DETERMINATION OF THE FREQUENCY OF FOUR PATHOGENIC VARIANTS CAUSING INBORN ERRORS OF METABOLISM IN THE WESTERN CAPE BLACK POPULATION, USING A MULTIPLEXED ARMS PCR APPROACH

by

RONALD DALMACIO

STUDENT NUMBER: DLMRON002

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN

In partial fulfilment of the requirements for the degree

MMED (Chemical Pathology)

Faculty of Health Sciences

UNIVERSITY OF CAPE TOWN

Date of Submission 28/04/2023

Supervisors: Mrs. S Meldau

Prof. AD Marais

Division of Chemical Pathology

University of Cape Town

DECLARATION

I, Ronald Dalmacio, hereby declare that the work on which this dissertation is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: ……28/04/2023..……..…….

CONTENTS

Referencing style: Vancouver Method

ACKNOWLEDGEMENTS AND CONTRIBUTIONS

Mrs. Surita Meldau: Thank you for everything you have done through this journey. Your passion for genetics is evident in your expertise and skill. The immense knowledge that I have had the opportunity to garner throughout this project would not have been possible without your constant guidance.

Prof. David Marais: It has been a privilege to gain insight into your process of the nuances of writing scientifically. Your guidance has always impressed upon me the need to effectively communicate in concise and organised language.

My parents (Leo and Judy Dalmacio) and parents in-law (Ian and Sue Oliver): Thank you for the constant motivation, especially when things were extremely difficult and the end seemed far away. Sue Oliver: Thank you for always being willing to review and edit the work. Your constant support has helped me through the years.

Kim Dalmacio: The journey has been long and at times difficult, but you have always had my back. All that I have been able to achieve has only been possible because you notice the blind spots that I miss. Thank you especially for the gift of our son Matthew, who teaches me that all children allow us to live the gift of unconditional love.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAT	alpha-1 antitrypsin
AD	autosomal dominant
AR	autosomal recessive
ARMS PCR	Amplification Refractory Mutation System polymerase chain reaction
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
ExAC	Exome Aggregation Consortium
Fe-S	iron-sulphur
FGF23	fibroblast growth factor 23
G6P	glucose-6-phosphate
G6Pase	Glucose-6-phosphatase
G6PC	glucose-6-phosphatase catalytic subunit
G6PT	glucose-6-phosphate transporter
GA1	glutaric aciduria type 1
GCS	glycine cleavage system
GLUT	glucose transporter
gnomAD	Genome Aggregation Database
GSD 1a	glycogen storage disease type 1a
H3Africa	Human Hereditary and Health in Africa
HFTC	hyperphosphataemic familial tumoral calcinosis
HRM	high resolution melting
IEM	Inborn errors of metabolism

IMD	Inherited metabolic disease
LDLR	Low density lipoprotein receptor
MMDS 2	multiple mitochondrial dysfunction syndrome type 2
NBS	newborn screening
NCBI	National Center for Biotechnology Information
NGS	Next-generation sequencing
NPT	sodium/phosphate cotransporters
NHLS	National Health Laboratory Service
PCR	Polymerase chain reaction
PDH	pyruvate dehydrogenase
PDCH	pyruvate dehydrogenase complex
RFLP	restriction fragment length polymorphism
SCOT	succinyl-CoA:3-ketoacid CoA transferase
SNV	single nucleotide variant
SSCP	single-strand conformation polymorphism
Taq	Thermus aquaticus
Tm	melting temperature
UCT	University of Cape Town
UV	ultraviolet
VUS	variant of unknown significance
XLR	x-linked recessive

ABSTRACT

Submitted to the

UNIVERSITY OF CAPE TOWN

In fulfilment of the requirements for the degree

MASTER OF MEDICINE IN CHEMICAL PATHOLOGY

Academic Departments: Pathology (Chemical Pathology)

TITLE:

Determination of the Frequency of Four Pathogenic Variants Causing Inborn Errors of Metabolism in the Western Cape Black Population, using a Multiplexed ARMS PCR Approach

Authorship

R Dalmacio1,2 A D Marais1 and S Meldau1,2

Affiliations

1 Division of Chemical Pathology, University of Cape Town and Groote Schuur Hospital, Cape Town, South Africa

2 National Health Laboratory Services, Cape Town, South Africa

Keywords

Amplification refractory mutation system PCR (ARMS PCR), Inborn error of metabolism (IEM), Carrier frequency, *GALNT3*, hyperphosphataemic familial tumoral calcinosis (HFTC), *OXCT1*, Succinyl-CoA:3-ketoacid CoA transferase (SCOT), *G6PC*, Glycogen storage disease type 1a (GSD 1a, von Gierke's disease), *BOLA3*, multiple mitochondrial dysfunction syndrome type 2 (MMDS 2).

ABSTRACT

Background:

Carrier frequency determination of repeatedly identified pathogenic variants causing inborn errors of metabolism will enable early diagnosis and treatment of illness, and counselling of prospective parents. Four single nucleotide variants (SNV) were identified in our black South African population, on two or more separate alleles, namely, c.484C>T(p.Arg162Ter) in the *GALNT3* gene causing hyperphosphataemic familial tumoral calcinosis (HFTC); c.803G>A(p.Arg268His) in the *OXCT1* gene causing Succinyl-CoA:3-ketoacid CoA transferase (SCOT) deficiency; c.189G>A(p.Trp62Ter) in the *G6PC* gene causing glycogen storage disease type 1a (GSD 1a), and c.159dupT(p.Asp54*) in the *BOLA3* gene causing multiple mitochondrial dysfunction syndrome type 2 (MMDS 2). Analysing large population cohorts for all four variants individually is time-consuming and expensive. Therefore, a simple, cost effective, and robust method like multiplexed ARMS PCR using standard PCR chemistry is attractive for use in resource constrained environments in which common population variants account for most of the disease burden.

Methods:

ARMS PCR primers were designed to detect the four variants of interest. Individual PCR methods were optimised for each primer pair, followed by an attempt to combine these reactions in a multiplex assay. A multiplex ARMS PCR method designed to detect both the *BOLA3* and *OXCT1* pathogenic variants listed above was used to screen a cohort of 750 samples, followed by Sanger sequencing to confirm findings in positive cases.

Results:

Individual PCR reactions performed well for all primer pairs at 54°C annealing temperature. Attempts to combine all four primer sets into a single multiplex reaction repeatedly failed. A smaller multiplex assay containing primers for the *BOLA3* and *OXCT1* variants showed promise initially, but Sanger sequencing failed to confirm the positive *OXCT1* results found in all 14 ARMS PCR positive cases identified.

Conclusions:

This study investigated the feasibility of using multiplexed ARMS PCR to screen for multiple variants simultaneously in a clinically unaffected cohort. This study highlights the challenges of combining PCR reactions. Troubleshooting is laborious, time-consuming and may delay obtaining frequencies. The carrier frequency of the four IEM causing variants investigated in this study requires individual PCR assays, unless multiplex assays are optimised, or other methods are used.

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Inborn errors of metabolism (IEM) are a broad group of monogenic, heterogenous disorders that affect the function of various metabolic pathways and systems. They arise from DNA variants leading to defects in expression or function of proteins involved in amino acid, lipid, and carbohydrate metabolism. The phenotypic manifestations of an IEM are determined by the accumulation of toxic precursors, metabolites generated by alternative pathways, or to the absence of essential metabolites downstream of the dysfunctional enzymes **(Figure 1.**). IEM should be considered in the differential diagnosis of many clinical problems. While individually rare, collectively they impact significantly on health, and their complications may be costly or cause disability and death. Pro-active identification of carriers for diseases where no effective treatment exists, will identify couples at risk of conceiving a child with fatal disease, thereby guiding decisions around pregnancy. Global prevalence estimations indicate approximately 50 affected individuals per 100 000 live births, when IEM is considered as an overarching entity (1,2).

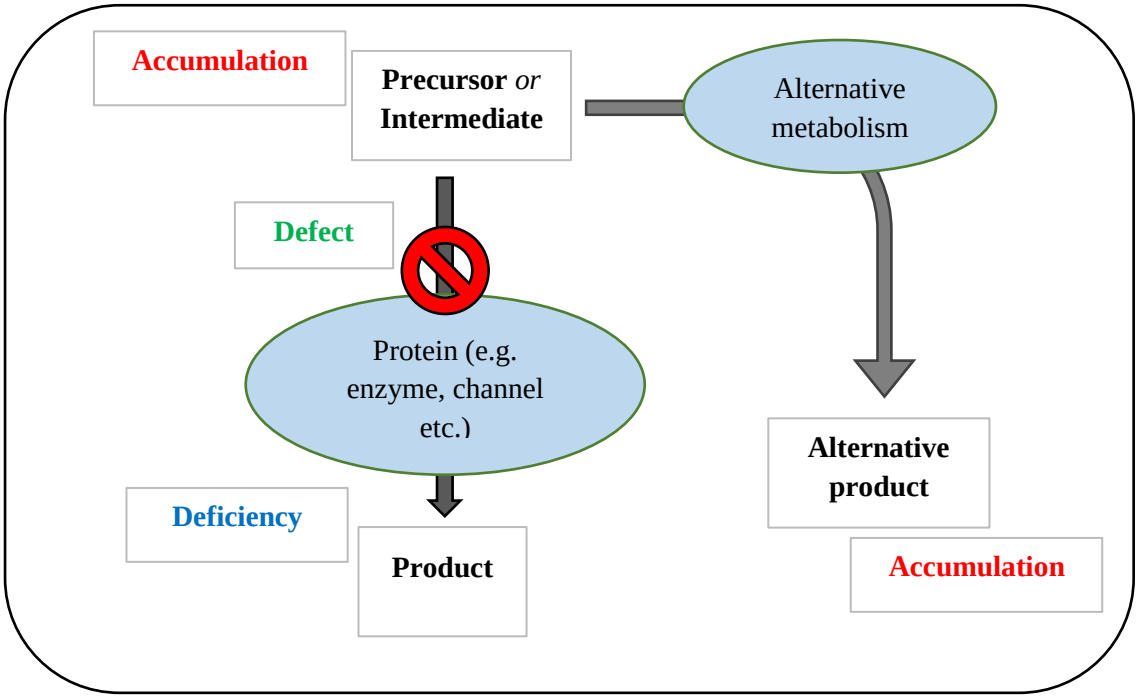

Figure 1. Basic representation of metabolic consequences of a defective pathway with either accumulation of precursors, deficiency of essential products, or the appearance of alternative products.

Metabolic medicine, focusing on IEM, is a rapidly developing field. The term IEM was first used in the early 20th century, by Sir Archibald Garrod (referring to Garrod's tetrad of albinism, alkaptonuria, cystinuria and pentosuria) (3). Since then, scientific and medical advancements expanded this field largely as a result of biochemical and genetic investigations. Furthermore, mass spectroscopy has advanced metabolomic understanding of pathological variations, while sequencing methodologies has expedited detection of genetic defects (4).

IEM have variable inheritance patterns. The most common mode of inheritance is autosomal recessive (AR) but autosomal dominant (AD), x-linked recessive (XLR) and maternal inheritance are also seen. Alternatively, *de novo* genetic variants may cause unexpected IEM. For a disease phenotype to manifest in AR disorders both alleles should be affected. This can either be the same pathogenic variant in both parental alleles (true homozygosity), or two different pathogenic variants on separate alleles (compound heterozygosity). The severity of an AR disorder is governed by the residual function, metabolic stress on the pathway, and the scope of alternative pathways.

Prevalence is defined as the proportion of individuals within a studied population at a specified period or point in time with a particular disease or characteristic (5). Various factors may affect the prevalence of carriers for variants resulting in AR IEM. Variants may arise by *de novo* mutations. Once present their prevalence may be enriched by genetic drift through selection of favourable traits. Founder effects increase the prevalence of variants in a population when a small group relocates or becomes isolated geographically or culturally. Incidence considers the amount of new disease cases occurring within a population over a specified period of time (5). Consanguinity greatly increases the incidence of AR IEM. The clinical impact on offspring of consanguineous partnerships includes fertility challenges, pregnancy loss and complications, and biochemical complications manifesting as IEM (5,6).

Population migration patterns to and across Africa has given rise to subsequent founder effects and genetic drift, potentially enhanced by bottleneck effects from climate challenges and infectious disease (8). Founder mutations have been reported in different ancestral populations within Southern Africa. Some examples include, the autosomal dominant p.R59W variant of the *PPOX* gene (encoding protoporphyrinogen oxidase) causing variegate porphyria predominantly in the Afrikaner population, traced back to Dutch settlers (9); the autosomal recessive p.S135L variant in the *GALT* gene (encoding galactose 1-phosphate uridyltransferase) causing galactosaemia in black South African individuals (10,11); and at least three variants in the low density lipoprotein receptor (LDLR) have been shown to give rise to familial hypercholesterolaemia (FH) in the Afrikaner population (12,13).

Detection of pathogenic variants in affected patients is helpful for confirmation of the diagnosis. Importantly, genetic testing can unambiguously identify members of the family who are carriers. These members could benefit from early intervention in AD disorders, or identify those at risk for AR disorders. This is especially relevant when an appreciable risk exists of homozygosity owing to a high prevalence of one or few single nucleotide variants (SNV) in a given population.

Population-based carrier frequency determination is challenging (14). A modern approach to determining the prevalence of a SNV involves retrospectively analysing existing genomic data of at-risk groups. These are generated through high-cost methods, such as next-generation sequencing (NGS) (15). This is currently not feasible in resource constrained environments due to the lack of existing large population genomic data sets. However, carrier frequencies have been determined in the South African context using less expensive methods.

Carrier frequency data for specific IEM in the sub-Saharan African population is available and includes the following: Gaucher disease, caused by the c.222-224delTAC(p.T36del) pathogenic variant in *GBA* (Carrier frequency of 1 in 66 random black individuals); Glutaric aciduria type 1 (GA1), caused by the c.877G>A(p.A293T) pathogenic variant in *GCDH* (carrier frequency of 1 in 36 black South Africans in the Western Cape); Cystinosis, caused by the c.971-12G>A pathogenic variant in *CTNS* (carrier frequency of 1 in 50 black South Africans in the Western Cape); Galactosaemia, caused by the c.404C>T(p.S135L) pathogenic variant in *GALT* (carrier frequency of 1 in 60 black South Africans in the Western Cape); Mitochondrial DNA depletion syndrome 6, caused by the c.106C>T(p.Gln36Ter) pathogenic variant in *MPV17* (carrier frequency of 1 in 68 black South Africans in the Western Cape) (11,16–19). See **Table 1**.

Table 1. Carrier Frequency data for pathogenic variants of IEM identified in sub-Saharan Africa. Each disorder was identified in a black South African population.

IEM Name	**Gene affected**	**Pathogenic Variant**	**Carrier Frequency**
Gaucher disease	*GBA* gene	c.222-224delTAC(p.T36del)	1:66
Glutaric aciduria type 1	*GCDH* gene	c.877G>A(p.Ala293Thr)	1:36
Cystinosis	*CTNS* gene	c.971-12G>A	1:50
Galactosaemia	*GALT* gene	c.404C>T(p.Ser135Leu)	1:60
Mitochondrial DNA depletion syndrome 6	*MPV17* gene	c.106C>T(p.Gln36Ter)	1:68

These IEM are related to local populations, but there is a possibility that they are relevant in the broader sub-Saharan African setting. Therefore, understanding the African migratory patterns may assist in identifying at risk groups.

1.2 Southern African Population

Three migrations of distinct indigenous populations contributed significantly to Southern African diversity, along with other lesser immigrations from India and the East. This admixture makes determining primordial variants, based on race and ethno-linguistic background difficult, although such division has had practical application.

The indigenous Khoe and San (Khoe-San) populations mark the earliest diversification event in Southern Africa with significant mitochondrial lineages and genomic diversity (20). Ancestry of this group arose from admixture, even though no geographical origin is identified. Despite differences within these groups, common ancestry can be traced back to the early divergence time ~110 000 years ago (21).

Another broad group in Southern Africa is identified linguistically (Bantu-speaking groups) (22). The Bantu expansion was the second migration, originating from West to Central Africa. Initial migration covered eastern and southern expansion across the rainforest. The southern migration further split into eastern and then southern expansions, with distinct linguistic and genetic characteristics (23). Continued southern migration of the eastern expansion gave rise to the south-eastern branch of the Bantu group. Of the eleven official South African languages, nine are from this south-eastern branch, with two distinct clusters being Nguni (Zulu, Xhosa, Swazi and Ndebele) and Sotho-Tswana (includes Sotho, Tswana and Pedi) (20,21).

The third migration was during the colonial era. This introduced European ancestral DNA to Southern Africa. However, this additionally caused inter-continental genomic admixture from the slave trade (including DNA from Indian populations) (23). Trans-Atlantic slave trading contributed significantly to this displacement, giving rise to African diasporas in Latin-America and spreading northward (24).

The predominant migratory group in Southern Africa is the Bantu-speaking group. Primordial variants from this group may occur frequently across many geographical locations. This is possibly true for other populations in the wider African diaspora, including dispersions to the Americas from slave trading. This dispersion of primordial variants is suggested by the identification of the same pathogenic variant causing HFTC in our population and in an African-American family (25).

