

Investigating gene mutations in a South African paediatric cohort diagnosed with mitochondrial disease

M Schoonen



orcid.org 0000-0001-6762-2232

Thesis submitted in fulfilment of the requirements for the degree
Doctor of Philosophy in Biochemistry at the North-West
University

Promoter:

Prof FH van der Westhuizen

Co-promoter:

Prof I Smuts

Graduation May 2019

20574495

"It is paradoxical, yet true, to say, that the more we know, the more ignorant we become in the absolute sense, for it is only through enlightenment that we become conscious of our limitations. Precisely one of the most gratifying results of intellectual evolution is the continuous opening up of new and greater prospects."

Nikola Tesla

*"Persistence is very important.
You should not give up unless you are forced to give up."*

Elon Musk

ACKNOWLEDGEMENTS

I would like to thank the following individuals and institutions, whom all played a pivotal part in completing this thesis:

First, I would like to express my utmost thanks to my promotor, **Prof. Francois H. van der Westhuizen**, and co-promoter, **Prof. Izelle Smuts**, for their kindness, support, guidance and expert knowledge throughout this study. I will always be grateful for your invaluable time.

Dr. **Marianne** Pretorius, who showed me unconditional love and support throughout this study. Thank you for all the “uppers and downers”, the care packages when times were tough, and the smile of encouragement every day.

Dr. **Laneke** Luies for the comprehensive and timely formatting of my thesis and your endless and unconditional support throughout this study. Thank you for being the most wonderful best friend a girl could ask for.

Prof. **Eugene** Engelbrecht, for comprehensive and prompt language editing.

My Mitochondrial Research Laboratory friends, in particular **John-Drew** Bosshoff, **Jaundrie** Fourie, **Michelle** Mereis, and **Liesel** Mienie, all of whom helped me, encouraged me, and made my life easier over the past five years. Thank you for all the laughs and craziness that kept me sane.

My dearest family-friends, Ds. **Leon** Geel, Oom **Nielen** and Tannie **Rita** Conradie. Thank you for the constant belief in me, your endless hugs and warm words of encouragements and above all, your sincere interest in my work.

I would like to express my deepest appreciation to my family, in particular my parents, **Johan** and **Joanie** Schoonen. Thank you for accepting nothing less than excellence from me. You are my pillars and my strength. I love you.

Bertie Seyffert, I would never have been able to complete this huge undertaking without your constant support, love and encouragements. I love you.

Finally, the financial assistance of the National Research Foundation (NRF), the South African Medical Research Council (MRC), and the North-West University (NWU) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF, MRC, and NWU.

SUMMARY

Mitochondrial diseases (MD) are a clinically heterogeneous group of genetic disorders that affect the neuromuscular system, the central nervous system (collectively known as encephalomyopathies), and other high-energy demanding organs, often with multi-system involvement. This involvement, whether single- or multi-system, can have extensive phenotypical features which could make the diagnosis of MD a daunting task. Over recent years, substantial progress has been made towards a better understanding of the most common mitochondrial encephalomyopathies including their clinical manifestation and underlying genetic cause. This progress has been made in mainly non-African patient populations. For these well-studied populations, routine clinical, biochemical, and genetic diagnostic approaches have consequently been established, improving diagnosis of MD in paediatric and adult patients alike. The limited diagnostic capacity in developing countries such as South Africa with its ethnically diverse populations, has hindered the advancement of diagnostic procedures for MDs in these countries. As a result, there is limited information available on clinical manifestations and genetic causation for the most common MDs in these understudied patient populations. However, over the last decade the following progress has been made towards an understanding of the underlying genetic cause of MD in South African patients: Firstly, since 1998 paediatric clinical referrals and assessments have been performed on paediatric patients with mitochondrial-disease-like signs and symptoms at the *Steve Biko Academic Hospital*, Pretoria (forming the cohort for this investigation). Secondly, biochemical analyses also commenced shortly afterwards and were performed for every patient who has been clinically diagnosed with an MD, which at the time of this investigation was 212. Thirdly, genetic investigations followed, starting with a small number of patients ($n = 71$) for whom whole mitochondrial genome next-generation sequencing was performed. This number has been expanded throughout recent years, and to date there are whole mitochondrial DNA sequencing data for 123 patients. Results from this initial mtDNA sequencing approach revealed relatively few pathogenic mutations. For the majority of these cases, a clear MD aetiology could not be established. It was collaboratively decided to perform targeted gene panel sequencing on nuclear-encoded genes associated with MD. Three panels were designed and used in this study, each consisting of genes directly involved with the mitochondrion. An in-house bioinformatics pipeline was developed during this study to analyse the sequence data.

The results for the nuclear gene investigations revealed that a clear genotype-phenotype correlation could be established in only two of the 85 selected cases. One of these patients was

extensively investigated and later published as a case-report. Due to the disappointingly low diagnostic yield of panel sequencing, whole exome sequencing was performed on a small number of African patients to probe the outcome of this approach. These initial results were promising as six of the eight sequenced cases presented with a pathogenic or likely pathogenic variant. Additional screening for two putative African-population-specific gene mutations was also performed on all of the African cases. Combined, from the results, the aetiology could be determined in ten cases; eight cases using next-generation sequencing, and two cases using selected gene mutation screening. In conclusion, this study was the first in size and scope to investigate the molecular genetics of MD in an understudied ethnically diverse population, providing new knowledge on these patients and insight into future strategies. Diagnosis of MD remains a daunting task in South Africa and will only improve once diagnostic capacity has been significantly enhanced.

Key words: African paediatric cohort; mitochondrial disease; bioinformatics; panel sequencing; whole exome sequencing; mitochondrial DNA; nuclear DNA

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	v
SUMMARY	vii
LIST OF FIGURES	xv
LIST OF TABLES.....	xxiii
LIST OF ABBREVIATIONS AND SYMBOLS.....	xxvii
CHAPTER 1: INTRODUCTION	1
1.1 BACKGROUND AND MOTIVATION.....	1
1.2 RESEARCH AIM AND OBJECTIVES	3
1.3 STRUCTURE OF THESIS AND RESEARCH OUTPUTS	6
1.4 AUTHOR CONTRIBUTIONS	8
CHAPTER 2: LITERATURE REVIEW	11
2.1 MITOCHONDRIAL BIOLOGY	11
2.1.1 Structure and function of mitochondria	11
2.1.2 Electron transport chain and oxidative phosphorylation	12
2.1.3 Oxidative phosphorylation dysfunction.....	13
2.1.3.1 Complex I.....	14
2.1.3.2 Complex II.....	15
2.1.3.3 Complex III.....	16
2.1.3.4 Complex IV	17
2.1.3.5 Complex V	18
2.1.3.6 Electron transfer proteins	19
2.2 MITOCHONDRIAL GENETICS.....	20
2.2.1 Mitochondrial genome	20
2.2.2 Mitochondrial DNA mutations of mitochondrial disease	21
2.2.3 Nuclear DNA mutations of mitochondrial disease	23
2.3 DIAGNOSIS OF MITOCHONDRIAL DISEASE	27
2.3.1 Clinical Diagnosis	27
2.3.2 Biochemical diagnosis	27

2.3.3	Molecular diagnosis	28
2.3.4	Classic “biopsy first” approach	28
2.3.5	New “genetics first” approach.....	28
2.3.6	South African diagnostic approach.....	30
CHAPTER 3: SOUTH AFRICAN MITOCHONDRIAL PAEDIATRIC COHORT		33
3.1	INTRODUCTION	33
3.2	MITOCHONDRIAL DISEASE COHORT IN SOUTH AFRICA	33
3.2.1	South African populations in general	33
3.2.2	South African diagnostic and clinical perspective	36
3.3	SAMPLE COLLECTION AND ETHICAL CONSIDERATIONS	37
3.3.1	Clinical overview of patient cohort with suspected mitochondrial disease.....	38
3.4	BIOCHEMICAL ASSESSMENTS	39
3.4.1	Metabolic screening	39
3.4.2	Muscle respiratory chain enzymology.....	40
3.5	STRATIFICATION FOR MOLECULAR WORK.....	41
3.5.1	Selection criteria.....	41
CHAPTER 4: NEXT-GENERATION SEQUENCING		43
4.1	INTRODUCTION	43
4.2	DNA ISOLATION.....	44
4.3	NUCLEAR TARGETED SEQUENCING	44
4.3.1	Panel sequencing.....	44
4.3.2	Library preparation using Agilent HaloPlex™ Target Enrichment for Panel 1	45
4.3.3	Library preparation using Ion AmpliSeq custom DNA Panels 2 and 3	50
4.3.4	Template preparation	54
4.3.5	Ion Torrent next-generation sequencing	55
4.4	MITOCHONDRIAL GENOME SEQUENCING	56
4.4.1	Amplification and sample preparation for next-generation sequencing	56
4.5	WHOLE EXOME SEQUENCING.....	57
4.6	MUTATION SCREENING IN TWO GENES, <i>GCDH</i> AND <i>MPV17</i>	57
4.7	PATIENTS SELECTED FOR SEQUENCING AND MUTATION SCREENING	58
CHAPTER 5: BIOINFORMATICS AND COMPUTATIONAL TOOLS		63
5.1	ABSTRACT	63
5.2	INTRODUCTION.....	64

5.3	BIOINFORMATICS PIPELINE	65
5.3.1	Primary data analysis	67
5.3.2	Secondary data analysis.....	68
5.4	ADDITIONAL INFORMATION ON GEMINI.....	71
5.5	MITOCHONDRIAL DNA ANALYSIS	73
5.6	CRITERIA FOR IDENTIFYING AND EVALUATING NUCLEAR VARIANTS	74
5.6.1	High confidence pathogenic variants	75
5.6.2	Moderately pathogenic variants	75
5.6.3	High confidence benign variants	76
5.7	CRITERIA FOR IDENTIFYING AND EVALUATING MITOCHONDRIAL DISEASE-CAUSING VARIANTS.....	78
5.8	CONCLUSION ON BIOINFORMATICS IN SOUTH AFRICA	79
 CHAPTER 6: PATHOGENICITY EVALUATION OF DETECTED mtDNA AND nDNA VARIANTS		81
6.1	OVERVIEW OF VARIANTS DETECTED	81
6.1.1	Panel sequencing	82
6.1.1.1	Panel 1: Agilent HaloPlex — complex I associated genes.....	82
6.1.1.2	Panel 2: Ion Ampliseq — complex I-IV associated genes.....	84
6.1.2	Whole exome sequencing	85
6.1.3	Mitochondrial DNA sequencing.....	87
6.2	NUCLEAR VARIANTS OF IMPORTANCE	89
6.2.1	Pathogenic variants	92
6.2.2	Likely pathogenic variants	92
6.2.3	Variants of uncertain significance	97
6.2.4	Likely benign and benign variants.....	98
6.3	DISCUSSION AND CONCLUSIONS	103
 CHAPTER 7: REPORTING OF A NOVEL <i>ETFDH</i> MUTATION — A CASE REPORT		105
7.1	ABSTRACT.....	105
7.2	INTRODUCTION	106
7.3	PATIENTS AND METHODS	107
7.3.1	Patient 1 (index patient).....	107
7.3.2	Patient 2	109
7.3.3	Patient 3	109

7.3.4	Metabolic and biochemical investigations.....	110
7.3.5	Mutation and protein analyses.....	110
7.4	RESULTS.....	111
7.4.1	Metabolic and biochemical investigations.....	111
7.4.2	Mutation and protein analysis.....	111
7.5	DISCUSSION AND CONCLUSION	113
 CHAPTER 8: MUTATION SCREENING OF TWO PUTATIVE POPULATION VARIANTS IN THE <i>GCDH</i> AND <i>MPV17</i> GENES.....		
8.1	INTRODUCTION.....	115
8.2	BACKGROUND ON FOUNDER MUTATION ON THE GENE <i>GCDH</i> , KNOWN TO BE INVOLVED WITH GLUTARIC ACIDEMIA TYPE I	116
8.2.1	Glutaric acidemia type I.....	116
8.2.2	Methods	118
8.2.2.1	Patient selection and samples	118
8.2.2.2	Metabolic and biochemical investigation	119
8.2.2.3	Genetic screening analysis	119
8.2.3	Results.....	121
8.2.3.1	Mutation analysis	121
8.2.4	Case descriptions.....	126
8.2.4.1	Case 1: Patient S013.....	127
8.2.4.2	Case 2: Patient S127	128
8.2.4.3	Case 3: Patient S094.....	129
8.2.4.4	Case 4: Patient S114.....	130
8.3	BACKGROUND ON FOUNDER MUTATION ON THE GENE <i>MPV17</i> , KNOWN TO BE INVOLVED WITH MITOCHONDRIAL DEPLETION SYNDROME.....	130
8.3.1	Mitochondrial DNA depletion syndrome	130
8.3.2	Methods	132
8.3.2.1	Patient selection and samples	132
8.3.2.2	Metabolic and biochemical investigation	132
8.3.2.3	Genetic screening analysis	132
8.3.3	Results.....	133
8.3.3.1	Mutation analysis	133
8.4	DISCUSSION	136
8.4.1	Glutaric acidemia type I.....	136

8.4.2	Mitochondrial DNA depletion syndrome	137
CHAPTER 9: SUMMARY, CONCLUSIONS AND FUTURE PROSPECTS		139
9.1	RATIONALE BEHIND AN INVESTIGATION INTO MITOCHONDRIAL DISEASE AETIOLOGY IN SOUTH AFRICAN POPULATIONS.....	139
9.2	RESEARCH AIMS AND OBJECTIVES.....	141
9.2.1	Objective 1: Nuclear gene selection for high throughput next-generation sequencing	141
9.2.2	Objective 2: Patient selection for next-generation sequencing	142
9.2.3	Objective 3: Generate high quality next-generation sequencing data.....	142
9.2.4	Objective 4: Design a bioinformatics pipeline with the focus on nuclear DNA sequencing data	143
9.2.5	Objective 5: Pathogenicity evaluation of identified variants	143
9.2.6	Objective 6: Follow up on a pathogenic mutation in a selected case.....	145
9.2.7	Objective 7: Population specific mutation screening	145
9.3	FINAL CONCLUSION AND FUTURE PROSPECTS	146
REFERENCES		151
APPENDIX A: SCIENTIFIC OUTPUTS FROM THIS INVESTIGATION		175
APPENDIX A.1		179
APPENDIX A.2		192
APPENDIX A.3		226
APPENDIX A.4		231
APPENDIX B: MITOCHONDRIAL COHORT CONSENT AND ASSENT FORM.....		237
APPENDIX B.1		238
APPENDIX B.2		240

LIST OF FIGURES

Chapter 1:

Figure 1.1: Illustration of the experimental design aimed at investigating the aetiology of MD in this understudied patient population.	5
---	----------

Chapter 2:

Figure 2.1: Simplified illustration of the mitochondria. These are ubiquitously found within almost all eukaryotic cells. They contain their own DNA, primarily found within the mitochondrial matrix and produce energy in the form of ATP via the oxidative phosphorylation system.	12
---	-----------

Figure 2.2: Oxidative phosphorylation in the mitochondria. Electrons are carried from CI to CIV (electron transport chain) in a step-wise manner. Ubiquinone carries electrons from CI and CII to CIII and cytochrome c carries the electrons from CIII to CIV. Hydrogen ions are expelled across the mitochondrial inner membrane into the intermembrane space by complexes I, III and IV. This proton motive force is a key component of ATP production via ATP synthase.....	13
--	-----------

Figure 2.3: The complete human mitochondrial genome of 16.6 kb double-stranded DNA. The heavy (H) strand is represented by the outer circle and the light (L) strand is represented by the inner circle. mtDNA encodes 24 ribosomal RNA (grey), 22 transfer RNA (white), and 13 RC structural genes, indicated by blue for CI genes, green for CIII genes, light blue for CIV genes, and orange for CV genes. Human mitochondrial genome map, courtesy of Emmanuel Douzery (https://commons.wikimedia.org/wiki/File: Map_of_the_human_mitochondrial_genome.svg last accessed 2018).....	22
---	-----------

Figure 2.4: Flowchart of diagnostic procedures followed for patient with suspected MD. A: “Biopsy first” approach that was followed until recently. B: “Genetics first” approach that is being followed more recently.....	29
---	-----------

Figure 2.5: Flowchart of diagnostic procedures followed in South Africa. White blocks represent the procedures that are routinely done while the grey blocks represent atypical procedures.	31
---	-----------

Chapter 3:

Figure 3.1: Simplified illustration of South African demographics. Panel A: Population density in South Africa. Panel B: Population ethnicity distribution in South Africa..... 34

Figure 3.2: Simplified illustration of South African ethnicity distribution. Panel A: Black African population distribution. Panel B: Coloured population distribution. Panel C: Caucasian population distribution. Legend applies to all three panels where lighter colour is lower density and darker colour is higher density..... 35

Figure 3.3: Provinces of South Africa. Orange dot: Paediatric Neurology Clinic of the Steve Biko Academic Hospital in Pretoria (Centre 1). Purple dot: National Health Laboratory Services (NHLS) in Cape Town (Centre 2) and red dot: Centre for Human Metabolomics, North-West University in Potchefstroom (Centre 3)..... 36

Figure 3.4: Demographical background of the paediatric patient cohort. Panel A: Distribution of ethnicity. Panel B: Age distribution at onset of symptoms. Caucasian, Asian, and Coloured patients are collectively referred to as “non-African patients” in the text. 38

Figure 3.5: Distribution of the clinical features of the patients. Percentages were calculated for 81 African and 46 non-African patients, respectively..... 39

Figure 3.6: Distribution of respiratory chain muscle deficiencies in the selected patient cohort. Distribution of 127 patients with combined or isolated respiratory chain enzyme deficiencies is shown in Panel A. Distribution of isolated CI–CIV deficiencies is shown in Panel B. 41

Chapter 4:

Figure 4.1: Process followed for HaloPlex Targeted Enrichment. Panel A: A schematic representation of a simplified HaloPlex™ Target Enrichment workflow. A1: Genomic DNA is fragmented using 16 restriction enzymes. A2: The probe library is added and designed so that both ends of the oligonucleotide. are hybridised to targeted DNA fragments. Circular DNA molecules are formed as a result. A unique sample barcode sequence is also incorporated during this step. The probes are biotinylated and can therefore be retrieved with magnetic streptavidin beads. A3: Ligations close the circular molecule only if these are perfectly hybridised fragments. A4: Circular DNA targets are amplified via PCR. This enriched barcoded product is ready for sequencing. Adapted from the application note: Agilent HaloPlex Target Enrichment and SureCall Data Analysis: Optimised for the Ion Torrent™ PGM Sequencer. Panel B: Agarose gel electrophoresis of digested samples with various restriction enzymes (Lanes 1–8), an undigested

enriched DNA control sample in Lane 9, with the ladder in lane far left. Panel C: Validation and quantification of enriched libraries using the 2100 Bioanalyzer High Sensitivity DNA Assay Kit. Successful target enrichment is indicated by a size distribution smear ranging from 175–625 bp, as can be seen for patient samples 1–7. The lower ladder marker is indicated by a green line (at 35 bp) and the upper ladder marker is indicated by a purple line (>10 000 bp). Panel D: The electropherogram for all samples (Sample 7 shown here) displayed a peak fragment size between 225–525 bp. The peak fragments observed at 35 bp and 10 380 bp indicate the lower and upper ladder markers, respectively. The peak fragments observed at ~75 bp and 125 bp are considered primer- and adaptor-dimer products, respectively. 49

Figure 4.2: Library preparation workflow for Ion AmpliSeq. Panel A: Targeted enrichment. A1: Selected target regions of interest from genomic DNA are amplified by specific primer pairs (red arrows) into ~200–300 bp amplicons, using standard PCR technique (A2). A3: Amplicons are partially digested and purified, followed by ligation of adaptors (light blue), barcodes (orange), and a P1 Adapter (green), necessary for binding to an ion sphere particle (A4). Adapted from the Ion AmpliSeq™ Library Kit 2.0 user guide. Panel B: Agarose gel electrophoresis of amplified samples (step 1 and 2 in A) in Lanes 1–5, with the ladder in the lane far left. Successful amplicon amplification is indicated by the smear ranging from ~150 bp–300 bp, as is the case in the five samples indicated here..... 54

Figure 4.3: A run report generated by the Ion PGM™ Torrent Suite. Panel A indicates the percentage of the chip loading with live ISPs in each of the wells. Higher percentages indicate better loading where deep red indicates highest loading and blue indicates lowest loading in each respective well. Panel B gives information on the ISPs. Percentages of enriched ISPs as well as the number of polyclonal ISPs are summarised. Panel C gives information on the mean, median and mode for read lengths..... 55

Figure 4.4: Validation of PCR amplification using agarose gel electrophoresis, with a 250–10 000 bp ladder in both cases. Panel A: Amplification of two fragments, where “I” is Fragment A and “II” is Fragment B. One sample was loaded in alternate lanes (Lanes 1–6). Panel B: An equimolar amount of Fragment A and Fragment B was used, where Fragment A is at the top and Fragment B is at the bottom in Lanes 1–4..... 57

Chapter 5:

Figure 5.1: Simplified illustration for the in-house developed next-generation sequencing bioinformatics pipeline used for identifying disease-causing variants from Ion Torrent sequencing data. The first component of this illustration is primary data analysis, a semi-

automated process done using the Ion Torrent Software (Torrent Suite). The second component illustrates software used for secondary data analysis..... 66

Figure 5.2: Example of a variant call format (VCF) generated on Torrent Suite. a: Metadata of sequencing run; b: VCF tag definitions; c: Sequencing contig information; d: Details on the variant detected. 68

Figure 5.3: Summary of mtDNA bioinformatics pipeline. Next-generation sequencing bioinformatics pipeline. This approach was followed for identifying disease-causing variants from Ion Torrent sequencing data using mtDNA-server..... 74

Chapter 6:

Figure 6.1: Overview of Panel 1 NGS findings for 32 selected patients. Panel A: The total number of variants detected in each of the 32 patients. Africans are marked with an asterisk. Panel B: The average number of variants detected for Africans and non-Africans. The black bar represents the average number of variants detected for both African and non-African patients. The orange bar indicates African patients, and the red bar indicates non-Africans..... 82

Figure 6.2: Novel and reported NGS variants detected in 32 patients. Panel A: The total number of novel (left vertical axis) and previously reported (right vertical axis) variants detected in each patient. Africans are marked with an asterisk. Panel B: Average number of novel (left) and reported (right) variants detected in African and non-African patients, respectively. The black bars represent the average number of variants detected for both African and non-African patients. The orange bar indicates African patients, and the red bar indicates non-Africans..... 83

Figure 6.3: Overview of Panel 2 NGS findings for 48 selected patients. Panel A: The total number of variants detected in each of the 48 patients. Africans are marked with an asterisk. Panel B: The average number of variants detected for Africans and non-Africans. The orange bar indicates African patients while the red bar indicates non-Africans. The black bar represents the average number of variants detected for both African and non-African patients..... 84

Figure 6.4: Novel and reported NGS variants detected in 48 patients. Panel A: The total number of novel (left vertical axis) and previously reported (right vertical axis) variants detected in each patient. Africans are marked with an asterisk. Panel B: Average number of novel (left) and reported (right) variants detected in African and non-African patients, respectively. The black bars represent the average number of variants detected for both African and non-African patients, the orange bars indicate African patients, and the red bars indicate non-Africans..... 85

Figure 6.5: Overview of WES findings for the selected eight African patients. Panel A: Total number of variants detected in each of the eight cases. Panel B: The total number of novel (left vertical axis) and previously reported (right vertical axis) variants detected in each patient. 86

Chapter 7:

Figure 7.1: Partial sequence alignment of sequence data (A) and RFLP analysis (B) of patients. In Panel A the sequence validation data for the three patients is shown in alignment with the reference sequence for the ETFDH gene (ENS00000171503) at the positions for the c.1067G>A and C.1448C>T mutations. R indicates a heterozygous G to A nucleotide base change and Y a heterozygous C to T nucleotide base change. In Panel B the results from restriction fragment length polymorphism (RFLP) analyses for the patients and appropriate controls are shown. For this, DNA was first amplified using PCR and selected primers for a 580 bp region covering the c.1067>A mutation (fwd: GCACATAGTGCTCCAAATAC; rev: CATGCCTGGCTAATCTTTCC) and a 568 bp region covering the c.1448>T mutation (fwd: CACACATTTGGGCAGTTTCG; rev: AAAGTGATCTGTCCATCGGG), respectively. The resulting fragments were digested with HpaI (GGTGA(N8)) and AluI (AGCT), as indicated in red for the two mutations in Panel A. In both cases the enzymes only cut at the positions when the mutations are present. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.) 112

Figure 7.2: Structural instability of muscle ETFDH in Patient 2. An enriched mitochondrial preparation from muscle of Patient 2 (P2, with c.1067G>A+ c.1448C>T variants) and two healthy controls (C1, C2) were separated on SDS-PAGE and immune-stained for ETFDH and β -actin (top, Panel A). The band intensities for ETFDH were normalized to total protein content in each lane (bottom, Panel A), of which the quantified results are shown in Panel B. 113

Chapter 8:

..... 117

Figure 8.2: Agarose RFLP results for BsaH1 digestion. Examples of a heterozygous (+/-) mutation in Lane 1, homozygous (+/+) mutation in Lane 2 and the wildtype (-/-) in Lane 3, using a 100 bp ladder. Three bands are visible for heterozygous mutations at 104 bp, 85 bp and 19 bp (the latter of which is not visible in this example). Only one band of 104 bp is visible for homozygous mutations since the recognition site is lost. The wildtype samples will result in two bands at 85 bp and 19 bp (the latter of which is not visible in this example). Abbreviations: blk: Blank; 1Kb+: 10 000bp ladder; bp: base pair. 120

Figure 8.3: Image of agarose gel with restriction enzyme digested in samples P006–P105 and S001–S078. A 100 bp DNA ladder was used. Abbreviations: Pos: positive control; Blk: blank.
.....122

Figure 8.4: Image of agarose gel with restriction enzyme digested samples P079–P195 and S080–S119. A 100 bp DNA ladder was used.....123

Figure 8.5: Image of agarose gel with restriction digested DNA samples P203–P218 and S120–S127. A 100 bp DNA ladder was used.....124

Figure 8.6: Image of agarose gel with restriction digested DNA samples that were repeated. The samples were P6–P19 and S001–S013. A 100 bp DNA ladder was used.....124

Figure 8.7: Image of agarose gel with digested DNA samples using a 100 bp ladder. Lane 1 is a repeat of sample P144 with no clear mutation. Lanes 2 and 3 are heterozygotes for the mutation (S094 and S114). Lane 4 (S127) is a homozygote for the mutation. Lane 5 is the positive control sample.. 125

Figure 8.8: Partial Sanger sequence validation data for the four patients is shown in alignment with the reference sequence for the gene GCDH (ENSG00000105607) at the position for the c.877G>A mutation. Patients S013 and S127 presented with homozygous mutations while S094 and S114 presented with heterozygous mutations. 126

Figure 8.9: World health organization head circumference-for-age. Patient S013 presents with a Z-score of 2.28 and is in the ~95th percentile. This indicates macrocephaly for this patient at her age (World Health Organization, 2018). 128

Figure 8.10: World health organization head circumference-for-age. Patient S127 presents with a Z-score of 0.59 indicating a normal head circumference for her age (World Health Organization, 2018). 129

Figure 8.11: Distributed consequences when Mpv17 is compromised due to a mutation in the gene MPV17. Mutations in the modulator responsible for reactive oxygen species (ROS) production cause accumulation of ROS that alters mtDNA content. The depleted amount of mtDNA affects the oxidative phosphorylation system directly, which in turn produces more ROS.
.....131

Figure 8.12: Agarose RFLP results for PvuII digestion. Example for heterozygous (+/-) mutation in Lane 1, homozygous (+/+) mutation in Lane 2 and the wildtype (-/-) in Lane 3, using a 100 bp ladder. Three bands are visible for heterozygous mutations at 336 bp, 244 bp and 96 bp.

