a difference in clinical outcomes despite the observed reduction in invasion of β-thalassaemia RBCs, might be attributed to the modest resistance of β-thalassaemia against malaria. Unlike other polymorphisms, such as HbAS and Dantu (Kariuki et al., 2020; Uyoga, Macharia, Ndila, et al., 2019), which provide more than 70% protection against clinical malaria, the modest resistance offered by β-thalassaemia may be reduced further or cancelled out by opposing factors such as differential expression of RBC membrane proteins and epistasis interaction with other protective polymorphisms. Consequently, detecting this modest effect on clinical malaria becomes challenging more so with reduced incidence of malaria as observed during the study period of this cohort.

Although I observed that β-thalassaemia does not confer any protection against clinical malaria, it is important to note that this does not rule out the possibility that β-thalassaemia may still offer some level of protection against clinical malaria. One of the limitations of our cohort studies was that it was conducted at a period when there was a decline in malaria prevalence (O'Meara et al., 2008) resulting in few cases of malaria events. This would make it difficult to detect a modest protective effect such as that identified through the recall-by-genotype study. This observation thus needs to be replicated in other cohorts in a region with higher prevalence of clinical malaria. More laboratory-based studies towards understanding the mechanisms of protection also need to be conducted. These studies should not only focus on invasion and growth but also look at other mechanisms that we did not investigate in this study such as enhanced clearance of infected β-thalassaemia RBC via opsonic phagocytosis and protection against severe malaria through interference of rosetting and cytoadhesion. To prevent variation in results more sensitive methods should be used and the cases matched as close as possible to

controls. As I have demonstrated in the current study careful matching of controls to cases is important as coinheritance of other malaria protective polymorphisms such as α-thalassaemia may results in an epistatic effect.

7.4 Health consequences associated with β-thalassaemia mutations in Kilifi

The last question I have addressed through this thesis relates to the health consequences of β-thalassaemia in Kilifi. Firstly, I found that β-thalassaemia accounts for 10% of SCD cases by causing HbS/β^0-thalassaemia. This is a condition that arises when one chromosome is affected by a sickle mutation and the contralateral chromosome is affected by a β^0-thalassaemia mutation. Although the clinical manifestations of HbS/β^0-thalassaemia are similar to those observed in individuals with HbSS, they can vary depending on the type of β-thalassaemia that is co-inherited (Serjeant et al., 1979; Yadav et al., 2016). In the recall by genotype study, I observed among individuals with HbA/β^0-thalassaemia these mutations result in a significant variation in RBC indices. It is possible that these mutations have a similar impact in individuals with HbS/β^0-thalassaemia. Future work should focus in understanding how these changes influence disease severity, which could help in better patient management.

Secondly, I have estimated the incidence of β-thalassaemia major to be 1 in 100,000. I also present, to the best of my knowledge, the first indigenous case of β-thalassaemia major to be reported from sub-Saharan Africa. The child in this report had an abnormal HPLC chromatogram and upon sequencing I found the child had β-thalassaemia major caused by homozygosity for the rs33941849 mutation. Because clinicians were initially unaware of the existence of this condition within this region, the child had been clinically diagnosed and managed as a case of SCD. The two conditions are however very different, as cases of β-thalassaemia major do not produce any β-globin and require life-long transfusion for survival (Weatherall & Clegg, 2001). This finding underscores the need to raise awareness about the existence of β-thalassaemia major within this region. Due to limited availability of diagnostic facilities many cases may be misdiagnosed as either SCD or iron deficiency anaemia, as was the

case with this child. It is important to develop local guidelines for early diagnosis and appropriate management of β-thalassaemia major cases. These measures could help improve outcomes of individuals affected by β-thalassaemia major.

In my final findings, I discovered that at steady state β-thalassaemia carriers (HbA/β-thalassaemia) had lower haemoglobin levels than normal children. However, when individuals experienced an illness requiring hospital admission, they were 5-times more likely to have severe anaemia compared to children without β-thalassaemia during the first 5 years of life. Interestingly, coinheritance of α-thalassaemia appeared to reverse this effect in a dose dependent manner. β-thalassaemia carriers who were also homozygous for α-thalassaemia had an incidence of severe anaemia that was equivalent to that of non-β-thalassaemia carriers. This suggests that coinheritance of α-thalassaemia might have a protective effect against severe anaemia in β-thalassaemia carriers. Previous studies have mainly been conducted in regions where malaria is absent (Gallerani et al., 1990; Graffeo et al., 2017; Weatherall & Clegg, 2001). In sub-Saharan Africa, however, the underlying causes for severe anaemia are probably very different, and include iron deficiency anaemia, malaria, hookworm and schistosomiasis (Kassebaum et al., 2014). Nevertheless, in our setting, the resulting severe anaemia was not associated with increased under-5 mortality. These findings have significant implications on clinical practise, not only in Kilifi but also in the wider East African region where these mutations may also be present but have not yet been properly documented. It is important to conduct further investigations in other malaria-endemic regions to validate our observations regarding susceptibility of β-thalassaemia carriers to developing severe anaemia. The investigations should prioritize the understanding of the underlying causes of severe anaemia and work towards developing prevention strategies.

In conclusion, β-thalassaemia is present in Kilifi albeit at a low prevalence and is characterised by four different mutations that are likely to be of different origins. It has significant health consequences including its contribution to sickle cell disease cases, β-thalassaemia major, and an increased risk of severe anaemia in carriers. However, β-thalassaemia does not appear to play

a major role in protection against malaria in this community. These findings have implications for clinical practice, emphasising the need for increased awareness and establishment of local guidelines for diagnosis and management of β-thalassaemia. Further research in malaria endemic regions is necessary to validate these findings.

7.5 Future directions based on findings from this thesis include

- Raising awareness and developing guidelines: Since the publication reports from this work (Macharia et al., 2022; Macharia et al., 2020), we have received enquiries from clinicians in various hospitals along the coast of Kenya seeking guidance on diagnosis of suspected cases of β-thalassaemia and on how they can manage the condition. This positive outcome aligns with the intended goal of this thesis which is to raise awareness of β-thalassaemia as a public health problem and ultimately lead to the development of locally appropriate diagnostic and treatment guidelines.
- Exploring the prevalence β-thalassaemia in other regions: It is possible that β-thalassaemia is also present in other regions within sub-Saharan Africa, and that as in this study, the children from these regions are misdiagnosed with other conditions such as SCD or iron deficiency anaemia. Further research should include identifying other regions where these mutations could be found. This thesis provides valuable clues for potential locations, where these mutations are likely to be found including regions that were involved in the Asian-Swahili trade in 1200CE.
- Understanding susceptibility of β-thalassaemia carriers to severe anaemia: Further investigations should be conducted to understand the observation made in this study on susceptibility of β-thalassaemia carriers to severe anaemia. This investigation should prioritise underlying causes of severe anaemia and work towards developing prevention strategies. In regions where malaria is present it would be important to understand impact of malaria on the development of severe anaemia in β-thalassaemia carriers.

- Investigating the relationship between β-thalassaemia and malaria: While laboratory and clinical investigations findings of this thesis suggest subtle effect on invasion with no effect on clinical malaria it is essential that similar investigations be conducted in a region where malaria is endemic as the study was conducted during a period of decline in malaria prevalence. Additionally, more laboratory investigations are necessary to not only confirm the mechanisms of invasion but also to investigate other mechanisms including enhanced clearance of infected β-thalassaemia RBC via opsonic phagocytosis and protection against severe malaria through interference of rosetting and cytoadhesion.

By undertaking these steps, we can enhance our understanding of β-thalassaemia in sub-Saharan Africa and provide locally-based interventions that could lead to improved healthcare outcomes for affected individuals.

REFERENCES

Abdaoui, W., Benouareth, D. E., Djenouni, A., Renoux, C., Grifi, F., Gouri, A., . . . Joly, P. (2019). Genetic Background of beta-Thalassemia in Northeast Algeria with Assessment of the Thalassemia Severity Score and Description of a new beta(0)-Thalassemia Frameshift Mutation (HBB: c.374dup; p.Pro126Thrfs*15). *Hemoglobin, 43*(4-5), 223-228. doi:10.1080/03630269.2019.1675689

Abuzenadah, A. M., Hussein, I. M., Damanhouri, G. A., FM, A. S., Gari, M. A., Chaudhary, A. G., . . . 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. doi:10.3109/03630269.2011.588508

Adekile, A. D., Azab, A. F., Al-Sharida, S. I., Al-Nafisi, B. A., Akbulut, N., Marouf, R. A., & Mustafa, N. Y. (2015). Clinical and Molecular Characteristics of Non-Transfusion-Dependent Thalassemia in Kuwait. *Hemoglobin, 39*(5), 320-326. doi:10.3109/03630269.2015.1053489

Adekile, A. D., Gu, L. H., Baysal, E., Haider, M. Z., al-Fuzae, L., Aboobacker, K. C., . . . Huisman, T. H. (1994). Molecular characterization of alpha-thalassemia determinants, beta-thalassemia alleles, and beta S haplotypes among Kuwaiti Arabs. *Acta Haematol, 92*(4), 176-181. doi:10.1159/000204216

Agouti, I., Badens, C., Abouyoub, A., Khattab, M., Sayah, F., Barakat, A., & Bennani, M. (2007). Genotypic correlation between six common beta-thalassemia mutations and the XmnI polymorphism in the Moroccan population. *Hemoglobin, 31*(2), 141-149. doi:10.1080/03630260701285050

Agouti, I., Badens, C., Abouyoub, A., Levy, N., & Bennani, M. (2008). Molecular basis of beta-thalassemia in Morocco: possible origins of the molecular heterogeneity. *Genet Test, 12*(4), 563-568. doi:10.1089/gte.2008.0058

Akhavan-Niaki, H., Derakhshandeh-Peykar, P., Banihashemi, A., Mostafazadeh, A., Asghari, B., Ahmadifard, M. R., . . . Elmi, M. M. (2011). A comprehensive molecular characterization of beta thalassemia in a highly heterogeneous population. *Blood Cells Mol Dis, 47*(1), 29-32. doi:10.1016/j.bcmd.2011.03.005

Al-Allawi, N. A., Al-Mousawi, B. M., Badi, A. I., & Jalal, S. D. (2013). The spectrum of beta-thalassemia mutations in Baghdad, Central Iraq. *Hemoglobin, 37*(5), 444-453. doi:10.3109/03630269.2013.810641

Al-Allawi, N. A., Jubrael, J. M., & Hughson, M. (2006). Molecular characterization of beta-thalassemia in the Dohuk region of Iraq. *Hemoglobin, 30*(4), 479-486. doi:10.1080/03630260600868097

Alberts, B. (2008). Molecular Biology of the Cell (5th ed.). W.W. Norton & Company. . doi:https://doi.org/10.1201/9781315735368

Aliyeva, G., Asadov, C., Mammadova, T., Gafarova, S., Guliyeva, Y., & Abdulalimov, E. (2020). Molecular and geographical heterogeneity of hemoglobinopathy mutations in Azerbaijanian populations. *Ann Hum Genet, 84*(3), 249-258. doi:10.1111/ahg.12367

Aljurf, M., Ma, L., Angelucci, E., Lucarelli, G., Snyder, L. M., Kiefer, C. R., . . . Schrier, S. L. (1996). Abnormal assembly of membrane proteins in erythroid progenitors of patients with beta-thalassemia major. *Blood, 87*(5), 2049-2056. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8634456

Allen, J. d. V. (1993). Swahili origins: Swahili culture and the Shungwaya phenomenon. *James Currey Publishers*.