According to the 2011 census data, the predominant South African population group is Black African, making up 79.2% (41 000 938 persons) (26). Subdivision of this population group by first language reveals the three largest groups as: AmaZulu 28.5% (11 519 234 persons), AmaXhosa 20.1% (8 104 752 persons) and Sepedi 11.4% (4 602 459 persons). Sesotho and Setswana have similar proportions at 9.45% (3 798 915 persons) and 9.9% (3 996 951 persons), respectively. Smaller groups in order of predominance are Xitsonga, Siswati, Tshivenda, IsiNdebele (26). The related Bantu ethnic groups of AmaZulu, AmaXhosa, Siswati and IsiNdebele are often classed as Nguni. The regional distribution of these groups based on language can be seen in **Figure 2.** Variable interactions between these groups contributed to the diverse admixture in the population. This increased with urbanization, especially relating to the establishment of mines.

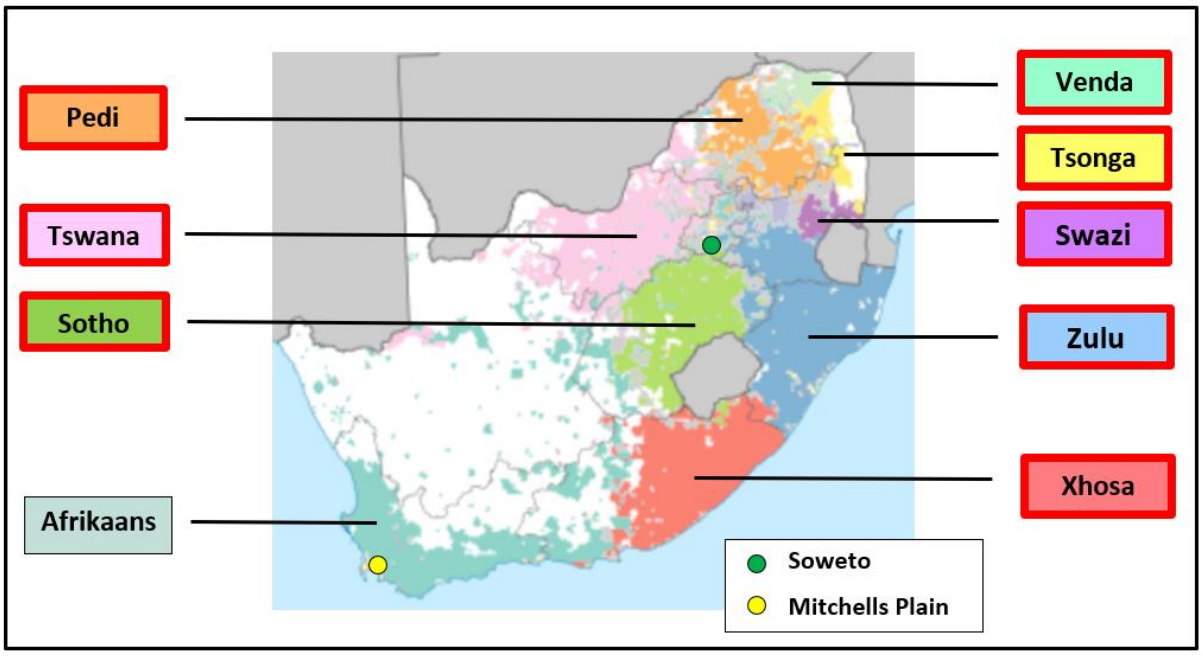

Figure 2. Map showing the dominant languages within regions of South Africa. Bantu-speaking groups from the south-eastern branch of the southern migration are labelled with a red border. Highly mixed venues are shown in coloured circles: Soweto in green, Mitchells Plain in yellow. The original map was obtained from: https://en.wikipedia.org/wiki/Languages_of_South_Africa.

Healthcare in South Africa is often sub-optimal amongst historically disadvantaged indigenous groups. Factors contributing to this include physical accessibility, financial accessibility, and levels of education (27,28). Improving knowledge of disorders affecting specific population groups may improve individual and societal care. This affirms the need to elucidate pathogenic variants transmitted within these groups to minimise further barriers to care.

Current identification of likely-pathogenic, and possibly variants of unknown significance (VUS), arises from referencing data of well-studied groups (29). Most available data on individuals of African ancestry are from West African rather than sub-Saharan African populations (30,31). This data is not always transferable to understudied populations. This paucity of knowledge highlights the need to research disorders identified within discrete regions. Pathogenic variants identified in black South African patients will be discussed in the next section. Thereafter, the analytical methodology and statistical analysis to determine carrier frequency will be described.

1.3 Description of Four IEM Identified in the Local Black Population

Health services in South Africa are struggling to investigate and diagnose uncommon, severe metabolic disorders. It is likely that only a small proportion of these presentations lead to a diagnosis. After identifying the pathogenic mutation in patients, establishing the prevalence of the carrier status can significantly influence health care planning and future diagnostic approaches.

In the following sections four IEM resulting from pathogenic variants, identified in a homozygous state in 1 or more patients, are described in more detail. These pathogenic variants were detected at the National Health Laboratory Service (NHLS) Inherited Metabolic Disease (IMD) molecular diagnostic laboratory at Groote Schuur Hospital in the Western Cape of South Africa.

1.3.1 Hyperphosphataemic Familial Tumoral Calcinosis (HFTC) (OMIM: 211900)

Fibroblast growth factor 23 (FGF23) maintains phosphate and calcitriol homeostasis by controlling their excretion and synthesis, respectively (32). Intact FGF23 (Bioactive form) down-regulates the membrane expression of two sodium/phosphate cotransporters (NPT), NPT2a and NPT2c, in the proximal convoluted tubule of the kidney. This results in decreased phosphate reabsorption, and thus phosphaturia (33). Dysregulation in this system leads to HFTC.

HFTC is a rare autosomal recessive disorder caused by defects in one of three genes involved in FGF23 physiology, namely, *FGF23*, *GALNT3* and *αKlotho*. Loss of function mutations in *FGF23* gene and *GALNT3* gene (GalNac-transferase 3) decreases the secretion of intact FGF23 and activation of precursor FGF23 to intact FGF23, respectively. Specifically, this impaired function in the UDP-N-acetyl-alpha-*D*-galactosamine:polypeptide N-acetylgalactoseaminyltransferase 3 (GALNT3) enzyme prevents O-glycosylation of the FGF23 precursor to its bioactive form, as illustrated in **Figure 3** (41)**.** Pathogenic variants in *αKlotho* gene prevent interaction of intact FGF23 with the FGF receptor preventing its end-organ effect (35).

HFTC is characterised by impaired phosphate regulation predisposing to calcific deposits. These deposits are found in skin and subcutaneous tissue, typically occuring around articular regions. The clinical findings appear after longstanding hyperphosphataemia. Other systems affected by these deposits include the opthalmologic, vascular and renal systems (37).

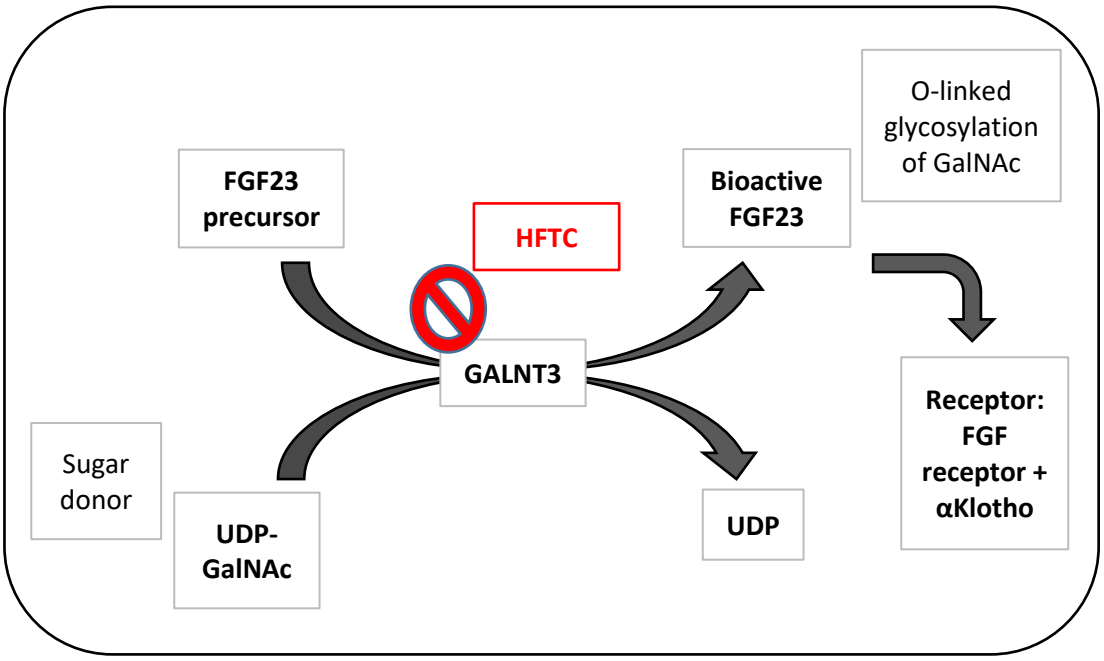

Figure 3. Pathophysiology of HFTC. GALNT3 catalyses the process of O-glycosylation of GalNAc from its sugar donor, UDP-GalNAc, onto serine or threonine residues of FGF23 precursors. The glycosylated form of FGF23 renders the protein bioactive. FGF23 binds to FGF receptor and the co-receptor αKlotho, controlling phosphate metabolism. Inactivation of *GALNT3* impairs this action leading to inadequate bioactive FGF23 (34,35). Figure adapted from Farrow et al. (36)

Two unrelated patients were diagnosed at the NHLS IMD laboratory, using NGS. Both patients were homozygous for the known c.484C>T(p.Arg162Ter) pathogenic variant in the *GALNT3* gene. The same variant was previously described in a multigenerational analysis of an African American family. The c.484C>T(p.Arg162Ter) *GALNT3* nonsense variant has been shown to terminate protein translation prematurely, leading to an absence of bioactive FGF23 in serum (25).

Based on the different geographical and cultural backgrounds of the two local patients (Ugandan and Eastern Cape Xhosa heritage) and the previously described African American family, the pathogenic c.484C>T(p.Arg162Ter) *GALNT3* variant likely arose from either a primordial ancestor, or is a recurring variant (38). The variant was noted by the 1000 Genomes project and was reported to have an allele frequency of 0.0008 in African Americans and a frequency of 0.0002 worldwide (rs137853086) (30).

1.3.2 Succinyl-CoA:3-Ketoacid CoA Transferase (SCOT) Deficiency (OMIM: 245050)

Ketoacidosis is a common metabolic complication of several disorders, often in individuals with type 1 diabetes. Increased blood ketones are a protective counter-regulatory response to depleted energy states. The SCOT enzyme catalyses the transfer of CoA between succinyl-CoA and acetoacetate, producing acetoacetyl-CoA and succinate. Acetoacetyl-CoA is further metabolised to form acetyl-CoA for mitochondrial energy generation (39).

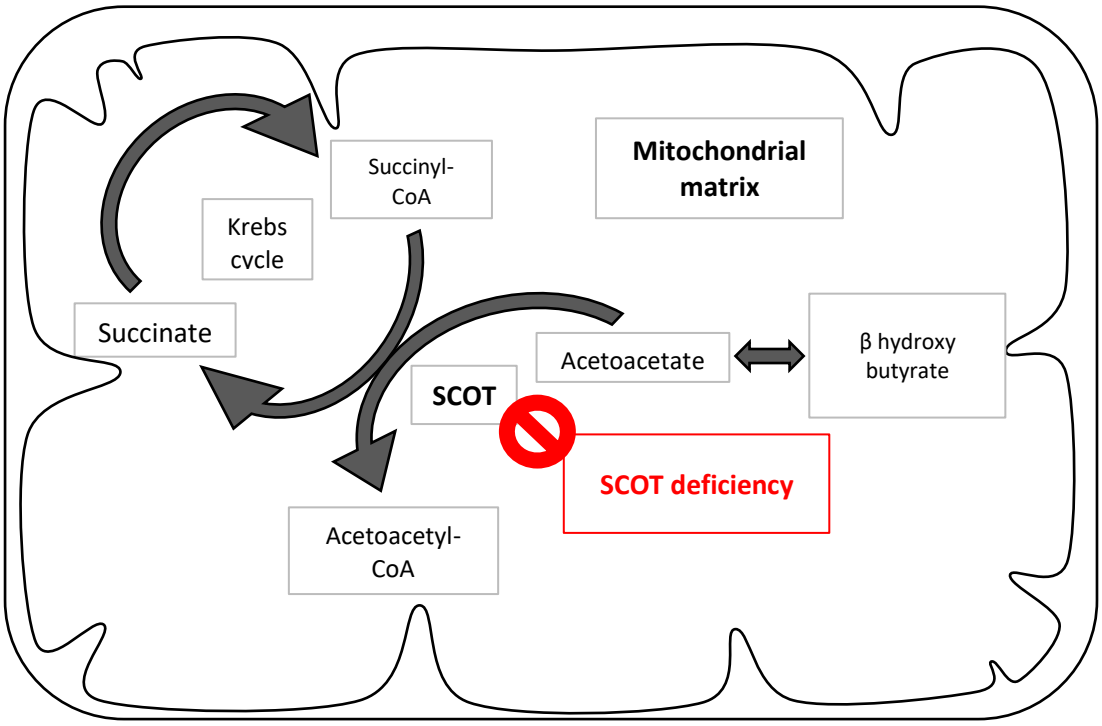

Figure 4. Pathway affected in SCOT deficiency. The succinyl CoA:3-ketoacid CoA transferase (SCOT) enzyme, within the mitochondrial matrix, converts β-hydroxybutyrate derived acetoacetate to acetoacetyl-CoA, for further metabolism by mitochondrial acetoacetyl-CoA thiolase to produce acetyl-CoA. SCOT deficiency prevents this CoA transfer, leading to accumulation of acetoacetate. Figure adapted from Fukao et al. (40,41).

SCOT deficiency is characterised by recurrences of ketoacidosis. However, persisting ketosis is rarely encountered (42). The accumulation of ketone bodies leads to an increased anion gap metabolic acidosis (43). Ongoing permanent ketosis, manifesting as tachypnoea, poor feeding, vomiting and lethargy, is a strong indicator of dysfunction in the SCOT enzyme (41).

Pretorius et al. described two South African siblings with SCOT deficiency in 1996. Further investigations identified the genetic cause as the c.803G>A(p.Arg268His) variant in *OXCT1* gene (44). This variant has not been investigated for carrier frequency in any population (41). More recently, a third unrelated patient was diagnosed through the NHLS IMD laboratory with the same homozygous pathogenic variant as the two index cases [unpublished data].

1.3.3 Glycogen Storage Disease Type 1A (GSD 1A, von Gierke's Disease) (OMIM:232200)

GSD comprises more than 15 genetically distinct causal entities that lead to impaired carbohydrate metabolism. The clinical phenotype arises due to cellular accumulation of glycogen from impaired breakdown, and impaired energy supply. The organs affected can be the liver, muscle or both (38,39).

Accumulation of glycogen in the liver and kidneys was first described in 1929 by von Gierke; while impaired glucose-6-phosphatase enzyme activity was demonstrated by Cori in 1959 (45). Four subtypes of GSD Type 1 (1a, 1b, 1c, and 1d) have been described, all of which are due to the inability to convert glucose-6-phosphate to glucose (47,48). These subtypes share similarities, with the most critical clinical manifestation being hypoglycaemia. Neutropenia is an important defining feature in GSD 1b (49). GSD type 1c and 1d are rare, with GSD type 1a occurring the most frequently of the subtypes (47).

The GSD 1a subtype, also known as Von Gierkes disease, is caused by pathogenic variants in the *G6PC* gene. This gene codes for the glucose-6-phosphatase enzyme, located on the endoplasmic reticulum membrane (50). This enzyme is responsible for converting glucose-6-phosphate to glucose at the end of the glycogenolysis and gluconeogenesis pathways, as illustrated in **Figure 5** (51). A notable clinical finding in GSD 1a is hepatomegaly (52). Biochemical manifestations accompanying hypoglycaemia include hypertriglyceridemia, lactic acidosis and hyperuricaemia (45,47,53).