Only one band of 336 bp is visible for homozygous mutations, since the recognition site is lost. The wildtype samples will result in two bands at 244 bp and 96 bp. 133

Figure 8.13: Image of agarose gel with restriction enzyme digested samples P006–P105 and S001–S078. A 100 bp DNA ladder was used. 134

Figure 8.14: Image of agarose gel with restriction enzyme digested samples P079–P195 and S080–S119. A 100 bp DNA ladder was used. 135

Figure 8.15: Image of agarose gel with restriction digested DNA samples P203–P218 and S120–S127. A 100 bp DNA ladder was used.. 136

Figure 8.16: Image of agarose gel with restriction digested DNA samples. Lanes 1–12 are samples that underwent RFLP a second time (repeated samples). All samples are wild-type and no heterozygous or homozygous mutations were detected. A 100 bp DNA ladder was used. 136

LIST OF TABLES

Chapter 2:

Table 2.1: Classification of the complex I structural and assembly factors and their respective inheritance patterns.....	14
Table 2.2: Classification of the complex II structural and assembly factors and their respective inheritance patterns.....	16
Table 2.3: Classification of the complex III structural and assembly factors and their respective inheritance patterns.....	17
Table 2.4: Classification of the complex IV structural and assembly factors and their respective inheritance patterns.....	18
Table 2.5: Classification of the complex V structural and assembly factors and their respective inheritance patterns.....	19
Table 2.6: Classification of mtDNA genes.....	21
Table 2.7: Genotype-phenotype correlations of well described human mtDNA pathogenic mutations.....	23
Table 2.8: Clinical manifestations for structural and non-structural RC subunits encoded by nDNA.	24

Chapter 4:

Table 4.1: Properties of three custom designed panels used for massively parallel sequencing.	45
Table 4.2: Properties of targeted Panel 1, arranged alphabetically according to the targeted genes involved with CI.	45
Table 4.3: Properties of targeted Panel 2, arranged alphabetically according to the genes involved with CI, CII, CIII and CIV.	50

Table 4.4: Properties of targeted Panel 3, arranged alphabetically according to the genes involved with CoQ ₁₀ biosynthesis.	53
Table 4.5: Primer pair sequences for Fragments A and B used in PCR amplification.....	56
Table 4.6: List of patient cohort with corresponding NGS approach followed.	58

Chapter 5:

Table 5.1: Description of the fields that are present in the output file generated by GEMINI. ...	71
Table 5.2: Summary of ACMG guidelines and criteria used for evaluating variants of importance, as adapted from Richards et al. (2015b).	77

Chapter 6:

Table 6.1: Summary of mitochondrial DNA variants of importance identified in cohort.	88
Table 6.2: Summary of nuclear DNA variants of interest identified in the cohort.	90
Table 6.3: Cohort phenotypes compared to reported ETFDH clinical phenotypes.	92
Table 6.4: Cohort phenotypes compared to reported SURF1 clinical phenotypes.	93
Table 6.5: Cohort phenotypes compared to reported COQ6 clinical phenotypes.	94
Table 6.6: Cohort phenotypes compared to reported RYR1 clinical phenotypes.	95
Table 6.7: Cohort phenotypes compared to reported STAC3 clinical phenotypes.	96
Table 6.8: Cohort phenotypes compared to reported ALAS2 clinical phenotypes.....	97
Table 6.9: Cohort phenotypes compared to reported TRIOBP clinical phenotypes.	98
Table 6.10: Cohort phenotypes compared to reported DARS2 clinical phenotypes.	99
Table 6.11: Cohort phenotypes compared to reported POLG clinical phenotypes.	100
Table 6.12: Cohort phenotypes compared to reported TAZ clinical phenotypes.	101
Table 6.13: Cohort phenotypes compared to reported ACADVL clinical phenotypes.	102
Table 6.14: Cohort phenotypes compared to reported GLUD2 clinical phenotypes.	103

Chapter 7:

Table 7.1: Summary of clinical, biochemical and genetic features of patients.....	108
---	-----

Chapter 8:

Table 8.1: Forward and reverse primers used for PCR amplification for the gene GCDH.	119
---	-----

Table 8.2: Summary of the clinical, metabolic and biochemical information of patients.....	127
--	-----

Table 8.3: Forward and reverse primers used for PCR amplification for the gene MPV17.....	132
---	-----

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviations:

Abbreviation	Meaning	Abbreviation	Meaning
[Fe-S]	iron-sulphur clusters	+/-	heterozygous
+/+	homozygous	3-OH-GA	3-hydroxyglutaric acid
A	African	A	alanine
ACAD9	acyl-CoA dehydrogenase family member 9	ADP	adenosine diphosphate
ALT	alanine aminotransferase	AST	aspartate aminotransferase
ATP	adenosine triphosphate	ATP5F1A/B/C/D/E	ATP synthase F1 subunit alpha/beta/gamma/delta/epsilon
ATP5IF1	ATP synthase inhibitory factor subunit 1	ATP5MC1/2/3	ATP synthase membrane subunit c locus 1/2/3
ATP5MD	ATP synthase membrane subunit DAPIT	ATP5ME	ATP synthase membrane subunit e/f/g
ATP5MPL	ATP synthase membrane subunit 6.8PL	ATP5PB/D	ATP synthase peripheral stalk-membrane subunit b/d
ATP5PF	ATP synthase peripheral stalk subunit F6	ATP5PO	ATP synthase peripheral stalk subunit OSCP
BAM	binary alignment/map	BCS1L	BCS1 homolog, ubiquinol-cytochrome c reductase complex chaperone
BE	behavioural and emotional involvement	Blk/BLK	blank
BN-PAGE	blue-native polyacrylamide gel electrophoresis	bp	base pair
cyt <i>b</i>	cytochrome <i>b</i>	cyt <i>c</i>	cytochrome <i>c</i>
C	cysteine	C4	butyryl-/isobutyrylcarnitine
C5	isovalerylcarnitine	C5-DC	glutarylcarnitine
C8	octanoylcarnitine	CAF	Centre for Analytical Facilities
Car	cardiac involvement	CBC	complete blood count
CHM	Centre for Human Metabolomics	CHR	chromosome
CI	complex I	CII	complex II
CIII	complex III	CIV	complex IV
CNS	central nervous system	COA1	cytochrome c oxidase assembly factor 1 homolog

COA3–7	cytochrome c oxidase assembly factor 3/4/5/6/7	CoQ₁₀	coenzyme Q ₁₀
COX	cytochrome c oxidase	COX10	haem A:farnesyltransferase cytochrome c oxidase assembly factor
COX11	cytochrome c oxidase copper chaperone	COX14	cytochrome c oxidase assembly factor 14
COX15–16	cytochrome c oxidase assembly homolog 15/16	COX18–20	cytochrome c oxidase assembly factor 18/19/20
COX23	cytochrome c oxidase assembly factor 23	COX17	cytochrome c oxidase copper chaperone
COX4–8C	cytochrome c oxidase subunit 4I1, 4I2, 5A, 5B, 6A, 6B, 6C, 7A, 7B, 7C, 8	CPEO	chronic progressive external ophthalmoplegia
CS	citrate synthase	CYC1	cytochrome C1
CRISPR	clustered regularly interspaced short palindromic repeats	CRISPR/Cas9	clustered regularly interspaced short palindromic repeats associated protein 9
N	aspartic acid	dbSNP	database for single nucleotide polymorphisms
DD	developmental delay	DR	developmental regression
Dys	dysmorphism (minor and major)	Q	glutamic acid
ECSIT	ECSIT signalling integrator	EMG	electromyography
End	endocrine involvement	ENT: Sens. Deaf.	Sensorineural deafness
ETC	electron transport chain	ETF	electron transfer flavoprotein
ETF-FAD	electron transfer flavoprotein-flavin adenine dinucleotide	ETF-FADH	electron transfer flavoprotein-flavin adenine dinucleotide — reduced
ExAC	exome aggregation consortium	Eye	vision involvement
F	female	F	phenylalanine
FAD	flavin adenine dinucleotide	FADH₂	flavin adenine dinucleotide — reduced
FASTKD2	FAST kinase domains 2	FBSN	familial bilateral striatal necrosis
FILA	fatal infantile lactic acidosis	FOXRED1	FAD-dependent oxidoreductase domain containing 1
G	glycine	GA	glutaric acid
GA-1	glutaric acidemia type I	GC-MS	gas chromatography mass spectroscopy
gDNA	genomic DNA	GEMINI	genome mining
GGT	gamma-glutamyltransferase	GIT	gastro intestinal tract involvement
H	histidine	H	hydrogen ions
I	isoleucine	IGV	integrative genomics viewer
IMD	inherited metabolic disease	IMS	intermembrane space
ISPs	Ion Sphere Particles	L	lysine
L	liver involvement	L1	leucine 1 (UUA/G)

L2	leucine 2 (CUN)	LC-MS	liquid chromatography mass spectroscopy
LHON	Leber's hereditary optic neuropathy	LoF	loss of function
LRPPRC	leucine-rich PPR motif-containing protein	LS	Leigh syndrome
LYRM7	LYR motif containing 7	M	male
M	methionine	MI	muscle involvement
MADD	multiple acyl-CoA dehydrogenase deficiency	MD	mitochondrial disease
MDC	mitochondrial disease criteria	MDDS	mitochondrial DNA depletion syndrome
MELAS	mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes	MERRF	myoclonic epilepsy with ragged red fibres
MIDD	maternally inherited diabetes and deafness	MILS	maternally inherited Leigh syndrome
MIM	mitochondrial inner membrane	MM	mitochondrial matrix
MOM	mitochondrial outer membrane	MPS	massively parallel sequencing
mRNA	messenger RNA	MT-ATP6/8	mitochondrially encoded ATP synthase membrane subunit 6/8
MTCO1–3	mitochondrially encoded cytochrome c oxidase I/II/III	MTCYB	mitochondrially encoded cytochrome <i>b</i>
mtDNA	mitochondrial DNA	MTND1–6, D4L	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 1–6 and D4L
MT-T	mitochondrially encoded tRNA	N	asparagine
N	neonatal	NA	non-African
N/A	not applicable	NAD+	nicotinamide adenine dinucleotide — oxidised
NADH	nicotinamide adenine dinucleotide — reduced	NaOH	sodium hydroxide
NARP	neuropathy, ataxia, and retinitis pigmentosa	NCV	nerve conduction velocities
ND	NADH-dehydrogenase	nDNA	nuclear DNA
NDUFA	NADH:ubiquinone oxidoreductase subunit	NDUFA1–13	NADH:ubiquinone oxidoreductase subunit A1/2/3/4/5/6/7/8/9/10/11/12/13
NDUFAB1	NADH:ubiquinone oxidoreductase subunit AB1	NDUFAF2–8	NADH:ubiquinone oxidoreductase complex assembly factor 2/3/4/5/6/7/8
NDUFB1–11	NADH:ubiquinone oxidoreductase subunit B1/2/3/4/5/6/7/8/9/10/11	NDUFC1–2	NADH:ubiquinone oxidoreductase subunit C1/2
NDUFS1–8	NADH:ubiquinone oxidoreductase core subunit S1/2/3/4/5/6/7/8	NDUFV1–3	NADH:ubiquinone oxidoreductase subunit V1/2/3
NHLS	National Health Laboratory Services	NGS	next-generation sequencing
OMIM	Online Mendelian Inheritance in Man	NWU	North-West University
OXPHOS	oxidative phosphorylation	OXA1	oxidase assembly 1-like protein
PCR	polymerase chain reaction	P	proline
Pi	inorganic phosphate	PGM	Personal Genome Machine

PNS	peripheral nervous system neuropathy	PLIEM	Potchefstroom Laboratory for Inborn Errors of Metabolism
Q	ubiquinone	Pos	positive control
R	arginine	Q	glutamine
RB	radiological changes in the brain	R	renal involvement
RCD	respiratory chain deficiency	RC	respiratory chain
RFLP	restriction fragment length polymorphism	rCRS	revised Cambridge Reference Sequence
ROS	reactive oxygen species	RNR1/2	RNA, ribosomal 45S cluster 1/2
S	skeletal involvement	rRNA	ribosomal RNAs
S2	serine 2 (AGU/C)	S1	serine 1 (UCN)
SDHA	succinate dehydrogenase complex flavoprotein subunit A	SCO1–2	cytochrome c oxidase assembly protein 1/2
SDHB	succinate dehydrogenase complex iron sulfur subunit B	SDHAF1–4	succinate dehydrogenase assembly factor1/2/3/4
SDHD	succinate dehydrogenase complex subunit D	SDHC	succinate dehydrogenase complex subunit C
SURF1	cytochrome c oxidase assembly factor	SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TACO1	translational activator of cytochrome c oxidase 1	T	threonine
TCA	tricarboxylic acid	TBE	Tris/Borate/EDTA
TIMMDC1	translocase of inner mitochondrial membrane domain containing 1	TE	Tris-EDTA
tRNA	transfer RNA	TMEM126B	transmembrane protein 126B
UQCC1–3	ubiquinol-cytochrome c reductase complex assembly factor 1/2/3	TTC19	tetratricopeptide repeat domain 19
UQCRB	ubiquinol-cytochrome c reductase binding protein	UQCR10/11	ubiquinol-cytochrome c reductase, complex III subunit X and XI
UQCRC1	ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1	UQCRC1–2	ubiquinol-cytochrome c reductase core protein 1/2
UQCRQ	ubiquinol-cytochrome c reductase complex III subunit VII	UQCRH	ubiquinol-cytochrome c reductase hinge protein
VCF	variant call format	V	valine
WB	Western-blot	W	tryptophan
Y	tyrosine	WES	whole exome sequencing

Symbols:

Symbol	Meaning	Symbol	Meaning
<	less than	↑	elevated
>	greater than	↓	decreased
%	percentage	°C	degrees Celsius
β	beta	kDa	kilo Dalton
μL	microliter	MDa	mega Dalton
M	molar	ng/μL	nanogram per microlitre
mmol/L	millimole per litre	μmol/L	micro molar per liter
nm	nanometer	pM	picomolar

CHAPTER 1:

INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

Mitochondrial diseases (MDs) are the most common clinically heterogeneous group of inherited metabolic disease (IMD), presenting in children with an overall prevalence of approximately 3–6.2 per 100 000 live births (Applegarth & Toone, 2000; Darin *et al.*, 2001; Gorman *et al.*, 2016; Skladal *et al.*, 2003). MDs can be caused by mutations in both mitochondrial and nuclear DNA (mtDNA and nDNA, respectively) as a result of the dual genetic control of oxidative phosphorylation (OXPHOS) (Alston *et al.*, 2017; Yaffe, 1999). Although some MDs affect only a single organ, many involve multiple organ systems with a predominance of neurological and myopathic features, collectively known as mitochondrial encephalomyopathies (DiMauro & Moraes, 1993). MDs have been extensively investigated in Caucasian populations, with well-established routine clinical, biochemical, and genetic approaches. In developed countries, genetic investigations have recently begun to be considered as a first-tier option for identifying pathogenic MD variations (Wortmann *et al.*, 2017). Diagnosis of MDs, however, still remains a daunting task due to the variation involved, including the wide range of clinical phenotypes and genetic heterogeneity of either the mitochondrial or nuclear genome (Alston *et al.*, 2017).

Selected glossary

Some terms used in this thesis could have more than one meaning and are used in the following manner:

1. Primary mitochondrial disease is caused by underlying genetic mutations directly affecting the mitochondria, more specifically the OXPHOS systems. The gene mutations can occur in both the mitochondrial genome and nuclear genes encoding OXPHOS proteins and other mitochondrial related proteins (Niyazov *et al.*, 2016).

2. Secondary mitochondrial disease is caused by mutations in genes that are not directly involved with OXPHOS functioning. Some disorders might present with “mitochondrial like phenotypes” in patients where no mitochondrial-associated gene mutations have been identified. Other non-genetic factors such as environment and toxins may also contribute to secondary MD (Niyazov *et al.*, 2016).

3. Variant vs. mutation: Variants, is the terminology used throughout this thesis for an alteration or nucleotide change observed in sequencing data when aligned to a reference genome. A variant can have no impact on cell function, or it can cause severe harm to the cell, resulting in clinical manifestations. The former is defined as single nucleotide variants and the latter is defined as mutations, where a clinical phenotype has been associated with the specific variant (Karki *et al.*, 2015).

In many developing countries, the absence of routine investigations necessary to make a basic diagnosis, such as histochemistry or biochemical assessment of biopsies, remains a consistent problem. This is glaringly evident when searching for listed professional diagnostic services for rare diseases using web resources such as Orphanet (www.orphanet.net). It can therefore be concluded that the majority of patients born with a MD in these countries remain undiagnosed.

South Africa is one of the few developing countries with the capacity to provide basic clinical and diagnostic services at selected centres for patients with IMDs, including MDs. With approximately 140 paediatric and adult neurologists in the country (equalling 2.5 per million of the population), the clinical expertise to recognize neuromuscular diseases is widely available. However, the major limiting factor — with MDs a good example — is the lack of specialised laboratory diagnostic services and information on aetiology in the population. Although the National Health Laboratory Services (NHLS) provide a number of tests, such as muscle histochemistry and screening for selected mutations (only recently full mtDNA sequencing), adequate routine investigations and other supporting resources to accurately diagnose MDs in an ethnically diverse population are not available.

Since 1998, clinical assessments of neuromuscular diseases for referred patients from the Northern provinces of South Africa have been done at the *Steve Biko Academic Hospital* at the University of Pretoria, by the paediatric neurologist Prof. I. Smuts and colleagues. Along with clinical data, urine and muscle biopsy samples were also collected (where possible) from patients with suspected MDs. As a research study (Smuts *et al.*, 2010), respiratory chain (RC) enzyme analysis on these muscle biopsies has been done at the NWU. A combined clinical-biochemical approach has been followed for over two decades to identify RC deficiencies, and subsequently a diagnosis of MDs could be made in these paediatric patients. Over time, this provided the first South African MD cohort with a biochemically-defined muscle RC deficiency that allowed for further investigations into the aetiology and other characteristics of MD in this population. One such additional investigation included a non-invasive metabolomics research approach using urine samples previously collected from this cohort (Reinecke *et al.*, 2012). Using various analytical techniques, a biosignature (group of biomarkers) for MD was identified in this cohort by Smuts *et al.* (2013), which included 13 metabolites frequently associated with this disorder.

Since 2006, a renewed set of data on clinical assessments, as well as RC enzyme analysis and metabolic screening for specific biomarkers, has been collected for this paediatric cohort. However, since the limited routine genetic services offered only a small number of well-documented mutations in mostly Caucasian studies, there remained a major shortage of genetic confirmation in these and other South African patients with MDs. Considering this, an initial molecular genetics investigation on this cohort was conducted between 2009 and 2012 by Van der Walt *et al.* (2012). The aim of this study was to determine the mtDNA genetic variants in 71

patients with an associated clinical and biochemical MD profile by identifying reported pathogenic as well as novel disease-causing variants. Van der Walt *et al.* (2012) observed that the clinical and biochemical profiles for patients of African descent did not correlate with those of their non-African (Caucasian) counterparts, and concluded that the significant differences observed in these diverse populations complicated the identification of pathogenic variants. In addition, they identified the lack of population genetic data as a key problem in the assessment of variants. These findings concurred with those of another centre involved in the mtDNA investigations of MD in South African populations, indicating a relatively low mutation diagnostic yield of 6% and less than 1% for referred patients in the Western Cape and Northern provinces, respectively (Van der Westhuizen *et al.*, 2015). Considering this, and especially the lack of nDNA information in these cases, it was clear that a more extensive investigation into the South African patient population was needed to expand our understanding of MDs on a molecular level, which would ultimately have a major impact on diagnostic approaches of MDs in the South African population.

The current investigation, was mainly conducted in response to the inconclusive molecular aetiological problems previously encountered (Smuts *et al.*, 2010; Van der Walt *et al.*, 2012; Van der Westhuizen *et al.*, 2015). Additionally, a number of aspects were taken into account at the start of this investigation to select the most suitable molecular diagnostic approach for this study. These included, but were not limited to, the available resources, funding and infrastructure, ethical considerations (e.g. incidental findings), limited sample availability (e.g. absence of fibroblasts), and the diverse population found in South Africa, for which there is limited genetic information available. Considering this, a collaborative decision was made to do panel sequencing of specific targeted genes known to be involved with MDs as a first step, in addition to full mtDNA genome sequencing on an extended number of patients in the cohort, in order to expand and continue the work set forth by Smuts *et al.* (2010) and Van der Walt *et al.* (2012). Furthermore, a subset of black African patients from this cohort was furthermore selected for whole exome sequencing (WES), as this is the next step towards a more complete understanding of mitochondrial aetiology in South Africa.

1.2 RESEARCH AIM AND OBJECTIVES

Given the problems one is faced with when identifying MDs in this paediatric cohort, it is clear that the molecular (genomic) information regarding MDs in the South African population is still lacking causative evidence. Since molecular confirmation is becoming the cornerstone of identifying MD in patients, this study aims to investigate both the mtDNA and nuclear genes known to be involved with MDs in a previously-diagnosed paediatric cohort.

The following objectives were set in order to achieve this aim:

1. Design a strategy whereby all the nuclear genes associated with complexes I, II, III, and IV, as well as those associated with coenzyme Q₁₀ (CoQ₁₀), can be sequenced in a high-throughput manner using target selection technology and next-generation sequencing (NGS).
2. Identify the patients to be included in this study according to specific clinical and biochemical criteria, and isolate DNA from muscle and/or blood samples from these selected patients.
3. Generate high quality NGS data from selected nuclear genes, as well as data from whole mitochondrial genome sequencing.
4. Design and program a bioinformatics pipeline to efficiently identify novel, reported and disease-causing/pathogenic variants from NGS data.
5. Assess the pathogenicity of novel, reported and previously reported disease-causing variants identified through NGS in African and non-African patients in this cohort.
6. Follow up on a pathogenic mutation in a selected case.
7. Screen for two putative founder mutations known to be involved with mitochondrial dysfunction.

An experimental design is illustrated in Figure 1.1 with the objectives indicated where applicable. The experimental design can be divided into two sections, the research conducted since 1998 in the first section, and the current research in the second section, as described in Sections 3.3 and 3.4 respectively. Each section forms part of the bigger investigations into the aetiology of MDs in these South African paediatric patients. Chapters 3–5 provide more detail on the methods and data processing used during this investigation and Chapters 6–8 provide detail on results obtained during this investigation. A full description of the structure of this thesis is given in the following section (Section 1.3).

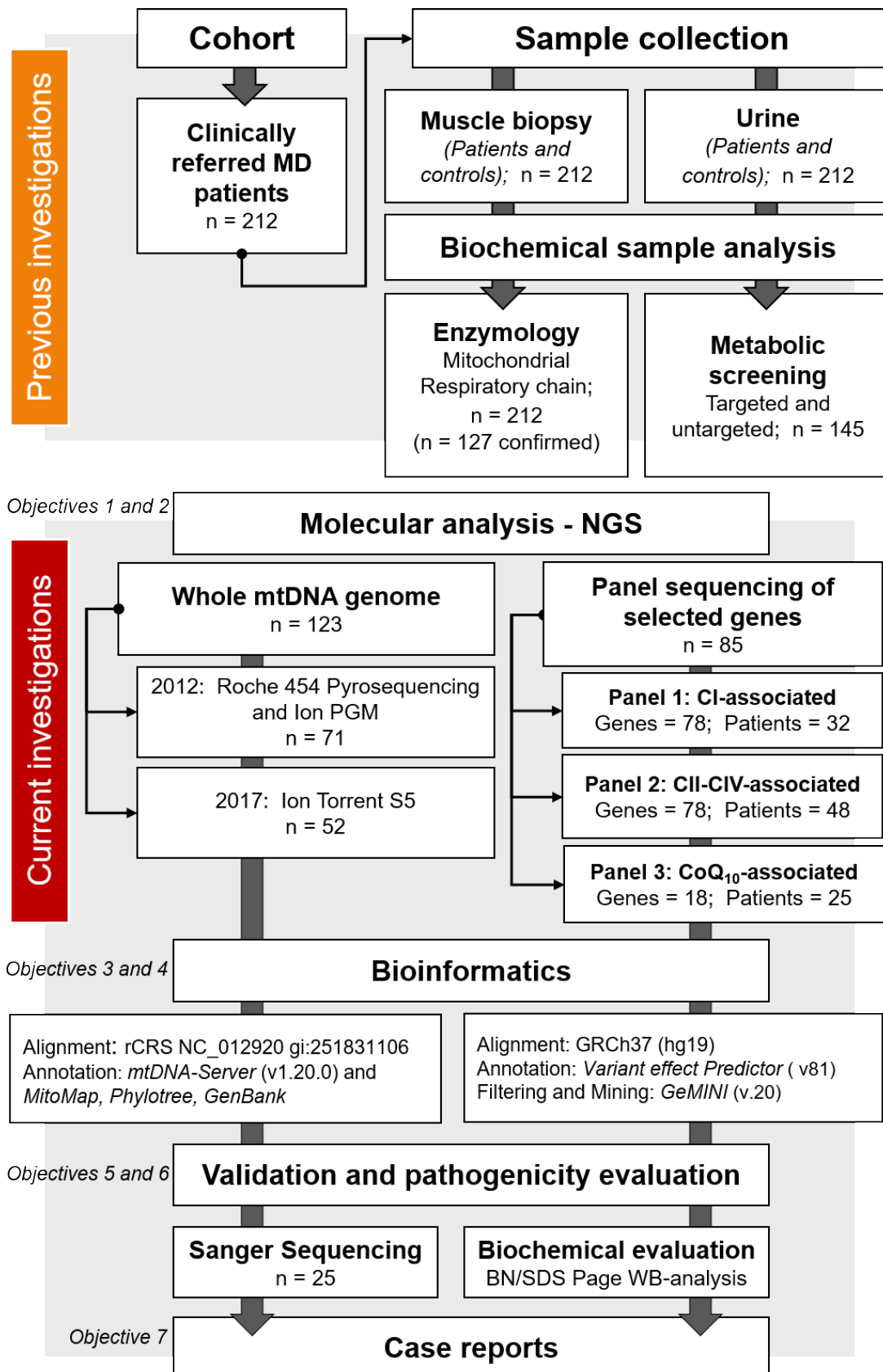


Figure 1.1: Illustration of the experimental design aimed at investigating the aetiology of MD in this understudied patient population.

1.3 STRUCTURE OF THESIS AND RESEARCH OUTPUTS

This thesis has been written in thesis format and complies with the requirements of the North-West University (NWU), South Africa, for the completion of the degree Philosophiae Doctor (Biochemistry). The thesis is presented in nine chapters and produced three scientific outputs (which are either published or currently submitted for publication — see Appendix A).

Chapter 1: Introduction

Here, the background and the motivation for this study is provided, and the research aim and objectives. Additionally, all of the primary author's outputs are listed.

Chapter 2: Literature review

This chapter consists of a relevant literature overview on energy production in mitochondria, and diseases associated with mitochondrial dysfunction. A discussion of the diagnosis of MDs in the South African population is also presented.

Chapter 3: Patient cohort and selection

Chapter 3 describes and provides relevant information on the South African paediatric cohort diagnosed with MDs. Furthermore, methods used for biochemical analyses (e.g. RC enzyme analysis) and the results so acquired are described. Using the clinical and biochemical information, the patient selection criteria for sequencing are formulated and described accordingly, thereby addressing Objectives 1 and 2.

Chapter 4: Next-generation sequencing

This chapter describes the methodology for NGS (addressing Objective 3), which was followed for whole mitochondrial genome sequencing and targeted sequencing of selected nDNA genes known to be involved with MDs. A section on WES is also given.

Chapter 5: Bioinformatics and computational tools

This chapter consists of peer-reviewed paper which has been published (Appendix A.1) in which Objective 4 is addressed, followed by a discussion of mtDNA bioinformatics used to identify disease-causing mtDNA variants.

- Schoonen, M., Seyffert, A.S., Van der Westhuizen, F.H. & Smuts, I. (2019). A bioinformatics pipeline for rare genetic diseases in South African patients. *South*

Chapter 6: Pathogenicity evaluation of detected mtDNA and nDNA variants

This chapter consist in part of an accepted paper (Appendix A.2) in which Objective 5 is addressed. Additional information on the overarching results obtained from NGS is provided.

- Schoonen, M., Smuts, I., Louw, R., Elson, J.L., Van Dyk, E., Jonck, L., Rodenburg, R.J.T. & Van der Westhuizen, F.H. (2019). Panel-based nuclear and mitochondrial next-generation sequencing outcomes of an ethnically diverse paediatric patient cohort with mitochondrial disease. *Journal for Molecular Diagnostics*, [DOI: <https://doi.org/10.1016/j.jmoldx.2019.02.002>] — with permission

Chapter 7: Reporting of a novel ETFDH mutation — A case report

This chapter consist of a peer-reviewed paper (Appendix A.3) in which Objective 6 is addressed.

- Van der Westhuizen F.H., Smuts, I., Honey, E., Louw, R., Schoonen, M., Jonck, L. & Dercksen, M. (2017). A novel mutation in ETFDH manifesting as severe neonatal-onset multiple acyl-CoA dehydrogenase deficiency. *Journal of the Neurological Sciences*, 384: 121–125 [DOI: [10.1016/j.jns.2017.11.012](https://doi.org/10.1016/j.jns.2017.11.012)] — with permission

Chapter 8: Mutation screening of two putative population variants in the GCDH and MPV17 genes

This chapter reports on the findings on two population-specific mutations in the genes *GCDH* and *MPV17* involved with glutaric acidemia type I and mtDNA depletion syndrome respectively. Objective 7 is addressed in this chapter.

Chapter 10: Final conclusions and future recommendations

This chapter includes a summary and evaluation of the data presented in this thesis. Final concluding remarks and recommendations for future studies are made.

1.4 AUTHOR CONTRIBUTIONS

Next-generation sequencing in Chapter 4: M. Schoonen was involved with the study design and was responsible for sample preparation for sequencing, which includes library building and template preparation. E. van Dyk was responsible for NGS on the Ion Personal Genome Machine (PGM™).

Manuscript submitted for publication in Chapter 5. This bioinformatics paper was submitted to a peer-reviewed journal. A.S. Seyffert was involved in computational script writing, manuscript writing, and revision. F.H. van der Westhuizen was involved in manuscript writing, revision and supervision. I. Smuts was involved in manuscript writing and supervision. M. Schoonen was responsible for study design, data processing, computational script writing and manuscript writing.

Manuscript submitted for publication in Chapter 7. I. Smuts was involved in study design, intellectual input, manuscript writing and supervision. R. Louw was involved with intellectual input and manuscript revision. J.L. Elson was involved in manuscript revision. E. van Dyk was responsible for NGS on the Ion PGM™. F.H. van der Westhuizen was involved with study design, manuscript writing and supervision. M. Schoonen was involved with study design, preparation of samples for sequencing including library building, template building and enrichment, data processing and bioinformatics, including data filtering, mining and interpretation thereof, and manuscript writing.

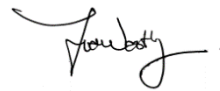
Peer-reviewed paper in Chapter 8. F.H. van der Westhuizen was involved in study design and manuscript writing. I. Smuts was involved in study design and revising. E. Honey was involved in manuscript writing and revision. R. Louw was involved in manuscript writing and revision. M. Schoonen was responsible for sample preparation for NGS which included library preparation, template building and enrichment, data processing as well as bioinformatics, which included data filtering, mining and interpretation, and manuscript writing and revision.

All authors involved signed the declaration on this page:

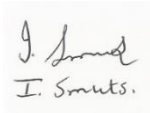
As a co-author/researcher, I hereby approve and give consent that the above-mentioned articles and data can be used for the Ph.D. of M. Schoonen. I declare that my role in the study, as indicated above, is a representation of my actual contribution.



M. Schoonen



F.H. van der Westhuizen



I. Smuts.

I. Smuts



A.S. Seyffert



E. Honey



E. van DYK.

E. van Dyk



J.L. Elson



L. Peters (née Jonck)



M. Dercksen



R. Louw



R. Rodenburg

CHAPTER 2:

LITERATURE REVIEW

2.1 MITOCHONDRIAL BIOLOGY

2.1.1 Structure and function of mitochondria

Mitochondria are dynamic intracellular organelles present in the cytosol of almost all eukaryotic cells. This organelle consists of a double membrane system, referred to as the mitochondrial outer and inner membranes (MOM and MIM) respectively, which defines the mitochondrial matrix (MM), as seen in Figure 2.1 (Sjöstrand, 2018). The MM contains mtDNA, enzymes, ribosomes and organic molecules, and typically proceeds via the integration of various metabolic pathways such as gluconeogenesis, ketogenesis, urea cycle, β -oxidation of fatty acids and the tricarboxylic acid (TCA) cycle (Dolezal *et al.*, 2006; Duchen, 2004; Hopper *et al.*, 2006). The two membranes forming the matrix each have distinctive functionalities. The MOM is selectively permeable due to the integral porins within its phospholipid bilayer, which allows metabolites to access the MM. These porins or voltage-dependant anion channels allow the free movement of molecules and substrates [(nutrients, ions, adenosine triphosphate (ATP), adenosine diphosphate (ADP)] of up to 10 kDa across the MOM (Blachly-Dyson & Forte, 2001; Wohlrab, 2009). The MIM is, however, semi-impermeable with a larger membrane surface due to its characteristic *cristae* feature (Mannella, 2006). It consists of highly specialised respiratory proteins and associated transport proteins [(ubiquinone (Q) and a soluble cytochrome *c* (cyt *c*)] responsible for energy production (Letts *et al.*, 2016). These membranes are separated by a rather small intermembrane space (IMS) with several components and small molecular weight proteins which fulfil important functions and processes (Craven *et al.*, 2017; Herrmann *et al.*, 2007). It also provides a space where hydrogen ions which are pumped from the MM during electron transporting build up and create a proton gradient that is necessary for ATP production via the OXPHOS system — the main function of mitochondria.