Altay, C. (2002). The Frequency and Distribution Pattern of ss-Thalassemia Mutations in Turkey. *Turk J Haematol, 19*(2), 309-315. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/27264774

Amin, S. S., Jalal, S. D., Ali, K. M., Mohammed, A. I., Rasool, L. K., & Osman, T. J. (2020). Beta-Thalassemia Intermedia: A Single Thalassemia Center Experience from Northeastern Iraq. *Biomed Res Int, 2020*, 2807120. doi:10.1155/2020/2807120

Angel, J. L. (1966). Porotic hyperostosis, anemias, malarias, and marshes in the prehistoric Eastern Mediterranean. *Science, 153*(3737), 760-763. doi:10.1126/science.153.3737.760

Arpaci, A., Gul, B. U., Ozcan, O., Ilhan, G., El, C., Dirican, E., . . . Kaya, H. (2021). Presentation of two new mutations in the 3'untranslated region of the beta-globin gene and evaluating the molecular spectrum of thalassemia mutations in the Mediterranean region of Turkey. *Ann Hematol, 100*(6), 1429-1438. doi:10.1007/s00277-021-04509-9

Auer, P. L., Reiner, A. P., Wang, G., Kang, H. M., Abecasis, G. R., Altshuler, D., . . . Leal, S. M. (2016). Guidelines for Large-Scale Sequence-Based Complex Trait Association Studies: Lessons Learned from the NHLBI Exome Sequencing Project. *Am J Hum Genet, 99*(4), 791-801. doi:10.1016/j.ajhg.2016.08.012

Ayi, K., Turrini, F., Piga, A., & Arese, P. (2004). Enhanced phagocytosis of ring-parasitized mutant erythrocytes: a common mechanism that may explain protection against falciparum malaria in sickle trait and beta-thalassemia trait. *Blood, 104*(10), 3364-3371. doi:10.1182/blood-2003-11-3820

Babameto-Laku, A., Mitre, A., Berisha, S., Mokini, V., & Roko, D. (2011). Molecular Genetic Characterization of beta-Thalassemia and Sickle Cell Syndrome in the Albanian Population. *Balkan J Med Genet, 14*(1), 45-50. doi:10.2478/v10034-011-0017-0

Baghernajad-Salehi, L., D'Apice, M. R., Babameto-Laku, A., Biancolella, M., Mitre, A., Russo, S., . . . Novelli, G. (2009). A pilot beta-thalassaemia screening program in the Albanian population for a health planning program. *Acta Haematol, 121*(4), 234-238. doi:10.1159/000226423

Bain, B. J. (2020). Haemoglobin and the genetics of haemoglobin synthesis. In *Haemoglobinopathy Diagnosis* (pp. 1-29).

Basak, A. N., Ozcelik, H., Ozer, A., Tolun, A., Aksoy, M., Agaoglu, L., . . . et al. (1992). The molecular basis of beta-thalassemia in Turkey. *Hum Genet, 89*(3), 315-318. doi:10.1007/BF00220549

Baum, J., Maier, A. G., Good, R. T., Simpson, K. M., & Cowman, A. F. (2005). Invasion by P. falciparum merozoites suggests a hierarchy of molecular interactions. *PLoS Pathog, 1*(4), e37. doi:10.1371/journal.ppat.0010037

Baysal, E., Indrak, K., Bozkurt, G., Berkalp, A., Aritkan, E., Old, J. M., . . . et al. (1992). The beta-thalassaemia mutations in the population of Cyprus. *Br J Haematol, 81*(4), 607-609. doi:10.1111/j.1365-2141.1992.tb03000.x

Beguin, Y., Lampertz, S., De Groote, D., Igot, D., Malaise, M., & Fillet, G. (1993). Soluble CD23 and other receptors (CD4, CD8, CD25, CD71) in serum of patients with chronic lymphocytic leukemia. *Leukemia, 7*(12), 2019-2025. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8255102

Belmokhtar, I., Lhousni, S., Elidrissi Errahhali, M., Ghanam, A., Elidrissi Errahhali, M., Sidqi, Z., . . . Benajiba, N. (2022). Molecular heterogeneity of beta-thalassemia variants in the Eastern region of Morocco. *Mol Genet Genomic Med, 10*(8), e1970. doi:10.1002/mgg3.1970

Bennani, C., Tamouza, R., Rouabhi, F., Benabadji, M., Malou, M., Elion, J., . . . Beldjord, C. (1993). The spectrum of beta-thalassaemia in Algeria: possible origins of the molecular heterogeneity and a tentative diagnostic strategy. *Br J Haematol, 84*(2), 335-337. doi:10.1111/j.1365-2141.1993.tb03075.x

Beris, P., Darbellay, R., Speiser, D., Kirchner, V., & Miescher, P. A. (1993). De novo initiation codon mutation (ATG-->ACG) of the beta-globin gene causing beta-thalassemia in a Swiss family. *Am J Hematol, 42*(3), 248-253. doi:10.1002/ajh.2830420303

Berniell-Lee, G., Plaza, S., Bosch, E., Calafell, F., Jourdan, E., Cesari, M., . . . Comas, D. (2008). Admixture and sexual bias in the population settlement of La Reunion Island (Indian Ocean). *Am J Phys Anthropol, 136*(1), 100-107. doi:10.1002/ajpa.20783

Bilgen, T., Arikan, Y., Canatan, D., Yesilipek, A., & Keser, I. (2011). The association between intragenic SNP haplotypes and mutations of the beta globin gene in a Turkish population. *Blood Cells Mol Dis, 46*(3), 226-229. doi:10.1016/j.bcmd.2011.01.004

Borges, E., Tchonhi, C., Couto, C. S. B., Gomes, V., Amorim, A., Prata, M. J., & Brito, M. (2019). Unusual beta-Globin Haplotype Distribution in Newborns from Bengo, Angola. *Hemoglobin, 43*(3), 149-154. doi:10.1080/03630269.2019.1647230

Boudrahem-Addour, N., Zidani, N., Carion, N., Labie, D., Belhani, M., & Beldjord, C. (2009). Molecular heterogeneity of beta-thalassemia in Algeria: how to face up to a major health problem. *Hemoglobin, 33*(1), 24-36. doi:10.1080/03630260802626061

Bouhass, R., Perrin, P., & Trabuchet, G. (1994). The spectrum of beta-thalassemia mutations in the Oran region of Algeria. *Hemoglobin, 18*(3), 211-219. doi:10.3109/03630269409043621

Bravo-Urquiola, M., Arends, A., Gomez, G., Montilla, S., Gerard, N., Chacin, M., . . . Krishnamoorthy, R. (2012). Molecular spectrum of beta-thalassemia mutations in the admixed Venezuelan population, and their linkage to beta-globin gene haplotypes. *Hemoglobin, 36*(3), 209-218. doi:10.3109/03630269.2012.674997

Brielle, E. S., Fleisher, J., Wynne-Jones, S., Sirak, K., Broomandkhoshbacht, N., Callan, K., . . . Kusimba, C. M. (2023). Entwined African and Asian genetic roots of medieval peoples of the Swahili coast. *Nature, 615*(7954), 866-873. doi:10.1038/s41586-023-05754-w

Brockelman, C. R., Wongsattayanont, B., Tan-ariya, P., & Fucharoen, S. (1987). Thalassemic erythrocytes inhibit in vitro growth of Plasmodium falciparum. *J Clin Microbiol, 25*(1), 56-60. doi:10.1128/jcm.25.1.56-60.1987

Busby, G. B., Band, G., Si Le, Q., Jallow, M., Bougama, E., Mangano, V. D., . . . Malaria Genomic Epidemiology, N. (2016). Admixture into and within sub-Saharan Africa. *Elife, 5*. doi:10.7554/eLife.15266

Cannata, M., Cassara, F., Vinciguerra, M., Licari, P., Passarello, C., Leto, F., . . . Giambona, A. (2019). Double Heterozygosity for Hb Durham-N.C. (HBB: c.344T>C) [beta114(G16)Leu-->Pro] and the IVS-I-110 (HBB: c.93-21G>A) Causing a Severe beta-Thalassemia Phenotype. *Hemoglobin, 43*(3), 210-213. doi:10.1080/03630269.2019.1655030

Carlson, J., Nash, G. B., Gabutti, V., al-Yaman, F., & Wahlgren, M. (1994). Natural protection against severe Plasmodium falciparum malaria due to impaired rosette formation. *Blood, 84*(11), 3909-3914. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/7949147

Carrocini, G. C. S., Venancio, L. P. R., Pessoa, V. L. R., Lobo, C. L. C., & Bonini-Domingos, C. R. (2017). Mutational Profile of Homozygous beta-Thalassemia in Rio de Janeiro, Brazil. *Hemoglobin, 41*(1), 12-15. doi:10.1080/03630269.2017.1289958

Cherry, L., Calo, C., Talmaci, R., Perrin, P., & Gavrila, L. (2016). beta-Thalassemia Haplotypes in Romania in the Context of Genetic Mixing in the Mediterranean Area. *Hemoglobin, 40*(2), 85-96. doi:10.3109/03630269.2015.1124113

Choi, Y., Chan, A. P., Kirkness, E., Telenti, A., & Schork, N. J. (2018). Comparison of phasing strategies for whole human genomes. *PLoS Genet, 14*(4), e1007308. doi:10.1371/journal.pgen.1007308

Chonat, S., & Quinn, C. T. (2017). Current Standards of Care and Long Term Outcomes for Thalassemia and Sickle Cell Disease. *Adv Exp Med Biol, 1013*, 59-87. doi:10.1007/978-1-4939-7299-9_3