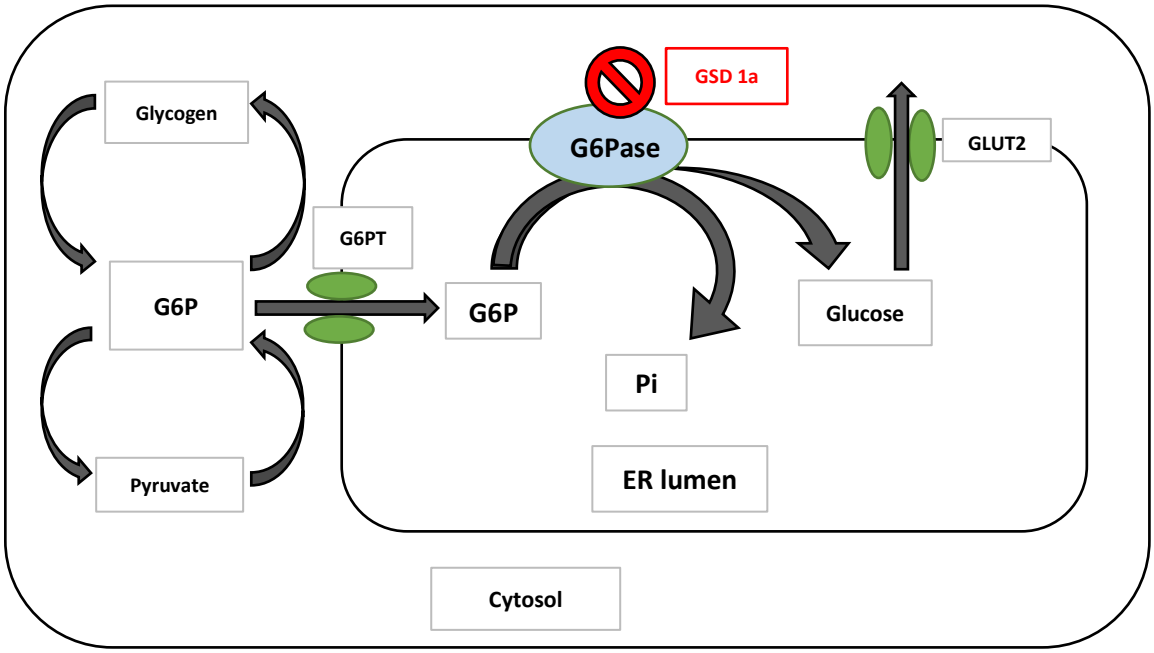

Figure 5. Pathway affected in GSD 1a. Glucose-6-phosphate (G6P) in the cytosol is transported by the glucose-6-phosphate transporter (G6PT) into the endoplasmic reticulum (ER) lumen. The ER membrane bound enzyme, glucose-6-phosphatase (G6Pase), catalyses the hydrolysis of G6P to glucose for further utilisation. GLUT2 (Glucose transporter 2) (51). Figure adapted from Chou et al. (54).

GSD 1a was confirmed in two unrelated black South African patients from non-consanguineous families. Both cases were homozygous positive for the known c.189G>A(p.Trp62Ter) pathogenic variant in the *G6PC* gene (rs764920787). The Genome Aggregation Database (gnomAD) analysed a cohort of 251420 samples and reported a worldwide allele frequency of 0.000004 (55).

1.3.4 Multiple Mitochondrial Dysfunction Syndrome Type 2 (MMDS 2) (OMIM: 614299)

Pyruvate dehydrogenase complex (PDHC) deficiency is a group of disorders related to pyruvate metabolism, manifesting clinically with neurodegenerative changes and hyperlactataemia. Five genes encoding the core subunits, along with at least 20 other genes, have been implicated in characteristic clinical presentations. Pathogenic variants in the *PDHA1* gene account for most cases in western populations but has never been reported in South African patients (56,57).

A PDHC deficiency gene panel (including *PDHA1, PDHB, DLAT, DLD, PDHX, BOLA3, GLRX5, IBA57, LIAS, LIPT1, LIPT2, NFU1, PDP1, PDP2, SLC19A2, SLC19A3, SLC25A19. SLC25A26, TPK1* and *FBXI4*) was analysed in five South African patients with low/absent PDHC enzyme activity. One patient was found to carry a novel homozygous pathogenic variant in the *BOLA3* gene, c.159dupT(p.Asp54*), causing a lipoic acid biosynthesis defect, MMDS type 2. Although a single patient was identified with this variant, true homozygosity in the setting of non-consanguinity suggests a higher presence of the pathogenic variant within the population (57).

The association between defects in PDHC and pathogenic variant in the *BOLA3* gene was first described by Haack et al. (58). MMDS 2 is a rare, autosomal recessively inherited IEM arising from dysfunctional biosynthesis of iron-sulphur (Fe-S) clusters. Fe-S clusters are required for lipoic acid biosynthesis and form an essential component of mitochondrial respiratory chain complexes (complex I, and II and III). The synthesis of lipoate or lipoylated proteins is affected by these defects (59,60). Lipoic acid is an integral part of various enzymes, including the 2-oxoacid dehydrogenase complexes: PDHC, 2-KGDH, 2-OADH, BCKDH, as well as the glycine cleavage system (GCS). The actual function of human BOLA3 protein is unclear, but inference is made from studies showing interactions between BOLA3 and the mono-thiol glutaredoxin family. The combined actions of BOLA3 and glutaredoxin 5 allow insertion of [2Fe-2S] and [4Fe-4S] clusters into apoproteins with the assistance of chaperones, illustrated in **Figure 6** (59,61).

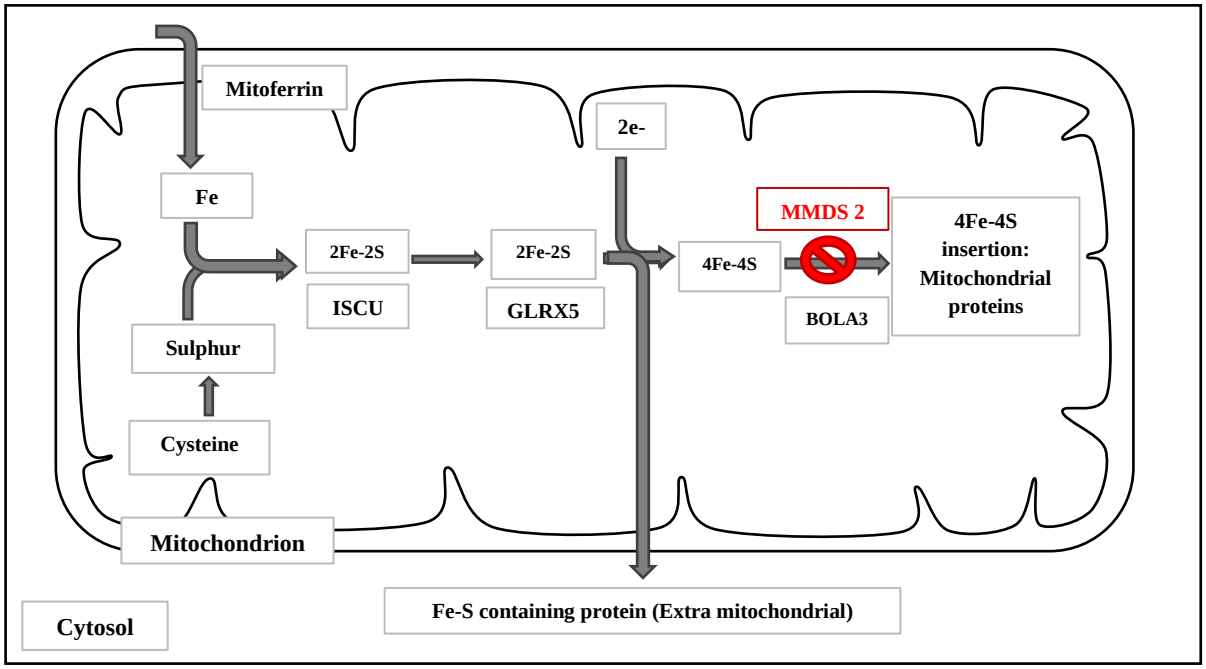

Figure 6. Pathway affected in MMDS 2. Fe-S protein maturation with *BOLA3* gene pathogenic variation causing MMDS 2. Mitochondrial entry of iron (Fe) through mitoferrin on the inner mitochondrial membrane with 2Fe-2S cluster assembly on the cluster scaffold protein (ISCU). The cluster transfers to glutaredoxin 5 (GLRX5). Transfer of 2 electrons (2e-) on 2Fe-2S cluster synthesizes 4Fe-4S cluster. BOLA3 protein action is involved in maturation of mitochondrial 4Fe-4S containing proteins. Figure adapted from Cameron et al. (59).

The clinical presentation Includes failure to thrive and varied neurological findings: severe encephalopathy, seizures, developmental delay, hypotonia and psychomotor delay with impaired movement. Additionally, pulmonary hypertension and cardiomyopathy may occur (62). The typical biochemical finding in these patients is a raised anion gap metabolic acidosis, as well as non-ketotic hyperglycinaemia (63).

1.4 Carrier Frequency and Associated Laboratory Approach

Determining the carrier frequencies of pathogenic variants giving rise to severe disease in a population can improve healthcare for the individual and society by guiding decisions on distribution of resources, and by implementing newborn screening (NBS) and genetic counselling programmes. Relying only on clinical suspicion to test for rare genetic variants may be costly and delay intervention (64). Progress has been made in describing carrier frequencies for other IEM in patients of mainly black African ancestry. As described previously, carrier frequencies have been described for Gaucher disease, galactosaemia, GA1, cystinosis and mitochondrial neurohepatopathy. Newly identified, true homozygous variants occurring in more than one unrelated individual with regional similarities justifies analysis for carrier frequency (8,11,17–19).

The four IEM, caused by the variants described in the previous section, have unique clinical features but include non-specific symptoms that overlap with more common disorders. These clinical features should be considered when formulating differential diagnoses. Education that enhances awareness and appropriate investigation of rare IEM requires that certain variables are established. These include demographics, ancestry, and carrier frequencies for pathogenic variants within the population. Another approach to determine frequency of disease alleles is the homozygosity index, and is useful in populations where the impact of consanguinity is known (65). This index requires the inbreeding coefficient of the population, which is not currently necessary for this study. Therefore, only carrier frequency will be considered further.

In AR disorders, the frequency at which a pathogenic variant is present in a heterozygous state in individuals from a population, is considered the carrier frequency (66). This can be calculated using the Hardy-Weinberg equation, represented as $p^2+2pq+q^2$. The parameters p^2 and q^2 represent the frequency of the homozygous wild type and pathogenic genotypes, respectively. The frequency of the heterozygous genotype is represented by 2pq (67). From the prevalence of the homozygotes, the carrier frequency can be calculated using this equation, assuming the population to be at equilibrium (72,69).

.

A review of carrier frequency data for the four IEM identified a paucity of information. An internet search for carrier frequency data on GSD 1a yielded few articles. Two articles were found on Pubmed, one of which identified the carrier frequency for a specific pathogenic variant in Ashkenazi Jews (69). The other reported the voluntary carrier screening of a Japanese cohort for a point mutation, in which one individual was detected out of 216 (70). However, neither article related to the black African population, or the c.189G>A(p.Trp62Ter) pathogenic variant identified in the patients described previously. Pubmed and Google searches failed to identify articles reporting carrier frequencies of HFTC, SCOT deficiency and MMDS 2. As described in the respective sections, analysis of data from large databases reported frequencies for HFTC and GSD. These where, however, reports on mainly African Americans. This scarcity in carrier frequency data for populations of sub-Sahran African ancestry thus requires further research.

Progress in technical expertise and cost reductions using NGS enables economically developed countries to study frequently encountered diseases, including those within established NBS programmes in the context of established population genomic databases (14). A challenge encountered with carrier frequency determination is the low rate at which AR inherited diseases occur. This low rate requires the testing of large numbers to accurately determine variant frequencies in low- and middle-income countries.

Large reference genomic data sets include the 1000 Genomes Project, 100 000 Genomes Project and the Exome Aggregation Consortium (ExAC) browser (31,71,72). These data sets comprise sequenced genomes from multiple populations of diverse sets of individuals. Despite African representation in these data sets, the investigated populations are often from discrete groups, such as those of African American descent in the 1000 Genomes project. Therefore, a genomic profile of predominantly west-African ancestry is likely, which may not be applicable in all African groups. More recent projects covering larger African populations exist, including the Human Hereditary and Health in Africa (H3Africa) initiative, but data generation is ongoing and may not be sufficient for IEM (73). The result is a paucity of IEM carrier frequency and prevalence data from sub–Saharan Africa.

Characteristics of effective screening tests to detect carrier frequency of single variants include high sensitivity, cost-effectiveness, and ease of interpretation (74). Methods capable of achieving some or most of these objectives include high resolution melt (HRM) curve analyses, single-strand conformation polymorphism (SSCP), restriction fragment length polymorphisms (RFLP), and amplification refractory mutations system (ARMS) PCR. Despite the need to test multiple variants simultaneously, the risk for assay failure exists especially in multiplexed assays. Therefore, analysing the four variants individually may still be required in follow-up studies.

HRM is considered sensitive and economical. PCR is performed with dyes that fluoresce when incorporated into double stranded DNA. Thereafter, DNA is analysed by evaluating melting profiles using real-time PCR. Mutations within different fragments of double stranded DNA causes variation in the temperature at which each fragment will separate. The incremental increases in temperature and dissociation will gradually decrease the fluorescence signal detected (75,76). Primer design, favouring short amplicons, mitigates interference from too many variants. However, this may require further investigations to confirm a specific variant of interest.

SSCP is used to detect differences in the way that single stranded DNA fragments fold back onto itself when allowed to quickly re-anneal. The fragments are obtained through denaturation. Differences in the folding patterns are introduced by sequence variations. Products are visualized on non-denaturing acrylamide gel electrophoresis (77). Parameters effecting the sensitivity of SSCP include mutation type, amplicon size, GC content of the fragment (affecting melting temperature), composition of the gel matrix (including percentage of polyacrylamide or the composition of other gel matrices), size of the gel used, electrophoresis temperature and time, DNA concentration, and composition of the buffers (including additives) used (77,78). Strictly controlled conditions are therefore needed for successful SSCP analysis. SSCP is technically simple and has a high sensitivity for detecting mutations, making it useful for screening (79). However, disadvantages to using SSCP include the availability of PCR methods better suited to our setting such as RFLP; complicated staining methods using silver; prolonged electrophoresis running times; and restrictions to amplicon sizes (80). This method is time-consuming, and outdated, and is therefore not a feasible option for this study. Furthermore, the method is not definitive and thus requires confirmatory testing.

RFLP relies on the ability of restriction endonucleases to recognise and cut specific palindromic sequences in double stranded DNA (81). DNA sequence variants can either create or abolish a restriction site, resulting in different DNA fragment lengths on gel electrophoresis (82). This method is widely used, simple, cost-effective, and sensitive for screening. Additionally, due to its frequent use in our setting, it is easy to implement. However, RFLP is disadvantaged by the slower turn-around time and tedious performance when compared to newer methods such as fragment analysis. These, however, require more advanced equipment (83). Furthermore, multiplexing can be difficult and performing individual reactions can be time consuming when screening large population cohorts.

The principle of ARMS PCR relies on the amplification of DNA using allele specific primers. The terminal 3'-nucleotide of one of a pair of primers is specific for a target allele. The absence of a complementary nucleotide on the template DNA prevents PCR amplification and absence of product on electrophoresis (84). Enhancements to the assay, such as inclusion of additional mismatches in the 3'-terminus may improve specificity (62,9). ARMS PCR requires less hands-on steps than RFLP and if well optimized can be multiplexed. Examples of the use of ARMS PCR for screening of specific genetic variants include work done on cystic fibrosis (*CFTR* gene) and galactosaemia (*GALT* gene) (11,86). It is potentially suited to resource constrained environments (86). This method can be incorporated into screening programs such as the development of diagnostic panels based on clinical phenotype.

1.5 Conclusion

Although individual IEM present with relatively low frequency, the collective incidence is estimated to be 1 in 2500-5000 live births (87,88). The significant clinical burden that these IEM place on medical systems, and the debilitating symptoms, often with fatal outcomes, make determining carrier frequencies of common pathogenic variants in specific population subgroups important for individuals and society as a whole (89).

Determining the carrier frequencies of specific variants in local populations may mitigate the impact of these disorders, allowing for additional interventions such as genetic counselling, prenatal testing, and carrier screening for family members. Awareness of pathogenic variants in the population allows timeous diagnosis and expedient intervention, avoiding disease complications and their management. Additionally, expenses are minimised by avoiding unnecessary diagnostics and expensive treatments, including ICU (3,90). The carrier frequency, once determined in this study, may be useful for the application of guidelines for screening couples from the given population to advise on the risk of conceiving affected offspring (91–93).

Although access to advanced methodologies and increasing availability of large genomic data sets for multiple populations are rapidly improving the ability to assess variant frequencies in well studied populations, their application or availability in South African populations is limited. A simple, cost-effective, and robust method such as ARMS PCR, is therefore attractive. It is useful in screening for known variants, while cost-effectively testing large cohorts to determine population carrier frequencies. This study will assess the feasibility of multiplex ARMS PCR assays for carrier frequency determination of 4 IEM causing variants in particular, but will indicate the potential for general application with any future variants of interest.

The IEM of interest, namely HFTC, SCOT deficiency, GSD 1A and MMDS 2, present with distinct clinical phenotypes. Therefore, if the causal variants are determined to be significantly common by this study, relevant targeted tests can be employed based on the clinical and/or biochemical features of presenting cases, and cascade testing can be advised. This may provide further benefit to healthcare, with additional research opportunities.

REFERENCES

1. Waters D, Adeloye D, Woolham D, Wastnedge E, Patel S, Rudan I. Global birth prevalence and mortality from inborn errors of metabolism: A systematic analysis of the evidence. J Glob Health. 2018;8(2).

2. Jeanmonod, Rebecca; Asuka, Edinen; Jeanmonod D. Inborn Errors Of Metabolism [Internet]. statpearls publishing LLC. 2020 [cited 2021 Apr 25]. Available from: https://www.ncbo.nlm.nih.gov/books/NBK459183/#_NBK459183_pubdet_

3. Arnold GL. Inborn errors of metabolism in the 21st century: past to present. Ann Transl Med. 2018;6(24):467–467.

4. Guerrero RB, Salazar D, Tanpaiboon P. Laboratory diagnostic approaches in metabolic disorders. Ann Transl Med. 2018;6(24):470–470.