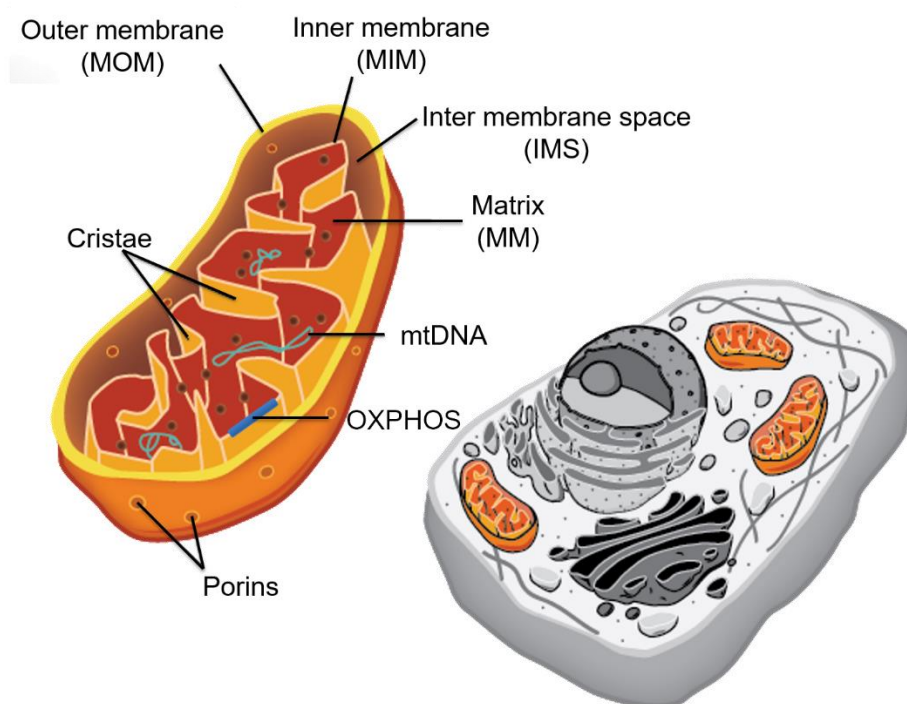


Figure 2.1: Simplified illustration of the mitochondria. These are ubiquitously found within almost all eukaryotic cells. They contain their own DNA, primarily found within the mitochondrial matrix and produce energy in the form of ATP via the oxidative phosphorylation system. Abbreviations: MOM: mitochondrial outer membrane; MIM: mitochondrial inner membrane; IMS: intermembrane space; MM: mitochondrial matrix; mtDNA: mitochondrial DNA; OXPHOS: oxidative phosphorylation.

Other functions and roles in cellular processes in which the mitochondria take part include (but are not limited) to biosynthesis of haem and iron-sulphur [Fe-S] clusters, amino acid and lipid metabolism, and calcium homeostasis (Friedman & Nunnari, 2014; Nunnari & Suomalainen, 2012). A simplified illustration of the mitochondria is given in Figure 2.1. As energy production via OXPHOS is considered to be the most important role of the mitochondria, this process is described in more detail in Section 2.1.2.

2.1.2 Electron transport chain and oxidative phosphorylation

The electron transport chain (ETC) is situated on the MIM, where four RC protein complexes; complex I [NADH:ubiquinone oxidoreductase; EC 1.6.5.3 (CI)], complex II [succinate-ubiquinone oxidoreductase; EC 1.3.5.1 (CII)], complex III [ubiquinol cytochrome c reductase; EC 1.10.2.2 (CIII)], and complex IV [cytochrome c oxidase; EC 1.9.3.1 (CIV)] as well as two electron carriers, ubiquinone and cyt c, function together with complex V (ATP synthase; EC 3.6.3.14 [CV]) to ultimately produce ATP. These five complexes and two electron carriers are collectively known as the OXPHOS system, as illustrated in Figure 2.2 (Gorman *et al.*, 2016; Hatefi, 1985; Koopman *et al.*, 2016). Under physiological conditions, electrons from the hydrogens on NADH and FADH₂

are transferred to oxygen starting at CI. Three complexes CI, CIII, and CIV convert the energy from the electron transfers into hydrogen ions and expel 4, 4 and 2 hydrogen ions across the MIM, respectively. The complexes are dynamic and can form supercomplexes and undergo other conformational changes to aid in proton pumping across the membrane. An electron proton gradient is created in the IMS as a result of the increased number of hydrogen ions. This proton motive force is utilised by ATP synthase to produce 38 ATP molecules per cycle, thereby concluding OXPHOS (Letts *et al.*, 2016). The RC protein complexes, CI to CV, including the transporter proteins, are described in more detail in Sections 2.1.3.1–2.1.3.6.

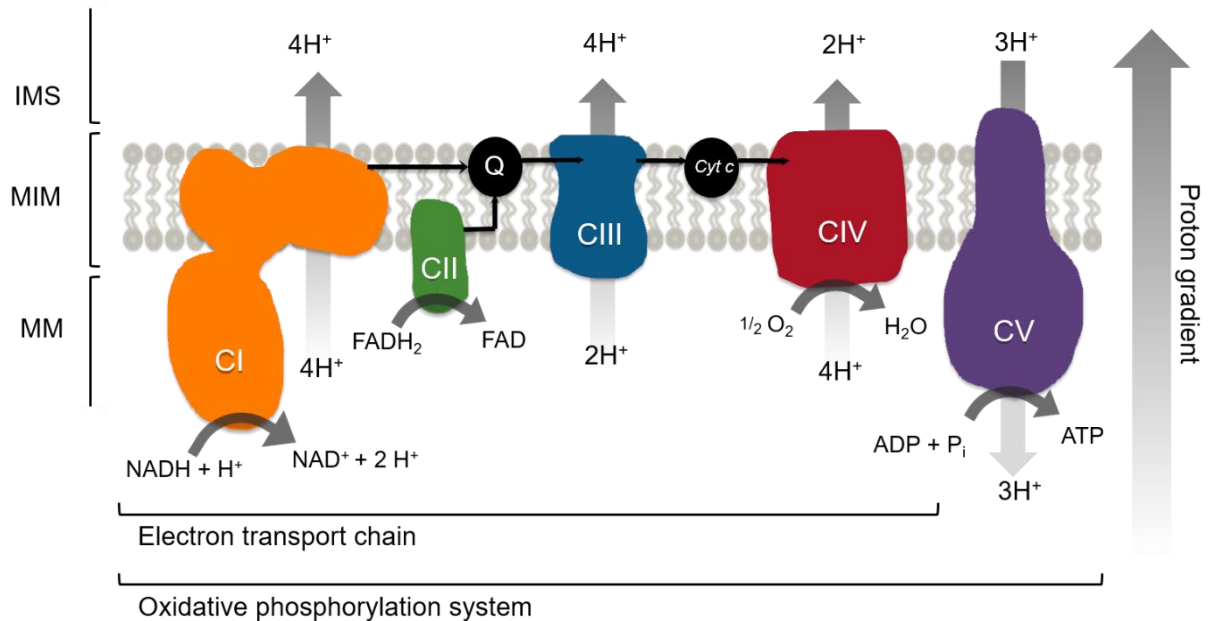


Figure 2.2: Oxidative phosphorylation in the mitochondria. Electrons are carried from CI to CIV (electron transport chain) in a step-wise manner. Ubiquinone carries electrons from CI and CII to CIII and cytochrome c carries the electrons from CIII to CIV. Hydrogen ions are expelled across the mitochondrial inner membrane into the intermembrane space by complexes I, III and IV. This proton motive force is a key component of ATP production via ATP synthase. Abbreviations: NADH: nicotinamide adenine dinucleotide — reduced; FAD: flavin adenine dinucleotide; CI: complex I; CII: complex II; CIII: complex III; CIV: complex IV; CV: complex IV; Q: Ubiquinone; Cyt c: cytochrome c; H: hydrogen ions; IMS: inter membrane space; MIM: mitochondrial inner membrane; MM: mitochondrial matrix. Adapted from Nijtmans *et al.* (2004).

2.1.3 Oxidative phosphorylation dysfunction

Oxidative phosphorylation is critical to energy production. Organs throughout the body have different energy requirements. The brain, skeletal muscle and heart have a high energy demand and therefore require larger amounts of ATP than other organs. Energy supply can easily be compromised if problems arise within OXPHOS which can lead to a mitochondrial disorder or disease. The MD term collectively refers to a group of heterogeneous disorders. Several factors contribute to MD, of which dysfunctional OXPHOS is most common (Craven *et al.*, 2017). The

RC complexes forming OXPHOS can be deficient in an isolated manner or combined, i.e. only one complex or more than one RC complex is defective (Gorman *et al.*, 2016). Mutations or alterations in either the mitochondrial genome or in several nuclear genes encoding RC protein complexes are largely responsible for MDs. Aspects of OXPHOS dysfunction are explained by means of the respiratory complexes and mitochondrial genetics throughout Sections 2.1.3, 2.2, and 2.3.

2.1.3.1 Complex I

Complex I, also known as NADH:ubiquinone oxidoreductase, is the largest enzyme complex in the RC (~1 MDa). This complex has characteristic L-shape with a hydrophobic membrane arm situated on the MIM, and a peripheral arm protruding into the MM (Wirth *et al.*, 2016) (see Figure 2.2). It can furthermore be subdivided into an N-module, where NADH binding and oxidation takes place, the Q-module, which transfers electrons to ubiquinone, and the P-module, responsible for proton pumping. The N- and Q-modules lie within the matrix arm, while the P-module forms part of the membrane arm (Wirth *et al.*, 2016). Moreover, CI comprises 44 subunits, of which 14 are conserved and necessary for its catalytic core functions (Kmita & Zickermann, 2013). Seven of these 14 conserved NADH-dehydrogenase (ND) core subunits are encoded by mtDNA and synthesised on mitochondrial ribosomes (Christian & Spremulli, 2012), while the remaining seven ND core subunits are encoded by nDNA (listed in Table 2.1). Of the remaining 30 accessory subunits, 16 are incorporated into the peripheral arm and 14 into the membrane arm. These 30 subunits are thought to be involved with the stability, regulation and assembly of CI (Andrews *et al.*, 2013). Additionally, at least 13 assembly factors involved with the assembly of CI have been identified, and are also listed in Table 2.1.

Table 2.1: Classification of the complex I structural and assembly factors and their respective inheritance patterns.

CI subunits	No. of subunits	Genes involved in coding of subunits ^a		Reference
		mtDNA	nDNA	
Structural and accessory subunits	44	<i>MTND1–6, MTND4L</i>	<i>NDUFS1–S8, NDUFV1–3, NDUFAB1, NDUFA1–3, NDUFA5–13, NDUFB1, NDUFB2, NDUFB4–11, NDUFC1, NDUFC2</i>	Alston <i>et al.</i> (2017); Giachin <i>et al.</i> (2016); Sánchez-Caballero <i>et al.</i> (2016)
Assembly factors	13		<i>NDUFAF2–8, ACAD9, ECSIT, FOXRED1, TMEM126B, TIMMDC1, NUBPL</i>	

^amtDNA: mitochondrial DNA; nDNA: nuclear DNA, MTND1–6; D4L: mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 1–6 and D4L; NDUFS1–8: NADH:ubiquinone oxidoreductase core subunit S1–8; NDUFV1–3: NADH:ubiquinone oxidoreductase core subunit V1–2; NDUFAB1: NADH:ubiquinone oxidoreductase subunit AB1;

NDUFA1–3 and A5–13: NADH:ubiquinone oxidoreductase subunit A1–3 and A5–13; NDUFB1–2 and B4–11: NADH:ubiquinone oxidoreductase subunit B1–2 and B4–11; NDUFC1–2: NADH:ubiquinone oxidoreductase subunit C1–2; NDUF2–8: NADH:ubiquinone oxidoreductase complex assembly factor 2–8; ACAD9: acyl-CoA dehydrogenase family member 9; ECSIT: ECSIT signalling integrator; FOXRED1: FAD-dependent oxidoreductase domain containing 1; TMEM126B: transmembrane protein 126B; TIMMDC1: translocase of inner mitochondrial membrane domain containing 1. The HUGO gene nomenclature was used here and elsewhere in this thesis for naming genes, as can be found at <https://www.genenames.org/>.

Complex I dysfunction is more frequently encountered in MDs than in any other complex defects (Smeitink *et al.*, 2001) where isolated CI is most prevalent among paediatric cases (Alston *et al.*, 2017). Causes for CI dysfunction can be ascribed to catalytic problems or instability in the CI assembly process, usually as a result of an underlying genetic mutation in either the mtDNA or nDNA-encoded structural subunits, or in assembly factors (Distelmaier *et al.*, 2009; Fassone & Rahman, 2012). The most prevalent clinical phenotypes associated with CI dysfunction include Leigh syndrome (LS), Leber's hereditary optic neuropathy (LHON), mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS), fatal infantile lactic acidosis (FILA), and neonatal cardiomyopathy with lactic acidosis (Chen *et al.*, 2017; Schrier & Falk, 2011; Smeitink *et al.*, 2001; Wong, 2013).

2.1.3.2 Complex II

Succinate-ubiquinone oxidoreductase, or CII, has four subunits, i.e. two hydrophilic proteins comprising [Fe-S] (Ip) and flavoprotein, and two hydrophobic transmembrane proteins comprising cytochrome *b* (cyt *b*), collectively forming this tetrameric structure. The subunits are entirely encoded by nDNA (see Table 2.2). CII is not only involved in RC, but also the TCA cycle, providing a link between these metabolic processes. The two main functions of CII are oxidation of succinate to fumarate during the TCA cycle, and the transfer of electrons through its [Fe-S] cluster directly to ubiquinone, reducing it to ubiquinol with the help of prosthetic groups (Bezawork-Geleta *et al.*, 2017; Sun *et al.*, 2005). The CII subunits are assembled by succinate dehydrogenase assembly factors 1 and 2 (SDHAF1 and SDHAF2, respectively), and two chaperone-like factors, namely succinate dehydrogenase assembly factors 3 and 4 (SDHAF3 and SDHAF4, respectively) (Bezawork-Geleta *et al.*, 2017).

Table 2.2: Classification of the complex II structural and assembly factors and their respective inheritance patterns.

CII subunits	No. of subunits	Genes involved in coding of subunits ^a		Reference
		mtDNA	nDNA	
Structural subunits	4		SDHA, SDHB, SDHC, SDHD	Alston et al. (2017)
Assembly factors	4		SDHAF1, SDHAF2, SDHAF3, SDHAF4	

^amtDNA: mitochondrial DNA; nDNA: nuclear DNA; SDHA: succinate dehydrogenase complex flavoprotein subunit A; SDHB: succinate dehydrogenase complex iron sulfur subunit B; SDHC: succinate dehydrogenase complex subunit C; SDHD: succinate dehydrogenase complex subunit D; SDHAF1–4: succinate dehydrogenase assembly factor 1–4.

Underlying recessive inherited genetic mutations in any of the four subunits and/or assembly factors cause CII deficiency. However, these are the least frequent in patients with MDs, with a prevalence of 2–8% for isolated CII deficiency (Alston et al., 2015; Alston et al., 2017). Mutations in the SDHA subunit and SDHAF1 are often the leading cause for isolated or combined CII deficiency in all MD cases. CII deficiency caused by homozygous mutations presents with congenital manifestations, mostly affecting the central nervous system (CNS) and heart, whereas compound heterozygous mutations cause cancerous manifestations, such as pheochromocytoma and paraganglioma (Jain-Ghai et al., 2013; Timmers et al., 2009; Wong, 2013).

2.1.3.3 Complex III

Complex III, also known as ubiquinol cytochrome c reductase, is a transmembrane protein comprised of 11 subunits, of which 10 are encoded by nDNA while only one subunit, cyt *b*, is encoded by mtDNA (see Table 2.3). These 10 subunits constitute three respiratory subunits (i.e. two haem groups, MTCYB and CYC1, the Rieske [Fe-S] protein, UQCRFS1), two core proteins (UQCRC1 and UQCRC2), and six accessory subunits (UQCRH, UQCRB, UQCRQ, UQCR10, UQCR11 and a cleavage product of Rieske). Together, all of these subunits form the systemic dimer (CIII₂), which contains the functionally active form of the enzyme (Fernández-Vizarra & Zeviani, 2015). The main function of CIII is transferring electrons from the active form of CoQ₁₀ (also referred to as ubiquinol) to cyt *b*, followed by electron transfer to cyt *c*, resulting in proton pumping across the membrane from the MM to the IMS. The latter is achieved through the Q-cycle mechanism, which refers to a series of reactions, such as oxidation and reduction of CoQ₁₀, between ubiquinol and ubiquinone, causing the proton-motive force that helps with movement of protons across the MIM (Crofts et al., 2008).

Table 2.3: Classification of the complex III structural and assembly factors and their respective inheritance patterns.

CIII subunits	No. of subunits	Genes involved in coding of subunits ^a		Reference
		mtDNA	nDNA	
Structural and accessory subunits	10	<i>MTCYB</i>	<i>CYC1, UQCRB, UQCRC1, UQCRC2, UQCRFS1, UQCRH, UQCRQ, UQCR10, UQCR11</i>	Alston et al. (2017); Fernández-Vizarra and Zeviani (2015)
Assembly factors	6		<i>UQCC1, UQCC2, UQCC3, BCS1L, TTC19, LYRM7</i>	

^amtDNA: mitochondrial DNA; nDNA: nuclear DNA, *MTCYB*: mitochondrially encoded cytochrome b; *UQCRB*: ubiquinol-cytochrome c reductase binding protein; *CYC1*: cytochrome C1; *UQCRB*: ubiquinol-cytochrome c reductase binding protein; *UQCRC1–2*: ubiquinol-cytochrome c reductase core protein 1–2; *UQCRFS1*: ubiquinol-cytochrome c reductase; Rieske iron-sulfur polypeptide 1; *UQCRH*: ubiquinol-cytochrome c reductase hinge protein; *UQCRQ*: ubiquinol-cytochrome c reductase complex III subunit VII; *UQCR10*: ubiquinol-cytochrome c reductase, complex III subunit X; *UQCR11*: ubiquinol-cytochrome c reductase, complex III subunit XI; *UQCC1–3*: ubiquinol-cytochrome c reductase complex assembly factor 1–3; *BCS1L*: BCS1 homolog; ubiquinol-cytochrome c reductase complex chaperone; *TTC19*: tetratricopeptide repeat domain 19; *LYRM7*: LYR motif containing 7.

Mutations encoded by both mtDNA and nDNA are well-known causes of mitochondrial CIII deficiencies, with mutations in the *MTCYB* and *BCS1L* genes accounting for more than 50% of these in all MD cases. Pathogenic mutations have also been found in the genes *UQCRB*, *UQCRQ*, *UQCR2* and *CYC1*. Clinical phenotypes include exercise intolerance, cardiomyopathy, encephalomyopathy, muscle weakness and in some cases lactic acidosis, liver dysfunction, renal tubulopathy, dystonia and ataxia (Alston et al., 2017; Fernández-Vizarra & Zeviani, 2015; Wong, 2013).

2.1.3.4 Complex IV

The final complex and electron acceptor of the ETC is cytochrome c oxidase (COX), or CIV, situated on the MIM. Similar to CI and CIII, the 14 CIV-associated subunits are encoded by both mtDNA and nDNA. The three largest hydrophobic subunits, cytochrome I–III (COXI–3), are translated from mtDNA within the organelle and form the functional catalytic core. The remaining 10 nuclear-encoded subunits and isoforms form around the core as a protective shield with the help of several ancillary proteins and assembly factors (listed in Table 2.4) (Rak et al., 2016; Soto et al., 2012). The assembled COX core (often as a dimer) contains the two haem groups, two copper centres (Cu_A and Cu_B), cytochrome a, and cytochrome a₃, all needed for catalysis of electron transferring from reduced cyt c to molecular oxygen (Rak et al., 2016; Tsukihara et al., 1995). Protons are transferred across the MIM which contributes to the proton motive force needed for ATP production. The 14th subunit of COX, namely NDUFA4, was detected and

confirmed by Balsa *et al.* (2012). It was found that a deletion of NDUFA4 did not affect CI, as it was thought to be CI subunit, but rather affected CIV. Furthermore, it was confirmed that this subunit plays a role in CIV function and biogenesis.

Table 2.4: Classification of the complex IV structural and assembly factors and their respective inheritance patterns.

CIV subunits	No. of subunits	Genes involved in coding of subunits ^a		Reference
		mtDNA	nDNA	
Structural and accessory subunits	17	<i>MTCO1, MTCO2, MTCO3</i>	<i>COX4I1, COX4I2, COX5A, COX5B, COX6A, COX6B, COX6C, COX7A, COX7B, COX7C, COX8, NDUFA4, LRPPRC, TACO1</i>	Alston <i>et al.</i> (2017)
Assembly factors	22		<i>OXA1, COA1, COA3, COA5–7, COX10, COX11, COX14, COX15–16, COX17, COX18–COX20, COX23, FASTKD2, PET100, SURF1, SCO1 SCO2</i>	

^amtDNA: mitochondrial DNA; nDNA: nuclear DNA; MTCO1–3: mitochondrially encoded cytochrome c oxidase I–III; COX4–8C: cytochrome c oxidase subunit 4I1, 4I2, 5A, 5B, 6A, 6B, 6C, 7A, 7B, 7C, 8; NDUFA4: NADH:ubiquinone oxidoreductase subunit A4; LRPPRC: leucine-rich PPR motif-containing protein; TACO1: translational activator of cytochrome c oxidase 1; OXA1: oxidase assembly 1-like protein; COA1: cytochrome c oxidase assembly factor 1 homolog; COA3 and 5–7: cytochrome c oxidase assembly factor 3 and 5–7; COX10: haem A:farnesyltransferase cytochrome c oxidase assembly factor; COX11: cytochrome c oxidase copper chaperone; COX14: cytochrome c oxidase assembly factor 14; COX15–16: cytochrome c oxidase assembly homolog 15 and 16; COX17: cytochrome c oxidase copper chaperone; COX18–20: cytochrome c oxidase assembly factor 18–20; COX23: cytochrome c oxidase assembly factor 23; FASTKD2: FAST kinase domains 2; SURF1: cytochrome c oxidase assembly factor; SCO1–2: cytochrome c oxidase assembly protein 1–2.

This enzyme complex is considered to be an OXPHOS regulation site, and thus mutations in any of the subunits are known to cause severe early-onset mitochondrial CIV deficiency, with mostly neuromuscular phenotypes (Baertling *et al.*, 2017). Other manifestations include LS, encephalomyopathy, cardiomyopathy and leukodystrophy (Tanigawa *et al.*, 2012). Mutations are predominantly found in nDNA assembly factors, especially in early assembly intermediates such as *SURF1*, *SCO1*, *SCO2* and the core mtDNA subunits.

2.1.3.5 Complex V

ATP-synthase (F_1F_0 -ATP synthase), or CV, has two functionally important domains, F_1 and F_0 , and consists of 20 subunits, of which two are encoded by mtDNA and three are nuclear-encoded assembly factors (listed in Table 2.5) that work together to phosphorylate ADP, and an inorganic phosphate (P_i) to produce ATP by a membrane rotary motor (Alston *et al.*, 2017). The electrochemical proton gradient created by the ETC is the driving force for this machinery. A hydrogen ion, currently in the IMS, enters CV, while another hydrogen ion leaves the complex

and enters the MM. Internal rotation of the γ subunit takes place with each new hydrogen entering from the IMS. Three hydrogen ions are necessary to produce one ATP molecule from ADP and Pi (Boyer, 1997; Stock *et al.*, 1999).

Table 2.5: Classification of the complex V structural and assembly factors and their respective inheritance patterns.

CV subunits	No. of subunits	Genes involved in coding of subunits ^a		Reference
		mtDNA	nDNA	
Structural and accessory subunits	20	MTATP6, MTATP8	ATP5A1, ATP5F1B, ATP5F1C, ATP5F1D, ATP5F1E, ATP5MC1, ATP5MC2, ATP5MC3, ATP5MD, ATP5ME, ATP5MF, ATP5MG, ATP5MPL, ATP5PB, ATP5PD, ATP5PF, ATP5PO, ATP5IF1	Alston <i>et al.</i> (2017)
Assembly factors	3		ATPAF1, ATPAF2, TMEM70	

^amtDNA: mitochondrial DNA; nDNA: nuclear DNA; MT-ATP6/8: mitochondrially encoded ATP synthase membrane subunit 6/8; ATP5F1A/B/C/D/E: ATP synthase F1 subunit alpha/beta/gamma/delta/epsilon; ATP5MC1/2/3: ATP synthase membrane subunit c locus 1/2/3; ATP5MD: ATP synthase membrane subunit DAPIT; ATP5ME/F/G: ATP synthase membrane subunit e/f/g; ATP5MPL: ATP synthase membrane subunit 6.8PL; ATP5PB/D: ATP synthase peripheral stalk-membrane subunit b/d; ATP5PF: ATP synthase peripheral stalk subunit F6; ATP5PO: ATP synthase peripheral stalk subunit OSCP; ATP5IF1: ATP synthase inhibitory factor subunit 1.

Complex V mutations can occur in both mtDNA and nDNA encoded subunits. Clinical phenotypes associated with mtDNA mutations manifest at a high heteroplasmic threshold of ~80–90% (Hejzlarová *et al.*, 2014), while manifestations due to nDNA mutations are less common (Alston *et al.*, 2017). These manifestations can include neuropathy, ataxia, retinitis pigmentosa (NARP), and neonatal encephalopathy. Symptoms include lactic acidosis, cardiomyopathy, encephalopathy, and dysmorphic features as reported in some cases (De Meirleir *et al.*, 2004; Hejzlarová *et al.*, 2014; Isohanni *et al.*, 2018; Jonckheere *et al.*, 2013; Jonckheere *et al.*, 2012).

2.1.3.6 Electron transfer proteins

Two electron transfer proteins are embedded in the phospholipid bilayer of the MIM, transferring electrons between RC enzyme complexes without directly taking part in OXPHOS. The two electron carriers are a lipophilic quinone referred to as CoQ₁₀, and a hydrophobic haem protein, cyt c, situated on the external surface of the MIM (Enriquez & Lenaz, 2014). CoQ₁₀ consists of three functional domains to form one structural domain. The first domain is a flavin adenine dinucleotide (FAD), buried deep within the functional domain, the second domain is an iron

cluster, and the third is an enzyme-specific binding site. CoQ₁₀ carries electrons from CI and CII to CIII, whereas cyt c carries electrons from CIII to CIV (see Figure 2.2).

Mutations in either of the electron transfer proteins can result in altered fatty acid and amino acid oxidation. Pathogenic mutations in CoQ₁₀ biosynthesis genes, namely *COQ2*, *PDSS1*, *PDSS2* and *ADCK3* have been associated with patients presenting with primary CoQ₁₀ deficiency with multisystemic phenotypes, including encephalomyopathy (Ogasahara *et al.*, 1989), childhood-onset cerebral ataxia and atrophy (Artuch *et al.*, 2006), infantile isolated myopathy (Horvath *et al.*, 2006), and nephrotic syndrome (Salviati *et al.*, 2005). Secondary CoQ₁₀ deficiency arise from mutations in genes indirectly involved with CoQ₁₀ biosynthesis, namely *APT*X, *BRAF* and *ETFDH* (Quinzii *et al.*, 2008). Multiple acyl-CoA dehydrogenase deficiency (MADD) is associated with the latter, resulting in the build-up of fats and proteins. Symptoms can occur at neonatal stage and include metabolic acidosis, cardiomyopathy, and liver disease (Liang *et al.*, 2009; Van der Westhuizen *et al.*, 2017).

2.2 MITOCHONDRIAL GENETICS

2.2.1 Mitochondrial genome

The maternally [or as recently shown, paternally inherited as well (Chinnery & Hudson, 2013; Luo *et al.*, 2018)] inherited mitochondrial genome is a circular, double-stranded molecule, consisting of a network of highly compact multiple copies of DNA within the MM (Giles *et al.*, 1980). The complete sequence of 16 569 bp forms the inner light strand (L strand, non-coding region), an outer heavy strand (H strand, coding region), and a 1.1 kbp D-loop (replication, initiation, and termination site) (Anderson *et al.*, 1981). It is comprised of 37 coding genes, of which 24 are involved with the translation of mtDNA and the remaining 13 are directly committed to subunits of RC. From these 24, there are two ribosomal RNAs (rRNA) and 22 transfer RNAs (tRNA) (listed in Table 2.6 and Figure 2.3).

Table 2.6: Classification of mtDNA genes.

	mtDNA genes ^a	Reference
Structural	CI: MT-ND1–6, MT-ND4L, CIII: MT-CYB, CIV: MT-CO1–3, CV: MT-ATP6, MT-ATP8	Alston <i>et al.</i> (2017); Anderson <i>et al.</i> (1981)
rRNA	RNR1 (12S rRNA), RNR2 (16S rRNA)	
tRNA	MT-NC2, MT-TA, MT-TC, MT-TD, MT-TE, MT-TF, MT-TG, MT-TH, MT-TI, MT-TK, MT-TL1, MT-TL2, MT-TM, MT-TN, MT-TP, MT-TQ, MT-TR, MT-TS1, MT-TS2, MT-TT, MT-TV, MT-TW, MT-TY	

^aMTND1–6, D4L: mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 1–6 and D4L; MTCYB: mitochondrially encoded cytochrome b; MTCO1–3: mitochondrially encoded cytochrome c oxidase I–III; MT-ATP6: mitochondrially encoded ATP synthase membrane subunit 6/88; RNR1: RNA, ribosomal 45S cluster 1; RNR2: RNA, ribosomal 45S cluster 2; MT-T: mitochondrially encoded tRNA; A: alanine; C: cysteine; D: aspartic acid; E: glutamic acid; F: phenylalanine; G: glycine; H: histidine; I: isoleucine; K: lysine; L1: leucine 1 (UUA/G); L2: leucine 2 (CUN); M: methionine; N: asparagine; P: proline; Q: glutamine; R: arginine; S1: serine 1 (UCN); S2: serine 2 (AGU/C); T: threonine; V: valine; W: tryptophan; Y: tyrosine.