Chouk, I., Daoud, B. B., Mellouli, F., Bejaoui, M., Gerard, N., Dellagi, K., & Abbes, S. (2004). Contribution to the description of the beta-thalassemia spectrum in Tunisia and the origin of mutation diversity. *Hemoglobin, 28*(3), 189-195. doi:10.1081/hem-120040305

Clark, B. E., & Thein, S. L. (2004). Molecular diagnosis of haemoglobin disorders. *Clin Lab Haematol, 26*(3), 159-176. doi:10.1111/j.1365-2257.2004.00607.x

Codrington, J. F., Li, H. W., Kutlar, F., Gu, L. H., Ramachandran, M., & Huisman, T. H. (1990). Observations on the levels of Hb A2 in patients with different beta-thalassemia mutations and a delta chain variant. *Blood, 76*(6), 1246-1249. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1698102

Colaco, S., & Nadkarni, A. (2021). Borderline HbA2 levels: Dilemma in diagnosis of beta-thalassemia carriers. *Mutat Res Rev Mutat Res, 788*, 108387. doi:10.1016/j.mrrev.2021.108387

Cowman, A. F., & Crabb, B. S. (2006). Invasion of red blood cells by malaria parasites. *Cell, 124*(4), 755-766. doi:10.1016/j.cell.2006.02.006

Crosnier, C., Bustamante, L. Y., Bartholdson, S. J., Bei, A. K., Theron, M., Uchikawa, M., . . . Wright, G. J. (2011). Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum. *Nature, 480*(7378), 534-537. doi:10.1038/nature10606

Curuk, M. A., Yuregir, G. T., Asadov, C. D., Dadasova, T., Gu, L. H., Baysal, E., . . . Huisman, T. H. (1992). Molecular characterization of beta-thalassemia in Azerbaijan. *Hum Genet, 90*(4), 417-419. doi:10.1007/BF00220470

da Silveira, Z. M., das Vitorias Barbosa, M., de Medeiros Fernandes, T. A., Kimura, E. M., Costa, F. F., de Fatima Sonati, M., . . . de Medeiros, T. M. (2011). Characterization of beta-thalassemia mutations in patients from the state of Rio Grande do Norte, Brazil. *Genet Mol Biol, 34*(3), 425-428. doi:10.1590/S1415-47572011005000032

Das, S., Hertrich, N., Perrin, A. J., Withers-Martinez, C., Collins, C. R., Jones, M. L., . . . Blackman, M. J. (2015). Processing of Plasmodium falciparum Merozoite Surface Protein MSP1 Activates a Spectrin-Binding Function Enabling Parasite Egress from RBCs. *Cell Host Microbe, 18*(4), 433-444. doi:10.1016/j.chom.2015.09.007

De Sanctis, V., Kattamis, C., Canatan, D., Soliman, A. T., Elsedfy, H., Karimi, M., . . . Angastiniotis, M. (2017). beta-Thalassemia Distribution in the Old World: an Ancient Disease Seen from a Historical Standpoint. *Mediterr J Hematol Infect Dis, 9*(1), e2017018. doi:10.4084/MJHID.2017.018

Derakhshandeh-Peykar, P., Akhavan-Niaki, H., Tamaddoni, A., Ghawidel-Parsa, S., Naieni, K. H., Rahmani, M., . . . Farhud, D. D. (2007). Distribution of beta-thalassemia mutations in the northern provinces of Iran. *Hemoglobin, 31*(3), 351-356. doi:10.1080/03630260701462030

Dharmawardena, P., Premaratne, R., Wickremasinghe, R., Mendis, K., & Fernando, D. (2022). Epidemiological profile of imported malaria cases in the prevention of reestablishment phase in Sri Lanka. *Pathog Glob Health, 116*(1), 38-46. doi:10.1080/20477724.2021.1951556

Dimovski, A., Efremov, D. G., Jankovic, L., Juricic, D., Zisovski, N., Stojanovski, N., . . . et al. (1990). Beta-thalassemia in Yugoslavia. *Hemoglobin, 14*(1), 15-24. doi:10.3109/03630269009002251

Douzi, K., Moumni, I., Zorai, A., Ben Mustapha, M., Ben Mansour, I. M., Dorra, C., & Salem, A. (2015). Two new beta+ -thalassemia mutation [beta -56 (G --> C); HBBc. -106 G --> C] and [beta -83 (G --> A); HBBc. -133 G --> A] described among the Tunisian population. *Am J Hum Biol, 27*(5), 716-719. doi:10.1002/ajhb.22695

Ebel, E. R., Kuypers, F. A., Lin, C., Petrov, D. A., & Egan, E. S. (2021). Common host variation drives malaria parasite fitness in healthy human red cells. *Elife, 10*. doi:10.7554/eLife.69808

Economou, E. P., Antonarakis, S. E., Dowling, C. C., Ibarra, B., de la Mora, E., & Kazazian, H. H., Jr. (1991). Molecular heterogeneity of beta-thalassemia in mestizo Mexicans. *Genomics, 11*(2), 474. doi:10.1016/0888-7543(91)90161-7

Edison, E. S., Shaji, R. V., Devi, S. G., Moses, A., Viswabandhya, A., Mathews, V., . . . Chandy, M. (2008). Analysis of beta globin mutations in the Indian population: presence of rare and novel mutations and region-wise heterogeneity. *Clin Genet, 73*(4), 331-337. doi:10.1111/j.1399-0004.2008.00973.x

Eissa, A. A., Kashmoola, M. A., Atroshi, S. D., & Al-Allawi, N. A. (2015). Molecular Characterization of beta-Thalassemia in Nineveh Province Illustrates the Relative Heterogeneity of Mutation Distributions in Northern Iraq. *Indian J Hematol Blood Transfus, 31*(2), 213-217. doi:10.1007/s12288-014-0369-1

el-Hazmi, M. A., al-Swailem, A. R., & Warsy, A. S. (1995). Molecular defects in beta-thalassaemias in the population of Saudi Arabia. *Hum Hered, 45*(5), 278-285. doi:10.1159/000154314

el-Kalla, S., & Mathews, A. R. (1997). A significant beta-thalassemia heterogeneity in the United Arab Emirates. *Hemoglobin, 21*(3), 237-247. doi:10.3109/03630269708997384

El-Latif, M. A., Filon, D., Rund, D., Oppenheim, A., & Kanaan, M. (2002). The beta+-IVS-I-6 (T-->C) mutation accounts for half of the thalassemia chromosomes in the Palestinian populations of the mountain regions. *Hemoglobin, 26*(1), 33-40. doi:10.1081/hem-120002938

El-Shanshory, M., Hagag, A., Shebl, S., Badria, I., Abd Elhameed, A., Abd El-Bar, E., . . . Sharaf, E. (2014). Spectrum of Beta Globin Gene Mutations in Egyptian Children with beta-Thalassemia. *Mediterr J Hematol Infect Dis, 6*(1), e2014071. doi:10.4084/MJHID.2014.071

Elmezayen, A. D., Kotb, S. M., Sadek, N. A., & Abdalla, E. M. (2015). beta-Globin Mutations in Egyptian Patients With beta-Thalassemia. *Lab Med, 46*(1), 8-13. doi:10.1309/LM1AYKG6VE8MLPHG

Esan, G. J. (1970). The thalassaemia syndromes in Nigeria. *Br J Haematol, 19*(1), 47-56. doi:10.1111/j.1365-2141.1970.tb01600.x

Etyang, A. O., Munge, K., Bunyasi, E. W., Matata, L., Ndila, C., Kapesa, S., . . . Scott, J. A. (2014). Burden of disease in adults admitted to hospital in a rural region of coastal Kenya: an analysis of data from linked clinical and demographic surveillance systems. *Lancet Glob Health, 2*(4), e216-224. doi:10.1016/S2214-109X(14)70023-3

Faraon, R., Daraghmah, M., Samarah, F., & Srour, M. A. (2019). Molecular characterization of beta-thalassemia intermedia in the West Bank, Palestine. *BMC Hematol, 19*, 4. doi:10.1186/s12878-019-0135-6

Farra, C., Abdouni, L., Souaid, M., Awwad, J., Yazbeck, N., & Abboud, M. (2021). The Spectrum of beta-Thalassemia Mutations in the Population Migration in Lebanon: A 6-Year Retrospective Study. *Hemoglobin, 45*(6), 365-370. doi:10.1080/03630269.2021.1920975

Fattoum, S., Messaoud, T., & Bibi, A. (2004). Molecular basis of beta-thalassemia in the population of Tunisia. *Hemoglobin, 28*(3), 177-187. doi:10.1081/hem-120040307

Fettah, A., Bayram, C., Yarali, N., Isik, P., Kara, A., Culha, V., & Tunc, B. (2013). Beta-globin Gene Mutations in Turkish Children with Beta-Thalassemia: Results from a Single Center Study. *Mediterr J Hematol Infect Dis, 5*(1), e2013055. doi:10.4084/MJHID.2013.055

Fonseca, S. F., Kerbauy, J., Escrivao, C., Figueiredo, M. S., Cancado, R., Arruda, V. R., . . . Costa, F. F. (1998). Genetic analysis of beta-thalassemia major and beta-thalassemia intermedia in Brazil. *Hemoglobin, 22*(3), 197-207. doi:10.3109/03630269809113134

Frischknecht, H., Troxler, H., Greiner, J., Hengartner, H., & Dutly, F. (2008). Compound heterozygosity for Hb S [beta6(A3)GluVal, GAG-->GTG] and a new thalassemic mutation [beta132(H10)Lys-->term, AAA-->TAA] detected in a family from West Africa. *Hemoglobin, 32*(3), 309-313. doi:10.1080/03630260701758866

Furbetta, M., Tuveri, T., Rosatelli, C., Angius, A., Falchi, A. M., Cossu, P., . . . Cao, A. (1983). Molecular mechanism accounting for milder types of thalassemia major. *J Pediatr, 103*(1), 35-39. doi:10.1016/s0022-3476(83)80771-9

Galanello, R., Sanna, S., Perseu, L., Sollaino, M. C., Satta, S., Lai, M. E., . . . Cao, A. (2009). Amelioration of Sardinian beta0 thalassemia by genetic modifiers. *Blood, 114*(18), 3935-3937. doi:10.1182/blood-2009-04-217901

Galehdari, H., Salehi, B., Azmoun, S., Keikhaei, B., Zandian, K. M., & Pedram, M. (2010). Comprehensive spectrum of the beta-Thalassemia mutations in Khuzestan, southwest Iran. *Hemoglobin, 34*(5), 461-468. doi:10.3109/03630269.2010.514153