5. Longmore M, Wilkinson IB, Rajagopalan S. Oxford Handbook of Clinical Medicine. 6th ed. New York: Oxford University Press Inc.; 2004. 23 p.

6. Oniya O, Neves K, Ahmed B, Konje JC. A review of the reproductive consequences of consanguinity. Vol. 232, European Journal of Obstetrics and Gynecology and Reproductive Biology. Elsevier Ireland Ltd; 2019. p. 87–96.

7. Afzal RM, Lund AM, Skovby F. The impact of consanguinity on the frequency of inborn errors of metabolism. Mol Genet Metab Reports. 2018 Jun 1;15:6–10.

8. Krause A, Seymour H, Ramsay M. Common and founder mutations for monogenic traits in sub-saharan African populations. Annu Rev Genomics Hum Genet. 2018;19:149–75.

9. Meissner PN, Dailey TA, Hift RJ, Ziman M, Corrigall A V, Roberts AG, et al. A R59W mutation in human protoporphyrinogen oxidase results in decreased enzyme activity and is prevalent in South Africans with variegate porphyria. Nat Genet. 13(1):95–7.

10. Hift R, Meissner P, Corrigall A, Ziman M, Petersen L, Meissner D, et al. Variegate porphyria in South Africa, 1688-1996--new developments in an old disease. S Afr Med J [Internet]. 87(6):722–31. Available from: https://pubmed.ncbi.nlm.nih.gov/9254745/

11. Henderson H, Leisegang F, Brown R, Eley R. The clinical and molecular spectrum of galactosemia in patients from the Cape Town region of South Africa. BMC Pediatr. 2002;2:7.

12. Kotze MJ, Langenhoven E, Warnich L, Du Plessis L, Marx MP, Oosthuizen CJJ, et al. The identification of two low-density lipoprotein receptor gene mutations in South African familial hypercholesterolaemia. South African Med J. 1989;76(8):399–401.

13. Leitersdorf E, Van Der Westhuyzen DR, Coetzee GA, Hobbs HH. Two common low density lipoprotein receptor gene mutations cause familial hypercholesterolemia in Afrikaners. J Clin Invest. 1989;84(3):954–61.

14. Yamaguchi-Kabata Y, Yasuda J, Uruno A, Shimokawa K, Koshiba S, Suzuki Y, et al. Estimating carrier frequencies of newborn screening disorders using a whole-genome reference panel of 3552 Japanese individuals. Hum Genet [Internet]. 2019;138(4):389–409. Available from: http://dx.doi.org/10.1007/s00439-019-01998-7

15. Collet C, Laplanche JL, Page J, Morel H, Woimant F, Poujois A. High genetic carrier frequency of Wilson's disease in France: Discrepancies with clinical prevalence. BMC Med Genet. 2018;19(1):1–6.

16. Arndt S, Heitner R, Lane A, Ramsay M. Glucocerebrosidase gene mutations in black South Africans with Gaucher disease. Blood Cells, Mol Dis [Internet]. 2009;43(1):129–33. Available from: http://dx.doi.org/10.1016/j.bcmd.2009.02.008

17. van der Watt G, Owen EP, Berman P, Meldau S, Watermeyer N, Olpin SE, et al. Glutaric aciduria type 1 in South Africa-high incidence of glutaryl-CoA dehydrogenase deficiency in black South Africans. Mol Genet Metab [Internet]. 2010;101(2–3):178–82. Available from: http://dx.doi.org/10.1016/j.ymgme.2010.07.018

18. Owen EP, Nandhlal J, Leisegang F, Van der Watt G, Nourse P, Gajjar P. Common mutation causes cystinosis in the majority of black South African patients. Pediatr Nephrol. 2015;30(4):595–601.

19. Meldau S, De Lacy RJ, Riordan GTM, Goddard EA, Pillay K, Fieggen KJ, et al. Identification of a single MPV17 nonsense-associated altered splice variant in 24 South African infants with mitochondrial neurohepatopathy. Clin Genet. 2018;93(5):1093–6.

20. Schlebusch CM, Skoglund P, Sjödin P, Gattepaille LM, Hernandez D, Jay F, et al. Genomic Variation in Seven Khoe-San Complex African History. 2012;1187(October):374–9.

21. Veeramah KR, Wegmann D, Woerner A, Mendez FL, Watkins JC, Destro-Bisol G, et al. An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. Mol Biol Evol. 2012;29(2):617–30.

22. Sengupta D, Choudhury A, Fortes-Lima C, Aron S, Whitelaw G, Bostoen K, et al. Genetic substructure and complex demographic history of South African Bantu speakers. Nat Commun. 2021 Dec 1;12(1).

23. Petersen DC, Libiger O, Tindall EA, Hardie RA, Hannick LI, Glashoff RH, et al. Complex Patterns of Genomic Admixture within Southern Africa. PLoS Genet. 2013;9(3):10–3.

24. Rotimi CN, Tekola-Ayele F, Baker J, Shriner D. The African Diaspora: History, Adaptation and Health. Curr Opin Genet Develoment. 2016;41:77–84.

25. Ichikawa S, Lyles KW, Econs MJ. Rapid communication: A novel GALNT3 mutation in a pseudoautosomal dominant form of tumoral calcinosis: Evidence that the disorder is autosomal recessive. J Clin Endocrinol Metab. 2005;90(4):2420–3.

26. Lehohla P. Census 2011: Census in brief [Internet]. Pretoria; 2022. Available from: http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf

27. Nyang'ori O, Ruth W, Dorothy J, Clive N, Kenneth G, Mugarura B. Health of indigenous people in Africa. Lancet. 2006;367(9526):1937–46.

28. Horrill T, McMillan DE, Schultz ASH, Thompson G. Understanding access to healthcare among Indigenous peoples: A comparative analysis of biomedical and postcolonial perspectives. Nurs Inq. 2018;25(3):1–9.

29. Choudhury A, Aron S, Botigué LR, Sengupta D, Botha G, Bensellak T, et al. High-depth African genomes inform human migration and health. Nature. 2020;586(7831):741–8.

30. The 1000 genomes Project Consortium. A Global reference for human genetic variation. Nature. (526):68–74.

31. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, et al. The ExAC browser: Displaying reference data information from over 60 000 exomes. Nucleic Acids Res. 2017;45(D1):D840–5.

32. Prié D, Friedlander G. Reciprocal control of 1,25-dihydroxyvitamin D and FGF23 formation involving the FGF23/Klotho system. Clin J Am Soc Nephrol. 2010;5(9):1717–22.

33. Jüppner H. Phosphate and FGF-23. Kidney Int. 2011;79(SUPPL. 121):25–8.

34. Topaz O, Shurman DL, Bergman R, Indelman M, Ratajczak P, Mizrachi M, et al. Mutations in GALNT3, encoding a protein involved in O-linked glycosylation, cause familial tumoral calcinosis. Nat Genet. 2004;36(6):579–81.

35. Farrow EG, Imel EA, White KE. Hyperphosphatemic familial tumoral calcinosis (FGF23, GALNT3 and αKlotho). Best Pract Res Clin Rheumatol [Internet]. 2011 Oct;25(5):735–47. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1521694211000957

36. Bennett EP, Hassan H, Clausen H. cDNA cloning and expression of a novel human UDP-N-acetyl-α-D- galactosamine. Polypeptide N-acetylgalactosaminyltransferase, GalNAc-T3. J Biol Chem [Internet]. 1996;271(29):17006–12. Available from: http://dx.doi.org/10.1074/jbc.271.29.17006

37. Boyce AM, Lee AE, Roszko KL, Gafni RI. Hyperphosphatemic Tumoral Calcinosis: Pathogenesis, Clinical Presentation, and Challenges in Management. Front Endocrinol (Lausanne). 2020;11(May):1–9.

38. Jorde LB, Wooding SP. Genetic variation, classification and "race." Nat Genet. 2004;36(11 SUPPL. 1):28–33.

39. Zhang M, Xu HY, Wang YC, Shi ZB, Zhang NN. Structure of succinyl-CoA:3-ketoacid CoA transferase from Drosophila melanogaster. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2013;69(10):1089–93.

40. Fukao T, Song X, Mitchell G, Yamaguchi S, Sukegawa K, Orii T, et al. Enzymes of Ketone Body Utilization in Human Tissues: Protein and Messenger RNA in Human Tissues: Protein and Messenger RNA Levels of Succinyl-Coenzyme A (CoA):3-Ketoacid CoA Transferase and Mitochondrial and Cytosolic Acetoacetyl-CoA Thioloases. Pediatr Res. 1997;42:498–502.

41. Fukao T. Succinyl-CoA : 3-ketoacid CoA transferase (SCOT) deficiency. 2001. p. 1–6.

42. Rare Disease InfoHub. Succinyl-CoA : 3-ketoacid CoA transferase deficiency [Internet]. [cited 2022 Apr 14]. Available from: https://rarediseases.oscar.ncsu.edu/disease/succinyl-coa3-ketoacid-coa-transferase-deficiency/about/#:~:text=Genetics Home Reference Succinyl-CoA%3A3-ketoacid CoA transferase %28SCOT%29 deficiency,appear within the first few years of life.

43. Kim YA, Kim SH, Cheon CK, Kim YM. A rare cause of life-threatening ketoacidosis: Novel compound heterozygous OXCT1 mutations causing succinyl-CoA:3-ketoacid coa transferase deficiency. Yonsei Med J. 2019;60(3):308–11.

44. Pretorius CJ, Son GGL, Bonnici F, Harley EH. Two siblings with episodic ketoacidosis and decreased activity of succinyl-CoA:3-ketoacid CoA-transferase in cultured fibroblasts. J Inherit Metab Dis. 1996;19(3):296–300.

45. Mahler R. Glycogen Diseases. J Clin Pathol [Internet]. 1969;22(2):32–41. Available from: http://www.wjgnet.com/1007-9327/full/v13/i18/2541.htm

46. Özen H, Bayraktar Y. Glycogen storage diseases: New perspectives Professor, Series Editor. World J Gastroenterol [Internet]. 2007;13(18):2541–53. Available from: www.wjgnet. comhttp://www.wjgnet.com/1007-9327/13/2541.asp

47. Hicks J, Wartchow E, Mierau G. Glycogen storage diseases: A brief review and update on clinical features, genetic abnormalities, pathologic features, and treatment. Ultrastruct Pathol. 2011;35(5):183–96.

48. Burchell A, Gibb L. Diagnosis of type 1B and 1C glycogen storage disease. J Inherit Metab Dis [Internet]. 14(3):305–7. Available from: https://doi.org/10.1007/BF01811688

49. Annabi B, Hiraiwa H, Mansfield BC, Lei KJ, Ubagai T, Polymeropoulos MH, et al. The gene for glycogen-storage disease type lb maps to chromosome 11q23. Am J Hum Genet. 1998;62(2):400–5.

50. Charbonneau DH, Healy AM. Genetics Home Reference [Internet]. Vol. 9, Journal of Consumer Health On the Internet. 2005 [cited 2022 Feb 10]. p. 61–8. Available from: https://medlineplus.gov/download/genetics/gene/g6pc.pdf

51. GeneCards The Human Gene Database. G6PC1 Gene - Glucose-6-Phosphatase Catalytic Subunit 1 [Internet]. [cited 2023 Mar 11]. Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=G6PC1

52. Smit G. The long-term outcome of patients with glycogen storage disease type Ia. Eur J Pediatr [Internet]. 1993;152:52–5. Available from: https://doi.org/10.1007/BF02072089

53. Ellingwood SS, Cheng A. Biochemical and clinical aspects of glycogen storage diseases. J Endocrinol. 2018;238(3):R131–41.

54. Chou JY, Mansfield BC. Mutations in the glucose-6-phosphatase-α (G6PC) gene that cause type Ia glycogen storage disease. Hum Mutat. 2008;29(7):921–30.

55. gnomAD. Single nucleotide variant : 17-41053082 - G-A (GRCh37) [Internet]. [cited 2023 Feb 11]. Available from: http://www.gnomad-sg.org/variant/17-41053082-G-A?dataset=gnomad_r2_1

56. Ganetzky R, McCormick E, Falk M. Primary Pyruvate Dehydrogenase Complex Deficiency Overview [Internet]. GeneReviews. [cited 2023 Mar 11]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK571223/pdf/Bookshelf_NBK571223.pdf

57. Meldau S, Fratter C, Bhengu LN, Sergeant K, Khan K, Riordan GT, et al. Pitfalls of relying on genetic testing only to diagnose inherited metabolic disorders in non-western populations - 5 cases of pyruvate dehydrogenase deficiency from South Africa. Mol Genet Metab Reports. 2020;24(May):6–8.

58. Haack TB, Rolinski B, Haberberger B, Zimmermann F, Schum J, Strecker V, et al. Homozygous missense mutation in BOLA3 causes multiple mitochondrial dysfunctions syndrome in two siblings. J Inherit Metab Dis. 2013;36(1):55–62.

59. Cameron JM, Janer A, Levandovskiy V, MacKay N, Rouault TA, Tong WH, et al. Mutations in iron-sulfur cluster scaffold genes NFU1 and BOLA3 cause a fatal deficiency of multiple respiratory chain and 2-oxoacid dehydrogenase enzymes. Am J Hum Genet [Internet]. 2011;89(4):486–95. Available from: http://dx.doi.org/10.1016/j.ajhg.2011.08.011

60. Conditions G. Genetics Home Reference multiple mitochondrial dysfunction syndrome. :1–6. Available from: https://ghr.nlm.nih.gov/condition/multiple-mitochondrial-dysfunctions-syndrome

61. Nasta V, Giachetti A, Ciofi-Baffoni S, Banci L. Structural insights into the molecular function of human [2Fe-2S] BOLA1-GRX5 and [2Fe-2S] BOLA3-GRX5 complexes. Biochim Biophys Acta - Gen Subj. 2017;1861(8):2119–31.

62. Conditions G. Genetics Home Reference multiple mitochondrial dysfunction syndrome [Internet]. [cited 2019 May 25]. p. 1–6. Available from: https://ghr.nlm.nih.gov/ condition /multiple-mitochondrial-dysfunctions-syndrome

63. Baker PR, Friederich MW, Swanson MA, Shaikh T, Bhattacharya K, Scharer GH, et al. Variant non ketotic hyperglycinemia is caused by mutations in LIAS, BOLA3 and the novel gene GLRX5. Brain. 2014;137(2):366–79.

64. Taruscio D, Groft SC, Cederroth H, Melegh B, Lasko P, Kosaki K, et al. Undiagnosed Diseases Network International (UDNI): White paper for global actions to meet patient needs. Mol Genet Metab. 2015;116(4):223–5.

65. Gialluisi A, Pippucci T, Anikster Y, Ozbek U, Medlej-Hashim M, Mégarbané A, et al. Estimating the Allele Frequency of Autosomal Recessive Disorders through Mutational Records and Consanguinity: The Homozygosity Index (HI). Ann Hum Genet. 2012;76(2):159–67.

66. Zaborowska L, Wooding S. Allele Frequency Calculator [Internet]. OMNI calculator. [cited 2023 Mar 12]. Available from: https://www.omnicalculator.com/biology/allele-frequency

67. Scitable. Hardy-Weinberg Equation [Internet]. NatureEducation. [cited 2023 Mar 13]. Available from: https://www.nature.com/scitable/definition/hardy-weinberg-equation-299/

68. Mayo O. A Century of Hardy – Weinberg Equilibrium. 2008;11(3):249–56.

69. Josef Ekstein, Berish Y. Rubin, Sylvia L. Anderson, David A. Weinstein, Gideon Bach, Dvorah Abeliovich, Michael Webb NR. Mutation frequencies for glycogen storage disease Ia in the Ashkenazi Jewish population. Am J Hum Genet. 2004;(129A):162–4.

70. Minoru O, Aoyama Y, Kishimoto M, Shishiba Y, Murase T. Identification of a point mutation (G727T) in the glucose-6-phosphatase gene in Japanese patients with glycogen storage disease type 1a, and carrier screening in healthy volunteers. Clin Genet [Internet]. 2008;51(3):179–83. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1399-0004.1997.tb02449.x

71. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68–74.

72. Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, et al. The 100 000 Genomes Project: Bringing whole genome sequencing to the NHS. BMJ [Internet]. 2018;361(April):1–7. Available from: http://dx.doi.org/doi:10.1136/bmj.k1687

73. Bentley AR, Callier SL, Rotimi CN. Evaluating the promise of inclusion of African ancestry populations in genomics. npj Genomic Med [Internet]. 2020;5(1):1–9. Available from: http://dx.doi.org/10.1038/s41525-019-0111-x

74. Arteche-López A, Ávila-Fernández A, Romero R, Riveiro-Álvarez R, López-Martínez MA, Giménez-Pardo A, et al. Sanger sequencing is no longer always necessary based on a single-center validation of 1109 NGS variants in 825 clinical exomes. Sci Rep [Internet]. 2021;11(1):1–7. Available from: https://doi.org/10.1038/s41598-021-85182-w

75. Er TK, Chang JG. High-resolution melting: Applications in genetic disorders. Clin Chim Acta [Internet]. 2012;414:197–201. Available from: http://dx.doi.org/10.1016/j.cca.2012.09.012

76. Wittwer C. High-Resolution DNA Melting Analysis: Advancementsand Limitations. Hum Mutat [Internet]. 2009;30(6):857–9. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1002/humu.20951

77. Konstantinos K V, Panagiotis P, Antonios VT, Agelos P, Argiris N. PCR-SSCP: A Method for the Molecular Analysis of Genetic Diseases. Mol Biotechnol. 2008;(38):155–63.