2.2.2 Mitochondrial DNA mutations of mitochondrial disease

The mitochondrial genome is relatively small compared to the human genome. However, mutations in the genome can cause lethal disorders, with a mutation rate 10–17-times higher than in nDNA (Tuppen *et al.*, 2010). This is mainly due to a lack of mtDNA repair systems, as well as reactive oxygen species (ROS) production by RC complexes in close proximity to mtDNA. The majority of mutations are polymorphisms, but more than 250 pathogenic mtDNA mutations have been reported. This includes point mutations as well as large scale deletions, rearrangement of nucleotides, and hotspots in mitochondrial tRNA genes and CI encoded genes (Carelli & La Morgia, 2018; Tuppen *et al.*, 2010). The large amount of mtDNA copies within a mitochondrion can either be mutant, wild-type, or a combination of both. This former is known as homoplasmy, where all the mtDNA are the same, and the latter as heteroplasmy, where both wild-type and mutant mtDNA are present. A threshold is found with heteroplasmic variants since a cell can tolerate a certain amount of defective mtDNA (Herbert & Turnbull, 2018). However, if the threshold load is exceeded, metabolic dysfunction can occur (Alston *et al.*, 2017). The threshold limits for heteroplasmic mutations are often only exceeded in adulthood, with a prevalence of 2.9–9.2 per 100 000 adults (Gorman *et al.*, 2015; Schaefer *et al.*, 2008).

Mutations cause a wide variety of clinical manifestations and can present at any age (Gorman *et al.*, 2015). The most common mutations are listed in Table 2.7. An extended description of known pathogenic mutations for mtDNA can be found on the MitoMAP database (Brandon *et al.*, 2005).

Table 2.7: Genotype-phenotype correlations of well described human mtDNA pathogenic mutations.

Mutation	Locus	Associated phenotypes and manifestations^a	Reference^b
m.1555A>G	MT-RNR1	Deafness	Usami <i>et al.</i> (1997)
m.1624C>T	MT-TV	LS	McFarland <i>et al.</i> (2002)
m.3243A>G	MT-TL	MELAS, MIDD, CPEO	Manouvrier <i>et al.</i> (1995); Moraes <i>et al.</i> (1993)
m.3271T>C	MT-TL	MELAS	Hayashi <i>et al.</i> (1993)
m.3460G>A	ND1	LHON	Howell <i>et al.</i> (1991)
m.4300A>G	MT-TI	Cardiomyopathy	Casali <i>et al.</i> (1995)
m.5545C>T	MT-TW	Multisystem disorder, RC deficiency	Sacconi <i>et al.</i> (2008)
m.7445A>G	MT-TS1	Deafness/hearing loss	Reid <i>et al.</i> (1994)
m.7472CinsC	MT-TS1	Myopathy	Jaksch <i>et al.</i> (1998)
m.8344A>G	MT-TK	MERRF	Shoffner <i>et al.</i> (1990)
m.8356T>C	MT-TK	MERRF	Masucci <i>et al.</i> (1995)
m.8993T>G/C	ATP6	MILS, NARP	De Vries <i>et al.</i> (1993); Harding <i>et al.</i> (1992)
m.9176T>G/C	ATP6	LS, FBSN	Carrozzo <i>et al.</i> (2000); Thyagarajan <i>et al.</i> (1995)
m.10158T>C	ND3	LS/Leigh-like syndrome	Crimi <i>et al.</i> (2004)
m.10191T>C	ND3	LS/Leigh-like syndrome	Taylor <i>et al.</i> (2001)
m.10197G>A	ND3	LS/Leigh-like syndrome	Sarzi <i>et al.</i> (2007)
m.11777C>A	ND4	Encephalopathy, LS	Deschauer <i>et al.</i> (2003); Komaki <i>et al.</i> (2003)
m.11778G>A	ND4	LHON	Wallace <i>et al.</i> (1988)
m.13513G>A	ND5	MELAS, LS	Santorelli <i>et al.</i> (1997b)
m.14484T>C	ND6	LHON	Brown <i>et al.</i> (1992)
m.14709T>C	MT-TE	Myopathy, weakness, and diabetes	Damore <i>et al.</i> (1999); Mancuso <i>et al.</i> (2005); McFarland <i>et al.</i> (2004b)

^a LS: Leigh syndrome; MELAS: mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes; MIDD: maternally inherited diabetes and deafness; CPEO: chronic progressive external ophthalmoplegia; LHON: Leber's hereditary optic neuropathy; MERRF: myoclonic epilepsy with ragged red fibres; MILS: maternally inherited Leigh syndrome; NARP: neuropathy, ataxia, and retinitis pigmentosa; FBSN: familial bilateral striatal necrosis.

^b References to journals articles and databases where the referred variants were previously reported on.

2.2.3 Nuclear DNA mutations of mitochondrial disease

NGS methods have advanced significantly over recent years. This progress has consequently increased the identification of pathogenic mitochondrial-encoded nuclear gene mutations. There

are roughly 1 300 proteins in the mitoproteome, of which ~150 nuclear encoded genes are directly involved with OXPHOS (Craven *et al.*, 2017). The nuclear genes are synthesised and translated onto cytosolic ribosomes prior to translocation through the mitochondrial membrane into the MM, via translocase of MIM and MOM complexes respectively (Truscott *et al.*, 2003). Mutations have been reported in more than 250 genes. Furthermore, DiMauro and Schon (2003) stated that mutations in nDNA genes are more common and severe in paediatric cases, compared to adults who present more with mtDNA mutations. It is important to note that in most cases these patients present with asymptomatic features, making it difficult to identify mitochondrial involvement.

A description of known pathogenic mutations for nDNA is listed on the MitoMAP database (Lott *et al.*, 2013). Clinical manifestations often associated with RC structural and non-structural subunits for each complex are listed in Table 2.8.

Table 2.8: Clinical manifestations for structural and non-structural RC subunits encoded by nDNA.

Gene	OMIM ^a	Clinical manifestation ^b	References ^c
Complex I			
NDUFS1	157655	LS	Martín <i>et al.</i> (2005)
NDUFS2	602985	Encephalopathy, cardiomyopathy	Loeffen <i>et al.</i> (2001)
NDUFS3	603846	LS	Benit <i>et al.</i> (2004)
NDUFS4	602694	LS	Petruzzella <i>et al.</i> (2001)
NDUFS6	603848	FILA	Spiegel <i>et al.</i> (2009)
NDUFS7	601825	LS	Smeitink and van den Heuvel (1999)
NDUFS8	602141	LS	Loeffen <i>et al.</i> (1998)
NDUFB3	603839	FILA	Calvo <i>et al.</i> (2012)
NDUFB9	601445	Hypotonia, lactic acidosis	Haack <i>et al.</i> (2012)
NDUFB10	603843	Lactic acidosis, cardiomyopathy	Friederich <i>et al.</i> (2015)
NDUFB11	300403	Intrauterine growth restriction, lactic acidosis	Van Rahden <i>et al.</i> (2015)
NDUFV1	161015	LS	Laugel <i>et al.</i> (2007)
NDUFV2	600532	Cardiomyopathy, Hypotonia, Encephalopathy	Smeitink and van den Heuvel (1999)
NDUFA1	300078	LS, progressive neurodegenerative disorder	Loeffen <i>et al.</i> (1998)
NDUFA2	602137	LS	Hoefs <i>et al.</i> (2008)
NDUFA9	603834	LS	Van den Bosch <i>et al.</i> (2012)
NDUFA10	603835	LS	Hoefs <i>et al.</i> (2011)
NDUFA11	612638	FILA, encephalocardiomyopathy	Berger <i>et al.</i> (2008)
NDUFAF12	609653	LS	Ostergaard <i>et al.</i> (2011)
NDUFA13	609435	Encephalopathy, optic atrophy	(Angebault <i>et al.</i> , 2015)

NDUFAF1	606934	Cardioencephalomyopathy	Dunning <i>et al.</i> (2007)
NDUFAF2	609653	Early onset progressive encephalopathy	Ogilvie <i>et al.</i> (2005)
NDUFAF3	612911	Neonatal encephalopathy	Saada <i>et al.</i> (2009)
NDUFAF4	611776	Infantile encephalopathy	Saada <i>et al.</i> (2008)
NDUFAF5	612360	LS	Gerards <i>et al.</i> (2010); Sugiana <i>et al.</i> (2008)
NDUFAF6	612392	LS	Bianciardi <i>et al.</i> (2016)
NUBPL	613621	Encephalomyopathy	Calvo <i>et al.</i> (2010)
FOXRED1	613622	LS	Calvo <i>et al.</i> (2010)
ACAD9	611103	Hypertrophic cardiopathy encephalopathy	Haack <i>et al.</i> (2010)
Complex II			
SDHA	600857	LS	Bourgeron <i>et al.</i> (1995)
SDHB	185470	Phaeochromocytoma and paraganglioma	Astuti <i>et al.</i> (2001)
SDHC	602413	Autosomal dominant paraganglioma type III	Niemann and Müller (2000)
SDHD	602690	Autosomal dominant paraganglioma type I, pheochromocytoma	Baysal <i>et al.</i> (2000)
SDHAF1	612848	Leukoencephalopathy	Ghezzi <i>et al.</i> (2009)
SDHAF2	613019	Autosomal dominant paraganglioma type II	Hao <i>et al.</i> (2009)
Complex III			
UQCRB	191330	Hypoglycaemia, lactic acidosis	Haut <i>et al.</i> (2003)
UQCRQ	612080	Severe neurological phenotype	Barel <i>et al.</i> (2008)
BCS1L	603647	encephalopathy, hepatic failure and tubulopathy, LS, Bjornstad syndrome, GRACILE syndrome	De Lonlay <i>et al.</i> (2001); Hinson <i>et al.</i> (2007); Visapää <i>et al.</i> (2002)
UQCC2	614461	Lactic acidosis and renal tubular dysfunction	Tucker <i>et al.</i> (2013)
UQCC3	616097	Lactic acidosis, hypoglycaemia, hypotonia	Wanschers <i>et al.</i> (2014)
Complex IV			
COX6A1	602072	Charcot-Marie-Tooth disease	Tamiya <i>et al.</i> (2014)
COX61B	124089	Encephalomyopathy	Massa <i>et al.</i> (2008)
COX7B	300885	Microphthalmia with linear skin lesions	Indrieri <i>et al.</i> (2012)
COX8A	25600	LS	Hallmann <i>et al.</i> (2015)
SURF1	185620	LS	Zhu <i>et al.</i> (1998)
SCO1	603644	Neonatal hepatic failure and encephalopathy	Sacconi <i>et al.</i> (2003)
SCO2	604272	Neonatal cardioencephalomyopathy	Papadopoulou <i>et al.</i> (1999); Sacconi <i>et al.</i> (2003)

COX10	602125	Neonatal tubulopathy and encephalopathy, LS, cardiomyopathy	Antonicka <i>et al.</i> (2003a)
COX14	614478	Neonatal lactic acidosis	Weraarpachai <i>et al.</i> (2012)
COX15	603646	Early-onset hypertrophic cardiomyopathy, LS	Antonicka <i>et al.</i> (2003b); Oquendo <i>et al.</i> (2004)
COX20	61498	Ataxia, muscle hypotonia	Szklarczyk <i>et al.</i> (2012)
COA3	614775	Neuropathy, exercise intolerance	Ostergaard <i>et al.</i> (2015)
COA5	613920	Cardioencephalomyopathy	Huigsloot <i>et al.</i> (2011)
COA6	614772	Cardioencephalomyopathy	Baertling <i>et al.</i> (2015)
LRPPRC	607544	French-Canadian LS	Debray <i>et al.</i> (2011)
FASTKD2	612322	Cncephalomyopathy	Ghezzi <i>et al.</i> (2008)
TACO1	612958	LS	Weraarpachai <i>et al.</i> (2009)
Complex V			
ATP5E	606153	Lactic acidosis, mental retardation, peripheral neuropathy	Mayr <i>et al.</i> (2010)
ATP5A1	164360	Neonatal encephalopathy	Hejzlarová <i>et al.</i> (2014)
ATP8A2	605870	Cerebellar ataxia, mental retardation	Onat <i>et al.</i> (2013)
ATPAF2	608918	Early-onset encephalopathy, lactic acidosis	De Meirleir <i>et al.</i> (2004)
TMEM70	612418	Neonatal encephalopathy, cardiomyopathy	Čížková <i>et al.</i> (2008)
CoQ₁₀			
COQ2	609825	Encephalomyopathy, nephropathy	Quinzii <i>et al.</i> (2006)
COQ4	612898	Encephalomyopathy, mental retardation	Salviati <i>et al.</i> (2012)
COQ5	616359	Encephalomyopathy, cerebellar ataxia	Malicdan <i>et al.</i> (2018)
COQ6	614647	Nephrotic syndrome, deafness	Heeringa <i>et al.</i> (2011)
COQ7	601683	Hypotonia, cardiac hypertrophy	Freyer <i>et al.</i> (2015)
COQ9	612837	Neonatal lactic acidosis, seizures, cardiomyopathy	Duncan <i>et al.</i> (2009)
ATPX	606350	Cerebellar ataxia, oculomotor apraxia	Quinzii <i>et al.</i> (2005)
PDSS1	607429	Deafness, valvulopathy, mental retardation	Mollet <i>et al.</i> (2007)
PDSS2	610564	LS, nephrotic syndrome	López <i>et al.</i> (2006)
CABC1	606980	Cerebellar ataxia, lactic acidosis	Mollet <i>et al.</i> (2008)
ETFDH	231675	MADD	Beard <i>et al.</i> (1993)

^a Numbers for specific phenotypes as listed on the OMIM database.

^b OMIM: Online Mendelian Inheritance in Man; LS: Leigh syndrome; FILA: fatal infantile lactic acidosis; GRACILE: growth retardation, amino aciduria, cholestasis, iron overload, lactic acidosis, and early death; MADD: multiple acyl-CoA dehydrogenation deficiency.

^c References to journals articles and databases where the referred genes were previously reported on.

2.3 DIAGNOSIS OF MITOCHONDRIAL DISEASE

Successful diagnosis of MD, primary or secondary, which is considered to be a daunting task, often relies on a multidisciplinary approach. The different techniques include clinical assessments, biochemical analyses, histopathological testing, and genetic screening. In many cases where the MD phenotypical features are clear and precise (well-studied genotype-phenotype correlations), it is sufficient to perform only one or two of these techniques to make a diagnosis (Taylor & Thorburn, 2018). However, for the majority of cases with a broad spectrum of features, [genotype-phenotype correlations are unclear or inconsistent from the expected well-studied correlations (Ham *et al.*, 2018)] these techniques are used in unison to make clear and precise diagnoses.

2.3.1 Clinical Diagnosis

Childhood-onset MD phenotypes are considered more severe when compared to adult-onset MD, and is furthermore diagnostically challenging since MDs have extremely heterogeneous features (DiMauro and Schon (2003). Clinical examination and assessment for MD should include full neurological evaluations, as both children and adults frequently present with multisystem involvements which include central nervous system involvements (encephalopathy, seizures, and stroke-like episodes) and neuromuscular features (myopathy and muscle weakness, sensory ataxia, exercise intolerance, fatigue, myalgia, and rhabdomyolysis) (Keogh *et al.*, 2018). Clinical investigations characterise the scope of multisystem involvement and several diagnostic techniques can be used such as electromyography and nerve conduction studies, audiological investigations, cardiac investigations, and brain magnetic resonance imaging to name a few.

2.3.2 Biochemical diagnosis

Biochemical phenotypes play an important part in MD diagnosis. A range of functional analysis and techniques are used to investigate patients with suspected MD for evidence of an RC deficiency (isolated or combined). Biochemical analysis using a muscle biopsy to evaluate the functional state of the RC through enzyme analysis of the respective RC complexes is considered a fundamental diagnostic component (Rodenburg, 2011). Unfortunately, enzyme analysis using frozen muscle biopsy will only reveal reliable results in CI-CIV and CoQ₁₀ (Taylor & Thorburn, 2018). Defects found throughout the mitochondria may lead to a number of elevated metabolites such as lactic acid, and also reduced metabolites such as pyruvate. Metabolite analysis using

urine and blood sample is usually performed prior to enzyme analysis, as it is often sufficient to reveal the primary metabolic pathway and location of disease (Rodenburg, 2011).

2.3.3 Molecular diagnosis

Next-generation sequencing has begun to be the preferred diagnostic technique. It is less invasive and is particularly effective in identifying disease-causing variants and genes in heterogeneous diseases such as MD. There are currently a variety of options available such as whole genome sequencing (WGS), whole exome sequencing (WES), and targeted gene panel sequencing. A brief description on each of these approaches is given in Section 4.1. Mitochondrial DNA mutations can be detected through these approaches, however, with limitations. For example, WGS has potential interference by nDNA sequences and WES gives incomplete coverage of mtDNA (Taylor & Thorburn, 2018). Whole mitochondrial genome sequencing is thus performed to eliminate the limitations from WGS and WES. A conclusive diagnosis requires proof of the existence of a disease-causing variant (mutation) associated with a MD phenotype in the patient, and ideally with information of genetic segregation within the family.

2.3.4 Classic “biopsy first” approach

Diagnosis of primary MD is often a daunting task. There are currently two approaches to follow when diagnosing MD. The first follows a “biopsy first” approach where a muscle or skin biopsy is collected from patients with clinical features suggestive of an MD. Biochemical analysis is performed on the muscle sample and includes enzyme analysis of the RC, and histopathological testing. Consequently, the patients are diagnosed with either an isolated or combined RC deficiency. Based on these evaluations, candidate genes known to cause MD are selected for sequencing. The genes can be encoded by either mtDNA or nDNA. Variants of interest detected through NGS are confirmed through Sanger sequencing, followed by stratified functional analysis to validate the pathogenicity of the detected variant (Rodenburg, 2011; Taylor *et al.*, 2004; Wolf & Smeitink, 2002; Wortmann *et al.*, 2017).

2.3.5 New “genetics first” approach

This classical diagnostic approach has been, and still remains, effective, but recent advances in NGS have shifted this diagnostic paradigm to a “genetics first” era (Wortmann *et al.*, 2017). NGS techniques, namely panel sequencing, WES, WGS, and whole mitochondrial genome sequencing, have become widely available, especially in developed countries for identification of

known disease-causing variants (Craven *et al.*, 2017; Wortmann *et al.*, 2015; Wortmann *et al.*, 2017). If novel/rare variants are identified in a patient with a clear mitochondrial phenotype, functional analyses are performed on fibroblast for the patient. Furthermore, parents and family of the patient are also genetically tested to confirm segregation of variants to aid in a more conclusive MD diagnosis. This is, however, not the case for many patients, as MDs are often asymptomatic or present with an atypical clinical profile. For these patients, genetic as well as functional tests are performed in parallel and it is thus often necessary to obtain a muscle, liver, or skin biopsy on which these functional tests are performed. However, obtaining muscle biopsies are highly invasive (Rodenburg, 2011; Taylor *et al.*, 2004). Wortmann *et al.* (2017) suggest the following guidelines for when to perform tissue biopsies: 1) the patient presents with a rapidly progressing disease (for example LS or MELAS), 2) the patient is in an unstable clinical condition (episodes of neonatal lactic acidosis), and 3) the patient presents with single organ involvement (isolated myopathy). Following a “genetics first” approach can prevent (to some extent) paediatric patients from undergoing a muscle or liver biopsy. A flowchart of the two diagnostic approaches, as suggested by Wortmann *et al.* (2017) is illustrated in Figure 2.4. With a “genetics first” approach, patients are often diagnosed based only on genetic results. However, this advanced approach relies on established and extensive capacity which most often does not exist in developing countries.

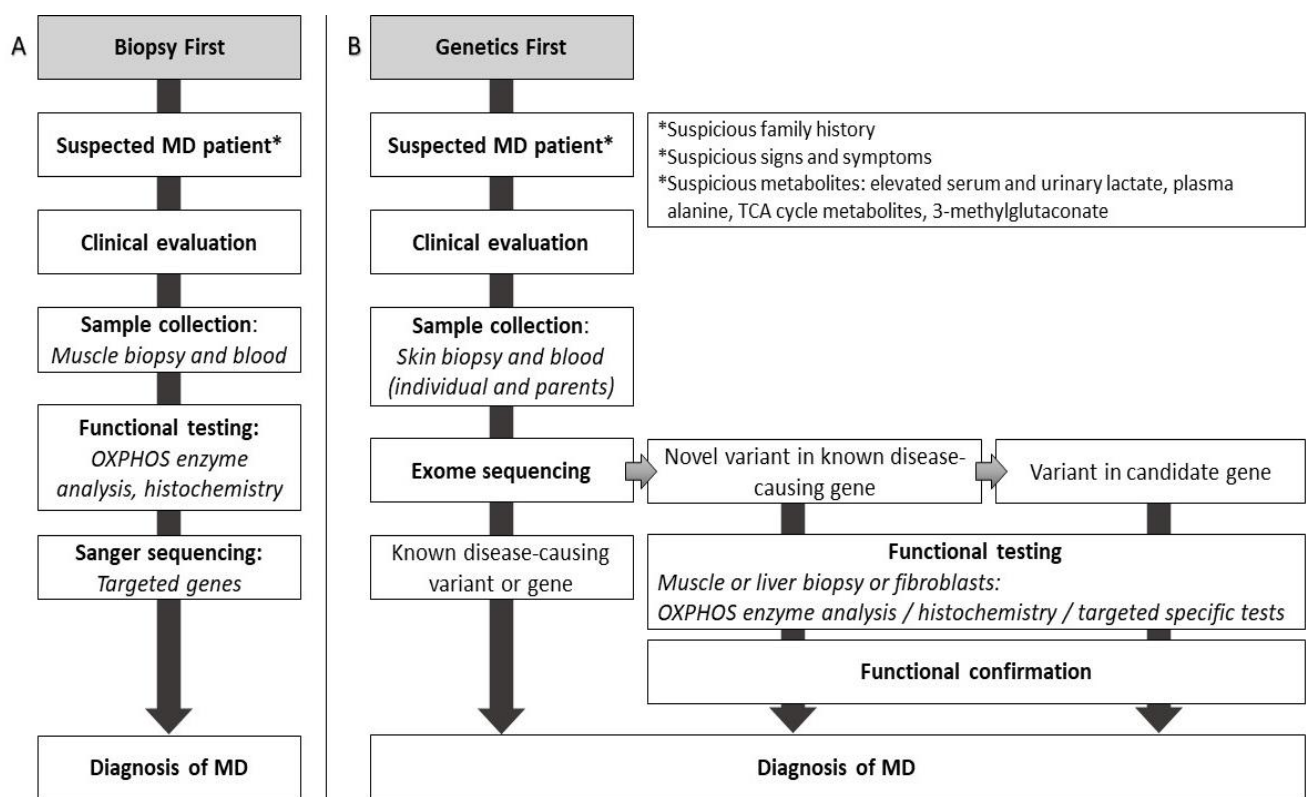


Figure 2.4: Flowchart of diagnostic procedures followed for patient with suspected MD. A: “Biopsy first” approach that was followed until recently. B: “Genetics first” approach that is being followed more recently. Abbreviations: MD: mitochondrial disease; OXPHOS: oxidative phosphorylation; TCA: tricarboxylic acid cycle. Adapted from Wortmann *et al.* (2017) with permission.

2.3.6 South African diagnostic approach

Diagnosis of MDs in developing countries such as South Africa is especially challenging due to a number of factors. Firstly, the population is highly heterogeneous, and the prevalence of MDs is still largely unknown (Meldau *et al.*, 2016). Secondly, information regarding the severity of MDs in the black African population groups is lacking when compared to developed countries (Van der Westhuizen *et al.*, 2015). Thirdly, infrastructure to perform routine diagnostic procedures, such as genetic testing and the interpretation of sequencing data is in short supply mainly due to the limited number of diagnostic centres (Smuts *et al.*, 2010; Van der Westhuizen *et al.*, 2015) (See Section 3.2.2 and Figure 3.3). With clinical expertise and training available, albeit localised at selected medical centres such as the Universities of Pretoria and Cape Town, there are only two laboratories involved with genetic diagnosis of MDs in SA, namely the IMD, NHLS in Cape Town (which focuses on genetic testing) and the Mitochondria Research Laboratory in the *Centre for Human Metabolomics* (CHM) at the NWU in Potchefstroom (which is performing biochemical, metabolic and selected genetic investigations) (Meldau *et al.*, 2016; Van der Westhuizen *et al.*, 2015). The current diagnostic research approaches which are being implemented in South Africa are illustrated in Figure 2.5.

Due to the challenges of limited information, infrastructure and diagnostic centres, a “biopsy first” approach is followed for research purposes. As with most inherited metabolic diseases, metabolic analyses in blood and urine are also done at either the NHLS laboratory or the Potchefstroom Laboratory for Inborn Errors of Metabolism (PLIEM), NWU. Patients with suspected MDs are clinically evaluated (see Section 3.3) after which biopsy samples are collected from all the patients. Frozen biopsy samples are then used for biochemical evaluation of the kinetics of the relevant RC enzyme. There is capacity for non-routine research-orientated functional tests that include histochemistry of the biopsy sample and metabolomics on collected urine samples. The ideal diagnostic procedure is then to do sequencing on selected patients with a clinically diagnosed and biochemically confirmed MD. Variants identified in patients can then be evaluated and investigated in more detail later on. Firstly, variants of interest need to be confirmed with Sanger sequencing, after which functional and structural analyses [(blue-native polyacrylamide gel electrophoresis (BN-PAGE), sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), and Western-blot (WB)] can be performed on patient samples. This is an important step in classifying a variant as disease-causing or pathogenic. Finally, it is important to note that this approach is an ideal one and not always easily implemented in different diagnostic settings in SA. Based on the problems with identifying and diagnosing MDs in developing countries, we aimed to investigate the mostly unknown molecular profile and aetiology for MDs in this cohort of predominantly African paediatric patients using NGS technology (see Chapter 1 for the full description of the aim and objectives).

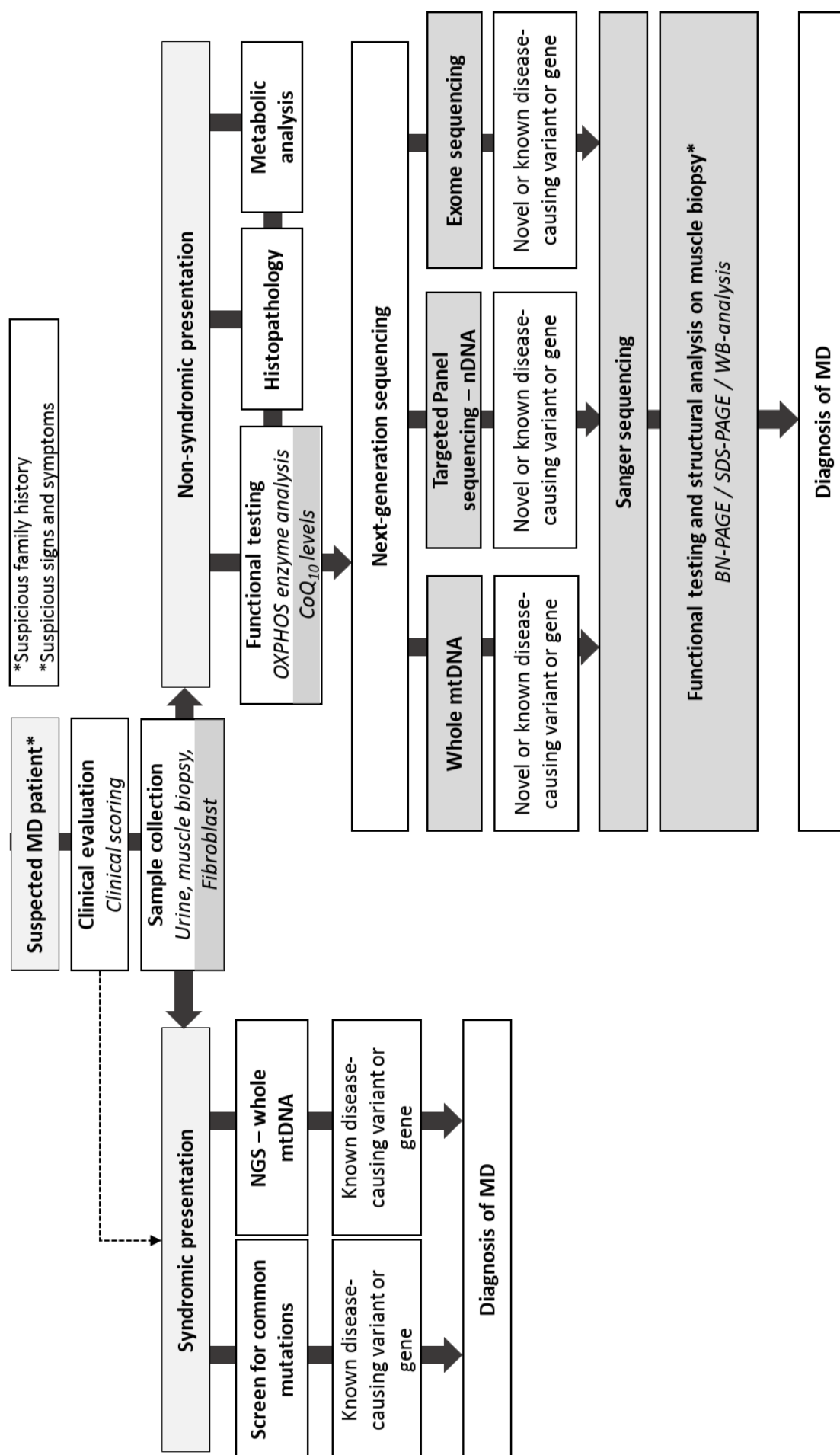


Figure 2.5: Flowchart of diagnostic procedures followed in South Africa. White blocks represent the procedures that are routinely done while the grey blocks represent atypical procedures. Abbreviations: MD: mitochondrial disease; mtDNA: mitochondrial DNA; nDNA: nuclear DNA; BN-PAGE: blue-native polyacrylamide gel electrophoresis; SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis; WB: Western-blot. Adapted from (Meldau et al., 2016; Smuts et al., 2010; Van der Westhuizen et al., 2015).

CHAPTER 3:

SOUTH AFRICAN MITOCHONDRIAL PAEDIATRIC COHORT

3.1 INTRODUCTION

Successful diagnosis of MDs relies heavily on clinical evaluation in conjunction with metabolic screening for known metabolites (although metabolic biomarkers of MD, such as lactate/pyruvate/alanine, are poorly selective or specific), biochemical evaluation of the RC enzymatic function, as well as molecular screening and/or sequencing of known MD disease-causing genes (Wortmann *et al.*, 2017). In this chapter, we describe the demographic background of the South African paediatric cohort of patients diagnosed with MD whom were the focus of this study. The information includes the clinical profiles of the patients and the biochemical outcomes, which have been documented since 2006 (Smuts *et al.*, 2010; Smuts *et al.*, 2013; Van der Walt *et al.*, 2012). Furthermore, the stratification and selection criteria for patient inclusion for NGS are described. At the time of commencement of this study, samples from the base cohort of 212 patients were available for continuing of biochemical and genetic analyses. The complete mitochondrial genomes of 71 patients were already sequenced previously (Van der Walt *et al.*, 2012). In the current study, mtDNA sequencing was done for an additional 54 patients, in conjunction with nDNA sequencing of genes known to be involved with MDs in 85 patients.