Gallerani, M., Cicognani, I., Ballardini, P., Savelli, S., Martinelli, L., Ricci, A., . . . Stabellini, G. (1990). Average life expectancy of heterozygous beta thalassemic subjects. *Haematologica, 75*(3), 224-227. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2227618

Gallo, V., Skorokhod, O. A., Schwarzer, E., & Arese, P. (2012). Simultaneous determination of phagocytosis of Plasmodium falciparum-parasitized and non-parasitized red blood cells by flow cytometry. *Malar J, 11*, 428. doi:10.1186/1475-2875-11-428

Gardikas, C. (1968). Modes of presentation of thalassaemia minor. *Acta Haematol, 40*(1), 34-36. doi:10.1159/000208881

Gasperini, D., Cao, A., Paderi, L., Barella, S., Paglietti, E., Perseu, L., . . . Galanello, R. (1993). Normal individuals with high Hb A2 levels. *Br J Haematol, 84*(1), 166-168. doi:10.1111/j.1365-2141.1993.tb03042.x

Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., . . . Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68-74. doi:10.1038/nature15393

Georgiou, I., Makis, A., Chaidos, A., Bouba, I., Hatzi, E., Kranas, V., . . . Bourantas, K. L. (2003). Distribution and frequency of beta-thalassemia mutations in northwestern and central Greece. *Eur J Haematol, 70*(2), 75-78. doi:10.1034/j.1600-0609.2003.00017.x

Gerald, P. S., & Diamond, L. K. (1958). The diagnosis of thalassemia trait by starch block electrophoresis of the hemoglobin. *Blood, 13*(1), 61-69. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/13499581

Ghanem, N., Girodon, E., Vidaud, M., Martin, J., Fanen, P., Plassa, F., & Goossens, M. (1992). A comprehensive scanning method for rapid detection of beta-globin gene mutations and polymorphisms. *Hum Mutat, 1*(3), 229-239. doi:10.1002/humu.1380010310

Giardine, B., Borg, J., Viennas, E., Pavlidis, C., Moradkhani, K., Joly, P., . . . Patrinos, G. P. (2014). Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. *Nucleic Acids Res, 42*(Database issue), D1063-1069. doi:10.1093/nar/gkt911

Gonzalez-Redondo, J. M., Stoming, T. A., Lanclos, K. D., Gu, Y. C., Kutlar, A., Kutlar, F., . . . et al. (1988). Clinical and genetic heterogeneity in black patients with homozygous beta-thalassemia from the southeastern United States. *Blood, 72*(3), 1007-1014. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2458145

Gorakshakar, A. C., Breganza, P. V., Colaco, S. P., Shaikh, R. F., Bohra, M. Y., Sawant, P. M., . . . Ghosh, K. K. (2018). Rare beta- and delta-Globin Gene Mutations in the Pathare Prabhus: Original Inhabitants of Mumbai, India. *Hemoglobin, 42*(5-6), 297-301. doi:10.1080/03630269.2018.1544909

Graffeo, L., Vitrano, A., Giambona, A., Scondotto, S., Dardanoni, G., Gluud, C., & Maggio, A. (2017). The heterozygote state for beta-thalassemia detrimentally affects health outcomes. *Am J Hematol, 92*(3), E23-E25. doi:10.1002/ajh.24619

Gunes, A. K., & Gozden, H. E. (2021). The Spectrum of Beta-Thalassemia Mutations in Syrian Refugees and Turkish Citizens. *Cureus, 13*(6), e15434. doi:10.7759/cureus.15434

Gupta, A., Hattori, Y., & Agarwal, S. (2002). Initiation codon mutation in an Asian Indian family. *Am J Hematol, 71*(2), 134-136. doi:10.1002/ajh.10189

Guzelgul, F., Seydel, G. S., & Aksoy, K. (2020). beta-Globin Gene Mutations in Pediatric Patients with beta-Thalassemia in the Region of Cukurova, Turkey. *Hemoglobin, 44*(4), 249-253. doi:10.1080/03630269.2020.1792489

Habib, Z., & Book, J. A. (1982). Thalassaemia in the egyptian population. *Hereditas, 96*(1), 149-158. doi:10.1111/j.1601-5223.1982.tb00045.x

Haghi, M., Khorshidi, S., Hosseinpour Feizi, M. A., Pouladi, N., & Hosseinpour Feizi, A. A. (2009). beta-Thalassemia mutations in the Iranian Kurdish population of Kurdistan and West Azerbaijan provinces. *Hemoglobin, 33*(2), 109-114. doi:10.1080/03630260902862020

Haldane, J. (1949). The rate of mutation of human genes. *Hereditas, 35*(s1). doi:https://doi.org/10.1111/j.1601-5223.1949.tb03339.x

Hammitt, L. L., Etyang, A. O., Morpeth, S. C., Ojal, J., Mutuku, A., Mturi, N., . . . Scott, J. A. G. (2019). Effect of ten-valent pneumococcal conjugate vaccine on invasive pneumococcal disease and nasopharyngeal carriage in Kenya: a longitudinal surveillance study. *Lancet, 393*(10186), 2146-2154. doi:10.1016/S0140-6736(18)33005-8

Harteveld, C. L., & Higgs, D. R. (2010). Alpha-thalassaemia. *Orphanet J Rare Dis, 5*, 13. doi:10.1186/1750-1172-5-13

Hill, A. V., Bowden, D. K., O'Shaughnessy, D. F., Weatherall, D. J., & Clegg, J. B. (1988). Beta thalassemia in Melanesia: association with malaria and characterization of a common variant (IVS-1 nt 5 G----C). *Blood, 72*(1), 9-14. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2898955

Hosseinpour Feizi, M. A., Hosseinpour Feizi, A. A., Pouladi, N., Haghi, M., & Azarfam, P. (2008). Molecular spectrum of beta-thalassemia mutations in Northwestern Iran. *Hemoglobin, 32*(3), 255-261. doi:10.1080/03630260802004145

Huang, S., Wong, C., Antonarakis, S. E., Ro-lien, T., Lo, W. H., & Kazazian, H. H., Jr. (1986). The same "TATA" box beta-thalassemia mutation in Chinese and US blacks: another example of independent origins of mutation. *Hum Genet, 74*(2), 162-164. doi:10.1007/BF00282081

Hussain, A., Ahmed, S., Ali, N., H, S. M., Anees, M., Chuahdry, A. H., & Ahmed, P. (2017). Rare beta-Globin Gene Mutations in Pakistan. *Hemoglobin, 41*(2), 100-103. doi:10.1080/03630269.2017.1339612

Hussein, I. R., Temtamy, S. A., el-Beshlawy, A., Fearon, C., Shalaby, Z., Vassilopoulos, G., & Kazazian, H. H., Jr. (1993). Molecular characterization of beta-thalassemia in Egyptians. *Hum Mutat, 2*(1), 48-52. doi:10.1002/humu.1380020109

Indrak, K., Brabec, V., Indrakova, J., Chrobak, L., Sakalova, A., Jarosova, M., . . . et al. (1992). Molecular characterization of beta-thalassemia in Czechoslovakia. *Hum Genet, 88*(4), 399-404. doi:10.1007/BF00215673

Ingram, V. M. (1957). Gene Mutations in Human Hæmoglobin: the Chemical Difference Between Normal and Sickle Cell Hæmoglobin. *Nature, 180*(4581), 326-328. doi:10.1038/180326a0

Introini, V., Marin-Menendez, A., Nettesheim, G., Lin, Y. C., Kariuki, S. N., Smith, A. L., . . . Penman, B. S. (2022). The erythrocyte membrane properties of beta thalassaemia heterozygotes and their consequences for Plasmodium falciparum invasion. *Sci Rep, 12*(1), 8934. doi:10.1038/s41598-022-12060-4

Jacob, H. S. (1970). Mechanisms of Heinz body formation and attachment to red cell membrane. *Semin Hematol, 7*(3), 341-354. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/5425759

Jalal, S. D., Al-Allawi, N. A., Bayat, N., Imanian, H., Najmabadi, H., & Faraj, A. (2010). beta-Thalassemia mutations in the Kurdish population of northeastern Iraq. *Hemoglobin, 34*(5), 469-476. doi:10.3109/01676830.2010.513591

Jalilian, M., Azizi Jalilian, F., Ahmadi, L., Amini, R., Esfehani, H., Sosanian, M., . . . Mahdieh, N. (2017). The Frequency of HBB Mutations Among beta-Thalassemia Patients in Hamadan Province, Iran. *Hemoglobin, 41*(1), 61-64. doi:10.1080/03630269.2017.1302468

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. doi:10.3109/03630269009046958

Jarjour, R. A., Murad, H., Moasses, F., & Al-Achkar, W. (2014). Molecular update of beta-thalassemia mutations in the Syrian population: identification of rare beta-thalassemia mutations. *Hemoglobin, 38*(4), 272-276. doi:10.3109/03630269.2014.912661

Jassim, N., Al-Arrayed, S., Al-Mukharraq, H., Merghoub, T., &, & Krishnamoorthy, R. (2002). Spectrum of β-thalassaemia mutations in Bahrain. *Bahrain Med Bull, 22*, 8-12.