78. Kakavas K V. Sensitivity and applications of the PCR Single-Strand Conformation Polymorphism method. Mol Biol Rep. 2021;48(4):3629–35.

79. Gasser RB, Hu M, Chilton NB, Campbell BE, Jex AJ, Otranto D, et al. Single-strand conformation polymorphism (SSCP) for the analysis of genetic variation. Nat Protoc. 2007;1(6):3121–8.

80. Hashim HO, Al-Shuhaib MBS. Exploring the potential and limitations of PCR-RFLP and PCR-SSCP for SNP detection: A review. J Appl Biotechnol Reports. 2019;6(4):137–44.

81. Williams RC. Restriction Fragment Length Polymorphism (RFLP). Yearb Phys Anthropol. 1989;(32):159–84.

82. NIH. Restriction Fragment Length Polymorphism (RFLP) [Internet]. [cited 2022 Nov 16]. Available from: https://www.ncbi.nlm.nih.gov/probe/docs/techrflp/#:~:text=Restriction Fragment Length Polymorphism %28RFLP%29 is a difference,DNA samples in question with specific restriction endonucleases.

83. ThermoFisher SCIENTIFIC. Restriction Fragment Length Polymorphism (RFLP) Analysis [Internet]. [cited 2022 Nov 16]. Available from: https://www.thermofisher.com/za/en/home/life-science/sequencing/fragment-analysis/restriction-fragment-length-polymorphism-rflp-analysis.html

84. Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res. 1989;17(7):2503–16.

85. Ye S. An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Res. 2002 Jul 26;29(17):88e – 88.

86. Figueredo Lago JE, Armas Cayarga A, González González YJ, Collazo Mesa T. A simple, fast

and inexpensive method for mutation scanning of CFTR gene. BMC Med Genet. 2017 May 25;18(1).

87. Martins AM. Inborn errors of metabolism: a clinical overview. Sao Paulo Med J. 1999;117(6):251–65.

88. Seymour C, Thomason M, Chalmers R, Addison G, Bain M, Cockburn F, et al. Newborn screening for Inborn errors of metabolism: a systematic review. Vol. 1, Health technology assessment. 1997.

89. Hutchesson ACJ, Bundey S, Preece MA, Hall SK, Green A. A comparison of disease and gene frequencies of inborn errors of metabolism among different ethnic groups in the West Midlands, UK. J Med Genet. 1998;35(5):366–70.

90. Held PK, Rice GM, Kuhl A, Drilias N, Baker M, Deline J, et al. Newborn Screening for Inherited Metabolic Disorders: Early Identification and Long-Term Care for Patients in the Plain Community, Wisconsin, 2011-2017. Public Health Rep. 2019;134(2_suppl):58S-63S.

91. Gregg A, Aarabi M, Klugman S, Leach N, Bashford M, Goldwaser T, et al. Screening for autosomal recessive and X-linked conditions during pregnancy and preconception: a practice resource of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 23(10):1793–806.

92. Schmidtke J, Krawczak M. Carrier detection probabilities for autosomal recessive variants in unrelated and consanguineous couples — an evaluation of the 86 genes of the ACMG 'Tier 3' panel. J Community Genet [Internet]. 2022;13(4):399–401. Available from: https://doi.org/10.1007/s12687-022-00593-0

93. Johansen Taber K, Ben-Shachar R, Torres R, Arjunan A, Muzzey D, Kaseniit KE, et al. A guidelines-consistent carrier screening panel that supports equity across diverse populations. Genet Med [Internet]. 2022;24(1):201–13. Available from: https://doi.org/10.1016/j.gim .2021.09.009

CHAPTER 2: METHODS

This chapter describes the cohort investigated, primer design, and development of the assays. ARMS PCR primer pairs were designed for detection of the four pathogenic variants described in the literature review. Experiments were performed to optimise individual PCR reactions, followed by attempts to combine primer pairs into multiplexed assays. Finally, multiplexed primer combinations were used to screen the population control cohort for carrier status of two chosen variants, followed by validation of findings using Sanger sequencing.

2.1 The Cohort and Ethical Approval

DNA samples were previously extracted from dried cord-blood spots on filter paper, for use in population studies (1). The 750 blood samples originated from a randomised, de-identified neonatal cohort from the Cape Town metropole, in the Western Cape province, with UCT human research ethics approval. The neonates were born to mothers who identified themselves to the healthcare system as black South Africans. Neither the ethnic group, nor the sex was recorded within this cohort which is known to be prodominantly Xhosa. The cohort formed part of the routine thyroid newborn screening programme at Red Cross War Memorial Children's Hospital. All the samples were stored at -20°C in the division of Chemical Pathology. The ethnic composition was mixed.

The cohort of 750 samples was available from previous studies and was therefore used for this project. It was not possible to set up a larger study sample size, based on a formal power calculation to detect an exact frequency. Nevertheless, the statistical power of this project can be calculated. The sample size required for a one sample, dichotomous outcome is 600. This is calculated from the equation (2):

$$N = P(1\text{-}P)(z / E)^2$$

N=Sample size

P=Proportion of disease presence in population

Z=Standard normal distribution

E=Desired margin of error

Due to the absence of data in black African populations for the four IEM a *P* value of 0.5 was used to attain the most conservative sample size. A standard normal distribution (*Z)* of 1.96 (95%) was used. A margin of error of 400 in 10000 (4%) was chosen as the desired margin of error (*E*). This (*E*) value has been arbitrarily chosen as there are no previous studies relevant to this population.

Therefore:

$$N = P(1\text{-}P)(z / E)^2 = 0.5(1\text{-}0.5)(1.96 / 0.04)^2 = 600$$

Ethics approval for this study was obtained from the University of Cape Town (UCT) Human Research Ethics Committee (HREC/REF: 157/2021).

2.2 Primer Design

ARMS PCR primers were designed using the freely accessible primer-BLAST software tool hosted on the National Center for Biotechnology Information (NCBI) website (3). Bases were manually edited at the site of the mutation, as well as one or two bases upstream to introduce a destabilising mismatch for increased specificity.

The primer pairs were designed to produce amplicons with significant size differences, for the different variants, to ensure discrimination when multiplexed on agarose gel electrophoresis. Additional parameters included primer melting temperature (Tm) between 52°C and 58°C (with a final Tm after adding terminal and additional mismatch ranging from 49.4°C to 57.8°C), guanine-cytosine (GC) content between 40% and 60%, and primer length between 20 and 25 bases.

Predicted amplicon sizes were 503 bp, 400 bp, 250 bp and 117 bp for the *GALNT3, BOLA3, G6PC* and *OXCT1* PCR products, respectively. Primer details are given in **Table 2**. All primers were synthesized by Integrated DNA Technologies (USA).

Table 2. The pathogenic variants with relevant ARMS PCR primers. Nucleotide bases annotated in red indicate the position of the pathogenic variant; blue annotated bases indicate the location of additional mismatch bases.

Disease	Gene	Exon	Targeted Pathogenic Variant	Product Length	Forward Primer	Reverse Primer
Succinyl CoA: 3-oxoacid CoA transferase (SCOT) deficiency	*OXCT1*	8	c.803G>A (p.Arg268His)	117	5'-CCTCAGATTTATGTACA**G**C**A**-3'	5'-GGGATGCACTATCTCTGTAA-3'
Glycogen storage disease 1a (GSD1a)	*G6PC*	1	c.189G>A (p.Trp62Ter)	250	5'-GGCTCTGCTGACATCTTCCTG-3'	5'-CCAGTCTCCAATCACAGCTA**G**T-3'
Multiple mitochondrial dysfunction syndrome type 2 (MMDS-2)	*BOLA3*	2	c.159 dupT (p.Asp54Ter)	400	5'-GCTACAGCTATAAAAGTCAC**G**T-3'	5'-ACTCCAGCTCTGATATCCTAG-3'
Hyperphosphataemic familial tumoral calcinosis (HFTC)	*GALNT3*	2	c.484C>T (p.Arg162Ter)	503	5'-GGCTCACCTAAAGCGACTAG-3'	5'-GAGTGTCTGGTCCAAGATC**A**C**A**-3'

2.3 PCR Optimisation

The individual ARMS PCR for each primer pair was optimised through a gradient annealing temperature approach using positive (known to carry the pathogenic variant of interest) and negative (wild type) control DNA samples.

Individual PCR mixes contained 0.5µM of each primer (forward and reverse), 0.2mM dNTPs, 0.025U Supertherm Gold (STG) Thermus aquaticus (Taq) DNA polymerase and 1X STG buffer (containing 1.5mM $MgCl_2$). Addition of 1X solution Q (Qiagen, Germany) was used for amplification of the BOLA3 amplicon.

Thermocycler parameters: initial denaturation at 95°C for 8 minutes and at 98°C for 2 minutes; followed by 40 cycles of denaturation at 95°C for 30 seconds, gradient temperatures for annealing temperature were 48°C, 50°C, 53°C and 54°C for 30 seconds, and elongation at 72°C for 30 seconds with the final elongation at 72°C for 2 minutes. Gradient PCR was performed on the G-Storm GS1 Thermal cycler (Gene technologies Ltd, UK).

Products were separated on a 3% SeaKem LE agarose (Lonza, USA) gel containing 1µg/ml Ethidium bromide (Bio-Rad, USA) and visualised under ultraviolet (UV) light using a GelDoc-lt T53 imager (UVP, Germany). An annealing temperature of 54°C was chosen for all further PCR reactions.

2.4 Multiplex Optimisation Experiments

Multiplexing of the four individual PCR reactions described in **Section 2.3** was examined next.

Various combinations of the four primer pairs at different concentrations were explored to optimise individual amplicon signals on agarose gel electrophoresis. Internal (housekeeping) control primer pairs were added (see **Table 3.**) to evaluate PCR efficiency in each individual reaction.

Table 3. ARMS-PCR primer and internal control primer combinations evaluated. Amplicon sizes of internal controls: *AGXT* exon 1-2 978 bp, *MT-ATP6* gene 477 bp, *GALT* exon 5 146 bp (Internal control details in **Appendix A.**).

ARMS PCR primers: Concentrations	**Internal control: Concentrations**
BOLA3, OXCT1, G6PC, GALNT3: 0.5µM	*No controls*
BOLA3, OXCT1, G6PC: 0.5µM; **GALNT3:** 0.15µM, 0.25µM, 0.5µM	
BOLA3, OXCT1, G6PC: 0.3µM	***AGXT* exon 1-2:** 0.15µM
BOLA3, OXCT1, G6PC: 0.5µM	***MT-ATP6* gene:** 0.125µM
BOLA3, OXCT1: 0.5µM	
GALNT3, G6PC: 0.5µM	***GALT* exon 5:** 0.25µM

The thermocycler programme was set as described in **Section 2.3**.

2.5 Cohort Screening for *BOLA3* and *OXCT1* Variants

Cohort samples were tested using the chosen multiplex combination of *BOLA3*, *OXCT1* and *MT-ATP6* (housekeeping PCR) to detect the pathogenic variants of interest in these genes. The assays for variants in *G6PC* and *GALNT3* genes were excluded because of difficulties encountered during the development of the multiplex assay. This is discussed in more detail in **Section 3.2**.

BOLA3 and *OXCT1* primers were combined in equimolar concentrations at 0.5 µM, with 0.25 µM *MT-ATP6* primers, and added to the PCR mix (Summarised in **Table 4.**). PCR were performed in a total volume of 15µl on 96 well reaction plates (Axygen Inc., USA).

Table 4. PCR reagent mix of individual sample runs for the *BOLA3*, *OXCT1* and *MT-ATP6* multiplex used for cohort screening. Final concentrations are given.

Primers: Forward/Reverse	*BOLA3/OXCT1*: 0.5 uM; *MT-ATP6*: 0.25 uM
STG buffer	1X
Solution Q	20%
TAQ STG	0.025U
DNA template	10-500 ng
Final volume 15ul	

A total of 455 samples from the study neonatal cohort was screened with the *BOLA3*, *OXCT1*, *MT-ATP6* multiplex method described in section 2.4, using thermocycler settings as described in **Section 2.3**. Two hundred and ninety five samples of the total 750 were excluded due to assay failure. Each reaction plate included a positive control containing combined DNA from known carriers of each of the two variants under investigation, a wild type control DNA, and a no-template control. Electrophoresis and UV visualisation were performed as described in **Section 2.3**.

The outcomes for *BOLA3* or *OXCT1* amplicons were reported as positive if the respective ARMS products were present or negative if the respective ARMS products were absent. The run was reported as failed if the housekeeping PCR failed or if the results were uninterpretable. Sanger sequencing was performed on all samples flagged as positive to confirm findings.

2.6 Sequencing Confirmations

Sequencing of cohort samples that were flagged as positive for the *OXCT1* pathogenic variant during the screening described in **Section 2.5** was performed to confirm the findings. Exon 8 of the *OXCT1* gene was amplified by PCR and sequenced using the previously optimised method described below.

PCR reactions were performed using 0.5µM of each M13 tagged primer (primer details are given in **Appendix A.**), 0.025U STG Taq DNA polymerase, 1X STG buffer (containing 1.5mM $MgCl_2$), 20% solution Q, and approximately 10 to 500ng DNA template, in a final reaction volume of 15 µL. Thermocycler parameters started with an initial denaturation at 95°C for 8 minutes and 98°C for 2 minutes, followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at 58°C for 30 seconds, and elongation at 72°C for 30 seconds, with the final elongation at 72°C for 2 minutes. Products were separated on a 2% agarose gel and visualised as before.

PCR products were purified using the NucleoSpin Gel and PCR clean-up kit (Machery-Nagel, Germany) as per the manufacturer's protocol. Sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (applied biosystems, USA). This method was optimised based on the manufacturer specifications to maintain sequence quality (4). Reactions contained 5µM M13 reverse primers, 5-20 ng cleaned PCR product, 1X BigDye sequencing buffer, and 0.25X BigDye Terminator v3.1 Ready Reaction Mix. The final volume was made up to 10 µl with nuclease free water. Cycle sequencing was performed according to the following protocol: initial denaturation at 96°C for 15 seconds, followed by 30 cycles of denaturation at 96°C for 15 seconds, annealing at 50°C for 15 seconds and elongation at 60°C for 4 minutes. Final elongation was at 60°C for 10 minutes.

Excess dye terminators, primers, dNTPs, salts, enzyme, and any impurities were removed from sequencing products during a purification step. An ethanol/EDTA/sodium acetate precipitation method was used according to the manufacturer's protocol (4). Purified sample was resuspended with 10 µl HiDi Formamide (Applied Biosystems, USA) as per the manufacturer's protocol (4). Samples were analysed on the ABI 3500 Genetic Analyzer (Applied Biosystems, USA). Resulting sequence data files were analysed using Bioedit v7.0.9.0 (5) and Chromas software v2.6.5 (Technelysium, Australia). Sequences were compared with the NCBI reference sequence for this gene (NM_000436.3).

2.7 Mathematical Calculation of Carrier Frequency

The Hardy-Weinberg equation will be used to determine the carrier frequency of variants confirmed on Sanger sequencing. The two equations used are represented below (6).This estimated carrier frequency will be calculated from the total sample size analysed during screening experiments. The principle for the Hardy-Weinberg equilibrium considers the following: no natural selection, No migration in or out of the population, no mutations, no genetic drift, and random mating occurs (7).

(1) The Hardy-Weinberg equation:

$$p^2 + 2pq + q^2 = 1$$

p^2 = Wild type homozygous frequency

$2pq$ = Heterozygous frequency (Carrier frequency)

q^2 = Pathogenic homozygous frequency

(2) allele frequency equation

$$p + q = 1$$

p = Dominant allele

q = Recessive allele

Recessive homozygous samples will require use of equations (1) and (2). Heterozygous samples (carriers) will require calculation of only the proportion between confirmed samples and the total number of samples tested.

CHAPTER 3: RESULTS

The results of optimisation experiments for the individual ARMS PCR assays are given followed by that of the multiplex optimisation experiments. Thereafter, the outcome of the carrier screening using multiplex assays are reported. Finally, results of Sanger sequencing analysis are provided.

3.1 PCR Optimisation

Individual ARMS PCR were tested across a range of annealing temperatures. A single acceptable temperature for all PCR was identified to perform the multiplex analysis. Samples containing the pathogenic variant DNA template, and wild type DNA template were used.

Templates were amplified at annealing temperatures ranging between 48°C and 54°C. All four fragments were successfully amplified between 53°C and 54°C, as shown in **Figure 7**.