3.2 MITOCHONDRIAL DISEASE COHORT IN SOUTH AFRICA

3.2.1 South African populations in general

The census from 2011 estimates that there are approximately 52 million people living in South Africa (https://en.wikipedia.org/wiki/South_African_National_Census_of_2011, <http://www.statssa.gov.za/>). As seen in Figure 3.1A, the vast majority of the population is concentrated in

Gauteng Province (~12.3 million citizens) followed by Kwazulu-Natal with ~10.3 million citizens. The different ethnic populations can be divided into five groups: Black African (80.2%), Coloured (8.8%), White/Caucasian (8.4%), and Indian or Asian (2.5%) (https://en.wikipedia.org/wiki/Ethnic_groups_in_South_Africa). Figure 3.1B illustrates the distribution of these five major ethnic groups throughout South Africa.

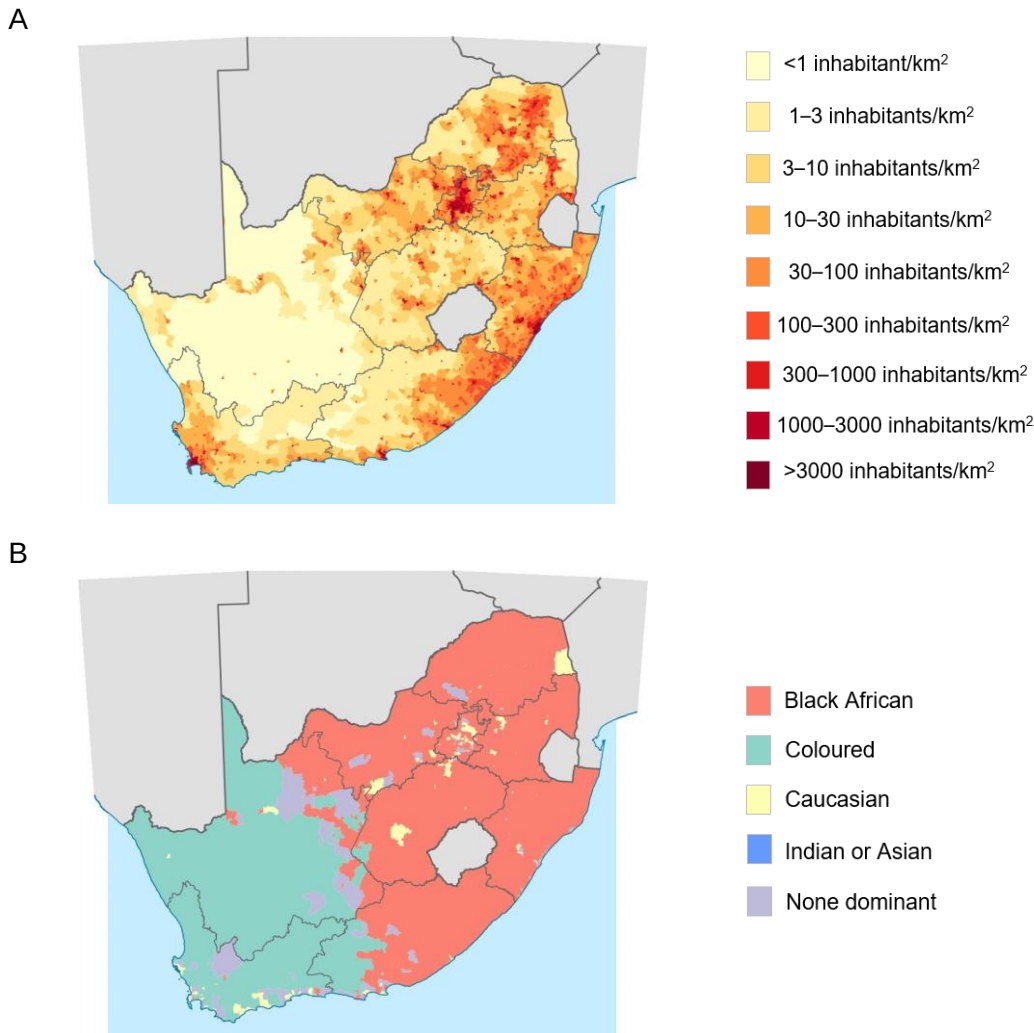


Figure 3.1: Simplified illustration of South African demographics. Panel A: Population density in South Africa. Panel B: Population ethnicity distribution in South Africa. Adapted from Statistics South Africa's Census 2011 (https://en.wikipedia.org/wiki/South_African_National_Census_of_2011).

The ethnicity distribution in South Africa is unique with a complex human history. Recent publications by Choudhury *et al.* (2017); Gurdasani *et al.* (2015), and Schlebusch *et al.* (2012) reported on the diverse genetic variations found among South Africans as a result of migrations, substantial stratification and interactions between population groups (about 2 000 years ago). As a result the Northern provinces are home to the vast majority of Black African Bantu-speakers [Zulu, Xhosa, Tsonga/Shangaan, Southern Sotho, Pedi, Tswana, and Venda (Lane *et al.*, 2002)] as seen in Figure 3.1B and Figure 3.2A, while the Cape provinces are home to the vast majority of Coloureds (Figure 3.1B and Figure 3.2B). The Coloured population group were considered

hunter-gatherers, and stem predominantly form an admixture of click-speaking Southern African Khoe-San population (32-43%), ethnic European migrants (21-28%) and Bantu-speaking Africans (20-36%) (De Wit *et al.*, 2010; Petersen *et al.*, 2013). Caucasian, or White South Africans migrated originally from Europe in the 1600s. Slave trading was consequently introduced to South Africa and as a result, Indian, South Asian, and Southeast Asian populations were brought to South Africa (Choudhury *et al.*, 2017; De Wit *et al.*, 2010). The Caucasian, Indian, and Asian populations are equally distributed across the country as seen in Figure 3.1B and Figure 3.2C (Caucasian distribution).

The migration of different ethnicity populations, the interactions between the Black Africans, the native Khoe-San, and European Caucasians, have given rise to the extreme ethnic diversity and genetic complexity found in South Africa.

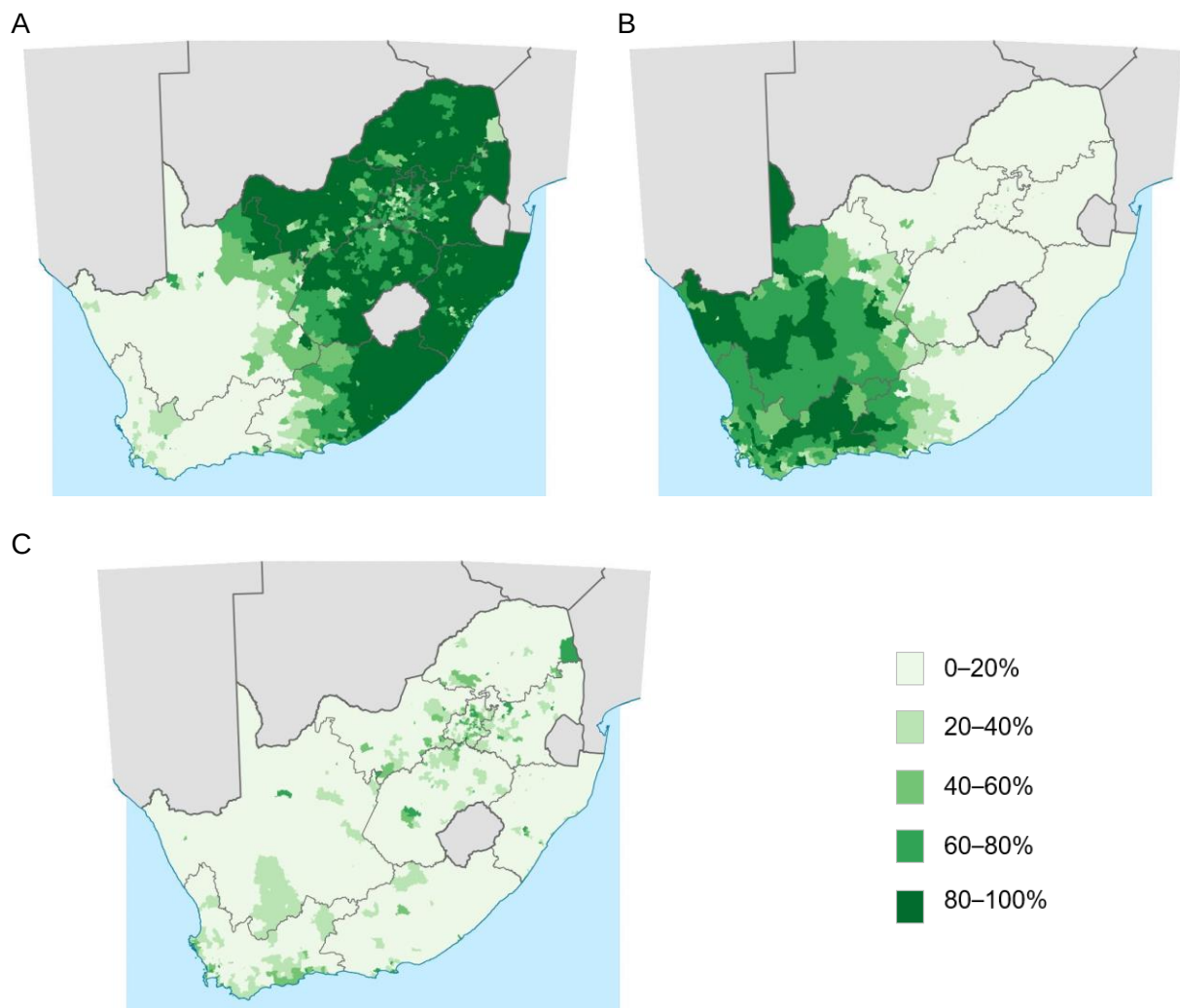


Figure 3.2: Simplified illustration of South African ethnicity distribution. Panel A: Black African population distribution. Panel B: Coloured population distribution. Panel C: Caucasian population distribution. Legend applies to all three panels where lighter colour is lower density and darker colour is higher density. Adapted from Statistics South Africa's Census 2011 (https://en.wikipedia.org/wiki/South_African_National_Census_of_2011).

3.2.2 South African diagnostic and clinical perspective

There are currently three institutions/centres in South Africa actively involved with the diagnostics of MD; 1) The Paediatric Neurology Clinic of the *Steve Biko Academic Hospital* in Pretoria, 2) *The National Health Laboratory Services* (NHLS) in Cape Town, Western Cape Province; and 3) CHM, NWU in Potchefstroom. The centres in Pretoria and Potchefstroom mostly support patients from the Northern Provinces of South Africa (Limpopo, Gauteng, and Mpumalanga), while the centre in Cape Town supports patients mostly from the Southern Provinces (Western and Eastern Cape). The different ethnically diverse population groups found in South Africa, as previously illustrated in Figure 3.2, are thus attended to at different centres. Patients from the Northern Provinces with neurological features are clinically evaluated at Centre 1 (the orange dot on Figure 3.3). As reported in this study, most of these patients have been biochemically diagnosed using muscle mitochondrial RC enzyme analysis, targeted gene selection sequencing and more recently WES at Centre 2 (the red dot on Figure 3.3). The NGS approaches done at NWU are, however, not routinely performed. The NHLS (the purple dot on Figure 3.3) use targeted mutation screening, and more recently, WES is being performed in selected cases to identify disease-causing mutations in patients from the Southern Provinces.

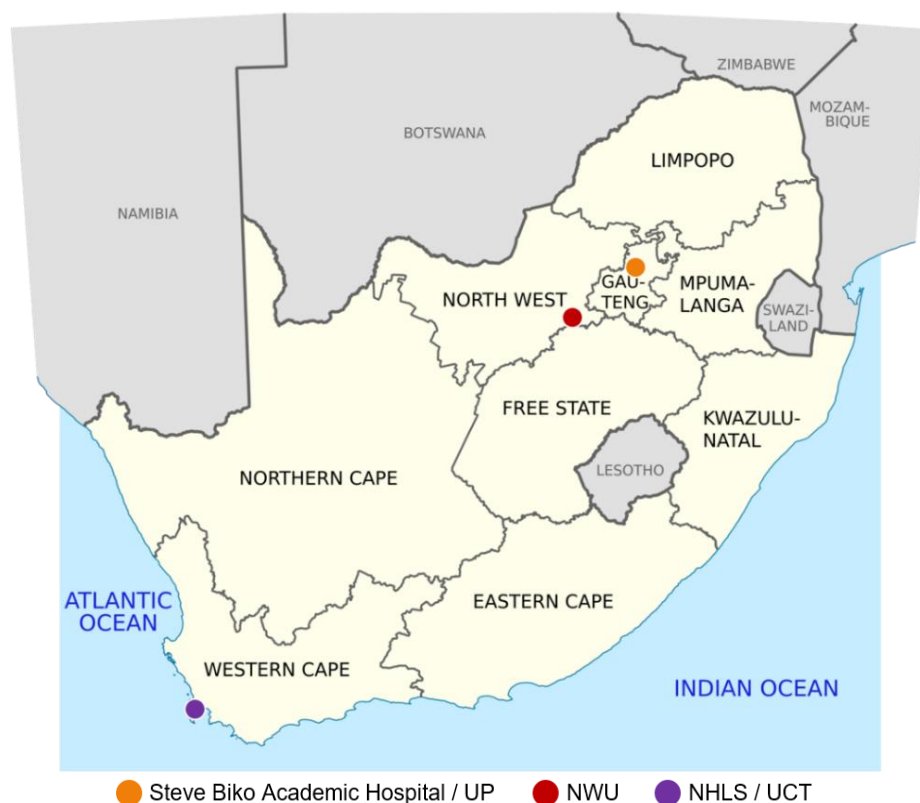


Figure 3.3: Provinces of South Africa. Orange dot: Paediatric Neurology Clinic of the Steve Biko Academic Hospital in Pretoria (Centre 1). Purple dot: National Health Laboratory Services (NHLS) in Cape Town (Centre 2) and red dot: Centre for Human Metabolomics, North-West University in Potchefstroom (Centre 3). Artwork adapted from Htonli; Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=5135568>.

From Figure 3.3 it is evident that there is a general shortage of MD diagnostic capacity for South Africans, and more specifically patients with neurological features. As only three facilities or institutions focus on MD diagnosis, the probability of failure to identify patients with MD or to misdiagnose patients with mitochondrial phenotypical features is high. Despite these limitations, more than 6 000 patients have been referred to Centre 1 for clinical evaluations. Since 2006, an MD cohort was formed and has been continuously expanded throughout the years. This cohort of MD patients consisted at the start of this study of 212 patients, all from the Northern Provinces of South Africa. As mentioned previously, patient clinical evaluations were done at Centre 1 followed by biochemical and genetic analysis at Centre 2 followed by mutation screening for selected genes at Centre 3. The cohort is described in more detail in Sections 3.3—3.4.

3.3 SAMPLE COLLECTION AND ETHICAL CONSIDERATIONS

This study was conducted according to the Declaration of Helsinki and the International Conference on Harmonisation Guidelines, and ethically approved by the Ethics Committees of the University of Pretoria (No. 91/98 and amendments) and the NWU (NWU-00170-13-A1). Informed parental or legal guardian consent and assent, when applicable, were obtained for all patients and controls (see Appendix B for an example of these forms). Muscle samples of patients with no clinical MD features, who underwent routine orthopaedic surgical procedures were obtained and included as control samples for RC enzyme analysis.

A clinical scoring system for the evaluation of paediatric patients with suspected MDs was done based on the mitochondrial disease criteria (MDC) set forth by Wolf and Smeitink (2002), and further refined by Smuts *et al.* (2010). These criteria were applied to patients mostly residing in the Northern provinces (i.e. Gauteng, Limpopo, Mpumalanga) of South Africa, who were referred to the Paediatric Neurology Clinic of the *Steve Biko Academic Hospital*, Pretoria, South Africa. These patients have been evaluated by Prof. I. Smuts and colleagues at this institution since 1998. These investigations include biochemical screening, performed at the NHLS or PLIEM, and genetic screening for disorders, as provided by the NHLS. Patients who were clinically and biochemically suspected to have a MD, with a supporting MDC score, were given the option to undergo a muscle biopsy for further evaluation and were included in this study's cohort after consent and assent were given. Since 2006, *vastus lateralis* muscle biopsies were obtained from 212 patients suspected to have MD, which were used for RC enzyme analysis, histology (where available), and structural and molecular investigations. Sections of these muscle biopsies, collected as described in Section 3.4.2, were frozen at -80°C and transported to the NWU's Mitochondria Research Laboratory at the CHM for further analysis. Supporting information available for this study included clinical profiles from examinations, baseline investigations

including lactate, pyruvate, creatine kinase and ammonia, as well as other patient-specific investigations following clinical findings.

3.3.1 Clinical overview of patient cohort with suspected mitochondrial disease

From the 212 clinically referred patients, only 127 were biochemically confirmed to have a MD (see Section 3.4). These patients were mainly of African descent ($n = 82$) and consist of roughly an equal number of males and females ($n = 64$ and $n = 63$, respectively), presenting mostly with clinical features of MD in their first year of life, including the neonatal period (66%). Muscle involvement (80%) was the main clinical feature in this cohort. It was also evident from this cohort that CNS involvement is comparatively less frequent in patients of African descent (51%), compared to non-African (mostly Caucasian) patients (72%). The non-African patients also presented more with behavioural and emotional involvement (22%), as well as gastrointestinal involvement (30%) when compared to African patients (5% and 8% respectively). A breakdown of the demographical and clinical features of the study's cohort is provided in Figure 3.4 and Figure 3.5.

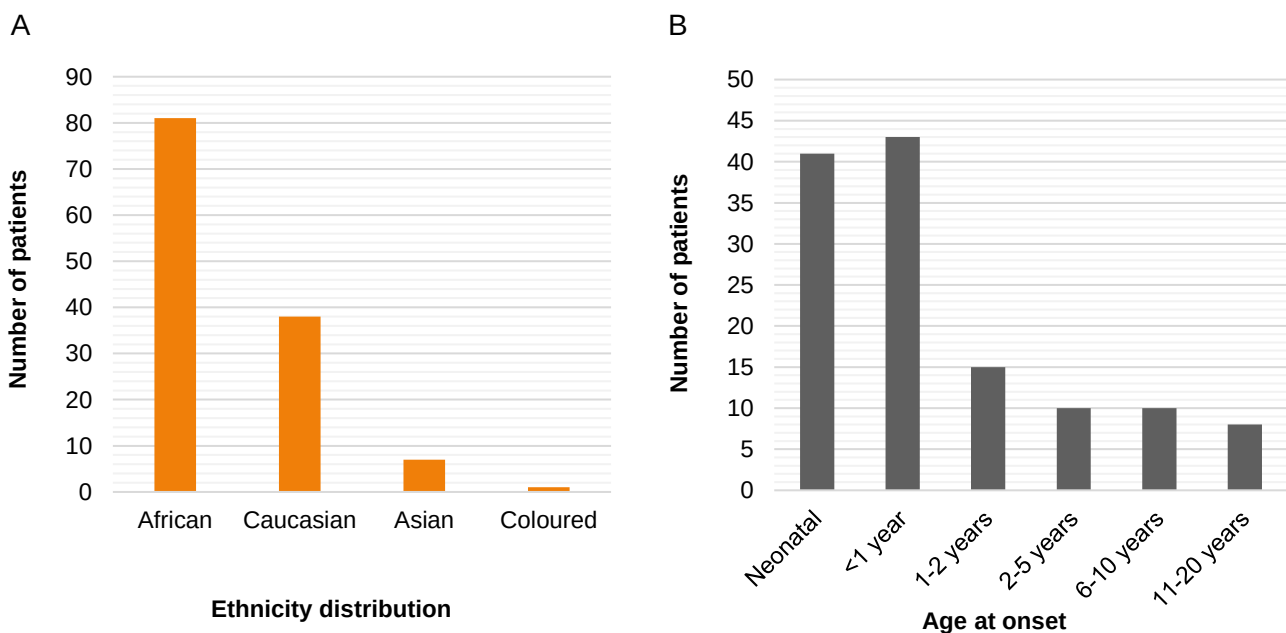


Figure 3.4: Demographical background of the paediatric patient cohort. Panel A: Distribution of ethnicity. Panel B: Age distribution at onset of symptoms. Caucasian, Asian, and Coloured patients are collectively referred to as “non-African patients” in the text.

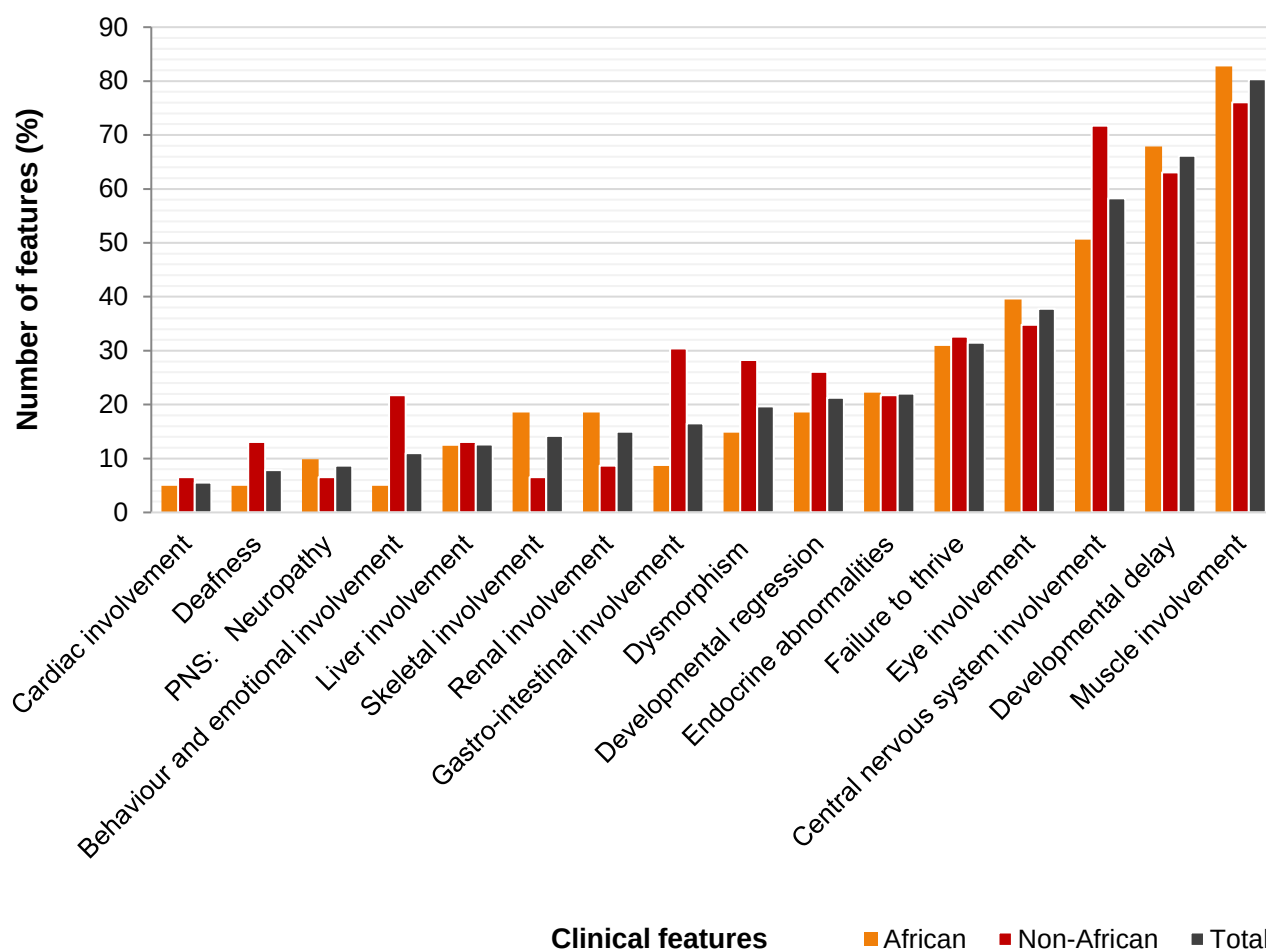


Figure 3.5: Distribution of the clinical features of the patients. Percentages were calculated for 81 African and 46 non-African patients, respectively.

3.4 BIOCHEMICAL ASSESSMENTS

3.4.1 Metabolic screening

Metabolic screenings to identify metabolite markers of IMD in blood, urine and/or cerebrospinal fluid were routinely performed on the specimens at the NHLS or PLIEM. This information was used to select patients suspected of having MD before they undergo muscle biopsies for further investigations. In a previous investigation, as mentioned in Chapter 1, targeted (organic acids, amino acids and fatty acylcarnitines, using gas and liquid chromatography mass spectroscopy [GC-MS and LC-MS, respectively]), as well as untargeted (using LC-MS and nuclear magnetic resonance spectroscopy) metabolomic analyses were performed on urine collected from this cohort of patients and controls. These investigations identified a number of elevated metabolite markers in this cohort, collectively referred to as a “biosignature”, which correlate with what is expected to occur in mitochondrial dysfunction. These metabolites include elevated lactate,

succinate, 2-hydroxyglutarate, 3-hydroxyisobutyrate, 3-hydroxyisovalerate, 3-hydroxy-3-methylglutarate, alanine, glycine, glutamate, serine, tyrosine, alpha-aminoadipate and creatine (Reinecke *et al.*, 2012; Smuts *et al.*, 2013; Venter *et al.*, 2015).

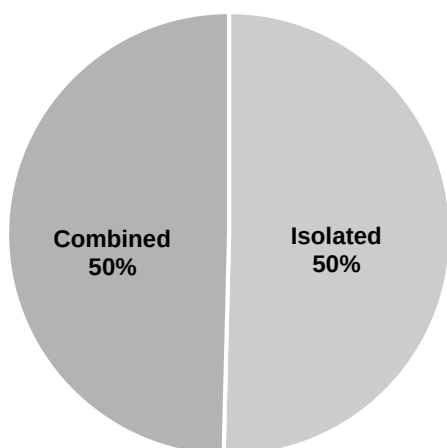
3.4.2 Muscle respiratory chain enzymology

Muscle biopsies from the *vastus lateralis* were obtained for 212 paediatric patients with suspected MDs and immediately frozen at -80°C. All surgical procedures were done at the *Steve Biko Academic Hospital* in Pretoria, at the Department of Paediatrics, after which the frozen biopsies were transferred to NWU.

To summarise, frozen muscle samples (± 100 mg) were prepared on ice and suspended in Zheng buffer, a homogenising buffer containing mannitol (210 mM), sucrose (70 mM), HEPES (5 mM), EGTA (0.1 mM), pH 7.2. A Potter-Elvehjem homogeniser was used to prepare a 10% (w/v) homogenate on ice, resulting in a homogenate with even dispersion and little to no remaining solid masses. Approximately 100 μ L of this homogenate was removed and stored at -80°C for DNA analysis. The remaining homogenate was centrifuged at 600 x g for 10 min at 4°C. The supernatant (referred to as “600 x g supernatant”) was removed and stored at -80°C to be used in downstream analyses such as protein quantification and RC enzyme analysis. Following this, the activities of RC enzyme CI–CIV and combined CII+III were measured and normalised against citrate synthase (CS). These analyses were done according to the standard operating procedures of the Mitochondria Research Laboratory (NWU), based on methods originally described by Rahman *et al.* (1996), Janssen *et al.* (2007), Shepherd and Garland (1969), and Luo *et al.* (2008), with reference values described by Smuts *et al.* (2010). Kinetic readings were done using the Synergy™ HT Microplate Reader (BioTek Instruments, Winooski, United States) and Gene5 software for data analyses (BioTek Instruments, v1.05). Quantification of proteins was done using the bicinchoninic acid method, as described by Smith *et al.* (1985).

Respiratory chain enzyme analysis resulted in a positive identification of a deficiency in 127 (60%) of the 212 referred patients. Figure 3.6 summarises the profile of RC deficiencies in this cohort, and shows an equal number of isolated ($n = 64$) and combined enzyme deficiencies ($n = 63$) (Figure 3.6A). Furthermore, as frequently reported, CI deficiency ($n = 75$, 59%) whether isolated ($n = 43$, 34%) or combined ($n = 32$, 25%), is the most prevalent, followed by patients with a CIII deficiency ($n = 54$, 42%, isolated: $n = 15$, 12%) or CIV deficiency ($n = 36$, 28%, isolated: $n = 4$, 3%) (See Figure 3.6B).

A



B

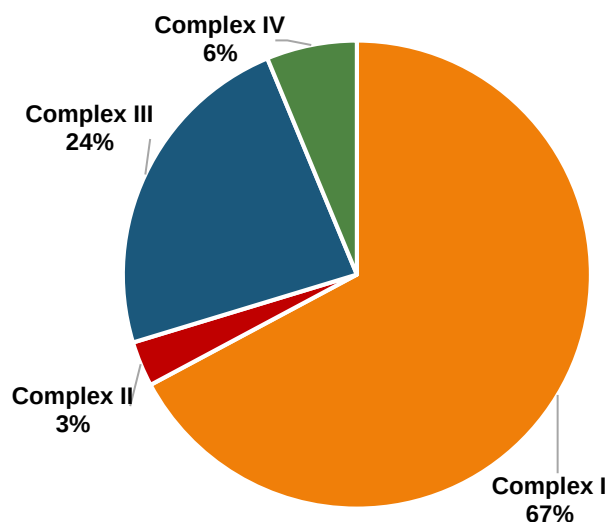


Figure 3.6: Distribution of respiratory chain muscle deficiencies in the selected patient cohort. Distribution of 127 patients with combined or isolated respiratory chain enzyme deficiencies is shown in Panel A. Distribution of isolated CI–CIV deficiencies is shown in Panel B.

3.5 STRATIFICATION FOR MOLECULAR WORK

Whole mtDNA NGS was only done on 123 of the 127 patients due to limited sample volumes. Paediatric patients were selected for nDNA NGS based on certain criteria, as there were limitations to both the number of genes (due to the kits used) and samples (due to funding limitations) which could be sequenced using massively parallel sequencing (MPS) in this study. Thus, nDNA NGS of targeted genes (Panels 1–3) was only done on 85 patients based on their biochemical profiles and the severity thereof, as will be discussed below.