Jha, A. N., Mishra, H., Verma, H. K., Pandey, I., & Lakkakula, B. (2018). Compound Heterozygosity of beta-Thalassemia and the Sickle Cell Hemoglobin in Various Populations of Chhattisgarh State, India. *Hemoglobin, 42*(2), 84-90. doi:10.1080/03630269.2018.1483946

Jiffri, E. H., Bogari, N., Zidan, K. H., Teama, S., & Elhawary, N. A. (2010). Molecular updating of beta-thalassemia mutations in the Upper Egyptian population. *Hemoglobin, 34*(6), 538-547. doi:10.3109/03630269.2010.526440

Kalle Kwaifa, I., Lai, M. I., & Md Noor, S. (2020). Non-deletional alpha thalassaemia: a review. *Orphanet J Rare Dis, 15*(1), 166. doi:10.1186/s13023-020-01429-1

Kalleas, C., Anagnostopoulos, K., Sinopoulou, K., Delaki, E., Margaritis, D., Bourikas, G., . . . Tentes, I. (2012). Phenotype and genotype frequency of beta-thalassemia and sickle cell disease carriers in Halkidiki, Northern Greece. *Hemoglobin, 36*(1), 64-72. doi:10.3109/03630269.2011.642489

Kamau, A., Nyaga, V., Bauni, E., Tsofa, B., Noor, A. M., Bejon, P., . . . Hammitt, L. L. (2017). Trends in bednet ownership and usage, and the effect of bednets on malaria hospitalization in the Kilifi Health and Demographic Surveillance System (KHDSS): 2008-2015. *BMC Infect Dis, 17*(1), 720. doi:10.1186/s12879-017-2822-x

Kaminsky, R., Kruger, N., Hempelmann, E., & Bommer, W. (1986). Reduced development of Plasmodium falciparum in beta-thalassaemic erythrocytes. *Z Parasitenkd, 72*(4), 553-556. doi:10.1007/BF00927899

Kanavakis, E., Wainscoat, J. S., Wood, W. G., Weatherall, D. J., Cao, A., Furbetta, M., . . . Sophocleous, T. (1982). The interaction of alpha thalassaemia with heterozygous beta thalassaemia. *Br J Haematol, 52*(3), 465-473. doi:10.1111/j.1365-2141.1982.tb03916.x

Kariuki, S. N., Marin-Menendez, A., Introini, V., Ravenhill, B. J., Lin, Y. C., Macharia, A., . . . Rayner, J. C. (2020). Red blood cell tension protects against severe malaria in the Dantu blood group. *Nature, 585*(7826), 579-583. doi:10.1038/s41586-020-2726-6

Kassebaum, N. J., Jasrasaria, R., Naghavi, M., Wulf, S. K., Johns, N., Lozano, R., . . . Murray, C. J. (2014). A systematic analysis of global anemia burden from 1990 to 2010. *Blood, 123*(5), 615-624. doi:10.1182/blood-2013-06-508325

Kattamis, C., Hu, H., Cheng, G., Reese, A. L., Gonzalez-Redondo, J. M., Kutlar, A., . . . Huisman, T. H. (1990). Molecular characterization of beta-thalassaemia in 174 Greek patients with thalassaemia major. *Br J Haematol, 74*(3), 342-346. doi:10.1111/j.1365-2141.1990.tb02593.x

Kattamis, C., Lagos, P., Metaxotou-Mavromati, A., & Matsaniotis, N. (1972). Serum iron and unsaturated iron-binding capacity in the -thalassaemia trait: their relation to the levels of haemoglobins A, A 2 , and F. *J Med Genet, 9*(2), 154-159. doi:10.1136/jmg.9.2.154

Kendall, A. G., & Bastomsky, C. H. (1981). Hemoglobin A2 in hyperthyroidism. *Hemoglobin, 5*(6), 571-577. doi:10.3109/03630268108991688

Keser, I., Sanlioglu, A. D., Manguoglu, E., Guzeloglu Kayisli, O., Nal, N., Sargin, F., . . . Luleci, G. (2004). Molecular analysis of beta-thalassemia and sickle cell anemia in Antalya. *Acta Haematol, 111*(4), 205-210. doi:10.1159/000077567

Khan, S. N., Riazuddin, S., & Galanello, R. (2000). Identification of three rare beta-thalassemia mutations in the Pakistani population. *Hemoglobin, 24*(1), 15-22. doi:10.3109/03630260009002269

Khelil, A. H., Laradi, S., Ferchichi, S., Carion, N., Bejaoui, M., Saad, A., . . . Perrin, P. (2003). [Diagnostic strategy of beta-thalassemic mutation in a Tunisian family, application in prenatal diagnosis]. *Ann Biol Clin (Paris), 61*(2), 229-233. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12702481

Kiani, A. A., Mortazavi, Y., Zeinali, S., & Shirkhani, Y. (2007). The molecular analysis of beta-thalassemia mutations in Lorestan Province, Iran. *Hemoglobin, 31*(3), 343-349. doi:10.1080/03630260701459382

Knott, M., Ramadan, K. M., Savage, G., Jones, F. G., El-Agnaf, M., McMullin, M. F., & Percy, M. J. (2006). Novel and Mediterranean beta thalassemia mutations in the indigenous Northern Ireland population. *Blood Cells Mol Dis, 36*(2), 265-268. doi:10.1016/j.bcmd.2005.12.005

Knox-Macaulay, H. H., & Weatherall, D. J. (1974). Studies of red-cell membrane function in heterozygous beta thalassaemia and other hypochromic anaemias. *Br J Haematol, 28*(3), 277-297. doi:10.1111/j.1365-2141.1974.tb00809.x

Koch, M., & Baum, J. (2016). The mechanics of malaria parasite invasion of the human erythrocyte - towards a reassessment of the host cell contribution. *Cell Microbiol, 18*(3), 319-329. doi:10.1111/cmi.12557

Kono, M., Kondo, T., Takagi, Y., Wada, A., & Fujimoto, K. (2009). Morphological definition of CD71 positive reticulocytes by various staining techniques and electron microscopy compared to reticulocytes detected by an automated hematology analyzer. *Clin Chim Acta, 404*(2), 105-110. doi:10.1016/j.cca.2009.03.017

Kossiva, L., Paterakis, G., Tassiopoulos, S., Papadhimitriou, S. I., Voukouti, E., Gligori, I., & Rombos, Y. (2003). Decreased expression of membrane alpha4beta1, alpha5beta1 integrins and transferrin receptor on erythroblasts in splenectomized patients with beta-thalassemia intermedia. Parallel assessment of serum soluble transferrin receptors levels. *Ann Hematol, 82*(9), 579-584. doi:10.1007/s00277-003-0708-z

Kountouris, P., Lederer, C. W., Fanis, P., Feleki, X., Old, J., & Kleanthous, M. (2014). IthaGenes: an interactive database for haemoglobin variations and epidemiology. *PLoS One, 9*(7), e103020. doi:10.1371/journal.pone.0103020

Kunkel, H. G., Ceppellini, R., Muller-Eberhard, U., & Wolf, J. (1957). Observations on the minor basic hemoglobin component in the blood of normal individuals and patients with thalassemia. *J Clin Invest, 36*(11), 1615-1625. doi:10.1172/JCI103561

Kyriacou, K., Al Quobaili, F., Pavlou, E., Christopoulos, G., Ioannou, P., & Kleanthous, M. (2000). Molecular characterization of beta-thalassemia in Syria. *Hemoglobin, 24*(1), 1-13. doi:10.3109/03630260009002268

Labie, D., Bennani, C., & Beldjord, C. (1990). Beta-thalassemia in Algeria. *Ann N Y Acad Sci, 612*, 43-54. doi:10.1111/j.1749-6632.1990.tb24289.x

Lam, V. M., Xie, S. S., Tam, J. W., Woo, Y. K., Gu, Y. L., & Li, A. M. (1990). A new single nucleotide change at the initiation codon (ATG----AGG) identified in amplified genomic DNA of a Chinese beta-thalassemic patient. *Blood, 75*(5), 1207-1208. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2306523

Lazarte, S. S., Monaco, M. E., Haro, A. C., Jimenez, C. L., Ledesma Achem, M. E., & Isse, B. A. (2014). Molecular characterization and phenotypical study of beta-thalassemia in Tucuman, Argentina. *Hemoglobin, 38*(6), 394-401. doi:10.3109/03630269.2014.968784

Lemsaddek, W., Picanco, I., Seuanes, F., Mahmal, L., Benchekroun, S., Khattab, M., . . . Osorio-Almeida, L. (2003). Spectrum of beta thalassemia mutations and HbF levels in the heterozygous Moroccan population. *Am J Hematol, 73*(3), 161-168. doi:10.1002/ajh.10358

Lopez-Escribano, H., Parera, M. M., Guix, P., Serra, J. M., Gutierrez, A., Balsells, D., . . . Picornell, A. (2013). Balearic archipelago: three islands, three beta-thalassemia population patterns. *Clin Genet, 83*(2), 175-180. doi:10.1111/j.1399-0004.2012.01864.x

Low, P. S., Waugh, S. M., Zinke, K., & Drenckhahn, D. (1985). The role of hemoglobin denaturation and band 3 clustering in red blood cell aging. *Science, 227*(4686), 531-533. doi:10.1126/science.2578228

Macharia , A. W., Mochamah, G., Makale, J., Howard, T., Mturi, N., Olupot-Olupot, P., . . . Williams, T. N. (2022). Case Report: β-thalassemia major on the East African coast [version 1; peer review: 1 approved]. . *Wellcome Open Res, 7:188*. doi:https://doi.org/10.12688/wellcomeopenres.17907.1

Macharia, A. W., Mochamah, G., Uyoga, S., Ndila, C. M., Nyutu, G., Tendwa, M., . . . Williams, T. N. (2020). beta-Thalassemia pathogenic variants in a cohort of children from the East African coast. *Mol Genet Genomic Med, 8*(7), e1294. doi:10.1002/mgg3.1294

Macharia, A. W., Uyoga, S., Ndila, C., Nyutu, G., Makale, J., Tendwa, M., . . . Williams, T. N. (2019). The population dynamics of hemoglobins A, A2, F and S in the context of the hemoglobinopathies HbS and alpha-thalassemia in Kenyan infants. *Haematologica, 104*(5), e184-e186. doi:10.3324/haematol.2018.199596

Makhoul, N. J., Wells, R. S., Kaspar, H., Shbaklo, H., Taher, A., Chakar, N., & Zalloua, P. A. (2005). Genetic heterogeneity of Beta thalassemia in Lebanon reflects historic and recent population migration. *Ann Hum Genet, 69*(Pt 1), 55-66. doi:10.1046/j.1529-8817.2004.00138.x

Martinez, G., & Menendez, R. (1983). Differences in affinity of beta and delta hemoglobin chains for alpha chains. A possible explanation for the variation in the percentages of hemoglobin A2 in thalassemia and other disorders. *Biochim Biophys Acta, 743*(2), 256-259. doi:10.1016/0167-4838(83)90222-4

Mbogo, C. N., Snow, R. W., Khamala, C. P., Kabiru, E. W., Ouma, J. H., Githure, J. I., . . . Beier, J. C. (1995). Relationships between Plasmodium falciparum transmission by vector populations and the incidence of severe disease at nine sites on the Kenyan coast. *Am J Trop Med Hyg, 52*(3), 201-206. doi:10.4269/ajtmh.1995.52.201

McGann, P. T., Williams, T. N., Olupot-Olupot, P., Tomlinson, G. A., Lane, A., Luis Reis da Fonseca, J., . . . Investigators, R. (2018). Realizing effectiveness across continents with hydroxyurea: Enrollment and baseline characteristics of the multicenter REACH study in Sub-Saharan Africa. *Am J Hematol, 93*(4), 537-545. doi:10.1002/ajh.25034

Mendilcioglu, I., Yakut, S., Keser, I., Simsek, M., Yesilipek, A., Bagci, G., & Luleci, G. (2011). Prenatal diagnosis of beta-thalassemia and other hemoglobinopathies in southwestern Turkey. *Hemoglobin, 35*(1), 47-55. doi:10.3109/03630269.2010.544607

Mettananda, S., Gibbons, R. J., & Higgs, D. R. (2015). alpha-Globin as a molecular target in the treatment of beta-thalassemia. *Blood, 125*(24), 3694-3701. doi:10.1182/blood-2015-03-633594

Michela, G., Raffaele, S., Stella, P., Maria Rosaria, S., & Paola, I. (2012). Molecular Basis of Thalassemia. In S. S. Donald (Ed.), *Anemia* (pp. Ch. 21). Rijeka: IntechOpen.