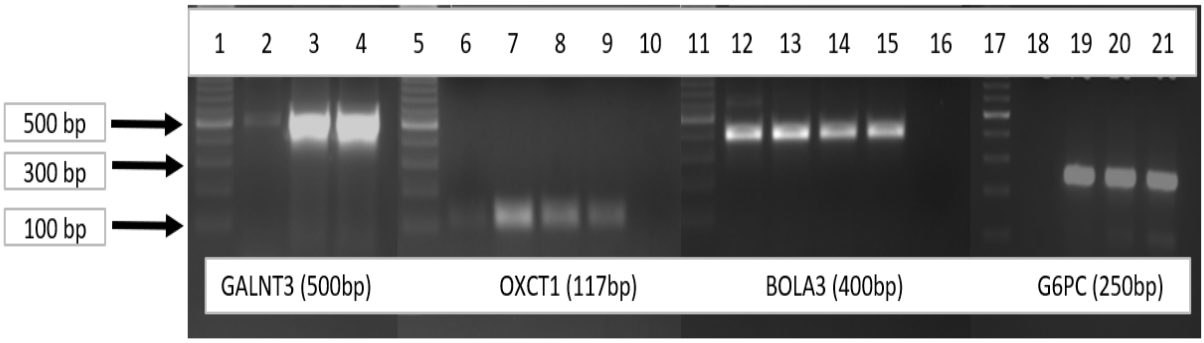

Figure 7. Agarose gel electrophoresis of the gradient PCR products separated on 2% agarose. Annealing temperatures: 48°C lanes 6 and 12; 50°C lanes 7,13, and 21; 52°C lane 20; 53°C lanes 4,8, and 14; 54°C lanes 3,9,15, and 19. Wild type control (54°C): lanes 2, 10, 16, and 18. Lanes 1,5,11, and 17 contain a 100bp incremental DNA ladder (Thermofisher, USA). Non-specific binding was seen at lower temperatures of the *BOLA3* (lane 11) and *G6PC* (lane 16) PCR. The *GALNT3* wild type control (lane 3) also showed extra bands.

3.2 Multiplex Optimisation

The four primer sets were tested in different combinations to assess PCR efficiency in multiplex runs, as well as adequate product separation upon electrophoresis.

Multiplexing of all four ARMS PCR using equal primer concentrations (0.5µM) failed to simultaneously amplify all products, with near complete loss of the *BOLA3* product in the presence of *GALNT3* primers. A gradient of lower concentration *GALNT3* primers at 0.15 µM, and 0.25 µM continually failed to improve BOLA3 amplification. The results of the GALNT3 primer concentration gradient can be seen in **Figure 8**.

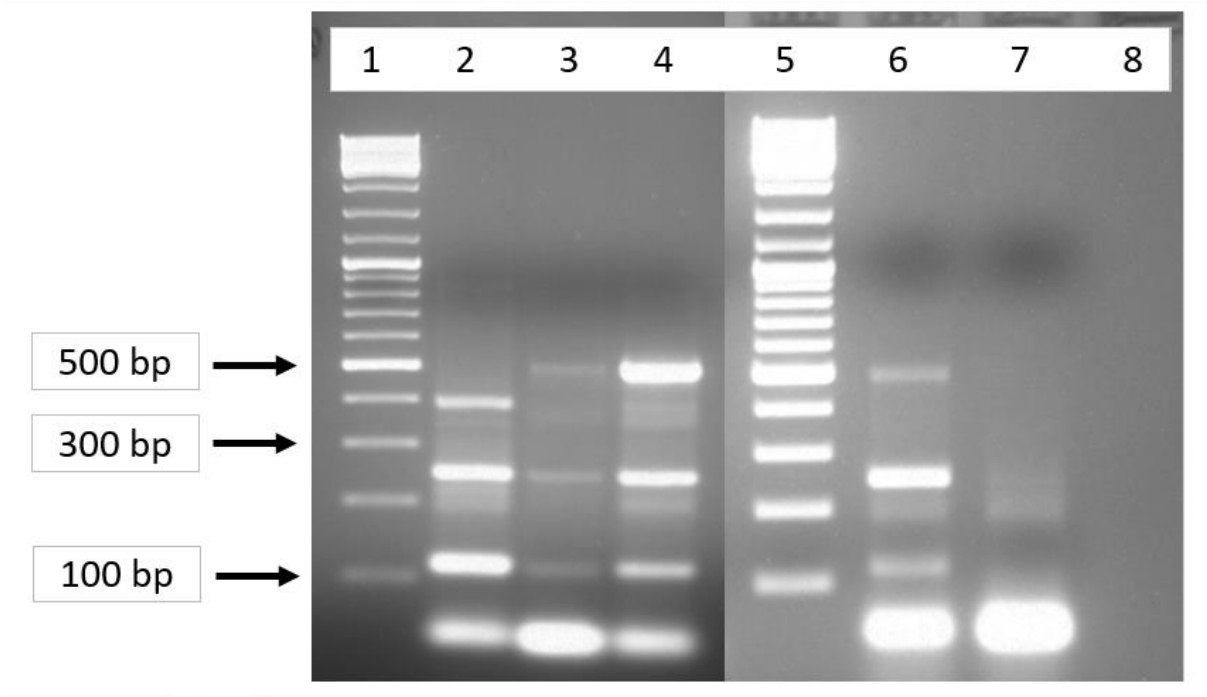

Figure 8. Agarose gel electrophoresis showing PCR products of multiplexed combinations of the ARMS PCR primer sets, using pathogenic control DNA as template. Lanes 1 and 5 contained a 100 bp incremental ladder (Thermofisher, USA). Lane 2 contained *BOLA3*, *G6PC*, and *OXCT1* primer pairs. Lane 3, 4, and 6 contained a concentration gradient of *GALNT3* primers (0.15 µM, 0.5 µM, 0.25 µM, respectively), combined with *BOLA3*, *G6PC* and *OXCT1* primer pairs. Lane 7 contained all four ARMS PCR primer pairs using wild type control DNA. Expected amplicon sizes: *GALNT3* 503 bp, *BOLA3* 400 bp, *G6PC* 250 bp, and *OXCT1* 117 bp.

The PCR efficiency of the different ARMS PCR primer pairs in combination with housekeeping primers was tested. The number of simultaneous ARMS PCR primer pairs combined in a single reaction was reduced. ARMS PCR product amplification had varying degrees of quality as shown in **Figure 9**. The housekeeping control primer pairs consistently amplified product.

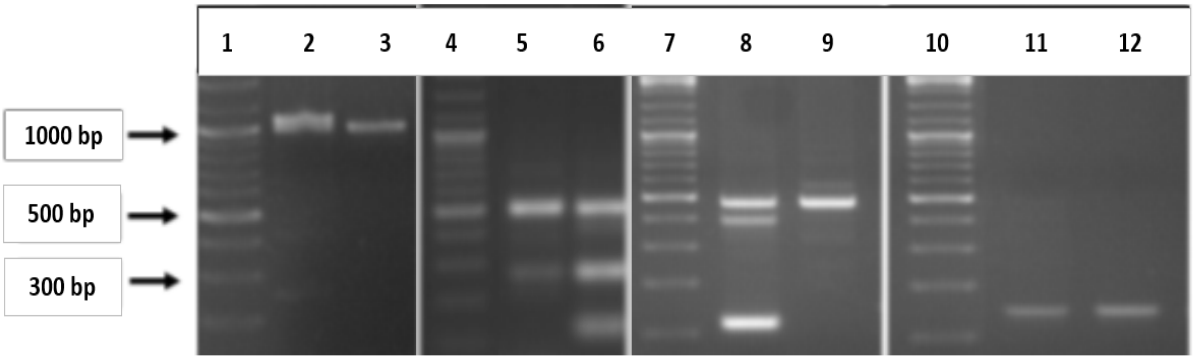

Figure 9. Agarose gel electrophoresis images showing different multiplexed PCR combinations at an annealing temperature of 54°C. Lanes 2 and 3: *AGXT, BOLA3, G6PC, OXCT1* for pathogenic and wild type DNA samples respectively. Lanes 5 and 6: *MT-ATP6, BOLA3, G6PC, OXCT1* for pathogenic and wild type DNA samples respectively. Lanes 8 and 9: *MT-ATP6, BOLA3, OXCT1* for pathogenic and wild type DNA samples respectively. Lanes 11 and 12: *GALNT3, G6PC, GALT* for pathogenic and wild type DNA samples respectively. Lanes 1, 4, 7, and 10: 100bp incremental DNA ladder (Thermofisher, USA). Expected amplicon sizes: *AGXT* 978 bp, *GALNT3* 503 bp, *MT-ATP6* 477 bp, *BOLA3* 400 bp, *G6PC* 250 bp, *GALT* 146, *OXCT1* 117 bp.

Amplicons in the multiplex reaction containing *BOLA3, OXCT1* and *MT-ATP6* primer pairs were distinguishable from each PCR product on 3% agarose gel electrophoresis (see **Figure 10.**). The band corresponding to the *MT-ATP6* product (477bp) could be unambiguously discriminated from *BOLA3* (400bp), despite the close band proximity. Initial attempts to add *G6PC* primer pairs into the multiplex reaction caused a failure of *BOLA3* PCR product to amplify, with non-specific amplification of product corresponding with *G6PC* pathogenic variant in the wild type control (**Figure 11.**).

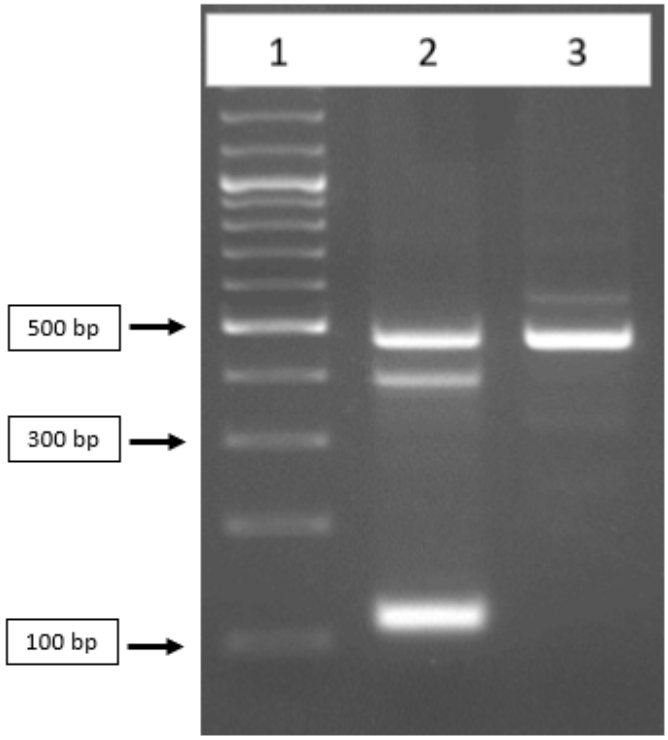

Figure 10. Agarose gel electrophoresis image showing a multiplex reaction containing *BOLA3*, *OXCT1*, and *MT-ATP6* primer pairs. Lane 2: PCR products of multiplex reaction containing positive samples. Lane 3: PCR product of multiplex reaction with wild type control. Lane 1: 100 bp incremental ladder. Expected amplicon sizes: *MT-ATP6* 477 bp, *BOLA3* 400 bp, *OXCT1* 117 bp.

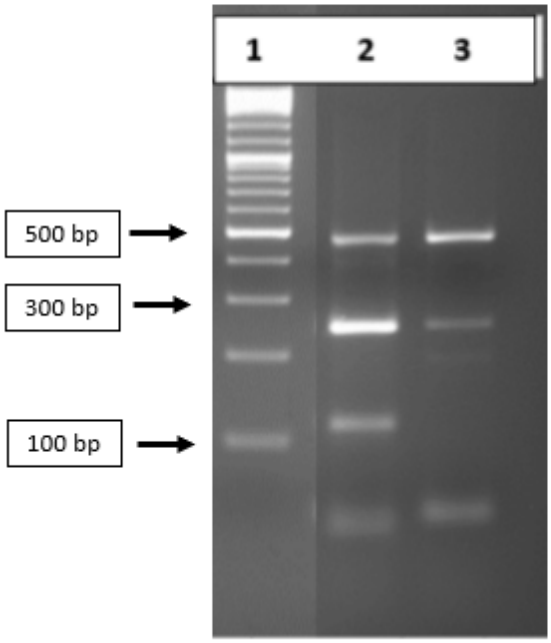

Figure 11. Agarose gel electrophoresis image showing a multiplex reaction containing *MT-ATP6 BOLA3, OXCT1* and *G6PC* primer pairs. Lane 2: PCR products of multiplex reaction containing pathogenic DNA samples. Lane 3: PCR products of multiplex reaction containing wild type control. Non-specific primer amplification of *G6PC* pathogenic variant in the presence of wild type DNA is noted. Lane 1: 100 bp incremental ladder. Expected amplicon sizes: *MT-ATP6* 477 bp, *BOLA3* 400 bp, *OXCT1* 117 bp, *G6PC* 250 bp.

A separate multiplex reaction using primer pairs for *GALNT3, G6PC* and *GALT* (internal control) was tested at a temperature gradient of 50°C, 52°C, 54°C, 56°C and 58°C. Persistent failures to amplify *GALNT3* and *G6PC* ARMS PCR products at different annealing temperatures. Time constraints prevented further exploration to achieve the desired outcomes.

3.3 Carrier Screening of *BOLA3* and *OXCT1* Variants

Carrier screening for the *BOLA3* and *OXCT1* pathogenic variants was performed using the method described in section 2.5. A total of 455 (61%) samples was tested from the cohort of 750.

OXCT1 variant analysis was deemed valid for 439 (96%) of the 455 samples tested, with 16 (4%) samples invalid for analysis due to inadequate amplification of the control product. PCR product corresponding to the *OXCT1* variant amplicon (117 bp) was seen in 14 (3.1%) of the 439 valid samples. The remaining 295 samples were not tested due to technical difficulties and time constraints.

BOLA3 ARMS PCR results were deemed valid for 246 (54%) of the 455 samples screened. The remaining 209 (46%) samples were considered invalid due to inadequate amplification of product corresponding to the 400 bp *BOLA3* amplicon in positive controls. None of the valid results were positive for the *BOLA3* pathogenic variant. The remaining 295 samples were not tested due to technical difficulties and time constraints.

3.4 Sequencing

Sanger sequencing was performed to confirm the presence of the c.803G>A(p.Arg268His) pathogenic variant in *OXCT1* in the 14 samples regarded as positive during *OXCT1* screening from **Section 3.3**.

Sequencing, to confirm the findings on ARMS PCR, did not reveal the presence of this variant in any of the samples tested.

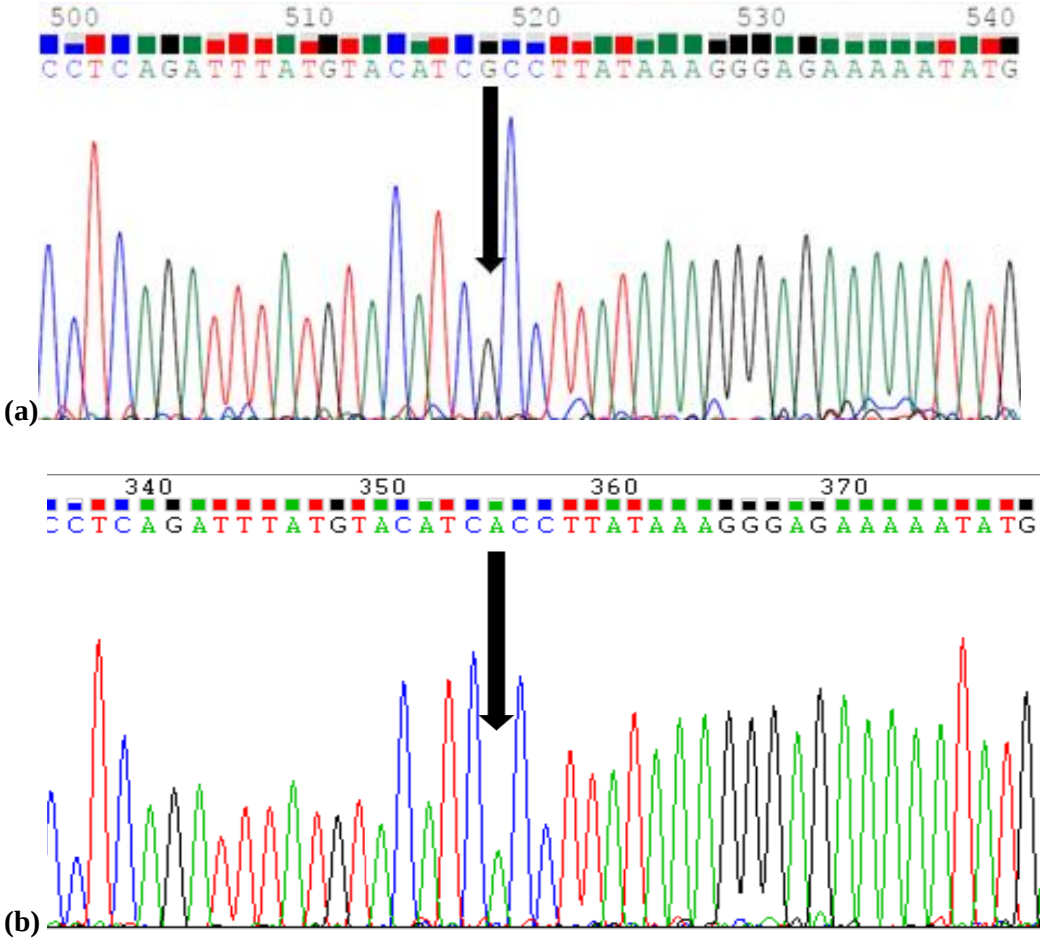

Figure 12. Sequence electropherogram of: (a) a representative sample which tested positive on the ARMS PCR screen, showing the site of the c.803G>A pathogenic variant in exon 8 of the *OXCT1* gene; and (b) the positive DNA control sample used in this study. The base of interest is indicated with an arrow.