3.5.1 Selection criteria

Muscle RC enzyme analysis data (described in Section 3.4.2) was used as a basis for NGS patient selection in this study. For nDNA MPS, a total of three sequencing panels, consisting of selected genes involved with MDs, were designed (described in Section 4.3.1). Panel 1, Panel 2, and Panel 3 patient selection were based on the deficiencies of the RC enzymes when expressed against CS. No biased choices were made regarding age, race or gender of the patients, or whether the patient had an isolated or a combined deficiency. The patients with the lowest enzyme activities compared to the lowest control value (arbitrarily set as 100%) were included as first choice for MPS and the clinical and other biochemical data available were also considered and discussed with the clinician before final inclusion.

For Panel 1, 32 patients were selected and for Panel 2 and Panel 3, 48 and 25 were selected respectively. In total, 85 patients (some patients were sequenced multiple times, see Table 4.6) were selected for MPS and consisted of an approximately equal number of male ($n = 43$) and female ($n = 42$) patients, of which 70% ($n = 60$) were of African descent. Age at onset ranged from the neonatal period to the second decade of life.

For whole mtDNA sequencing, 127 patients were selected, all of whom had an enzyme deficiency and therefore the criteria using the type of enzyme deficiency as used for nDNA were not applied here. Limited sample availability reduced the number of patients to be sequenced to 123, and furthermore a large section ($n = 71$) of this group of samples were already sequenced by Van der Walt *et al.* (2012). The 123 patients were of an approximately equal number of males ($n = 62$) and females ($n = 61$), of which 64% ($n = 79$) were of African descent. Age at onset ranged from the neonatal period to the second decade of life.

CHAPTER 4:

NEXT-GENERATION SEQUENCING

4.1 INTRODUCTION

Next-generation sequencing has widely become the preferred method for identifying genes and rare variants as a cause for an IMD, and especially so for MDs (Wortmann *et al.*, 2017). Generally, for heterogeneous diseases, such as MDs, the currently available options include targeted panels of candidate genes using MPS, unbiased WES, and WGS. Targeted panel sequencing focuses on specific genes or gene regions of interest known to be involved with, or which have been linked to, a specific phenotype or spectrum of phenotypes. That is to say, a panel can consist of as little as two genes, for example the breast cancer gene panel which consist of the genes *BRCA1* and *BRCA2*, or up to couple hundred genes, for example the inborn errors of metabolism panel consist of ~600 genes (<https://www.thermofisher.com>, last accessed Aug 2018). Using panel sequencing, either only exon regions of the specific gene are targeted and sequenced, or only the intron, or intergenic regions of a gene(s). Panel sequencing is often ideal to use in situations where there is already a foundation of candidate genes associated with a well-defined clinical phenotype. Data generated from panel NGS are easily manageable compared to the large amounts of data produced from WES. The latter includes all the protein-coding regions of the genome, i.e. only exomes are sequenced instead of the whole genome. This approach is often followed where the underlying genetic cause of disease is still unknown. The data produced using WES require computationally intensive bioinformatics analysis and a large data storage capacity. This is also the case for WGS. WGS is performed for sequencing of the entire genome of a species, including all intron as well as exon regions. In clinical environments, the latter is usually avoided due to the large amount of data generated and the high cost for one sample, even though this cost is decreasing as time goes by. Panel sequencing is the most cost-effective approach, especially when a large number of patients need to be sequenced. WES is, however, becoming the norm and standard when considering NGS, due to the broad spectrum of data obtained for each individual. However, it is not as readily available in developing countries.

As mentioned in the previous chapter, after extensive considerations were taken into account (see Sections 1.1 and 3.5), it was decided to do MPS or panel sequencing of selected nDNA genes known to be involved with MDs, along with whole mtDNA sequencing on an increased number of patients, as an expansion to the work done by Van der Walt *et al.* (2012). In addition to panel sequencing, WES was done on a subset of Black-African patients. The methodologies for both MPS, mtDNA NGS, and WES are described in this chapter.

4.2 DNA ISOLATION

DNA from the previously discussed 127 samples selected for sequencing, was either isolated from muscle biopsy tissue (126 cases) or whole blood sample (for a single case only), according to the manufacturer's protocol (NucleoSpin® Tissue, Macherey-Nagel, CAT #740952.250). Briefly, 25–30 mg of muscle tissue, or 50 µL whole blood, was lysed using Proteinase K and lysis buffer, followed by overnight incubation at 56°C. Samples then underwent a second lysis step, followed by incubation at 70°C for 10 min. The samples were transferred to a NucleoSpin® Tissue Column, and placed in a collection tube. The binding conditions of DNA to the NucleoSpin® Tissue Column were adjusted by adding absolute ethanol to the sample, followed by vigorous vortex. Hereafter, the samples were centrifuged at high speed (12 000 x g for 1 min at room temperature) and the membrane was washed with wash buffer, followed by another round of high-speed centrifugation for 1 min. This washing step, including the centrifugation step was repeated. The membrane was dried via centrifugation at high speed for 1 min in order to ensure that the residual wash buffer was discarded into the collection tube. Highly pure DNA was eluted twice (25 µL x 2) in low-Tris-EDTA (TE) buffer (pH 8.0) to ensure optimal yield. Spectrophotometric quantification and evaluation of quality (260:280nm ratio) of DNA was done using a Nanodrop® ND-1000 Spectrophotometer (Thermo Fisher Scientific, Massachusetts, United States).

4.3 NUCLEAR TARGETED SEQUENCING

4.3.1 Panel sequencing

Panel sequencing is ideal for identifying specific mutations in genes that have known, or suspected, association with disease. The focus is therefore on individual genes or gene regions that allow for sequencing at a higher depth, which also allows for the accurate identification of rare variants. In populations where little to no molecular information exists for IMDs, more specifically MDs, this is of great value when identifying disease-causing variants. This is indeed one of the objectives for this investigation. With these factors and other findings as described in

Section 3.5 in mind, it was decided to use panel sequencing. The genes in this study were selected in collaboration with Dr RJT Rodenburg (Department of Pediatrics, Radboud Center for Mitochondrial Medicine, Radboud University Medical Center, Nijmegen, The Netherlands) and previously reported publications.

Three custom panels were designed: one from Agilent HaloPlex™ Target Enrichment System and the remaining two from Ion AmpliSeq™ Library Preparation, and are described in detail below. There is no specific reason or criteria on for using two different types of kits. Table 4.1 lists properties and information about the three panels used while each of the three panel's genes and their corresponding mitochondrial involvement is listed in Table 4.2–Table 4.4.

Table 4.1: Properties of three custom designed panels used for massively parallel sequencing.

Panel name	Kit used	Target region size	No. of amplicons	No. of genes	Targeted genes
Panel 1	Agilent HaloPlex™ Target enrichment	360 091 kbp	24 861	78	CI-associated
Panel 2	Ion AmpliSeq™ Custom DNA Panels	157 834 kbp	1 115	78	CI–CIV-associated
Panel 3	Ion AmpliSeq™ Custom DNA Panels	61 kbp	304	18	CoQ ₁₀ -associated

Abbreviations: CI: complex I; CII: complex II; CIV: complex IV; CoQ₁₀: coenzyme Q₁₀.

4.3.2 Library preparation using Agilent HaloPlex™ Target Enrichment for Panel 1

Targeted amplicon libraries for the human genome GRCh37 were designed using the HaloPlex™ Advance Design Wizard (<http://www.genomics.agilent.com/en/HaloPlex-DNA/HaloPlex-Custom-Kits>). The final design (Panel 1) consisted of 78 CI structural, functional, assembly, and associated genes (listed in Table 4.2) and had a target region size of 360 091 kbp, with a total of 24 861 amplicons.

Table 4.2: Properties of targeted Panel 1, arranged alphabetically according to the targeted genes involved with CI.

Gene	Chromosome	No. of amplicons	Mitochondrial involvement	Reference
AARS2	CHR6	21	Aminoacyl-tRNA synthetases	Gorman <i>et al.</i> (2016) Rendón <i>et al.</i> (2016) Sugiana <i>et al.</i> (2008) Alston <i>et al.</i> (2017) Rorbach <i>et al.</i> (2008) Khan <i>et al.</i> (2016)
ACAD9	CHR3	21	RC assembly, CI	
AGK	CHR7	20	Phospholipid metabolism	
C12ORF65	CHR12	4	Release factor	
C20ORF7	CHR20	5	RC assembly, CI	
DARS2	CHR1	18	Aminoacyl-tRNA synthetases	

DGUOK	CHR2	8	Deoxynucleotide triphosphate synthesis	Janer <i>et al.</i> (2012) Miller <i>et al.</i> (2011)
EARS2	CHR16	13	Aminoacyl-tRNA synthetases	
ECSIT	CHR19	9	RC assembly factor, CI	
FOXRED1	CHR11	8	RC assembly, CI	
GFM1	CHR3	20	Elongation factor	
GFM2	CHR5	24	Elongation factor	
HARS2	CHR5	12	Aminoacyl-tRNA synthetases	
IARS2	CHR1	25		
LARS2	CHR3	25		
MARS2	CHR2	1		
MPV17	CHR2	9	Toxic compound metabolism	
MRPL3	CHR3	12	Mitoribosomal proteins	
MRPL40	CHR22	4		
MRPL44	CHR2	4		
MRPS16	CHR10	6		
MRPS22	CHR3	15		
MRRF	CHR9	10		
MTFMT	CHR15	9	tRNA modification	
MTO1	CHR6	14	tRNA modification	
NDUFA1	CHRX	3	RC subunit, CI	
NDUFA10	CHR2	17		
NDUFA11	CHR19	6		
NDUFA12	CHR12	11		
NDUFA13	CHR19	10		
NDUFA2	CHR5	3		
NDUFA3	CHR19	8		
NDUFA4	CHR7	5	RC subunit, CIV	
NDUFA4L2	CHR12	6	RC subunit, CI	
NDUFA5	CHR7	6		
NDUFA6	CHR22	3		
NDUFA7	CHR19	4		
NDUFA8	CHR9	5		
NDUFA9	CHR12	16		
NDUFAB1	CHR16	7		
NDUFAF1	CHR15	6	RC assembly, CI	
NDUFAF2	CHR5	6		
NDUFAF3	CHR3	3		
NDUFAF4	CHR6	4		
NDUFAF5	CHR20	15		
NDUFAF6	CHR8	30		
NDUFB1	CHR14	5	RC subunit, CI	
NDUFB10	CHR16	2		
NDUFB11	CHRX	3		
NDUFB2	CHR7	14		
NDUFB3	CHR2	5		

NDUFB4	CHR3	2	
NDUFB5	CHR3	6	
NDUFB6	CHR9	4	
NDUFB7	CHR19	3	
NDUFB8	CHR10	10	
NDUFB9	CHR8	7	
NDUFC1	CHR4	12	
NDUFC2	CHR11	4	
NDUFS1	CHR2	20	
NDUFS2	CHR1	15	
NDUFS3	CHR11	6	
NDUFS4	CHR5	7	
NDUFS5	CHR1	3	
NDUFS6	CHR5	5	
NDUFS7	CHR19	6	
NDUFS8	CHR11	6	
NDUFV1	CHR11	5	
NDUFV2	CHR18	11	
NDUFV3	CHR21	10	
NUBPL	CHR14	25	RC assembly, CI
PARS2	CHR1	2	Aminoacyl-tRNA synthetases
PEO1	CHR10	5	mtDNA replication
PNPT1	CHR2	26	mRNA processing
POLG	CHR15	22	Deoxynucleotide triphosphate synthesis
POLG2	CHR17	12	Deoxynucleotide triphosphate synthesis
PUS1	CHR12	6	tRNA modification
RARS2	CHR6	22	Aminoacyl-tRNA synthetases
RMND1	CHR6	13	Mitochondrial translation
RRM2B	CHR8	10	Deoxynucleotide triphosphate synthesis
SARS2	CHR19	17	Aminoacyl-tRNA synthetases
SERAC1	CHR6	18	Phospholipid metabolism
SUCLA2	CHR13	17	Deoxynucleotide triphosphate synthesis
SUCLG1	CHR2	13	Deoxynucleotide triphosphate synthesis
SUCLG2	CHR3	13	Succinyl-CoA synthetase
TAZ	CHRX	7	Phospholipid metabolism
TK2	CHR16	16	Deoxynucleotide triphosphate synthesis
TRMU	CHR22	11	tRNA modification
TSFM	CHR12	13	Elongation factors
TUFM	CHR16	9	
TYMP	CHR22	6	Deoxynucleotide triphosphate synthesis
YARS2	CHR12	6	Aminoacyl-tRNA synthetases

The HUGO gene nomenclature was used here and elsewhere in this thesis for naming genes, as can be found at <https://www.genenames.org/>.

Abbreviations: CHR: chromosome; RC: respiratory chain; CI: complex I; CIV: complex IV; tRNA: transfer RNA; mtDNA: mitochondrial DNA; mRNA: messenger RNA.

The targeted amplicon-enriched libraries were constructed according to the manufacturer's protocol (HaloPlex™ Target Enrichment System vD.3, March 2013, Agilent Technologies, California, United States). Reagents included the HaloPlex™ Target Enrichment System Kit (Agilent Technologies, CAT #G9902C), Herculase II Fusion Enzyme with dNTPs (Agilent Technologies, CAT #600677), nuclease-free water, Agencourt™ AMPure™ XP Reagent Kit (Beckman Coulter Genomics, CAT #A63881, part of Thermo Fisher Scientific, Massachusetts, United States), 10 M molecular grade sodium hydroxide (NaOH) (Sigma-Aldrich, CAT #72068, Darmstadt, Germany), 2 M acetic acid (Sigma-Aldrich, CAT #A8976), 10 mM Tris-HCl (Trizma hydrochloride) with pH 8.0, absolute ethanol (Sigma-Aldrich, CAT #E7023), and the Quant-iT dsDNA High Sensitivity Assay Kit, for use with the Qubit™ Fluorometer (Thermo Fisher Scientific, CAT #Q32866).

A simplified illustration of the HaloPlex™ workflow can be seen in Table 4.1A. Briefly, previously extracted DNA samples were diluted to a final concentration of 5 ng/μL, and digested into fragments using 16 unique restriction enzymes. An enriched control DNA sample (ECD, supplied by HaloPlex™ Target Enrichment System Kit) was used as a validation control sample during enzyme digestion. The restriction enzyme digestion reactions were validated using a 2.0% (w/v) agarose gel in Tris/Borate/EDTA (TBE) buffer stained with 0.5 μg/mL ethidium bromide (EtBr). Three predominant bands at approximately 125, 225 and 450 bp indicated successful digestion, an example of which is illustrated in Table 4.1B. The digested fragments were hybridised to the HaloPlex™ probe capture library. During this process, unique barcode sequences were incorporated into the targeted fragments to ensure successful multiplex polymerase chain reaction (PCR) for multiple samples. These targeted DNA hybrids contain biotin and are captured on streptavidin beads, followed by ligation to close any nicks in the circularised target DNA hybrids, but only if these are perfectly hybridised. The captured DNA was eluted by adding freshly prepared 50 mM NaOH, followed by PCR amplification according to protocol specifications. Libraries were purified using AMPure™ XP beads and washed with 70% freshly prepared ethanol, followed by elution and storing of the prepared libraries in low-TE buffer (pH 8.0). The enrichment process was validated using the 2100 Bioanalyzer High Sensitivity DNA Assay Kit (Agilent Technologies). The electropherogram obtained showed a peak fragment size range between 175–625 bp, with the majority of the library products between 225–525 bp as illustrated in Table 4.1C. The concentration was determined for each sample according to the area under the peak, and ranged from 150 to 550 pM (see Table 4.1D). Each sample containing a unique barcode, necessary for parallel sequencing, was diluted to ~100 pM, pooled, and stored at 4°C in low-TE buffer until used for downstream sequencing reactions.

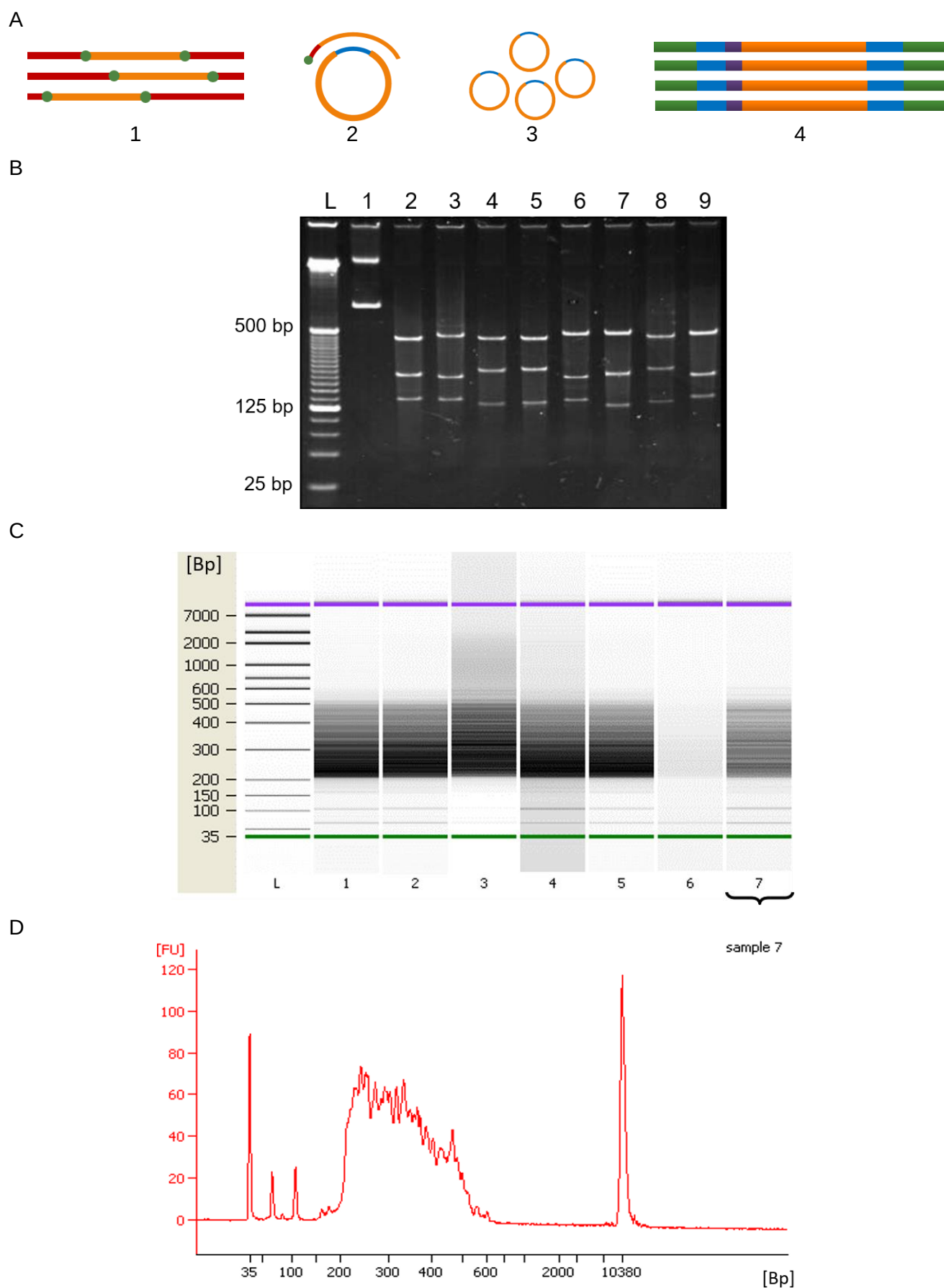


Figure 4.1: Process followed for HaloPlex Targeted Enrichment. Panel A: A schematic representation of a simplified HaloPlex™ Target Enrichment workflow. A1: Genomic DNA is fragmented using 16 restriction enzymes. A2: The probe library is added and designed so that both ends of the oligonucleotide.

are hybridised to targeted DNA fragments. Circular DNA molecules are formed as a result. A unique sample barcode sequence is also incorporated during this step. The probes are biotinylated and can therefore be retrieved with magnetic streptavidin beads. A3: Ligations close the circular molecule only if these are perfectly hybridised fragments. A4: Circular DNA targets are amplified via PCR. This enriched barcoded product is ready for sequencing. Adapted from the application note: Agilent HaloPlex Target Enrichment and SureCall Data Analysis: Optimised for the Ion Torrent™ PGM Sequencer. Panel B: Agarose gel electrophoresis of digested samples with various restriction enzymes (Lanes 1–8), an undigested enriched DNA control sample in Lane 9, with the ladder in lane far left. Panel C: Validation and quantification of enriched libraries using the 2100 Bioanalyzer High Sensitivity DNA Assay Kit. Successful target enrichment is indicated by a size distribution smear ranging from 175–625 bp, as can be seen for patient samples 1–7. The lower ladder marker is indicated by a green line (at 35 bp) and the upper ladder marker is indicated by a purple line (>10 000 bp). Panel D: The electropherogram for all samples (Sample 7 shown here) displayed a peak fragment size between 225–525 bp. The peak fragments observed at 35 bp and 10 380 bp indicate the lower and upper ladder markers, respectively. The peak fragments observed at ~75 bp and 125 bp are considered primer- and adaptor-dimer products, respectively.

4.3.3 Library preparation using Ion AmpliSeq custom DNA Panels 2 and 3

A second set of targeted amplicons for 78 genes (GRCh37) was designed using the online designer from the Ion AmpliSeq™ webpage (<https://ampliseq.com>, Ion AmpliSeq™ Custom DNA Panels, Thermo Fisher Scientific, last accessed August 2018). The second design (Panel 2, see Table 4.1 and Table 4.3) consisted of 1 115 amplicons, with a target region size of 157 834 kbp. A third set (Panel 3, see Table 4.1 and Table 4.4) of targeted amplicons for 18 genes (GRCh37) known to be involved with primary and secondary CoQ₁₀ deficiency, was also designed for Ion AmpliSeq™ Library Enrichment. This panel had a target region size of 61.37 kbp, with 304 amplicons.

Table 4.3: Properties of targeted Panel 2, arranged alphabetically according to the genes involved with CI, CII, CIII and CIV.

Gene	Chromosome	No. of amplicons	Mitochondrial involvement	References
ACAD9*	CHR3	25	RC assembly, CI	Abu-Libdeh <i>et al.</i> (2017); Alston <i>et al.</i> (2017); Craven <i>et al.</i> (2017); Gorman <i>et al.</i> (2016); Gorman <i>et al.</i> (2015); Koopman <i>et al.</i> (2016); Rendón <i>et al.</i> (2016); Sugiana <i>et al.</i> (2008)
BCS1L	CHR2	15	RC subunit, CIII	
COX10	CHR17	21	RC subunit, CIV	
COX15	CHR10	43		
COX4I1	CHR16	8		
COX4I2	CHR20	7		
COX5A	CHR15	6		
COX5B	CHR2	5		
COX6A1	CHR12	5		

COX6A2	CHR16	5		
COX6B1	CHR19	6		
COX6B2	CHR19	12		
COX6C	CHR8	8		
COX7A1	CHR19	8		
COX7A2	CHR6	8		
COX7B	CHRX	4		
COX7B2	CHR4	4		
COX7C	CHR5	5		
COX8A	CHR11	5		
COX8C	CHR14	3		
CYC1	CHR8	11	RC subunit, CIII	
FASTKD2	CHR2	45	RC subunit, CIV	
FOXRED1*	CHR11	18	RC assembly, CI	
LRPPRC	CHR2	64	mRNA processing	
MTFMT*	CHR15	18	tRNA modification	
NDUFA1*	CHRX	4	RC subunit, CI	
NDUFA10*	CHR2	32		
NDUFA11*	CHR19	22		
NDUFA12*	CHR12	6		
NDUFA13*	CHR19	9		
NDUFA2*	CHR5	7		
NDUFA7*	CHR19	5		
NDUFA8*	CHR9	6		
NDUFA9*	CHR12	16		
NDUFAF1*	CHR15	12		
NDUFAF2*	CHR5	8		
NDUFAF3*	CHR3	15		
NDUFAF4*	CHR6	18		
NDUFAF5*	CHR20	23		
NDUFAF6*	CHR8	17		
NDUFAF7*	CHR2	18		
NDUFB3*	CHR2	7	RC subunit, CI	
NDUFB6*	CHR9	8		
NDUFB9*	CHR8	7		
NDUFS1*	CHR2	37		
NDUFS2*	CHR1	18		
NDUFS3*	CHR11	9		
NDUFS4*	CHR5	8		
NDUFS5*	CHR1	5		
NDUFS6*	CHR5	7		
NDUFS7*	CHR19	12		
NDUFS8*	CHR11	10		
NDUFV1*	CHR11	14		
NDUFV2*	CHR18	10		

NDUFV3*	CHR21	13		
NUBPL*	CHR14	26	RC assembly, CI	
POLG*	CHR15	41	Deoxynucleotide triphosphate synthesis	
RRM2B*	CHR8	38		
SCO1	CHR17	13	RC subunit, CIV	
SCO2	CHR22	8		
SDHA	CHR5	22	RC subunit, CII	
SDHAF1	CHR19	9	RC assembly factor, CII	
SDHAF2	CHR11	10		
SDHB	CHR1	10	RC subunit, CII	
SDHC	CHR1	20		
SDHD	CHR11	13		
SUCLA2*	CHR13	24	Deoxynucleotide triphosphate synthesis	
SURF1	CHR9	8	RC assembly, CIV	
TACO1	CHR17	10	mRNA processing	
TRMU*	CHR22	16	tRNA modification	
UQCR10	CHR22	7	RC subunit, CIII	
UQCR11	CHR19	10		
UQCRB	CHR8	33		
UQCRC1	CHR3	16		
UQCRC2	CHR16	16		
UQCRFS1	CHR19	8		
UQCRH	CHR1	5		
UQCRQ	CHR5	10		

*Overlapping genes in Panel 1 and Panel 2.

Abbreviations: CHR: chromosome; RC: respiratory chain; CI: complex I; CIII: complex III; CIV: complex IV; tRNA: transfer RNA; mRNA: messenger RNA.

Library preparation as well as the enrichment of targeted regions, was done according to the Ion AmpliSeq™ Library Kit manufacturer's protocol (Ion AmpliSeq™ DNA and RNA Library Preparation, Revision B.0, Thermo Fisher Scientific). Reagents supplied by Thermo Fisher Scientific included an Ion AmpliSeq™ Library Kit 2.0 (CAT #4475345), Agencourt™ AMPure™ XP Reagent Kit (CAT #A63881), Qubit™ 2.0 Fluorometer, and Qubit™ dsDNA HS Assay Kit (CAT #Q32854).

Table 4.4: Properties of targeted Panel 3, arranged alphabetically according to the genes involved with CoQ₁₀ biosynthesis.

Gene	Chromosome	No. of amplicons	Mitochondrial involvement	Reference
ADCK3	CHR1	30	CoQ ₁₀ biosynthesis	Alston <i>et al.</i> (2017); Doimo <i>et al.</i> (2014); Duncan <i>et al.</i> (2009); Gorman <i>et al.</i> (2016); Jonassen and Clarke (2000); Lagier-Tourenne <i>et al.</i> (2008); Liang <i>et al.</i> (2009); Mancuso <i>et al.</i> (2010); Quinzii <i>et al.</i> (2008)
APOE	CHR19	9	-	
APTIX	CHR9	22	Secondary CoQ ₁₀ biosynthesis	
BRAF	CHR7	28		
COQ10A	CHR12	13	CoQ ₁₀ biosynthesis	
COQ10B	CHR2	11		
COQ2	CHR4	15		
COQ3	CHR6	11		
COQ4	CHR9	13		
COQ5	CHR12	10		
COQ6	CHR14	15		
COQ7	CHR16	19		
COQ9	CHR16	12		
ETFA	CHR15	16		
ETFB	CHR19	14		
ETFDH	CHR4	22	CoQ ₁₀ biosynthesis	
PDSS1	CHR10	11		
PDSS2	CHR6	24		

Abbreviations: CHR: chromosome; CoQ₁₀: coenzyme Q₁₀.

A simplified illustration of an AmpliSeq™ workflow can be seen in Figure 4.2. Briefly, the first part of enrichment included standard PCR amplification of the targeted region primer pools using 20 ng high quality genomic DNA (gDNA) and 5X Ion AmpliSeq™ HiFi Mix (5X; 4 µL) in 20 µL PCR reactions. Optimal annealing temperature for the primers was at 60°C and extension time was 4 min for all three panels. The number of cycles differed; Panels 1 and 2 had a total of 15 cycles and Panel 3 had a total of 17 cycles. As an added quality control step, the PCR products were validated with 1.0% (w/v) agarose gel in TBE buffer stained with 5 µg/mL EtBr. Following standard PCR amplification, the amplicons were partially digested by FuPa reagent (2 µL, supplied). A mixture of Ion P1 Adapter (2 µL), Ion Xpress™ Barcode (2 µL), and nuclease-free water (4 µL) was prepared and added to each amplified sample in a 1:4 ratio (2 µL). P1 Adapters allow the amplicons to clonally bind to Ion Sphere™ Particles (ISPs). Ligation of the barcode and P1 Adapter mixture was done using a supplied “Switch” solution and DNA ligase. The amplicons for each sample containing a unique barcode adapter mixture were purified with Agencourt™ AMPure™ XP Reagent (1.5X sample volume), followed by two wash steps using 70% freshly prepared ethanol. Libraries were quantified using a Qubit™ Fluorometer and diluted to 100 pM. These constructed libraries were pooled in equimolar volumes and stored at 4°C in low-TE buffer for later use in downstream sequencing reactions.