Miri-Moghaddam, E., Bahrami, S., Naderi, M., Bazi, A., & Karimipoor, M. (2016). Molecular Characterization of beta-Thalassemia Intermedia in Southeast Iran. *Hemoglobin, 40*(3), 173-178. doi:10.3109/03630269.2016.1167735

Moghadam, M., Karimi, M., Dehghani, S. J., Dehbozorgian, J., Montazeri, S., Javanmardi, E., . . . Afrasiabi, A. (2015). Effectiveness of beta-thalassemia prenatal diagnosis in Southern Iran: a cohort study. *Prenat Diagn, 35*(12), 1238-1242. doi:10.1002/pd.4684

Mohanty, D., Gorakshakar, A. C., Colah, R. B., Patel, R. Z., Master, D. C., Mahanta, J., . . . Muthuswamy, V. (2014). Interaction of iron deficiency anemia and hemoglobinopathies among college students and pregnant women: a multi center evaluation in India. *Hemoglobin, 38*(4), 252-257. doi:10.3109/03630269.2014.913517

Molchanova, T. P., Postnikov Yu, V., Gu, L. H., & Huisman, T. H. (1998). Historical note: the beta-thalassemia allele in the noble Russian family Lermontov is identified as the ATG->ACG change in the initiation codon. *Hemoglobin, 22*(3), 283-286. doi:10.3109/03630269809113143

Moradi, K., Aznab, M., Tahmasebi, S., Omidniakan, L., Bijari, N., & Alibakhshi, R. (2020). Distribution of HBB Gene Mutations in the Kurdish Population of Ilam Province, West Iran. *Hemoglobin, 44*(4), 244-248. doi:10.1080/03630269.2020.1797772

Mosca, A., Paleari, R., Ivaldi, G., Galanello, R., & Giordano, P. C. (2009). The role of haemoglobin A(2) testing in the diagnosis of thalassaemias and related haemoglobinopathies. *J Clin Pathol, 62*(1), 13-17. doi:10.1136/jcp.2008.056945

Muniz, A., Martinez, G., Lavinha, J., & Pacheco, P. (2000). Beta-thalassaemia in Cubans: novel allele increases the genetic diversity at the HBB locus in the Caribbean. *Am J Hematol, 64*(1), 7-14. doi:10.1002/(sici)1096-8652(200005)64:1<7::aid-ajh2>3.0.co;2-v

Murad, H., Moassas, F., Ghoury, I., & Mukhalalaty, Y. (2018). Haplotype Analysis of Three Common beta-Thalassemia Mutations in Syrian Patients. *Hemoglobin, 42*(5-6), 302-305. doi:10.1080/03630269.2018.1553789

Murad, H., Moassas, F., Jarjour, R., Mukhalalaty, Y., & Al-Achkar, W. (2014). Prenatal molecular diagnosis of beta-thalassemia and sickle cell anemia in the Syrian population. *Hemoglobin, 38*(6), 390-393. doi:10.3109/03630269.2014.978455

Muramatsu, T. (2016). Basigin (CD147), a multifunctional transmembrane glycoprotein with various binding partners. *J Biochem, 159*(5), 481-490. doi:10.1093/jb/mvv127

Muszlak, M., Pissard, S., Badens, C., Chamouine, A., Maillard, O., & Thuret, I. (2015). Genetic Modifiers of Sickle Cell Disease: A Genotype-Phenotype Relationship Study in a Cohort of 82 Children on Mayotte Island. *Hemoglobin, 39*(3), 156-161. doi:10.3109/03630269.2015.1023897

Najmabadi, H., Karimi-Nejad, R., Sahebjam, S., Pourfarzad, F., Teimourian, S., Sahebjam, F., . . . Karimi-Nejad, M. H. (2001). The beta-thalassemia mutation spectrum in the Iranian population. *Hemoglobin, 25*(3), 285-296. doi:10.1081/hem-100105221

Najmabadi, H., Pourfathollah, A. A., Neishabury, M., Sahebjam, F., Krugluger, W., & Oberkanins, C. (2002). Rare and unexpected mutations among Iranian beta-thalassemia patients and prenatal samples discovered by reverse-hybridization and DNA sequencing. *Haematologica, 87*(10), 1113-1114. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12368169

Ndila, C. M., Nyirongo, V., Macharia, A. W., Jeffreys, A. E., Rowlands, K., Hubbart, C., . . . Malaria, G. E. N. C. (2020). Haplotype heterogeneity and low linkage disequilibrium reduce reliable prediction of genotypes for the -alpha (3.7I) form of alpha-thalassaemia

using genome-wide microarray data. *Wellcome Open Res, 5*, 287. doi:10.12688/wellcomeopenres.16320.2

Nezhad, F. H., Nezhad, K. H., Choghakabodi, P. M., & Keikhaei, B. (2018). Prevalence and Genetic Analysis of alpha- and beta-Thalassemia and Sickle Cell Anemia in Southwest Iran. *J Epidemiol Glob Health, 8*(3-4), 189-195. doi:10.2991/j.jegh.2018.04.103

Njuguna, P., Maitland, K., Nyaguara, A., Mwanga, D., Mogeni, P., Mturi, N., . . . Bejon, P. (2019). Observational study: 27 years of severe malaria surveillance in Kilifi, Kenya. *BMC Med, 17*(1), 124. doi:10.1186/s12916-019-1359-9

Nokes, D. J., Okiro, E. A., Ngama, M., Ochola, R., White, L. J., Scott, P. D., . . . Medley, G. F. (2008). Respiratory syncytial virus infection and disease in infants and young children observed from birth in Kilifi District, Kenya. *Clin Infect Dis, 46*(1), 50-57. doi:10.1086/524019

Novelletto, A., Hafez, M., Deidda, G., Di Rienzo, A., Felicetti, L., el-Tahan, H., . . . et al. (1990). Molecular characterization of beta-thalassemia mutations in Egypt. *Hum Genet, 85*(3), 272-274. doi:10.1007/BF00206744

O'Meara, W. P., Bejon, P., Mwangi, T. W., Okiro, E. A., Peshu, N., Snow, R. W., . . . Marsh, K. (2008). Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya. *Lancet, 372*(9649), 1555-1562. doi:10.1016/S0140-6736(08)61655-4

Oner, R., Altay, C., Gurgey, A., Aksoy, M., Kilinc, Y., Stoming, T. A., . . . Huisman, T. H. (1990). Beta-thalassemia in Turkey. *Hemoglobin, 14*(1), 1-13. doi:10.3109/03630269009002250

Orkin, S. H., Sexton, J. P., Goff, S. C., & Kazazian, H. H., Jr. (1983). Inactivation of an acceptor RNA splice site by a short deletion in beta-thalassemia. *J Biol Chem, 258*(12), 7249-7251. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/6190800

Otieno, G. P., Murunga, N., Agoti, C. N., Gallagher, K. E., Awori, J. O., & Nokes, D. J. (2020). Surveillance of endemic human coronaviruses (HCoV-NL63, OC43 and 229E) associated with childhood pneumonia in Kilifi, Kenya. *Wellcome Open Res, 5*, 150. doi:10.12688/wellcomeopenres.16037.2

Ozkinay, F., Onay, H., Karaca, E., Arslan, E., Erturk, B., Ece Solmaz, A., . . . Vergin, C. (2015). Molecular Basis of beta-Thalassemia in the Population of the Aegean Region of Turkey: Identification of A Novel Deletion Mutation. *Hemoglobin, 39*(4), 230-234. doi:10.3109/03630269.2015.1038354

Pagnier, J., Mears, J. G., Dunda-Belkhodja, O., Schaefer-Rego, K. E., Beldjord, C., Nagel, R. L., & Labie, D. (1984). Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. *Proc Natl Acad Sci U S A, 81*(6), 1771-1773. doi:10.1073/pnas.81.6.1771

Park, E. S., Jung, H. L., Kim, H. J., Park, S. S., Bae, S. H., Shin, H. Y., . . . Hah, J. O. (2013). Hereditary hemolytic anemia in Korea from 2007 to 2011: A study by the Korean Hereditary Hemolytic Anemia Working Party of the Korean Society of Hematology. *Blood Res, 48*(3), 211-216. doi:10.5045/br.2013.48.3.211

Pasvol, G., Weatherall, D. J., & Wilson, R. J. (1980). The increased susceptibility of young red cells to invasion by the malarial parasite Plasmodium falciparum. *Br J Haematol, 45*(2), 285-295. doi:10.1111/j.1365-2141.1980.tb07148.x

Patrinos, G. P., Giardine, B., Riemer, C., Miller, W., Chui, D. H., Anagnou, N. P., . . . Hardison, R. C. (2004). Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. *Nucleic Acids Res, 32*(Database issue), D537-541. doi:10.1093/nar/gkh006

Pavlovic, S., Urosevic, J., Poznanic, J., Perisic, L. J., Petrucev, B., Tosic, N., . . . Bunjevacki, G. (2005). Molecular basis of thalassemia syndromes in Serbia and Montenegro. *Acta Haematol, 113*(3), 175-180. doi:10.1159/000084447

Perrin, P., Bouhassa, R., Mselli, L., Garguier, N., Nigon, V. M., Bennani, C., . . . Trabuchet, G. (1998). Diversity of sequence haplotypes associated with beta-thalassaemia mutations in Algeria: implications for their origin. *Gene, 213*(1-2), 169-177. doi:10.1016/s0378-1119(98)00200-5

Petkov, G. H., & Efremov, G. D. (2007). Molecular basis of beta-thalassemia and other hemoglobinopathies in Bulgaria: an update. *Hemoglobin, 31*(2), 225-232. doi:10.1080/03630260701290316