CHAPTER 4: DISCUSSION AND CONCLUSION

The study was undertaken, with due process of ethics approval, to improve insight into the prevalence of the four severe IEM conditions identified in the local black population. A brief summary of the rationale for the study will be followed by discussion of the results. Thereafter, the methodology used in the study will be analysed, highlighting the difficulties multiplexing the individual ARMS PCR assays. The non-specific binding encountered, and inconsistent performance will be summarised. Alternative methods will be considered, with a discussion on outcomes for future work.

4.1 Study Rationale

Autosomal recessive disorders usually only manifest in the context of homozygosity or compound heterozygosity, while true heterozygotes are phenotypically normal (8). Disease prevalence due to specific variants requires knowledge of heterozygote prevalence or carrier frequency in the population at large. This information may improve estimates of the impact of the relevant IEM (9).

The black population of Southern Africa comprises several ethnic groups that share primordial ancestral links. Regional variations may occur as a result of African ancestral migratory patterns, potentially causing genetic drift and founder effects. The possibility of these IEM being founder effects within this predominant group underscores the need to determine their prevalence. Furthermore, ancestry determination is important as presence of variants within closely related migratory group clusters may direct targeted analysis (15,16). This group makes up 79.2% of the South African population, amounting to more than 41 000 000 people (12). Given the large population, it is possible that even rare variants can occur in patients with recessive disorders, either as true homozygous or compound heterozygous. The impact of these variants on health and disease in sub-Saharan black Africans is uncertain.

Previously, access to healthcare and diagnosis of IEM may not have been well supported, but IEM are now increasingly recognized (13). In many of these disorders complications can be limited by proactive detection and intervention of affected persons. However, if not definitively diagnosed, many unnecessary treatments and investigations may be performed, with delay in diagnosis (14).

Awareness of heterozygous carriers may focus partner testing for pregnancies, thus proactively identifying individuals wherein a 1:4 chance of birthing a homozygous neonate is possible. Furthermore, the criteria identified by the American College of Medical Genetics and Genomics (ACMG) and American College of Obstetricians and gynecologists (ACOG) will assist in developing an approach to carrier screening. These criteria include carrier frequency, relationship between genotype and phenotype, disease severity, age of onset, capability of diagnosing the conditions prenatally, and analytical methods available for screening (15,16).

Knowing the prevalence of heterozygous carriers in the population allows use of the Hardy-Weinberg equation to calculate the number of putative homozygous individuals. This can improve planning and medical attention, in the form of specialized clinics and laboratories to address IEM comprehensively (13,14). However, investigating these variations requires analysis of large cohorts. Carrier frequency studies require appropriate sample size of a representative cohort for statistical analysis.

Consideration must be made regarding the cohort of 750. As determined in **Section 2.1**, a conservative estimate of the required sample size is 600, at a 4% margin of error. However, a narrower margin of error may be necessary if no carriers are identified in this cohort. Based on the equation in **Section 2.1**, a margin of error less than 4% may be used. As an example, a 1% margin of error will require a sample size of 9604. However, the feasibility of obtaining a greater sample size must be considered. Identifying carriers in this proof-of-concept study will provide the necessary variables to calculate an appropriate population size in subsequent research.

Despite the lack of data available for carrier or allele frequency, the data that exists indicate very low frequencies. This is especially true for HFTC and GSD 1a data presented in **Sections 1.3.1** and **1.3.3**, respectively. To reiterate, HFTC allele frequencies are 0.0008 for African Americans and 0.0002 worldwide, while GSD 1a allele frequency is 0.000004 worldwide (17,18). Furthermore, these data are generated from large numbers. This study was performed to assess for the frequencies of the pathogenic variants within our population, to determine the likelihood of these being common locally.

This proof-of-concept study aimed to determine carrier frequencies of four pathogenic variants identified by genetic investigation of severe IEM, causing HFTC, SCOT deficiency, GSD 1a and MMDS 2. To achieve this in a resource constrained setting, ARMS PCR was chosen as a cost-effective, high throughput methodology.

4.2 Outcome of the Study

The study encountered many problems associated with the multiplex ARMS PCR method. However, the importance of determining variant prevalence of disorders potentially arising from founder effects within a representative cohort, and the need for a cost-effective strategy to diagnose severe IEM, must be emphasised.

The project set out to test a total of 750 DNA samples for the presence of defined and clinically relevant pathogenic variants by PCR. In 14 of 455 samples screened, the *OXCT1* amplicon suggested a pathogenic variant was present. Sanger sequencing did not confirm the presence of the c.803G>A(p.Arg268His) pathogenic variant in the *OXCT1* gene in any of these samples. This discordance between screening and sequencing highlights the need for confirmation of the ARMS PCR procedure, especially when undertaking bulk processing. Technical problems in performing the assays and the need for further optimisation are the likely cause. The remaining 295 of the original 750 DNA samples could not be tested due to assay failure due to lack of amplification in the positive control. A total of 246 samples tested negative for the *BOLA3* variant with the *OXCT1, BOLA3* multiplex. Of the 455 samples tested, 209 were removed from the analysis due to lack of amplification in the control samples.

The persistent assay failures due to non-specific binding and inconsistencies with assay performance, which led to the eventual omission of 295 samples as described previously, could not be improved upon, despite multiple troubleshooting attempts, in the available time. Furthermore, the 455 samples tested is not sufficiently powered as this is below the required sample size of 600, calculated in **Section 2.1**. The prevalence of the variants thus remains unknown. The necessity for prevalence analysis using affordable and convenient methods persists. This highlights the need for alternative methods or a robust optimisation of the current approach. In the interim, should the clinical need arise, individual cases can be tested using individual assays and sequencing methods.

4.3 Analysis of Methodology

South Africa has difficulties in providing adequate healthcare for all its citizens. There are limitations to clinical expertise as well as laboratory expertise, staffing and equipment (19). For these reasons it is preferable that strategies are devised for rapid and simple diagnosis and to establish carrier frequencies.

Utilising methods that are easily implemented at the majority of laboratories is preferred. ARMS PCR has previously been used to test carrier frequency of the pathogenic c.404C>T(p.S135L) variant in the *GALT* gene coding for galactose-1-phosphate uridyltransferase, causing galactosaemia in the black South African population (1). The study screened 725 samples using ARMS PCR and assessed known transferase-deficiency galactosaemia patients using RFLP. Different forward primers were designed to amplify the normal or variant allele of the *GALT* gene in separate runs. Both forward primers were coupled to the same reverse primer. This method successfully screened pathogenic from wild type *GALT* alleles. Subsequently, population specific carrier frequency was elucidated with extrapolation of newborn incidence, demonstrating the utility of ARMS PCR in SNV analysis in the local resource constrained setting.

The ability to assess multiple variants simultaneously is crucial due to the increasing identification of local variants. Disorders where ARMS PCR has been combined to identify multiple pathogenic variants for a single disease include cystic fibrosis and cancer (20,21). A similar technique to the galactosaemia study was used by Newton et al. in *SERPINA1* gene analysis for Alpha-1 antitrypsin (AAT) deficiency. Duplicate genomic samples with either wild type or variant allele specific primers were assayed, coupled with primers for an unrelated gene to confirm the presence of DNA (22). Amplified products thus proved the feasibility of co-amplifying multiple PCR products for multiplex analysis.

Despite ARMS PCR being an attractive and facile strategy, problems were encountered when multiplexing was attempted in this study. The poor assay performance could be attributed to many reasons. The three chief reasons to consider are: technical difficulties when combining functioning single ARMS PCR assays; inherent difficulties of multiplexed assays, such as primer dimer formation; and operator inexperience and the time constraints experienced resulting in the need to perform laboratory work at staggered intervals. The first and second issues will be discussed in more detail in upcoming chapters. Other undetermined factors may also have affected the assay.

Individual ARMS PCR reactions at different temperatures, as shown in **Figure 7.**, showed promise but could not be reproduced upon multiplexing. The multiplexed ARMS PCR reactions in our study yielded inconsistent results. Factors that may have hindered the multiplex performance include primer dimer formation, and the disproportionate use of PCR reagents when primer pairs are combined (23).

Multiplex PCR are at a higher risk for primer dimer formation than individual PCR at standard primer concentrations. This risk increases with larger multiplex panel sizes (24). Additionally, primer dimers reduce PCR efficiency by reducing primer availability for binding onto DNA template, and consumption of available PCR reagents to maintain PCR efficiency. Primer dimer formation can be reduced through careful primer design; avoiding complementarity, especially at the 3' end of the primer; and optimisation of PCR mixtures, specifically for multiplex assays.

Combining multiple ARMS PCR primer pairs failed to amplify products reliably in this project. Initial multiplexing of the four primer pairs, with an additional internal control, into a single run proved difficult and time consuming. Additionally, adding equimolar primer concentrations in the multiplex reaction caused variable band intensities, with some amplicons failing altogether. Primer concentrations were therefore adjusted to balance the individual PCR efficiencies within the multiplex reaction. This failed to provide adequate benefit.

Importantly, the *GALNT3* product seemed to interfere with other ARMS reactions in the combined assays. This deduction was based on the relatively increased intensity of the *GALNT3* band, compared to the decreased intensity, or absence, of the other bands. This may be due to the large amplicon size, or the primer efficiency relative to other primers.

Recurring assay failures and time constraints to troubleshoot each multiplex assay resulted in the discontinuation of *GALNT3* and *G6PC* screening. Therefore, in the interest of time, a reduced multiplex assay containing *MT-ATP6* (internal control), *BOLA3 and OXCT1* primers was explored. Initially, this yielded seemingly consistent results. This combination amplified the correct products during PCR and multiplexing optimisation runs but was unreliable during screening tests.

Before considering alternative methods, using the available ARMS PCR assays should still be explored. Firstly, the multiplex assay may benefit from better optimisation. Multiple factors affect the performance of PCR assays, notably, PCR chemistry, primer design, PCR controls and cycling temperatures (25). These will be discussed in further detail in **Section 4.5**. Secondly, cohort analysis may continue due to the acceptable performance of the individual ARMS PCR primers. However, similar to the Galactosaemia study (1), a wild type ARMS primer will be required to perform a concurrent ARMS PCR run with the pathogenic variant's respective ARMS primer. Both these options will likely be successful but require additional time. Therefore, with these options considered and the inconsistent multiplex performance, assessing the feasibility of other methods is beneficial.

4.4 Alternative Methods

ARMS PCR assays are cost effective and quick to perform once validated (20). However, this study showed that optimisation and troubleshooting can be time consuming. Therefore, alternative methods for SNV analysis, such as RFLP which requires very little optimisation with better specificity, although more costly and time-consuming than ARMS PCR to run, may be more fit for purpose in the long run (26). NGS, fragment analysis and probe-based methods are additional options used in our setting but are more expensive to perform.

RFLP is an established molecular method using restriction endonucleases to identify variants by generating differently sized fragments. The resources required to perform these assays and interpret results are readily available. Since these enzymes recognize very specific palindromic sequences to cut DNA, enzymes may be selected to identify the wild type or the pathogenic variant in a reliable way. Planning is required to design primers with the goal of producing fragments depending on the restriction endonucleases. RFLPs are robust and cost effective compared to newer generation methods (27), and is a good alternative to ARMS PCR. Despite the cost and simplicity to perform, this method is not as easily multiplexed (28,29). This may require that cohorts are analysed for a single variant, which is laborious.

Next generation sequencing (NGS) and high-resolution melting (HRM) are popular diagnostic methods. NGS is capable of simultaneously testing multiple genes and variants at low cost per individual run (30). However, the cost benefit is only achieved with high sample volumes or analysis of many genes. Additionally, library preparation and equipment required for NGS is expensive. HRM is a reliable and less expensive method that has been used for SNV detection in various settings (31). This post-PCR analytical method can be used to scan for variants by generating recognisable patterns of amplicons with known variant alleles, before submitting for sequencing (32).

Both methods require advanced expertise and resources to perform, therefore, the benefits offered by simpler methods are appropriate, and have been achieved in our setting. Therefore, optimising multiplexed ARMS PCR will be beneficial.

4.5 Future Work

The method used should be revisited with meticulous attention to detail and appropriate time allocation. It may be necessary to perform the assays individually in this cohort to determine prevalence. It may be worth developing multiplex assays of fewer but better compatible assays, while still using ARMS PCR. Henegariu et al. assessed PCR combinations for over 50 loci, identifying guidelines for successful multiplex assays (25). Factors considered include PCR chemistry, primer design, PCR controls and cycling temperatures. Adjusting these factors for multiplex assays promotes optimal performance of simultaneous reactions.

The first consideration for multiplex assays is PCR chemistry, in particular the impact on amplicon production by buffer solution, additives, dNTPs and $MgCl_2$. Constituent concentrations of the master mix for this study uses established values for robust assays in the laboratory, specifically RFLP PCR. These may not be optimal for ARMS or multiplex reactions. The concentrations of constituents in the ARMS PCR assay were 1X PCR buffer, 0.2 mM dNTPs and 1.5 mM $MgCl_2$. In contrast, Henegariu et al. recommend in their concentrations 1.6X PCR buffer and 1.8-2mM $MgCl_2$ per 0.2mM dNTPs. These recommendations highlight the need for increased concentrations of PCR buffer and $MgCl_2$ in multiplex PCR assays.

Various buffer types allow for acceptable PCR performance, specifically if they are ammonium or potassium chloride (KCl) based (25,33). PCR additives, such as betaine, DMSO, formamide, and DTT can be considered. These additives improve PCR performance through multiple mechanisms, including preventing secondary structures seen with high GC content. Additionally, DNA melting behaviour is modified, thus optimizing PCR amplification (34,35). In our study, adding Qiagen Q-solution to the mastermix improved amplification of the BOLA3 product when compared to reactions without the additive.

The balance between dNTPs and $MgCl_2$ was highlighted by Henegariu et al. (25). They noted that increasing dNTPs while maintaining $MgCl_2$ impedes the PCR, while exclusively increasing $MgCl_2$ causes no inhibition. This relationship arises from two functions of $MgCl_2$: free $MgCl_2$ is a co-factor for *Taq*-polymerase; and magnesium binds to dNTPs and the DNA template. These combined effects on a multiplexed assay and the starting $MgCl_2$ concentration of 1.5 mM likely harms this balance, preventing polymerase activity. Fine-tuning PCR chemistry may therefore improve assay performance, particularly in the context of multiplexing.

In order to reduce cost for the purposes of this study, available resources were used where possible. As a result, an STG PCR kit was chosen. Specialised, commercially available kits designed specifically for multiplexing, although costly, may avoid cumbersome troubleshooting. Given the difficulties faced with optimising this assay, it may be worthwhile to assess such kits for future studies. An example is the QIAGEN Multiplex PCR kit, which is designed to increase primer availability at the DNA template and to stabilise primers bound to a specific site (33).

Another consideration involves re-assessing the primer pairs before undertaking future studies. Refractory primer extension varies depending on the complementary base used in the ARMS PCR primer. In Newton's work, some mismatches continued to amplify, specifically primer/template combinations guanine/thymine, adenine/cytosine, cytosine/adenine and thymine/guanine (22). Further research identified improved template discrimination with primers containing thymine, guanine or cytosine on the terminal 3'-end, but not adenine. The author determined that different sequences may affect discrimination and elongation (36). Additionally, further adjustments to the 3'-end enhances the primers ability to amplify in specific circumstances.

Introducing additional mismatches near the 3'-end destabilises the primer/template complex, which improves specificity. Destabilising factors include proximity to the 3'-terminal with mismatches closer to the terminal end having more instability; and purine/purine or pyrimidine/pyrimidine mismatches having better refractory ability than purine/pyrimidine mismatches. Additionally, primer stability may be manipulated by adjusting PCR chemistry, as discussed previously in this chapter. However, in addition to optimized ARMS PCR primers, the effect of combining multiple primers into a single assay must be considered.

Primer factors influencing multiplex assays include primer length, GC content and annealing temperature (25). The 20-25 base range used in the ARMS PCR primers corresponds with the commonly recommended range of 18-24 bases. Longer primers increase annealing temperatures, minimising non-specific products. GC content recommendations range between 35% and 60%, agreeing with the 40%-60% range used for our primers. This affects annealing temperature of PCR. Therefore, adjusting annealing temperature and duration to prevent competitive annealing is required (37). Owing to the complicated nature of multiplex ARMS PCR analysis, computer software could assist in primer development.

Primer designing software, Primer-BLAST, allows standardized parameters to be set (3). Advantageous to ARMS PCR is the ability to develop novel primers with a variant allele. Additionally, software assists in identifying risks for primer failure, such as non-specific binding. However, some programmes may not be compatible for factors impacting multiplex assays. Assessing these primer combinations through specific multiplex primer design software may reduce the likelihood of non-specific binding and secondary structure formation. An example is the free, online multiplex primer designing tool, Ultiplex (38). Other software capable of multiplex primer design are PrimerPlex and MPD. This type of software saves time and money.

The third consideration is the addition of controls to confirm PCR performance. This is commonly achieved by incorporating an internal control into each run, thus confirming function of the system. The galactosaemia study performed separate assays for normal and variant alleles rather than incorporating an internal control (1). This approach requires that a product is amplified in at least one of the paired runs. Despite similarities to the aforementioned study, Newton et al. performed an additional confirmatory step by incorporating primers for a region in the *APOB* gene to be used as an internal control (22). In contrast to these two experiments, another study using an ARMS PCR multiplex experiment, produced an internal control located in the unaffected intronic region of the gene investigated (39). However, systems incorporating internal controls to specifically enhance ARMS PCR have been developed.