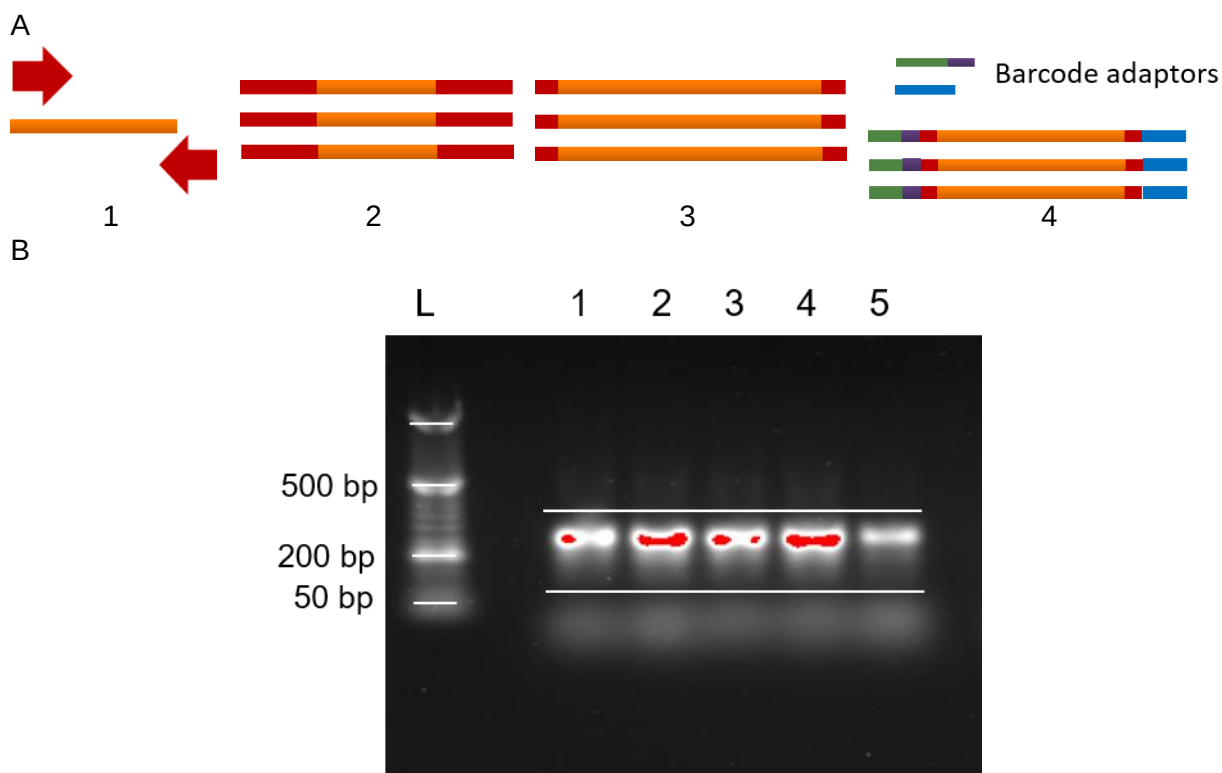


Figure 4.2: Library preparation workflow for Ion AmpliSeq. Panel A: Targeted enrichment. A1: Selected target regions of interest from genomic DNA are amplified by specific primer pairs (red arrows) into ~200–300 bp amplicons, using standard PCR technique (A2). A3: Amplicons are partially digested and purified, followed by ligation of adaptors (light blue), barcodes (orange), and a P1 Adapter (green), necessary for binding to an ion sphere particle (A4). Adapted from the Ion AmpliSeq™ Library Kit 2.0 user guide. Panel B: Agarose gel electrophoresis of amplified samples (step 1 and 2 in A) in Lanes 1–5, with the ladder in the lane far left. Successful amplicon amplification is indicated by the smear ranging from ~150 bp–300 bp, as is the case in the five samples indicated here.

4.3.4 Template preparation

The same procedure for template preparation was followed for all three panels. The template was prepared using the Ion PGM™ Template OT2 Solutions 200 Kit (CAT #4480974) and the Ion OneTouch™ 2 instrument (Thermo Fisher Scientific) according to manufacturer's protocol for sequencing on the Ion PGM™ platform.

The template preparation of positive ISPs was prepared from previously constructed amplicon libraries. The ISPs and quality of the template were assessed according to the Ion PGM™ Template OT2 200 Kit user guide, using the Ion Sphere™ Quality Control Assay (CAT #4468656) and the Qubit™ 2.0 Fluorometer. The template with positive ISPs was enriched using the Ion OneTouch™ 2 instrument. This enriched template was stored at 4°C for up to three days prior to chip loading and sequencing.

4.3.5 Ion Torrent next-generation sequencing

Massively parallel sequencing of enriched targeted regions was performed based on Ion Torrent sequencing technology, using the Ion PGM™ platform. The reagents and kits were all supplied by Thermo Fisher Scientific, which included Ion PGM™ Sequencing 200 Kit v2 (CAT #4482006) and the Ion 318™ Chip Kit v2 (CAT #4484355). The PGM™ was initialised for a sequencing run, followed by manual loading of the enriched template with positive ISPs on an Ion 318™ Chip according to the manufacturer's protocol. No more than eight samples were loaded per chip, ensuring an average coverage depth of ~300X. In total, 15 templates for the three panels were sequenced and analysed. A summary report was generated for each sequencing run and was used to assess the quality of chip loading, number of usable reads, and the average depth coverage for each sample. An example of a sequencing run report is illustrated in Figure 4.3.

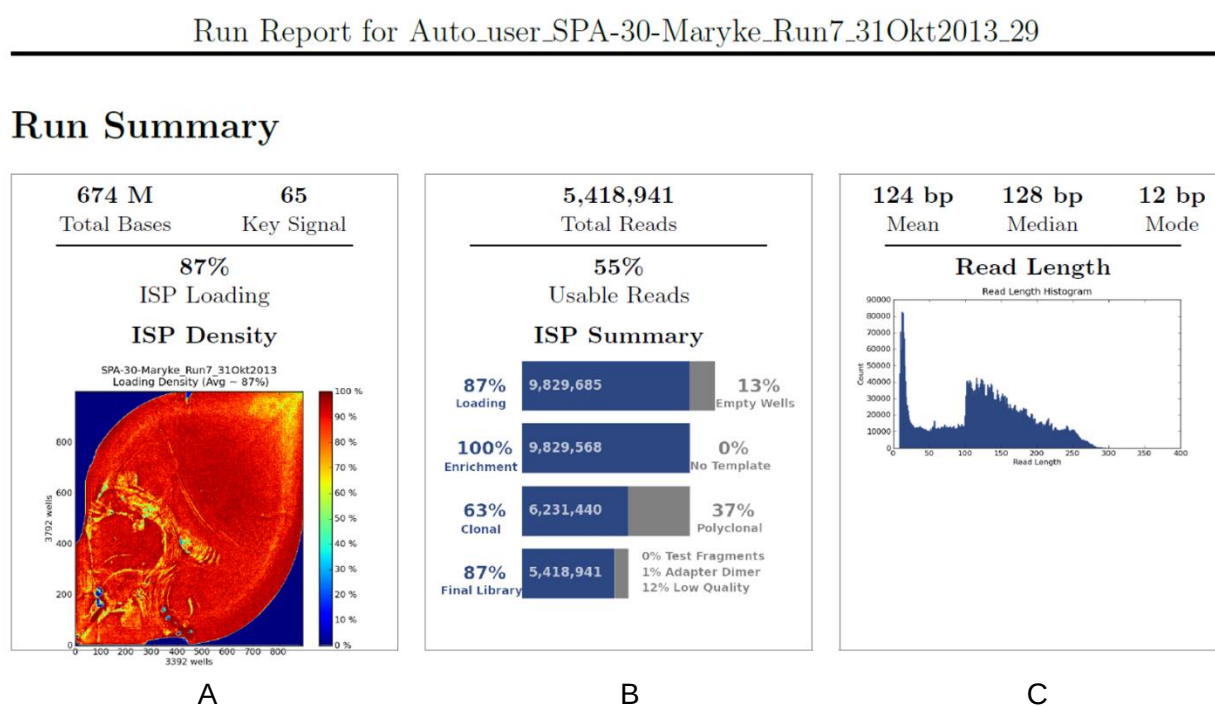


Figure 4.3: A run report generated by the Ion PGM™ Torrent Suite. Panel A indicates the percentage of the chip loading with live ISPs in each of the wells. Higher percentages indicate better loading where deep red indicates highest loading and blue indicates lowest loading in each respective well. Panel B gives information on the ISPs. Percentages of enriched ISPs as well as the number of polyclonal ISPs are summarised. Panel C gives information on the mean, median and mode for read lengths.

4.4 MITOCHONDRIAL GENOME SEQUENCING

4.4.1 Amplification and sample preparation for next-generation sequencing

The first mitochondrial sequencing for patients S001–S061 was previously done and described in detail by Van der Walt *et al.* (2012). The samples of that study were sequenced on a Roche 454 GS-FLX platform at Inqaba Biotech (Pretoria, South Africa). For this investigation, as an extension of work done by Van der Walt *et al.* (2012), an additional set of patients (S067–S127) were sequenced (whole mtDNA genome) according to the methods followed by Van der Walt *et al.* (2012) with minor adjustments. This included optimisation of PCR reaction and post-PCR clean-up steps. These samples were sequenced using the Ion Ion S5™ System (CAT #A27212, ThermoFisher Scientific) at the *Centre for Analytical Facilities* (CAF; Stellenbosch, South Africa), according to the protocols provided by Ion Torrent. The first part of sample preparation was done at the NWU and included DNA isolation (as previously described in Section 4.2) and mtDNA amplification and purification. To summarise the latter, the complete human mtDNA genome was amplified from gDNA in two overlapping fragments, namely Fragments A and B (refer to Table 4.5 for detail), via long template PCR, using AccuPrime™ High Fidelity enzyme (Thermo Fisher Scientific, CAT #12346-086) (1 unit; 5 µL) in 50 µL PCR reaction. The optimal annealing temperature for the primers was at 56.5°C. For the first 10 cycles, the extension step was kept constant for 10 min, after which it was periodically extended by 10s after each cycle, for the next 20 cycles.

Table 4.5: Primer pair sequences for Fragments A and B used in PCR amplification.

	Primer	Primer sequence	Position on mtDNA	Size (bp)
Fragment A	SC-H-Forward	5'-ATCATACACAAACGCCTGAGC-3'	13539-13559	9250
	SC-C-Reverse	5'-GGTAAGAGTCAGAAGCTTATG-3'	6200-6220	
Fragment B	SC-D-Forward	5'-AATACCCATCATAATCGGAGG-3'	6115-6135	7546
	SC-G-Reverse	5'-TTGACCTGTTAGGGTGAGAAGA-3'	13640-13660	

A 0.7% (w/v) agarose gel in TBE buffer stained with 0.5 µg/ml EtBr was used to separate fragments, which confirmed successful initial PCR amplifications (refer to Figure 4.4A). Furthermore, the fragments were purified using a PureLink™ Pro 96 PCR Purification Kit from Thermo Fisher Scientific (CAT #K3100-96A) and quantified using their Quant-iT™ PicoGreen® dsDNA Assay Kit (CAT #P11496). Equimolar amounts of Fragments A and B, in a ratio of 5.6:4.9 for each sample, were combined, with a final concentration of 100 ng/µL. These pooled fragments

were again separated using a 0.7% (w/v) agarose gel in TBE buffer stained with 0.5 µg/mL EtBr. Visual inspection confirmed that the fragments were indeed pooled into approximately equimolar amounts (refer to Figure 4.4B). Library preparation, template preparation and enrichment were done at CAF (Stellenbosch, South Africa) according to the manufacturer's protocols for sequencing using the Ion S5™ System.

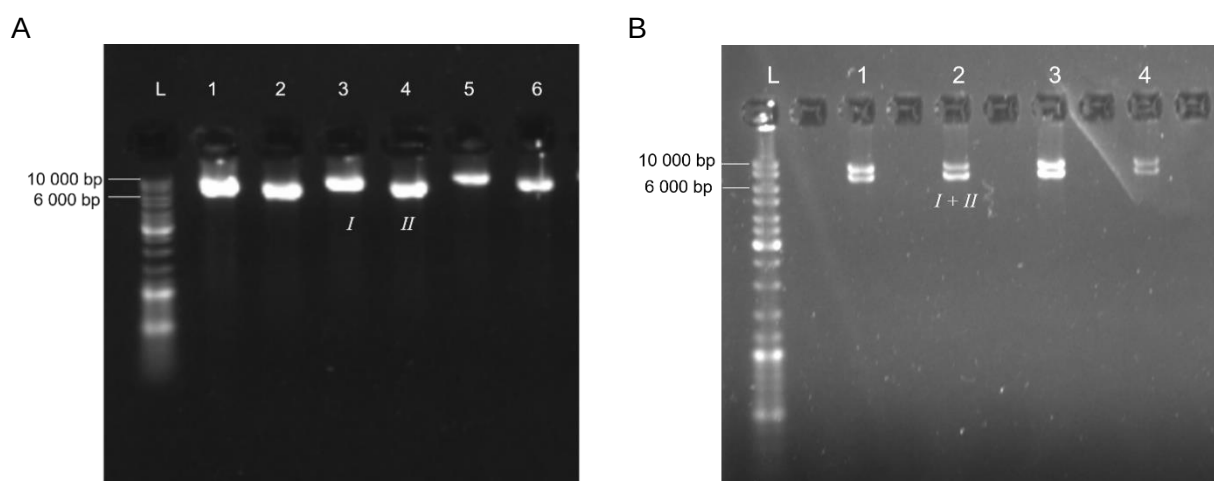


Figure 4.4: Validation of PCR amplification using agarose gel electrophoresis, with a 250–10 000 bp ladder in both cases. Panel A: Amplification of two fragments, where “I” is Fragment A and “II” is Fragment B. One sample was loaded in alternate lanes (Lanes 1–6). Panel B: An equimolar amount of Fragment A and Fragment B was used, where Fragment A is at the top and Fragment B is at the bottom in Lanes 1–4.

4.5 WHOLE EXOME SEQUENCING

In addition to panel sequencing and whole mitochondrial genome sequencing, a subset of 10% of African patients from the cohort of 127 was randomly selected for WES. The eight patients are listed in Table 4.6. This subset of patients was sequenced at CAF (Stellenbosch, South Africa). Briefly, DNA was diluted to 25 ng/µL after which several quality control (QC) steps were executed prior to library construction. The QC steps include DNA quantification using the Qubit fluorometer, DNA quantification and integrity check using the Nanodrop, and genome quality control using the PerkinElmer LabChip. Manual library preparation, template preparation and enrichment were done according to manufacturer's protocols for sequencing using Ion Proton™ System (CAT #4476610, Thermo Fisher Scientific).

4.6 MUTATION SCREENING IN TWO GENES, *GCDH* AND *MPV17*

Two founder mutations were detected in South African patients of African descent by Meldau *et al.* (2018) and Van der Watt *et al.* (2010). The cohort investigated here was also selected for

screening for these two mutations. All the patients who were selected for screening, permitting sample was sufficient and available, are listed in Table 4.6. The methods, results and discussion for the screening of the two mutations are fully described in Chapter 8.

4.7 PATIENTS SELECTED FOR SEQUENCING AND MUTATION SCREENING

In total, 127 ethnically diverse, unrelated patients were sequenced. For sequencing of the whole mtDNA genome, only 123 patients were selected as insufficient sample prevented sequencing of the following four patients, S007, S008, S078, and S107. For Panels 1–3, 86 patients were selected. In addition, eight African patients were randomly selected for WES. Table 4.6 provides a list of the entire paediatric cohort with their corresponding biochemical defects, in other words the RC enzyme analysis data on which the selection criteria was based.

Table 4.6: List of patient cohort with corresponding NGS approach followed.

Patient ID	Age, race, gender	Biochemical defect	NGS approach ¹					Screening
			Whole mtDNA	Panel 1	Panel 2	Panel 3	Exome	GCDH and MPV17
S001	N, A, F	CI, CII+CIII	X					X
S002	N, A, F	CII, CII+CIII	X		X	X	X	X
S003	N, A, F	CII+CIII	X					
S004	N, A, F	CIII	X			X		X
S005	N, A, F	CI	X		X			X
S006	N, NA, F	CIV	X					
S007	N, A, M	CII, CIII, CII+CIII			X			X
S008	<1, NA, M	CI		X				X
S009	N, NA, M	CI, CIII	X	X	X			X
S010	<1, A, F	CI, CII+CIII	X			X		X
S011	N, A, F	CIII	X				X	X
S012	10–20, NA, M	CII+CIII	X					X
S013	<1, A, F	CIII, CII+CIII	X		X			X
S014	1–2, A, M	CI	X	X				X
S015	N, A, M	CI, CII, CIII, CIV, CII+CIII	X		X	X		X
S016	2–5, A, M	CII	X					
S017	10–20, A, F	CI, CII, CIII, CIV, CII+CIII	X	X		X	X	X
S018	10–20, NA, F	CIV, CII+CIII	X					

S019	<1, NA, F	CI, CII+CIII	X					X
S020	N, A, M	CII+CIII	X			X		X
S021	<1, NA, M	CI	X		X			X
S022	10–20, A, M	CI, CIII	X					X
S023	1–2, A, M	CI, CIII, CIV, CII+CIII	X	X		X		X
S024	2–5, NA, M	CI	X					X
S025	1<, A, F	CI, CIII, CIV, CII+CIII	X	X	X	X		X
S026	1<, A, F	CIII, CII+CIII	X	X		X		X
S027	1<, A, M	CIII, CIV, CII+CIII	X			X	X	X
S028	1–2, A, M	CI, CII+CIII	X	X				X
S029	<1, A, M	CII, CIII, CII+CIII	X		X	X		X
S030	N, A, M	CIII, CII+CIII	X					X
S031	N, NA, F	CIII, CIV, CII+CIII	X					X
S032	N, A, F	CIV, CII+CIII	X			X	X	X
S033	N, A, F	CI	X	X		X	X	X
S034	6–10, A, F	CIII	X		X			X
S035	N, A, F	CI, CII+CIII	X	X				
S036	<1, NA, F	CI, CIII	X					X
S037	6–10, NA, M	CIII	X					X
S038	<1, NA, F	CI, CIII	X	X	X			X
S039	<1, A, F	CIII	X					
S040	<1, A, F	CIII	X		X			X
S041	6–10, A, M	CIII	X					X
S042	<1, A, F	CIII, CIV	X		X			X
S043	<1, A, F	CIII, CIV	X		X			X
S044	6–10, NA, M	CI, CII, CIII, CIV, CII+CIII	X	X	X	X		X
S045	N, A, M	CIII, CIV	X		X			X
S046	2–5, NA, M	CIII, CIV	X					X
S047	6–10, A, M	CII+CIII	X			X		X
S048	N, A, M	CIII, CIV, CII+CIII	X		X			X
S049	2–5, A, M	CIII	X		X			X
S050	<1, NA, F	CIII, CIV	X		X			
S051	6–10, A, M	CII+CIII	X					X
S052	<1, NA, M	CI	X	X				X
S053	10–20, NA, M	CIII	X					X

S054	2–5, A, F	CIII, CIV	X		X			X
S055	<1, A, F	CI, CIII, CIV	X	X	X			X
S056	<1, A, M	CI, CIII	X		X			X
S057	1–2, NA, F	CI, CIII, CIV, CII+CIII	X		X	X		X
S058	1–2, A, M	CI, CIII, CIV	X		X			X
S059	1–2, A, M	CI, CIII, CIV, CII+CIII	X		X		X	X
S060	N, NA, F	CIII, CIV	X		X			X
S061	<1, A, F	CI	X		X			X
S062	1–2, NA, F	CII+CIII	X			X		X
S063	1–2, A, M	CI, CIII, CIV, CII+CIII	X		X	X		X
S064	N, NA, F	CIII, CII+CIII	X		X	X		X
S065	N, NA, M	CIII, CII+CIII	X			X		X
S066	N, NA, M	CI, CIV	X					X
S067	<1, A, M	CIII	X					X
S068	<1, NA, F	CI, CIII	X					X
S069	<1, A, M	CII+CIII	X			X		X
S070	<1, A, F	CIII, CIV, CII+CIII	X		X	X		X
S071	<1, A, F	CIII	X					X
S072	<1, A, M	CIII	X		X			X
S073	<1, A, F	CII+CIII	X			X		X
S074	<1, NA, M	CIII, CII+CIII	X			X		X
S075	1–2, A, F	CIII	X		X			X
S076	N, A, F	CIII, CII+CIII	X		X	X		X
S077	N, A, F	CI	X	X				X
S078	1–2, A, F	CII			X			X
S079	1–2, A, M	CI, CIV	X		X			X
S080	<1, A, F	CI	X	X				X
S081	<1, A, F	CI	X					
S082	<1, NA, M	CI	X					X
S083	<1, A, F	CI	X	X				X
S084	<1, A, M	CI	X	X				X
S085	1–2, A, F	CIV	X		X			X
S086	N, A, F	CI	X					X
S087	N, NA, M	CI, CII+CIII	X					X
S088	N, NA, M	CI	X					X
S089	2–5, A, M	CI	X	X				X
S090	N, A, M	CI	X					X
S091	N, NA, M	CI	X					X
S092	1–2, NA, M	CI	X	X				X
S093	N, A, F	CIII	X					X

S094	1–2, A, F	CI	X		X			X
S095	6–10, NA, M	CI	X	X				X
S096	<1, NA, M	CI	X					X
S097	N, NA, F	CI	X	X				X
S098	N, NA, F	CI	X	X				X
S099	<1, A, M	CI	X	X		X		X
S100	<1, A, M	CI	X	X				X
S101	N, A, F	CI	X	X				X
S102	N, A, F	CI	X					X
S103	N, NA, M	CIV	X		X			X
S104	10–20, NA, F	CI	X					X
S105	6–10, NA, M	CI	X	X				X
S106	2–5, A, F	CI, CIV	X					X
S107	10–20, NA, F	CI, CIV		X				X
S108	<1, NA, M	CI	X	X				X
S109	<1, A, F	CI, CIV	X	X				X
S110	10–20, A, F	CI	X					X
S111	6–10, A, M	CI	X		X			X
S112	<1, A, M	CI, CII, CIV	X		X			X
S113	N, NA, M	CI, CIV	X		X			X
S114	6–10, A, F	CI, CIV	X		X			X
S115	N, NA, M	CI, CII, CIII, CIV, CII+CIII	X		X			X
S116	N, NA, M	CI	X		X			X
S117	2–5, A, M	CI, CIII, CIV	X		X		X	X
S118	N, NA, M	CI	X					X
S119	2–5, A, M	CI	X					X
S120	N, NA, F	CI	X					X
S121	<1, A, M	CIV	X		X			
S122	1–2, NA, M	CI	X					X
S123	1–2, A, F	CI	X		X			
S124	2–5, NA, F	CI	X					X
S125	<1, A, F	CI	X					X
S126	N, NA, M	CI	X					
S127	1–2, A, F	CI	X		X			X

Abbreviations: N: neonatal; <1: first year of life; 1–2: 1–2 years; 2–5: 2–5 years; 6–10: 6–10 years; 10–20: second decade; A: African; NA: non-African; F: female; M: male; CI–CIV: complex I–CIV; +: combined deficiency between the indicated complexes; NGS: next-generation sequencing.

¹Grey blocks indicate a positive marker for the corresponding NGS approach and white/empty blocks indicate a negative marker for the NGS approach used, i.e. no sequencing was done for that specific patient.

CHAPTER 5:

BIOINFORMATICS AND COMPUTATIONAL TOOLS

As discussed in the previous chapter, selected nuclear gene panels, and the whole mtDNA genome were sequenced for paediatric patients with MDs using Ion Torrent technology. For nDNA MPS, three panels were designed, constituting the RC complex genes as well as genes associated with mitochondria and their functioning, while the whole mitochondrial genome was sequenced in two overlapping fragments. WES was additionally done on a subset of patients. Sequencing data from MPS, whole mtDNA and WES were processed and analysed to identify novel and disease-causing variants using a bioinformatics pipeline developed in-house. Efficient bioinformatics pipelines form a fundamental component in identifying disease-causing variants in patients with rare disease such as MDs and have a direct impact on the diagnosis of these patients. As it was a specific objective and notable contribution of this study, this chapter will document the bioinformatics pipeline developed for nDNA NGS data analysis and the previously developed mtDNA bioinformatics pipeline used for mtDNA NGS data (Weissensteiner *et al.*, 2016).

A part of this chapter has been published in the *South African Journal of Science* (see Appendix A.1):

Schoonen, M., Seyffert, A.S., Van der Westhuizen, F.H. & Smuts, I. (2019). A bioinformatics pipeline for rare genetic diseases in South African patients. *South African Journal of Science*. 2019;115(3/4), Art 4876. <https://doi.org/10.17159/sajs.2019/4876>

5.1 ABSTRACT

The research fields of bioinformatics and computational biology are growing rapidly in South Africa. Bioinformatics pipelines play an integral role in handling sequencing data used to investigate common and rare disease aetiology. Bioinformatics platforms for common disease aetiology are well-supported and under continuous development in South Africa. However, this is not the case for rare disease aetiology research. Investigations into the latter rely on international

cloud-based tools for data analyses and ultimate confirmation of a genetic disease. These tools are not necessarily optimised for ethnically diverse population groups. We present a bioinformatics pipeline developed in-house to enable researchers to annotate and filter variants in either exome or amplicon NGS data. This pipeline was developed using NGS data from a predominantly African cohort of patients diagnosed with rare disease.

5.2 INTRODUCTION

The research fields of bioinformatics and computational biology are both growing rapidly in South Africa, with an ever-increasing number of both small and large laboratories having access to NGS technologies. This increase in sequencing capability continues to stimulate the development of a variety of platforms, databases, and initiatives, such as H3Africa (<https://www.h3africa.org/>), the South African Human Genome Programme (<http://sahgp.sanbi.ac.za>), and the South African Bioinformatics Institute (<https://www.sanbi.ac.za/>), to support sequencing data analysis. To date, though, the majority of these developments have been focussed on common-disease-related research, since diseases like HIV, fibromyalgia, tuberculosis, and malaria are major health challenges in Southern Africa (Mulder *et al.*, 2016).

Southern African researchers performing rare-disease related research (diseases affecting 6–8% of the global population) (Dawkins *et al.*, 2017), however, still rely on flexible, international cloud-based tools, such as Bystro, BrowseVCF, and RD-Connect. This is because their analysis needs have not yet been fully met locally (Kotlar *et al.*, 2018; Salatino & Ramraj, 2016; Thompson *et al.*, 2014). In European populations these tools allow researchers to leverage well-established genotype-phenotype correlations to guide investigations into rare disease aetiology. However, these correlations do not always hold for African populations significantly reducing the power of and the degree to which these online tools can be relied on in the South African research context.

Selected glossary

Some terms used in this thesis could have more than one meaning and are used in the following manner:

- 1. Bioinformatics:** The complex analysis of genetic data using several dependencies and tools. Analysis includes collection and classification of sequencing data, the storage thereof, and extraction of biological information using specially developed tools. This is effectively achieved by using a specialised pipeline.
 - 2. Script and pipeline** (computer science): Plain text files containing the series of commands that are executed to perform a specific function or process.
 - 3. Bioinformatics pipeline:** A computer science term which involves a series of processes or functions (file transformations), where a number of commands are executed in a chaining manner. Scripts are written to execute these commands.
-

A need therefore exists for a tool that is both sensitive to population heterogeneity and general enough in nature to enable effective domestic rare disease research.

African population groups are heterogeneous, with large genetic variety and limited information available on genotype-phenotype correlations (Van der Westhuizen *et al.*, 2015). Even though data are processed using the same reference genome, some aspects such as allele frequency for disease-causing variants and genotype-phenotype correlation could differ between ethnically diverse population groups, and must be taken into account (Shearer *et al.*, 2014).

In this paper a bioinformatics pipeline developed in-house to address these limitations is presented. This pipeline processes NGS data from ethnically diverse population groups without any strong prior assumptions regarding genotype-phenotype correlations. This offline pipeline, suitable for the analysis of both exome and amplicon sequencing data, is written in Bash (or the Bourne-Again SHell) and uses only open-source software. To allow other researchers to benefit from this work, the pipeline has been made available on GitHub under the GNU GPLv3 software license (Seyffert, 2018). The Ensembl Variant Effect Predictor (VEP) offline script is used for variant annotation, and the Genome Mining (GEMINI) command line database management tool for variant filtering (McLaren *et al.*, 2016; Paila *et al.*, 2013). The pipeline is easily adjustable with regard to what annotations are made and how they are filtered, which is especially useful when working with NGS data from ethnically diverse patients. The bioinformatics pipeline for variant annotation and filtering of amplicon and exome sequencing data presented here has been successfully used in research forming part of the project: *Investigating the aetiology of South African paediatric patients diagnosed with mitochondrial disorders* (Louw *et al.*, 2018; Smuts *et al.*, 2010; Van der Walt *et al.*, 2012; Van der Westhuizen *et al.*, 2017).

5.3 BIOINFORMATICS PIPELINE

Workflows leveraging NGS technology consist mainly of two parts, consisting of a primary and secondary analysis. Figure 5.1 shows what this workflow looks like when our newly-developed pipeline is incorporated. During primary analysis, patient samples are prepared and sequenced on a specialised platform such as Ion Torrent (used in our research case) or Illumina. These platforms perform the requisite signalcalling, basecalling, reference sequence alignment, and variant calling. The output from this primary analysis, for a single sample, is a Variant Call Format (VCF) file listing all the variants for that sample (*variants_patientX.vcf.gz* in Figure 5.1) (Danecek *et al.*, 2011). Secondary analysis, often with the identification of disease-causing variants in mind, entails variant annotation and filtering. Variant annotation is typically done using purpose-built tools such as ANNOVAR (Wang *et al.*, 2010) and VEP runner (used in our research case) that annotate variants with relevant metadata such as variant type, variant allele frequency, and

predicted impact. Variant filtering, where variants of interest are identified, can be done using tools like BrowseVCF, RD-Connect, and GEMINI (used in our case). These tools filter variants based on the metadata associated with each, which allows a researcher to find the variants most relevant to their investigation.

The pipeline presented in this paper focusses exclusively on secondary analysis, since the problematic potential population biases stemming from variant annotations made based on European-population-centric research results influence only the secondary analysis results. It utilises the offline VEP script's flexibility to annotate variants in a sufficiently comprehensive manner such that population-sensitive variant filtering can be achieved using GEMINI's high-specificity querying capabilities. Our pipeline is intended to form part of a larger NGS data analysis workflow, and is only responsible for variant annotation and filtering.