Petkov, G. H., Efremov, G. D., Efremov, D. G., Dimovski, A., Tchaicarova, P., Tchaicarov, R., . . . et al. (1990). Beta-thalassemia in Bulgaria. *Hemoglobin, 14*(1), 25-33. doi:10.3109/03630269009002252

Piel, F. B., & Weatherall, D. J. (2014). The alpha-thalassemias. *N Engl J Med, 371*(20), 1908-1916. doi:10.1056/NEJMra1404415

Powars, D. R. (1991). Beta s-gene-cluster haplotypes in sickle cell anemia. Clinical and hematologic features. *Hematol Oncol Clin North Am, 5*(3), 475-493. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1713910

Premawardhena, A., Arambepola, M., Katugaha, N., & Weatherall, D. J. (2008). Is the beta thalassaemia trait of clinical importance? *Br J Haematol, 141*(3), 407-410. doi:10.1111/j.1365-2141.2008.07071.x

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical

Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Ragusa, A., Amata, S., Lombardo, T., Castiglia, L., Maier-Redelsperger, M., Labie, D., & Bernini, L. (2003). Asymptomatic and mild beta-thalassemia in homozygotes and compound heterozygotes for the IVS2+1G-->A mutation: role of the beta-globin gene haplotype. *Haematologica, 88*(10), 1099-1105. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14555304

Roberts, J. (2007). The New Penguin History of the World (5th ed). *London, UK: Penguin Books.*

Roth, E. F., Jr., Raventos-Suarez, C., Rinaldi, A., & Nagel, R. L. (1983). Glucose-6-phosphate dehydrogenase deficiency inhibits in vitro growth of Plasmodium falciparum. *Proc Natl Acad Sci U S A, 80*(1), 298-299. doi:10.1073/pnas.80.1.298

Routy, J. P., Monte, M., Beaulieu, R., Toma, E., St-Pierre, L., & Dumont, M. (1993). Increase of hemoglobin A2 in human immunodeficiency virus-1-infected patients treated with zidovudine. *Am J Hematol, 43*(2), 86-90. doi:10.1002/ajh.2830430203

Rowe, A., Obeiro, J., Newbold, C. I., & Marsh, K. (1995). Plasmodium falciparum rosetting is associated with malaria severity in Kenya. *Infect Immun, 63*(6), 2323-2326. doi:10.1128/iai.63.6.2323-2326.1995

Sadiq, M. F., & Huisman, T. H. (1994). Molecular characterization of beta-thalassemia in north Jordan. *Hemoglobin, 18*(4-5), 325-332. doi:10.3109/03630269408996198

Sahli, C. A., Ben Salem, I., Jouini, L., Laouini, N., Dabboubi, R., Hadj Fredj, S., . . . 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. doi:10.1002/jcla.21867

Sajadpour, Z., Amini-Farsani, Z., Motovali-Bashi, M., Yadollahi, M., & Khosravi-Farsani, N. (2020). Association between Different Polymorphic Markers and beta-Thalassemia Intermedia in Central Iran. *Hemoglobin, 44*(1), 27-30. doi:10.1080/03630269.2019.1709204

Saleh-Gohari, N., Khademi Bami, M., Nikbakht, R., & Karimi-Maleh, H. (2015). Effects of alpha-thalassaemia mutations on the haematological parameters of beta-thalassaemia carriers. *J Clin Pathol, 68*(7), 562-566. doi:10.1136/jclinpath-2014-202825

Samha, L., Sirdah, M. M., Reading, N. S., Karmi, B., & Agarwal, A. M. (2020). Molecular Understanding of Severe Cases of beta-Thalassemia in the Nablus Region, West Bank, Palestine. *Hemoglobin, 44*(2), 128-130. doi:10.1080/03630269.2020.1763398

Schiliro, G., Di Gregorio, F., Samperi, P., Mirabile, E., Liang, R., Curuk, M. A., . . . Huisman, T. H. (1995). Genetic heterogeneity of beta-thalassemia in southeast Sicily. *Am J Hematol, 48*(1), 5-11. doi:10.1002/ajh.2830480103

Scott, J. A., Bauni, E., Moisi, J. C., Ojal, J., Gatakaa, H., Nyundo, C., . . . Williams, T. N. (2012). Profile: The Kilifi Health and Demographic Surveillance System (KHDSS). *Int J Epidemiol, 41*(3), 650-657. doi:10.1093/ije/dys062

Senok, A. C., Li, K., Nelson, E. A., Yu, L. M., Tian, L. P., & Oppenheimer, S. J. (1997). Invasion and growth of Plasmodium falciparum is inhibited in fractionated thalassaemic

erythrocytes. *Trans R Soc Trop Med Hyg, 91*(2), 138-143. doi:10.1016/s0035-9203(97)90200-5

Serjeant, G. R., Sommereux, A. M., Stevenson, M., Mason, K., & Serjeant, B. E. (1979). Comparison of sickle cell-beta0 thalassaemia with homozygous sickle cell disease. *Br J Haematol, 41*(1), 83-93. doi:10.1111/j.1365-2141.1979.tb03684.x

Shaeffer, J. R. (1980). Evidence for a difference in affinities of human hemoglobin beta A and beta S chains for alpha chains. *J Biol Chem, 255*(6), 2322-2324. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/7358673

Silva, A. N., Cardoso, G. L., Cunha, D. A., Diniz, I. G., Santos, S. E., Andrade, G. B., . . . Guerreiro, J. F. (2016). The Spectrum of beta-Thalassemia Mutations in a Population from the Brazilian Amazon. *Hemoglobin, 40*(1), 20-24. doi:10.3109/03630269.2015.1083443

Siniscalco, M., Bernini, L., Latte, B., & Motulsky, A. G. (1961). Favism and Thalassæmia in Sardinia and their Relationship to Malaria. *Nature, 190*(4782), 1179-1180. doi:10.1038/1901179a0

Sirdah, M. M., Sievertsen, J., Al-Yazji, M. S., Tarazi, I. S., Al-Haddad, R. M., Horstmann, R. D., & Timmann, C. (2013). The spectrum of beta-thalassemia mutations in Gaza Strip, Palestine. *Blood Cells Mol Dis, 50*(4), 247-251. doi:10.1016/j.bcmd.2012.12.004

Snow, R. W., Kibuchi, E., Karuri, S. W., Sang, G., Gitonga, C. W., Mwandawiro, C., . . . Noor, A. M. (2015). Changing Malaria Prevalence on the Kenyan Coast since 1974: Climate, Drugs and Vector Control. *PLoS One, 10*(6), e0128792. doi:10.1371/journal.pone.0128792

Spritz, R. A., Jagadeeswaran, P., Choudary, P. V., Biro, P. A., Elder, J. T., deRiel, J. K., . . . 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. doi:10.1073/pnas.78.4.2455

Sripichai, O., Munkongdee, T., Kumkhaek, C., Svasti, S., Winichagoon, P., & Fucharoen, S. (2008). Coinheritance of the different copy numbers of alpha-globin gene modifies severity of beta-thalassemia/Hb E disease. *Ann Hematol, 87*(5), 375-379. doi:10.1007/s00277-007-0407-2

Steinberg, M. H., & Embury, S. H. (1986). Alpha-thalassemia in blacks: genetic and clinical aspects and interactions with the sickle hemoglobin gene. *Blood, 68*(5), 985-990. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/3533181

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. *Br J Haematol, 170*(6), 781-787. doi:10.1111/bjh.13570

Stephens, A. D., Angastiniotis, M., Baysal, E., Chan, V., Davis, B., Fucharoen, S., . . . International Council for The Standardisation of, H. (2012). ICSH recommendations for the measurement of haemoglobin F. *Int J Lab Hematol, 34*(1), 14-20. doi:10.1111/j.1751-553X.2011.01367.x

Stuart, M. J., & Nagel, R. L. (2004). Sickle-cell disease. *Lancet, 364*(9442), 1343-1360. doi:10.1016/S0140-6736(04)17192-4

Sturm, A., Amino, R., van de Sand, C., Regen, T., Retzlaff, S., Rennenberg, A., . . . Heussler, V. T. (2006). Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science, 313*(5791), 1287-1290. doi:10.1126/science.1129720

Suh, D. D., Krauss, J. S., & Bures, K. (1996). Influence of hemoglobin S adducts on hemoglobin A2 quantification by HPLC. *Clin Chem, 42*(7), 1113-1114. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8674201

Tadmouri, G. O., Garguier, N., Demont, J., Perrin, P., & Basak, A. N. (2001). History and origin of beta-thalassemia in Turkey: sequence haplotype diversity of beta-globin genes. *Hum Biol, 73*(5), 661-674. doi:10.1353/hub.2001.0075

Telen, M. J. (2000). Red blood cell surface adhesion molecules: their possible roles in normal human physiology and disease. *Semin Hematol, 37*(2), 130-142. doi:10.1016/s0037-1963(00)90038-6

Thanomsub, B. W., Fucharoen, S., Brockelman, C. R., & Bhisutthibhan, J. (1989). Effects of thalassaemic serum on the in vitro development of the malarial parasite Plasmodium falciparum. *Parasitol Res, 75*(3), 212-217. doi:10.1007/bf00931278

Thein, S. L. (2013). The molecular basis of beta-thalassemia. *Cold Spring Harb Perspect Med, 3*(5), a011700. doi:10.1101/cshperspect.a011700

Thein, S. L. (2018). Molecular basis of beta thalassemia and potential therapeutic targets. *Blood Cells Mol Dis, 70*, 54-65. doi:10.1016/j.bcmd.2017.06.001

Theron, M., Cross, N., Cawkill, P., Bustamante, L. Y., & Rayner, J. C. (2018). An in vitro erythrocyte preference assay reveals that Plasmodium falciparum parasites prefer Type O over Type A erythrocytes. *Sci Rep, 8*(1), 8133. doi:10.1038/s41598-018-26559-2

Thomson-Luque, R., Wang, C., Ntumngia, F. B., Xu, S., Szekeres, K., Conway, A., . . . Jiang, R. H. Y. (2018). In-depth phenotypic characterization of reticulocyte maturation using mass cytometry. *Blood Cells Mol Dis, 72*, 22-33. doi:10.1016/j.bcmd.2018.06.004

Thornburg, C. D., Zimmerman, S. A., Schultz, W. H., & Ware, R. E. (2001). An infant with homozygous hemoglobin D-Iran. *J Pediatr Hematol Oncol, 23*(1), 67-68. doi:10.1097/00043426-200101000-00017