Tetra-primer ARMS PCR is a modification to ARMS PCR assays that utilizes the same four primers to simultaneously generate target allele and internal control (40). A large fragment generated from the outer primers is produced in all genotypes, and acts as the internal control. Furthermore, allele specificity occurs from variable combinations of inner and outer primers. The advantage in this assay is that a single PCR and gel electrophoresis is sufficient for analysis. However, optimisation may be laborious and time-consuming (41).

A final consideration for multiplex assays is thermocycler settings to enhance annealing on DNA templates and sufficient product amplification. Reproducible thermocycler parameters for single product amplification are often consistent but may vary for multiplex reactions. Settings have been recommended when co-amplifying multiple loci, with these being affected by primer design (25). Recommended multiplex assay thermocycler settings for elongation time and temperature, annealing temperature, and cycling times, have been reported by Henegariu et al. and are discussed below (25).

Experiments with lower elongation temperatures of 65°C for 1 minute had a better yield of amplified product than elongation temperatures of 72°C for 1 minute 20 seconds. Furthermore, elongation times of 4 minutes rather than 1 minute improved the yield of longer products. This likely allows prolonged polymerase activity to amplify the product normally affected by rate limiting quantities of enzymes and nucleotides. Our chosen elongation temperature and time was 72°C for 30 seconds, respectively, but benefit may arise by adjusting both.

Annealing temperature of 54°C used in our assay corresponds to recommendations for multiplex analysis. While temperatures between 56°C and 60°C are usually recommended to avoid nonspecific binding seen in lower annealing temperatures, this outcome is likely offset by co-amplifying multiple loci. Additionally, variable efficiencies of primer pairs cause more efficient primers to negatively affect less efficient primers. Both effects, unique to multiplex assays, likely prevent the overt presence of unwanted products.

Recommended PCR cycles of 28-30 are sufficient for adequate amplification with more than 60 cycles having no added benefit (25). The decision to use 40 cycles in our study was to increase product amplification. Therefore, subsequent assay adjustments may benefit from only increasing annealing and elongation parameters, while decreasing PCR cycles. Although the benefit may be negligible, adjusting PCR cycles would save time.

In summary, these considerations are practicable and should be implemented when troubleshooting. However, as this study was performed piecemeal due the specific time and resource constraints presented by the Chemical Pathology MMed programme, a dedicated prolonged duration in subsequent research to implement these recommendations would improve the chance of gaining and sustaining the necessary skills to reliably perform the assay.

4.6 Conclusion

This study set out to develop a multiplex ARMS PCR assay to determine carrier frequencies of the locally detected pathogenic variants causing HFTC, SCOT deficiency, GSD 1a and MMDS2 within the black population in Southern Africa, but that may also be relevant to sub-Saharan Africa. The goal was to obtain a cost-effective method, as a proof-of-concept for potentially wider analysis of this population. As explained, by choosing cost-effective and simple methods, such as ARMS PCR, laboratories could minimise expenses and workload in a constrained setting. The resulting data may provide a guide for the clinical management and distribution of resources. Based on such information, panels for certain clinical phenotypes can justifiably include a given variant for diagnosis.

Although the study failed to achieve the outcome described, wider reading of the literature shows that using multiplex ARMS PCR can be achieved, but alternative approaches should also be explored. As discussed, multiplex ARMS PCR undoubtedly remains a potentially valuable method especially in a setting of limited available resources. Despite the multiplex of all four variants being unsuccessful, omission of two primer pairs showed potential. The findings highlight the need for robust optimisation before approaching the method, and focused troubleshooting afterwards. Alternatively, by exploring other cost-effective methods, such as RFLP, or advanced methods, such as NGS or HRM, the outcomes could be achieved more expediently. However, as stated, this requires a cost analysis due to the expensive resources needed.

In conclusion, the necessity to determine carrier frequencies of the locally identified pathogenic variants of the black sub-Saharan population remains. Despite the failures of the study, the potential to troubleshoot or explore alternative methods is encouraging. To reiterate, further research is required, but the recommendations found within this study should be considered. This is especially important for the vastly growing field of IEM, and the frequency with which pathogenic variants within our cohort are being identified.

REFERENCES

1. Henderson H, Leisegang F, Brown R, Eley R. The clinical and molecular spectrum of galactosemia in patients from the Cape Town region of South Africa. BMC Pediatr. 2002;2:7.

2. Sullivan L. Power and Sample Size Determination [Internet]. Boston University School of Public Health. [cited 2023 Mar 13]. Available from: https://sphweb.bumc.bu.edu/otit/mph-modules/bs/bs704_power/bs704_power_print.html

3. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012;13:134.

4. BigDye Terminator v3.1 Cycle Sequencing Kit Protocol [Internet]. Applied Biosystems. 2010 [cited 2022 Jul 4]. p. 1–72. Available from: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_081527.pdf

5. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and program for Windows 95/98/NT. nucleic acid Symp Ser. 1999;41:95–8.

6. Scitable. Hardy-Weinberg Equation [Internet]. NatureEducation. [cited 2023 Mar 13]. Available from: https://www.nature.com/scitable/definition/hardy-weinberg-equation-299/

7. Mayo O. A Century of Hardy – Weinberg Equilibrium. 2008;11(3):249–56.

8. Martins AM. Inborn errors of metabolism: a clinical overview. Sao Paulo Med J. 1999;117(6):251–65.

9. Yamaguchi-Kabata Y, Yasuda J, Uruno A, Shimokawa K, Koshiba S, Suzuki Y, et al. Estimating carrier frequencies of newborn screening disorders using a whole-genome reference panel of 3552 Japanese individuals. Hum Genet [Internet]. 2019;138(4):389–409. Available from: http://dx.doi.org/10.1007/s00439-019-01998-7

10. Gouveia MH, Borda V, Leal TP, Moreira RG, Bergen AW, Kehdy FSG, et al. Origins, admixture dynamics, and homogenization of the african gene pool in the americas. Mol Biol Evol. 2020;37(6):1647–56.

11. Donovan BM, Semmens R, Keck P, Brimhall E, Busch KC, Weindling M, et al. Toward a more humane genetics education: Learning about the social and quantitative complexities of human genetic variation research could reduce racial bias in adolescent and adult populations. Sci Educ. 2019 May 1;103(3):529–60.

12. Lehohla P. Census 2011: Census in brief [Internet]. Pretoria; 2022 [Cited 2022 May 1]. Available: http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf

13. Bateman C. Inborn Errors of Metabolism: Heartache - Saving Test To Go National. SAMJ. 93(10):734.

14. Seymour C, Thomason M, Chalmers R, Addison G, Bain M, Cockburn F, et al. Newborn screening for Inborn errors of metabolism: a systematic review. Vol. 1, Health technology assessment. 1997.

15. Gregg A, Aarabi M, Klugman S, Leach N, Bashford M, Goldwaser T, et al. Screening for autosomal recessive and X-linked conditions during pregnancy and preconception: a practice resource of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 23(10):1793–806.

16. Johansen Taber K, Ben-Shachar R, Torres R, Arjunan A, Muzzey D, Kaseniit KE, et al. A guidelines-consistent carrier screening panel that supports equity across diverse populations. Genet Med [Internet]. 2022;24(1):201–13. Available from: https://doi.org/10.1016/j.gim.2021.09.009

17. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434–43.

18. The 1000 genomes Project Consortium. A Global reference for human genetic variation. Nature. (526):68–74.

19. Maphumulo W, Bhengu B. Challenges of quality improvement in the healthcare of South Africa post-apartheid: A critical review. Curationis [Internet]. 2019;42(1):a1901. Available from: https://doi.org/10.4102/curationis

20. Figueredo Lago JE, Armas Cayarga A, González González YJ, Collazo Mesa T. A simple, fast and inexpensive method for mutation scanning of CFTR gene. BMC Med Genet. 2017 May 25;18(1).

21. Hamfjord J, Stangeland AM, Skrede ML, Tveit KM, Ikdahl T, Kure EH. Wobble-enhanced ARMS method for detection of KRAS and BRAF mutations. Diagn Mol Pathol. 2011;20(3):158–65.

22. Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res. 1989;17(7):2503–16.

23. Santalucia J. The Four Most Commonly Encountered Problems in Multiplex Panel Design [Internet]. dnasoftware. [cited 2023 Mar 13]. Available from: https://www.dnasoftware.com /resources /whitepapers/

24. Santalucia J. The Four Most Commonly Encountered Problems in Multiplex Panel Design [Internet]. dnasoftware. [cited 2022 Oct 14]. Available from: www.dnasoftware.com

25. Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH. Multiplex PCR: Critical parameters and Step-by-step protocol. Biotechniques. 1997;(23):504–11.

26. Costa IR, Santos TR, Bento IA, Siqueira BO, Barbosa AM, Silva KSF, et al. RFLP-PCR is more efficient than ARMS-PCR for identifying CYP2C19*2 polymorphism in atherosclerotic patients. Genet Mol Res. 2019;18(3):1–5.

27. Williams RC. Restriction Fragment Length Polymorphism (RFLP). Yearb Phys Anthropol. 1989;(32):159–84.

28. Jawaheer D, Ollier WER, Thomson W. Multiplex ARMS-RFLP: A Simple and Rapid Method for HLA-DR4 Subtyping. Int J Immunogenet [Internet]. 1993;20(3):175–87. Available from: https:// doi.org/10.1111/j.1744-313X.1993.tb00108.x

29. Ogouma-Aworet L, Rabes JP, De Mazancourt P. A Simple RFLP-Based Method for HFE Gene Multiplex Amplification and Determination of Hereditary Hemochromatosis-Causing Mutation C282Y and H63D Variant with Highly Sensitive Determination of Contamination. Biomed Res Int. 2020;2020.

30. Guo K, Zhou X, Chen X, Wu Y, Liu C, Kong Q. Expanded newborn screening for inborn errors of metabolism and genetic characteristics in a Chinese population. Front Genet. 2018;9(APR):1–7.

31. Er TK, Chang JG. High-resolution melting: Applications in genetic disorders. Clin Chim Acta [Internet]. 2012;414:197–201. Available from: http://dx.doi.org/10.1016/j.cca.2012.09.012

32. ThermoFisher SCIENTIFIC. High Resolution Melting (HRM) [Internet]. [cited 2022 Oct 14]. Available from: https://www.thermofisher.com/za/en/home/life-science/pcr/real-time-pcr/real-time-pcr-applications/genetic-variation-analysis-using-real-time/high-resolution-melting-hrm.html

33. Engel H, Kupper C, Loffert D. Highly efficient multiplex PCR using novel reaction chemistry. QIAGEN News. 2003;(2):41–3.

34. qiagen. Taq DNA Polymerase [Internet]. [cited 2022 Apr 2]. Available from: https:// www. qiagen.com/us/products/discovery-and-translational-research/pcr-qpcr-dpcr/pcr-enzymes-and-kits/end-point-pcr/taq-dna-polymerase/

35. QIAGEN. Is Q-solution required for PCR with QIAGEN's PCR kits? [Internet]. Available from: https://www.qiagen.com/no/resources/faq?id=66b24579-8104-44be-8754-550624868882&lang=en

36. Ayyadevara S, Thaden JJ, Shmookler Reis RJ. Discrimination of primer 3'-nucleotide mismatch by Taq DNA polymerase during polymerase chain reaction. Anal Biochem. 2000;284(1):11–8.

37. Mamedov TG, Pienaar E, Whitney SE, TerMaat JR, Carvill G, Goliath R, et al. A fundamental study of the PCR amplification of GC-rich DNA templates. Comput Biol Chem. 2008;32(6):452–7.

38. Yuan J, Yi J, Zhan M, Xie Q, Zhen TT, Zhou J, et al. The web-based multiplex PCR primer design software Ultiplex and the associated experimental workflow: up to 100- plex multiplicity. BMC Genomics. 2021;22(1):1–17.

39. Dastsooz H, Imanieh MH, Dehghani SM, Haghighat M, Moini M, Fardaei M. Multiplex ARMS PCR to detect 8 common mutations of ATP7B gene in patients with Wilson disease. Hepat Mon. 2013 May;13(5).

40. Vamvakopoulos JE. Multiplex Universal Genotyping Using a Modified ARMS-PCR Protocol. Biotechniques. 2002;33(5):1110–2.

41. Felipe R, Medrano V, Andréa De Oliveira C. Guidelines for the Tetra-Primer ARMS-PCR Technique Development. [cited 2022 Aug 18]; Available from: http://primer1.soton.ac.uk/pri

APPENDIX A:

List of primers used in ARMS PCR assay used for Chapter 2 section 2.4

		Sequence (5’ to 3’)			
	Gene	Forward	Reverse	Size	AT
Internal control primers	*AGXT* exon 1-2	M13-CCCATCCACCAATCCTCACC	M13-AACCCCACCCTGTCACAAAG	978 bp	62°C
	MT-ATP6	GGCACAGTGATTATAGGCTTTCGC	GGAGCGTTATGGAGTGGAAGTGAA	477 bp	56°C
	GALT exon 5	ATGTGCTTCCGACCCTGGTC	GGAGGGGCGACCTCACAAAC	146 bp	55°C
Sequencing primers	OXCT1	M13-GCTCTTCAATTGACCAAGCAGA	M13-CCTGCACAGAGGCATAGACC		58°C

AT: Annealing temperature

APPENDIX B:

M13 location on OXCT1 gene

Intron 7-8 41,840,450 41,807,439 33,012

...tattaattaatatttcagttttgaagtttccatattggaattaattatctgtctttaagtaaattatttcagcaattgcagttgtatcacgatcctag
gaagctaacaagaataatttacaactgttacaaatagcatttgtatctatatttaaaaagctcttcaattgaccaagcagatgttattacagatttt
tgtctttgatatgataaatttcccattaatttagaaaatctgtgtttaaggaatgtgctatcattgcatagttgaactgcaatttgtgaaagttcaag
ctgctgtactgacaaaccatttgacaaatttgtaactgagactcttttaagcaaagactaacttctggagcctaatggtctgtattttctgttcagac
gcatacactttaatttcttctcttttgtgggtccaggaggtgtgatagacattacctatatgtatacatatacatacatagatatatgtccatattccg
tcatatgtttgcagcctaagttgtatgtgaagttgtgtatacttttttaaaaaattaaaaatgttcactttaagctttcactaactttgaaagaaatttc
tctttgtctag

8 ENSE00000742478 41,807,438 41,807,331 0 0 108

GTTGAAGAAATTGTGGATATTGGAGCATTTGCTCCAGAAGACATCCATATTCCTCAGATTTATGTACATC**G**CCT
TATAAAGGGAGAAAAATATGAGAAAAGAATTGAG

Intron 8-9 41,807,330 41,805,682 1,649

Gtaattgacttagctgctttgtcagtattcatggatccagttacagagatagtgcatccctggggccattcgagatgatggggaactagagctga
gcagaaagtctggtcttggatcttagtaagagcttctcatgtgtctatcatcatggggatatgatggtctatgcctctgtgcaggctgacagtgcct
tgtcatcacaagatatgtatcaccatgccatgagattctctaggctgtgtctactaggcttctactgtcctaatttatgatctgtagtttttctggctgt
cttagaggggagataagaataatcatgtctttggtggttttatccctgctgtagtcttatcttaatgt...

OXCT1 exon 8 with adjacent introns. Primer sites are highlighted in yellow. SNV variant is in bold red. Adapted from Ensembl.org (29)

APPENDIX C

Official ethics approval letter from the Human Research Ethics Committee

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room G50- Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 6492
Email: hrec-submissions@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

23 March 2021

HREC REF: 157/2021

Mrs S Meldau
Division of Chemical Pathology
FHS
Email: Surita.meldau@uct.ac.za

Dear Mrs Meldau

PROJECT TITLE: CARRIER FREQUENCY DETERMINATION OF PATHOGENIC VARIANTS IN FOUR INBORN ERRORS OF METABOLISM IN A WESTERN CAPE SOUTH AFRICAN POPULATION-MMED CANDIDATE-DR RONALD DALMACIO-SUB-STUDY LINKED TO 196/2018

Thank you for submitting your study to the Faculty of Health Sciences Human Research Ethics Committee (HREC) for review.

It is a pleasure to inform you that the HREC has **formally approved** the above-mentioned study.

This approval is subject to strict adherence to the HREC recommendations regarding research involving human participants during COVID -19, dated 17 March 2020 & 06 July 2020.

Approval is granted for one year until the 30 March 2022.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

The HREC acknowledge that the student: Dr Ronald Dalmacio will also be involved in this study.

Please quote the HREC REF 157/2021 in all your correspondence.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate institutional approval, where necessary, before the research may occur.

HREC/REF 157/2021sa

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, FACULTY OF HEALTH SCIENCES HUMAN RESEARCH ETHICS COMMITTEE
Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938
NHREC-registration number: REC-210208-007

This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use: Good Clinical Practice (ICH GCP), South African Good Clinical Practice Guidelines (DoH 2006), based on the Association of the British Pharmaceutical Industry Guidelines (ABPI), and Declaration of Helsinki (2013) guidelines. The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

HREC/REF 157/2021sa