Traeger-Synodinos, J., Kanavakis, E., Vrettou, C., Maragoudaki, E., Michael, T., Metaxotou-Mavromati, A., & Kattamis, C. (1996). The triplicated alpha-globin gene locus in beta-thalassaemia heterozygotes: clinical, haematological, biosynthetic and molecular studies. *Br J Haematol, 95*(3), 467-471. doi:10.1046/j.1365-2141.1996.d01-1939.x

Tsatalas, C., Chalkia, P., Pantelidou, D., Margaritis, D., Bourikas, G., & Spanoudakis, E. (2009). Pregnancy in beta-thalassemia trait carriers: an uneventful journey. *Hematology, 14*(5), 301-303. doi:10.1179/102453309X439791

Udomsangpetch, R., Sueblinvong, T., Pattanapanyasat, K., Dharmkrong-at, A., Kittikalayawong, A., & Webster, H. K. (1993). Alteration in cytoadherence and rosetting of Plasmodium falciparum-infected thalassemic red blood cells. *Blood, 82*(12), 3752-3759. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8260712

Ulasli, M., Oztuzcu, S., Kirkbes, S., Bay, A., Igci, Y. Z., Bayraktar, R., . . . Arslan, A. (2015). Novel Betaeta (beta)-Thalassemia Mutation in Turkish Children. *Indian J Hematol Blood Transfus, 31*(2), 218-222. doi:10.1007/s12288-014-0380-6

Uyoga, S., Macharia, A. W., Mochamah, G., Ndila, C. M., Nyutu, G., Makale, J., . . . Williams, T. N. (2019). The epidemiology of sickle cell disease in children recruited in infancy in Kilifi, Kenya: a prospective cohort study. *Lancet Glob Health, 7*(10), e1458-e1466. doi:10.1016/S2214-109X(19)30328-6

Uyoga, S., Macharia, A. W., Ndila, C. M., Nyutu, G., Shebe, M., Awuondo, K. O., . . . Williams, T. N. (2019). The indirect health effects of malaria estimated from health advantages of the sickle cell trait. *Nat Commun, 10*(1), 856. doi:10.1038/s41467-019-08775-0

Van Delft, P., Lenters, E., Bakker-Verweij, M., de Korte, M., Baylan, U., Harteveld, C. L., & Giordano, P. C. (2009). Evaluating five dedicated automatic devices for haemoglobinopathy diagnostics in multi-ethnic populations. *Int J Lab Hematol, 31*(5), 484-495. doi:10.1111/j.1751-553X.2009.01158.x

Vattanaviboon, P., Sriklup, C., Polkaew, C., Prawatmurng, P., & Ittarat, W. (1999). The Effect of Thalassemic Serum on Plasmodium Falciparum Growth in Vitro. *SLAS Technology, 4*(4), 78-79. doi:https://doi.org/10.1016/S1535-5535(04)00025-5

Veten, F., Ghaber, S., Habti, N., & Houmeida, A. (2015). Occurrence of the Codon 24 (A > T) Mutation in the Mauritanian Population. *Hemoglobin, 39*(4), 296-297. doi:10.3109/03630269.2015.1043060

Vetter, B., Schwarz, C., Kohne, E., & Kulozik, A. E. (1997). Beta-thalassaemia in the immigrant and non-immigrant German populations. *Br J Haematol, 97*(2), 266-272. doi:10.1046/j.1365-2141.1997.342674.x

Vettore, L., Falezza, G. C., Cetto, G. L., & De Matteis, M. C. (1974). Cation content and membrane deformability of heterozygous beta-thalassaemic red blood cells. *Br J Haematol, 27*(3), 429-437. doi:10.1111/j.1365-2141.1974.tb06809.x

Walliker, D., Quakyi, I. A., Wellems, T. E., McCutchan, T. F., Szarfman, A., London, W. T., . . . Carter, R. (1987). Genetic analysis of the human malaria parasite Plasmodium falciparum. *Science, 236*(4809), 1661-1666. doi:10.1126/science.3299700

Waterfall, C. M., & Cobb, B. D. (2001). Single tube genotyping of sickle cell anaemia using PCR-based SNP analysis. *Nucleic Acids Res, 29*(23), E119. doi:10.1093/nar/29.23.e119

Waye, J. S., Borys, S., Eng, B., Patterson, M., Chui, D. H., Badr El-Din, O. M., . . . Afify, Z. (1999). Spectrum of beta-thalassemia mutations in Egypt. *Hemoglobin, 23*(3), 255-261. doi:10.3109/03630269909005706

Waye, J. S., Eng, B., Got, T., Hanna, M., Hohenadel, B. A., Nakamura, L. M., & Walker, L. (2015). Sudanese (deltabeta)0-Thalassemia: Identification and Characterization of a Novel 9.6 kb Deletion. *Hemoglobin, 39*(5), 368-370. doi:10.3109/03630269.2015.1057736

Weatherall, D. J. (1964). Biochemical Phenotypes of Thalassemia in the American Negro Population. *Ann N Y Acad Sci, 119*, 450-462. doi:10.1111/j.1749-6632.1965.tb54046.x

Weatherall, D. J. (1998). Pathophysiology of thalassaemia. *Baillieres Clin Haematol, 11*(1), 127-146. doi:10.1016/s0950-3536(98)80072-3

Weatherall, D. J. (2001). Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat Rev Genet, 2*(4), 245-255. doi:10.1038/35066048

Weatherall, D. J., & Clegg, J. B. (2001). *The thalassaemia syndromes* (4th ed.). Oxford ; Malden, MA: Blackwell Science.

Weiss, G. E., Gilson, P. R., Taechalertpaisarn, T., Tham, W. H., de Jong, N. W., Harvey, K. L., . . . Crabb, B. S. (2015). Revealing the sequence and resulting cellular morphology of receptor-ligand interactions during Plasmodium falciparum invasion of erythrocytes. *PLoS Pathog, 11*(2), e1004670. doi:10.1371/journal.ppat.1004670

Westaway, D., & Williamson, R. (1981). An intron nucleotide sequence variant in a cloned beta +-thalassaemia globin gene. *Nucleic Acids Res, 9*(8), 1777-1788. doi:10.1093/nar/9.8.1777

WHO. (1989). Guidelines for the control of haemoglobin disorders . *WHO Working Group on Haemoglobinopathies. World Health Organization, Geneva*(94).

Wildmann, C., Larondelle, Y., Vaerman, J. L., Eeckels, R., Martiat, P., & Philippe, M. (1993). An initiation codon mutation as a cause of beta-thalassemia in a Belgian family. *Hemoglobin, 17*(1), 19-30. doi:10.3109/03630269308998882

Willcox, M., Bjorkman, A., Brohult, J., Pehrson, P. O., Rombo, L., & Bengtsson, E. (1983). A case-control study in northern Liberia of Plasmodium falciparum malaria in haemoglobin S and beta-thalassaemia traits. *Ann Trop Med Parasitol, 77*(3), 239-246. doi:10.1080/00034983.1983.11811704

Willcox, M. C. (1975). Thalassaemia in northern Liberia. A survey in the Mount Nimba area. *J Med Genet, 12*(1), 55-63. doi:10.1136/jmg.12.1.55

Williams, T. N., Mwangi, T. W., Wambua, S., Peto, T. E., Weatherall, D. J., Gupta, S., . . . Marsh, K. (2005). Negative epistasis between the malaria-protective effects of alpha+-thalassemia and the sickle cell trait. *Nat Genet, 37*(11), 1253-1257. doi:10.1038/ng1660

Williams, T. N., Wambua, S., Uyoga, S., Macharia, A., Mwacharo, J. K., Newton, C. R., & Maitland, K. (2005). Both heterozygous and homozygous alpha+ thalassemias protect against severe and fatal Plasmodium falciparum malaria on the coast of Kenya. *Blood, 106*(1), 368-371. doi:10.1182/blood-2005-01-0313

Williams, T. N., & Weatherall, D. J. (2012). World distribution, population genetics, and health burden of the hemoglobinopathies. *Cold Spring Harb Perspect Med, 2*(9), a011692. doi:10.1101/cshperspect.a011692

Wood, W. G. (1976). Haemoglobin synthesis during human fetal development. *Br Med Bull, 32*(3), 282-287. doi:10.1093/oxfordjournals.bmb.a071376

Yadav, R., Lazarus, M., Ghanghoria, P., Singh, M., Gupta, R. B., Kumar, S., . . . Shanmugam, R. (2016). Sickle cell disease in Madhya Pradesh, Central India: A comparison of clinical profile of sickle cell homozygote vs. sickle-beta thalassaemia individuals. *Hematology, 21*(9), 558-563. doi:10.1080/10245332.2016.1148893

Yiannikourides, A., & Latunde-Dada, G. O. (2019). A Short Review of Iron Metabolism and Pathophysiology of Iron Disorders. *Medicines (Basel), 6*(3). doi:10.3390/medicines6030085

Yilmaz, S. (2019). The Spectrum of beta-Thalassemia Mutations in Siirt Province, Southeastern Turkey. *Hemoglobin, 43*(3), 174-181. doi:10.1080/03630269.2019.1647852

Yuan, J., Kannan, R., Shinar, E., Rachmilewitz, E. A., & Low, P. S. (1992). Isolation, characterization, and immunoprecipitation studies of immune complexes from membranes of beta-thalassemic erythrocytes. *Blood, 79*(11), 3007-3013. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1586745

Yuzbasioglu Ariyurek, S., Yildiz, S. M., Yalin, A. E., Guzelgul, F., & Aksoy, K. (2016). Hemoglobinopathies in the Cukurova Region and Neighboring Provinces. *Hemoglobin, 40*(3), 168-172. doi:10.3109/03630269.2016.1155156

Zenonos, Z. A., Dummler, S. K., Muller-Sienerth, N., Chen, J., Preiser, P. R., Rayner, J. C., & Wright, G. J. (2015). Basigin is a druggable target for host-oriented antimalarial interventions. *J Exp Med, 212*(8), 1145-1151. doi:10.1084/jem.20150032

Zhang, J., He, J., Zeng, X. H., Ge, S. J., Huang, Y., Su, J., . . . Zhu, B. S. (2015). Genetic heterogeneity of the beta-globin gene in various geographic populations of Yunnan in southwestern China. *PLoS One, 10*(4), e0122956. doi:10.1371/journal.pone.0122956

Zlotogora, J., Hujerat, Y., Barges, S., Shalev, S. A., & Chakravarti, A. (2007). The fate of 12 recessive mutations in a single village. *Ann Hum Genet, 71*(Pt 2), 202-208. doi:10.1111/j.1469-1809.2006.00308.x