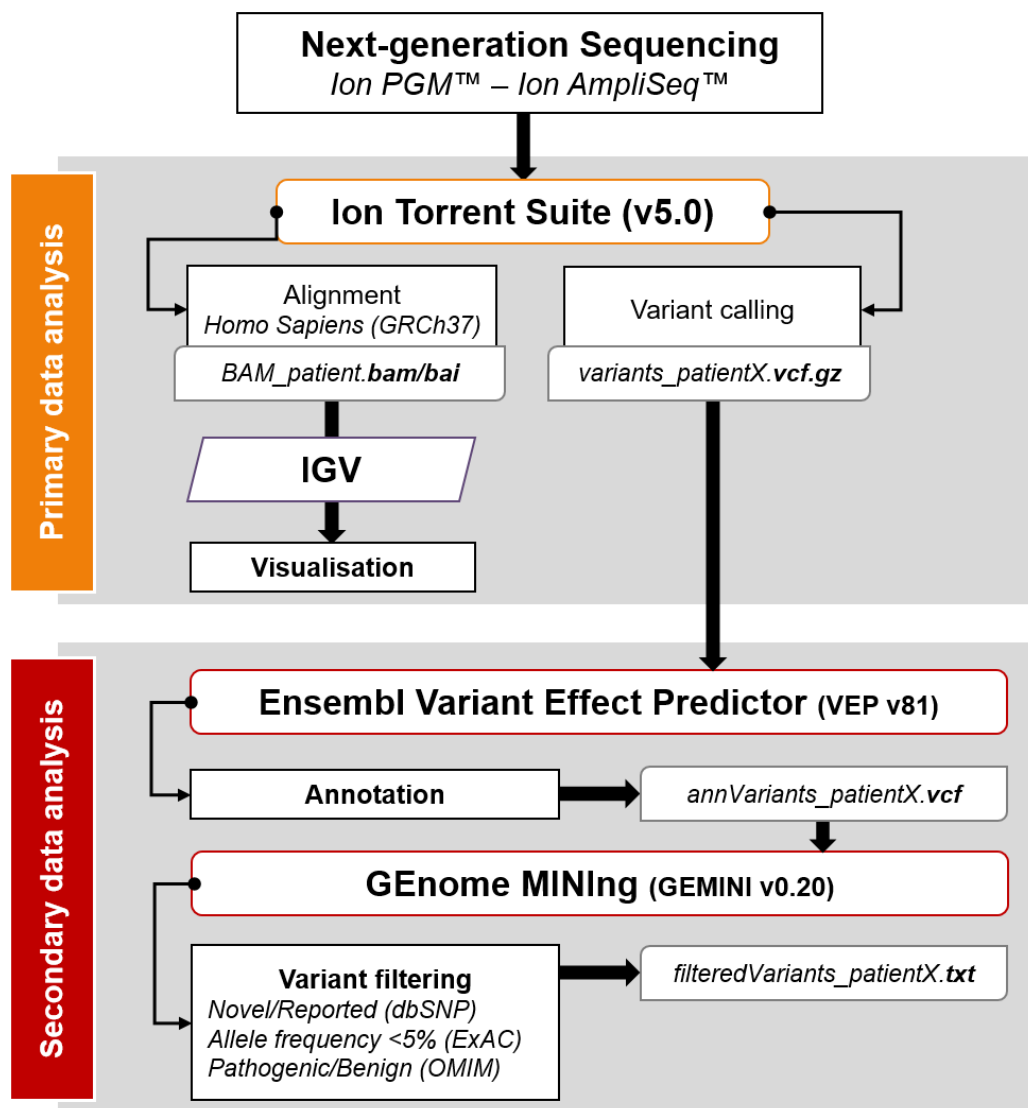


Figure 5.1: Simplified illustration for the in-house developed next-generation sequencing bioinformatics pipeline used for identifying disease-causing variants from Ion Torrent sequencing data. The first component of this illustration is primary data analysis, a semi-automated process done using the Ion Torrent Software (Torrent Suite). The second component illustrates software used for secondary data analysis. Abbreviations: PGM: Personal Genome Machine; BAM: binary alignment/map; IGV:

integrative genomics viewer; VCF: variant call format; dbSNP: database for single nucleotide polymorphisms; ExAC: exome aggregation consortium; OMIM: Online Mendelian Inheritance in Man.

5.3.1 Primary data analysis

The Ion PGM™ or Ion S5™ generates sequencing data via signal processing and base calling, which is then aligned to a reference genome and stored in Binary Alignment Map (BAM) format [*.bam* files; (Li *et al.*, 2009)], using Ion Torrent Suite (ITS). To enable visualisation of the sequencing data, BAM index (*.bam.bai*) files are also generated.

Several automated processes are performed by the ITS. These include a quality control (QC) check of the sequence runs, filtering out of poor quality reads, and trimming of the primer, barcode and adapter sequences. Reads with low quality signals due to mixed DNA templates on Ion Sphere Particles, are classified as poor quality reads. The reads, normally less than 250 bp read length, are assigned a Phred-scale mapping quality control score, which is defined as the error probability of reads containing no more than one error per 100 bases (Liu *et al.*, 2012). For this investigation, a score of 20 was chosen as it reflects 99% accuracy when aligned to the reference, which was GRCh37 in this case. Furthermore, the mapped reads and depth/coverage were calculated for each sequence run. The aligned BAM/BAI files can be visually inspected using integrative genomic viewer (IGV) to confirm sequencing integrity (Acevedo *et al.*, 2011).

Variant calling, i.e. identifying nucleotide differences when aligned to the reference genome, is done using the Torrent Variant Caller (TVC, version 5.0) plugin for ITS. The TVC takes the previously generated BAM file as input and compiles a list of the variants detected. As an output, it generates a Variant Call Format (VCF) file (*.vcf.gz*), as illustrated in Figure 5.2. Possible variant types identified include single nucleotide variations (SNVs), insertions and/or deletions (INDELs), and large structural variants. The VCF file has two parts, namely a header section and a data section. As indicated in Figure 5.2a–c, the header contains metadata about the sequencing run, VCF tag definitions, and contig information. On the other hand, each line in the data section details a single observed variant, providing, among others, its position, nucleotide change, and a quality score (Figure 5.2d) (Danecek *et al.*, 2011).

```

a  ##fileformat=VCFv4.1
   ##fileDate=20131017
   ##phasing=none
   ##reference=/results/referenceLibrary/tmap-f3/hg19/hg19.fasta
   ##reference=file:///results/referenceLibrary/tmap-f3/hg19/hg19.fasta
   ##source=Torrent Unified Variant Caller (Extension of freeBayes)
   ##INFO=<ID=OID,Number=.,Type=String,Description="List of original Hotspot IDs">
   ##INFO=<ID=OPOS,Number=.,Type=Integer,Description="List of original allele positions">
   ##INFO=<ID=OREF,Number=.,Type=String,Description="List of original reference bases">
   ##INFO=<ID=OALT,Number=.,Type=String,Description="List of original variant bases">
   ##INFO=<ID=AO,Number=A,Type=Integer,Description="Alternate allele observations">
   ##INFO=<ID=DP,Number=1,Type=Integer,Description="Total read depth at the locus">
   ##INFO=<ID=FAO,Number=A,Type=Integer,Description="Flow Evaluator Alternate allele observations">
   ##INFO=<ID=FDP,Number=1,Type=Integer,Description="Flow Evaluator read depth at the locus">
   ##INFO=<ID=FR,Number=1,Type=String,Description="Reason why the variant was filtered.">
   ##INFO=<ID=FRO,Number=1,Type=Integer,Description="Flow Evaluator Reference allele observations">
   ##INFO=<ID=FWDB,Number=A,Type=Float,Description="Forward strand bias in prediction.">
   ##INFO=<ID=FXF,Number=1,Type=Float,Description="Flow Evaluator failed read ratio">
   ##INFO=<ID=HS,Number=0,Type=Flag,Description="Indicate it is at a hot spot">
b  ##INFO=<ID=LEN,Number=A,Type=Integer,Description="allele length">
   ##INFO=<ID=MLLD,Number=A,Type=Float,Description="Mean log-likelihood delta per read.">
   ##INFO=<ID=NR,Number=1,Type=String,Description="Reason why the variant is a No-Call.">
   ##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of samples with data">
   ##INFO=<ID=QD,Number=1,Type=Float,Description="QualityByDepth as 4*QUAL/FDP (analogous to GATK)">
   ##INFO=<ID=RBI,Number=A,Type=Float,Description="Distance of bias parameters from zero.">
   ##INFO=<ID=REFB,Number=A,Type=Float,Description="Reference Hypothesis bias in prediction.">
   ##INFO=<ID=REVB,Number=A,Type=Float,Description="Reverse strand bias in prediction.">
   ##INFO=<ID=RO,Number=1,Type=Integer,Description="Reference allele observations">
   ##INFO=<ID=SAF,Number=A,Type=Integer,Description="Alternate allele observations on the forward strand">
   ##INFO=<ID=SAR,Number=A,Type=Integer,Description="Alternate allele observations on the reverse strand">
   ##INFO=<ID=SSEN,Number=A,Type=Float,Description="Strand-specific-error prediction on negative strand.">
   ##INFO=<ID=SSEP,Number=A,Type=Float,Description="Strand-specific-error prediction on positive strand.">
   ##INFO=<ID=SSSB,Number=A,Type=Float,Description="Strand-specific strand bias for allele.">
   ##INFO=<ID=STB,Number=A,Type=Float,Description="Strand bias in variant relative to reference.">
c  ##contig=<ID=chr1,length=249250621,assembly=hg19>
   ##contig=<ID=chr10,length=135534747,assembly=hg19>
   ##contig=<ID=chr11,length=135006516,assembly=hg19>
   ##contig=<ID=chr12,length=133851895,assembly=hg19>
d  #CHROM POS ID REF ALT QUAL FILTER INFO FORMAT None
   Chr1 39491927 . G C 1727.97 PASS
   Chr2 55861399 . T C 15.83 PASS
   Chr3 45533224 . G A 1733.84 PASS
   Chr3 45537973 . C T 764.32 PASS

```

Figure 5.2: Example of a variant call format (VCF) generated on Torrent Suite. a: Metadata of sequencing run; b: VCF tag definitions; c: Sequencing contig information; d: Details on the variant detected.

5.3.2 Secondary data analysis

Bash is the default command shell on most modern Unix-like systems (and Unix-like tools for Windows), and incorporates useful features from the Korn shell and the C shell (Ramey, 2011; Stevens & Rago, 2013). Bash users can write powerful scripts to automate analysis, which has made it a popular choice for research implementations, typically in the form of command line tools and analysis pipelines. Many of these command line tools are available as free software, giving researchers who use the bash shell access to a plethora of bioinformatics tools with which to do their research.

The main advantage of using bash is that it allows for the automation of established tools in a natural way, while minimising both the number of introduced software dependencies and the need for more advanced analysis infrastructure. These are important advantages considering the rare-disease focus of this pipeline. For researchers who focus exclusively on rare diseases whose

expertise and research interest often lie outside bioinformatics, a pipeline based on bash allows for flexible analyses that remain easily repeatable over long periods of time, while necessitating minimal skills development and infrastructure investment.

Here, we present the bioinformatics pipeline developed using these tools during our research. Scripts were run on Slackware v14.12 with GNU Bash, version 4.4.12.

Since primary analysis using the Ion Torrent system delivers sequencing results in batches, it seemed prudent to write the secondary analysis scripts to operate in a batch fashion. This, coupled with the built-in automation scripting provides, ensures consistency across the samples in each batch and across different batches.

In the secondary analysis pipeline, the scripts implement two steps: variant annotation and data mining. Variant annotation is done using the VEP script, which can be downloaded from Ensembl's website (McLaren *et al.*, 2016). Data mining is done using the command line tool GEMINI, which uses the SQLite relational database management system to enable effective sequencing data mining (Paila *et al.*, 2013). Each of these steps has a bash script dedicated to it: **vep_single.sh** and **gemini_single.sh**, respectively (the “.sh” extension indicates that these text files are bash scripts). These scripts are embedded into the **vep_batch.sh** and **gemini_batch.sh** scripts, respectively, which run them for each *.vcf.gz* file in a given directory. The **hetero_annotate.sh** script binds these two batch scripts into a full pipeline, calling each in sequence and managing their inputs and outputs. See Supplementary Files A–F (Appendix A.1) for the full contents of these scripts. A brief description of what each script does is given below.

The **vep_single.sh** script takes an Ion Torrent-generated input *.vcf.gz* file as its first argument and runs the VEP for it. It takes an output file as its second argument (a *.vcf* file), to which the annotated output of the VEP is written. Alongside these two arguments a number of additional arguments are passed to the VEP script, with the most notable being the *--fields [list]* argument, which allows for the specification of the required annotation fields included in the annotated output *.vcf* file. This argument (as well as the specification of the VEP arguments) is handled in an extensible way by the **vep_single.sh** script, and the list of desired fields can be built up over multiple lines or in multiple groups. A list of all possible fields and a list of all possible additional arguments can be found in the VEP's documentation (<https://www.ensembl.org/>) (McLaren *et al.*, 2016; The Ensembl genome database project - NCBI - NIH, 2018). The VEP also generates a “statistic run report” (*.html* file) when run, containing general statistics that give information on, among other things, the number of variants processed, the number of overlapping genes, and the number of novel/reported variants.

The **gemini_single.sh** script takes a VEP-annotated input *.vcf* file as its first argument and loads it into a SQLite database (a *.db* file). This file is then mined for relevant variants using user-defined

SQLite database query specifications. Each line in the **queries_spec.txt** auxiliary text file contains one such query specification, and consists of two comma-separated fields: the query's name and the relevant SQLite query snippet. Queries can easily be added to or removed from these files, or a different file can be specified in the **gemini_single.sh** script. An example of such a query, which filters based on variant's allele frequency in African populations, is: *rareAFR,aaf_1kg_afr <= 0.01*.

What information these queries should return for each variant is controlled by the *cols* variable, which is used to build the full SQLite query. Some examples of columns that can be included in a query are *is_coding* (which is true if the variant is in a coding region), *rs_ids* (which lists the rsIDs associated with each variant), and *aaf_1kg_afr* (which stores the allele frequency of the variant in the AFR population as reported in the 1000 Genomes Project). For a list of all possible columns, see GEMINI's documentation (Genome Mining, 2017.). This variable is handled similarly to the *--fields [list]* argument to the VEP, and is similarly extensible. The output of each query is a list of variants, with information from the database columns specified in the *cols* variable, stored in a *.txt* file that has the query's name as filename. These files constitute the main output of the pipeline.

Once all the queries have been performed, a *meta_info.txt* file is generated that summarises the lengths of the lists contained in the query output files. This file, along with the generated query outputs, is saved in the folder specified as the second argument to the **gemini_single.sh** script.

The **vep_batch.sh** and **gemini_batch.sh** scripts each run their counterpart script, as discussed above, for all the *.vcf.gz* files in a given directory. Both of these scripts take the directory containing the relevant input files as their first argument, with the second argument being the directory to which output should be written.

Finally, the **hetero_annotate.sh** script administers the operation of the two batch scripts described above. The first argument of this script is the directory containing the Ion Torrent output *.vcf.gz* files (the first argument for the **vep_batch.sh** and **vep_single.sh** scripts), and the second argument is the directory to which the GEMINI output should be written (the second argument to the **gemini_batch.sh** and **gemini_single.sh** scripts).

From here the user can further prioritise and filter the variants according to the criteria of their choice. For our African dataset, variants were first filtered based on their novelty. Secondly, variants were filtered based on African allele frequencies, as reported by Exome aggregation consortium [ExAC (Karczewski *et al.*, 2016)]. In addition, for mostly non-African population groups, known disease-causing variants can be identified from the dataset using databases such as the Online Mendelian Inheritance in Man [OMIM (Amberger *et al.*, 2014)] and ClinVar (Landrum *et al.*, 2013).

5.4 ADDITIONAL INFORMATION ON GEMINI

The text output files generated by GEMINI, as discussed in Section 5.3, can be directly imported into third party software such as RStudio or Microsoft Excel for viewing. The data is displayed in fields, in the software. Each of the fields display different types of information. Descriptions of fields displayed in the output text file generated by the scripts are listed in Table 5.1. A complete list of fields and their respective descriptions can be found at <https://gemini.readthedocs.io/en/latest/> (last accessed April 2018).

Table 5.1: Description of the fields that are present in the output file generated by GEMINI.

Field name	Description	Type
chrom	The chromosome on which the variant resides	STRING ¹
start	Start position of variant	INTERGER ²
end	End position of variant	INTERGER
ref	Reference allele	STRING
alt	Alternate allele for variant	STRING
gene	Corresponding gene name	STRING
exon	Information on exon	STRING
qual	Quality score for the assertion made in ALT	INTERGER
type	Type of variant, example, <i>snp</i> or <i>indel</i>	STRING
sub_type	Sub type of the variant, example, transition/transversion (ts/tv) or insertion or deletion (ins/del)	STRING
cytoband	Chromosomal cytobands that variant overlaps	STRING
num_hom_ref	The total number of homozygotes for the reference allele 0 = false, 1 = true	INTERGER
num_hom_alt	The total number of homozygotes for the alternate allele 0 = false, 1 = true	INTERGER
num_het	The total number of heterozygotes observed 0 = false, 1 = true	INTERGER
is_exonic	Does the variant affect an exon for ≥ 1 transcript? 0 = false, 1 = true	BOOL ³
is_coding	Does the variant fall within a coding regions 0 = false, 1 = true	BOOL
codon_change	To what does the codon change Ccc/Gcc: reference nucleotide changed from C to G	STRING
aa_change	To what does the amino acid change A/P means an alanine changed to proline	STRING
impact	What is the consequences for the affected transcript Examples: Missense_variant, synonymous_variant, frameshift_variant, inframe_deletion, inframe_insertion, stop_gained, stop_lost, stop_retained	STRING
impact_severity	Severity of the highest order observed for the variant Examples: LOW, MED, HIGH	STRING

cadd_raw	Raw CADD scores for scoring deleteriousness of SNV's in the human genome	FLOAT ⁴
cadd_scaled	Scaled CADD scores (Phred-like) for scoring deleteriousness of SNV's 0–10: less likely to be disease-causing, 10–20: more likely to be disease-causing 20–30+: highly likely to be disease-causing	FLOAT
is_lof	Based on the value of the impact field, does the amino acid change of variant have a loss-of-function effect 0 = false, 1 = true	BOOL
is_conserved	Does the variant overlap a conserved region? Based on the 29-way mammalian conservation study 0 = false, 1 = true	BOOL
depth	The number of aligned sequence reads that led to this variant call	INTERGER
qual_depth	Variant confidence or quality by depth 0 = low confidence, 50+ = high confidence	FLOAT
rs_id	A comma-separated list of refsequence id's for variants present in dbSNP Example: rs2270004	STRING
in_omim	Is the variant reported in OMIM: 0 = false and absent, 1 = true and present	BOOL
clinvar_sig	The clinical significance scores for each of the variant according to ClinVar: unknown, untested, non-pathogenic, probable-non-pathogenic, probable-pathogenic, pathogenic, drug-response, histocompatibility, other	STRING
clinvar_origin	The type of variant: unknown, germline, somatic, inherited, paternal, maternal, de-novo, biparental, uniparental, non-tested, tested-inconclusive, other	STRING
clinvar_disease_name	The name of the disease to which the variant is relevant	STRING
clinvar_gene_phenotype	' ' delimited list of phenotypes associated with this gene (includes any variant in the same gene in clinvar not just the current variant).	STRING
poly_pred	Polyphen predictions for the SNPs Benign, possibly_damaging, probably_damaging	STRING
poly_score	Polyphen scores for the SNPs prediction 0–0.4 = benign, >0.4–0.9 = possibly_damaging, >0.9–1.0 = probably_damaging	FLOAT
sift_pred	SIFT predictions for the SNPs Tolerated and deleterious	STRING
sift_score	SIFT scores for the predictions 0–0.05 = deleterious, >0.05–1.0 = tolerated	FLOAT
pfam_domain	Pfam (protein families) protein domain that the variant affects	STRING
in_hm3	Whether the variant was part of HapMap3	BOOL
in_esp	Presence/absence of the variant in the ESP project data	BOOL
in_1kg	Presence/absence of the variant in the 1000 genome project data (phase 3)	BOOL
aaf	The observed allele frequency for the alternate allele	FLOAT
aaf_1kg_all	Global allele frequency (based on AC/AN) (1000g project; phase 3)	FLOAT

aaf_1kg_eur	Allele frequency of the variant in European population based on AC/AN (1000g project; phase 3)	FLOAT
aaf_1kg_afr	Allele frequency of the variant in African population based on AC/AN (1000g project; phase 3)	FLOAT
gts	A compressed binary vector of sample genotypes A/A (homozygous) or A/G (heterozygous)	BLOB ⁵
gts_phases	A compressed binary vector of sample genotype phases A/A means unphased genotype and will display FALSE, A/G means phased genotype and will display TRUE	BLOB

¹ A string is any finite sequence of characters (i.e., letters, numerals, symbols and punctuation marks).

² An integer is a whole number (not a fraction) that can be positive, negative, or zero.

³ Boolean data type is a data type that has one of two possible values (usually denoted true and false).

⁴ A Float is a variable with a fractional value.

⁵ BLOBs are used to store programs or even fragments of code.

5.5 MITOCHONDRIAL DNA ANALYSIS

Analysis of mtDNA sequencing data was done using a previously developed bioinformatics pipeline described by Weissensteiner *et al.* (2016). The bioinformatics pipeline, called mtDNA-server, is fully automated and is freely accessible at <https://mtdna-server.uibk.ac.at> (last accessed March 2018). A simplified workflow for analysis of mtDNA is illustrated in Figure 5.3.

Next-generation sequencing raw BAM files generated by Ion PGM™ and Ion S5™ during this investigation was used as input files for the mtDNA-server. Several automatic steps are performed by the mtDNA-server that includes firstly a validation step for input data followed by alignment against the revised Cambridge Reference Sequence [rCRS, NC_012920.1 (Andrews *et al.*, 1999)], secondly a quality control step which gives feedback on the quality of the mapped reads, and lastly homoplasmic and heteroplasmic variants are detected. Reads are filtered with a mapping quality score of <20 (Phred-score) and reads below 25 bp. Several filters were applied for heteroplasmic detection. Mitochondrial hotspot around positions 309, 315 and 3107 were excluded and reads with coverage of 10 or less bases per strand are filtered. If a variant allele frequency of >1% was detected with an allele coverage of three bases per strand, the maximum likelihood heteroplasmy model was applied. Output files are generated and include the annotated variant files and HaploGrep files for identifying the haplogroup for each sample. A contamination check is applied to the previously generated files to prevent incorrect interpretation of variants, as found in a QC report. For a full description of this pipeline, refer to Weissensteiner *et al.* (2016).

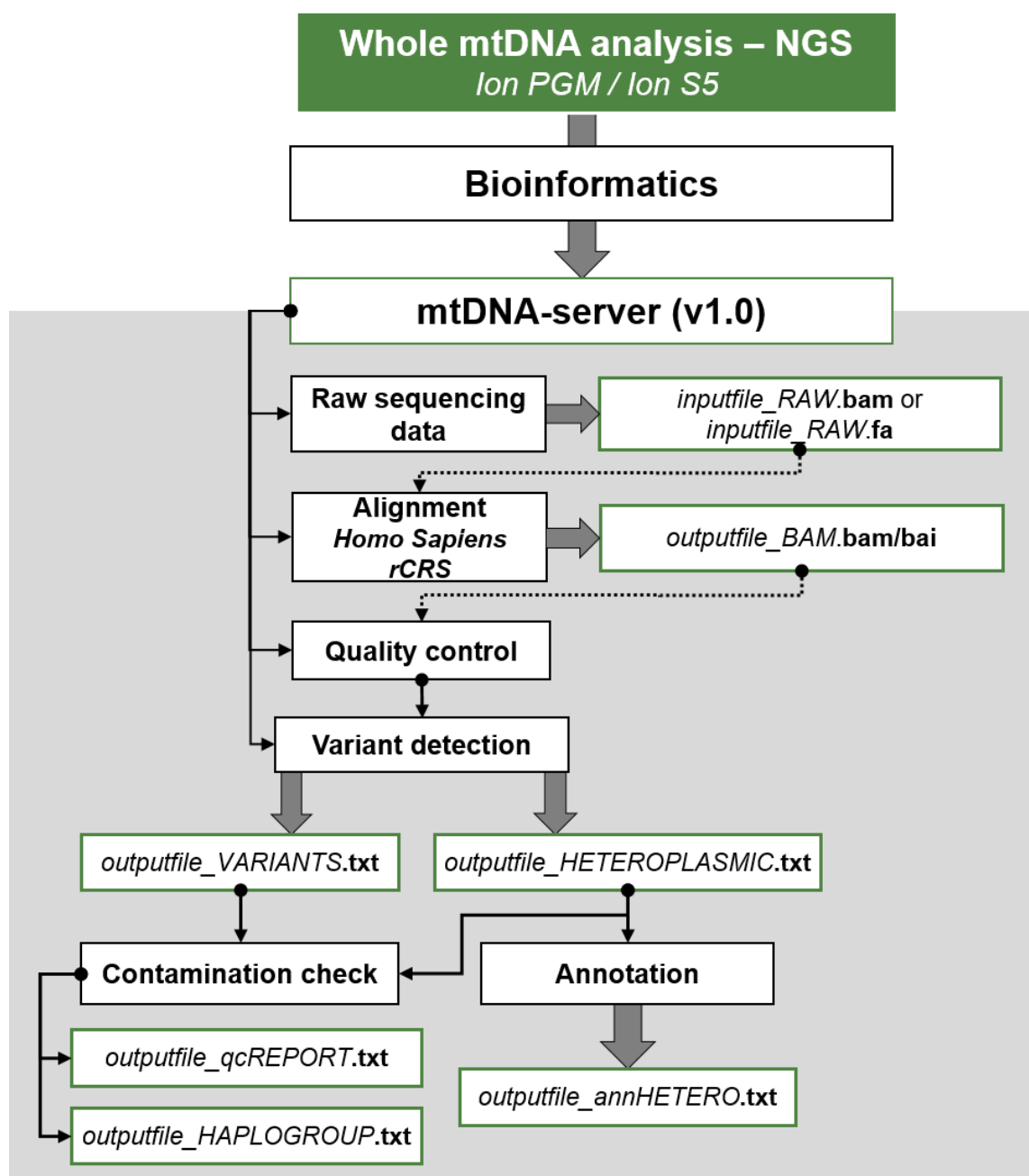


Figure 5.3: Summary of mtDNA bioinformatics pipeline. Next-generation sequencing bioinformatics pipeline. This approach was followed for identifying disease-causing variants from Ion Torrent sequencing data using mtDNA-server.

5.6 CRITERIA FOR IDENTIFYING AND EVALUATING NUCLEAR VARIANTS

Variants of importance were further evaluated using a set of standards and guidelines set forth by Richards *et al.* (2015). Variants were considered as pathogenic, likely-pathogenic, benign, likely-benign, or variant with uncertain significance. These classifications are formulated from combined criteria, and are briefly explained in the following paragraphs from Sections 5.6.1–5.6.3.

5.6.1 High confidence pathogenic variants

- Different types of variants will have different outcomes on a protein level. Higher impacting variants such as **nonsense, frameshift canonical splice site variants, deletions or insertion** disrupt gene functioning and often cause a complete or partial loss of function (LoF) of a gene. These high-impacting variants weigh towards pathogenic with strong evidence.
- Missense variants with a **known disease-causing mechanism** are also considered with high confidence to be pathogenic. Often a different nucleotide change that results in the same amino acid change at the same position of the known disease-causing mutation can also be assumed pathogenic.
- **De novo variants** should be handled with more care. A variant can only be classified as pathogenic if the patient has a family history with a known *de novo* inheritance pattern and the phenotype for that patient correlates with reported cases.
- **Functional test** should be performed on all novel or reported variants which are suspected to be disease-causing. If the functional tests confirm impact on RNA level, it can be extremely informative and will weigh more towards a high confidence of pathogenicity for that mutation. Certain functional tests can however be ineffective, if the full biological function of the mutation is not taken into account.
- **Population frequency** is of the essence when evaluating a gene variant. Several publically available population databases (e.g. 1000 Genomes Project, Exome Aggregation Consortium, Genome Aggregation Database) are used to assess the variants population allele frequency. It is expected that in the general population (control) an allele frequency for a variant is far greater than that expected for the frequency of a disease-causing variant. In other words, a variant with a high allele frequency (>5%) weighs more towards high confidence to be benign. Variants that are absent from these databases are considered with high confidence to be pathogenic, provided substantial evidence for a high read depth exists.

5.6.2 Moderately pathogenic variants

- **Mutations known to be in so-called “hot spots”** or in a well-established domain which causes serious non-functioning proteins, provided that there are no benign variants in these regions, are considered to be moderate evidence of pathogenicity.
- Novel missense variants which occur at the same position as a **well-known pathogenic** variant, but with different amino acid changes, can be considered to be moderate evidence but cannot be assumed to be pathogenic. Different phenotypes can, however, be associated with different amino acid changes on the same position and should be taken into account.

- **Segregation analysis** is a key component in evaluating variants, especially novel variants that are compound heterozygous. Moderate confidence of pathogenicity is considered if family members also harbour a specific variant mutation found in the patient. It is important not to assume that the variant(s) are pathogenic, as functional analysis should be done on these variants to confirm this. Segregation analysis that reveals no mutation with the specific phenotype in the parents provides strong evidence against pathogenicity.
- Variants that cause **changes in the protein length** (e.g. insertions or deletions), or disruption of protein function (e.g. stop gain or stop loss variants), are considered to be moderate evidence of pathogenicity. Larger highly conserved regions that are affected by insertion or deletions are more likely to be pathogenic than smaller repetitive regions in less well-conserved regions.
- **Computational algorithms** are used as supporting evidence to validate possible disease-causing variants as different algorithms might have different outcomes. The algorithms used include Sorting Intolerant from Tolerant (SIFT), Polyphen-2 (Adzhubei *et al.*, 2010), Combined Annotation Dependent Depletion [CADD, (Kircher *et al.*, 2014)] and MutationTaster (Schwarz *et al.*, 2014).

5.6.3 High confidence benign variants

- A synonymous variant detected in a patient is considered to be benign with high confidence, as the amino acid residues do not change. This should, however, not be assumed and should be handled with caution, especially if the variant is a splice variant that can disrupt the consensus sequence in highly conserved regions. When evaluating synonymous variants, it is suggested that computational algorithms be used to help with pathogenicity predictions, as these tools can provide more clarity as to the specific variant of interest.

The criteria explained in the previous sections (Section 5.6) are organized in Table 5.2, whereby the strength of evidence for pathogenic variants and benign variants is visually presented. Nuclear variants are throughout this thesis classified as pathogenic, likely pathogenic, likely benign, benign, variants of uncertain significance (VUS).

Table 5.2: Summary of ACMG guidelines and criteria used for evaluating variants of importance, as adapted from Richards et al. (2015b).

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
Population data	MAF is too high for disorder (>5%)			Absent in population databases	Prevalence in affected case statistically increased over controls	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene/gene product	Multiple lines of computational evidence support a deleterious effect on the gene /gene product	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before. Protein length changing variant	Same amino acid change as an established pathogenic variant	Predicted null variant in a gene where LoF is a known mechanism of disease
Functional data	Well-established functional studies show no deleterious effect		Missense in gene with low rate of benign missense variants	Mutational hot spot or well-studied functional domain without benign variation	Well-established functional studies show a deleterious effect	
Segregation data	Non-segregation with disease		Co-segregation with disease in multiple affected family members			
De Novo data		Observed in cis with a pathogenic variant		De novo (without paternity & maternity confirmed)	De novo (paternity and maternity confirmed)	
Allelic data				For recessive disorders, detected in trans with a pathogenic variant		

5.7 CRITERIA FOR IDENTIFYING AND EVALUATING MITOCHONDRIAL DISEASE-CAUSING VARIANTS

The evaluation of mtDNA variants is in some ways comparable to that of nDNA variants, but as mtDNA has unique genetic features, identifying pathogenic variants in mtDNA should be considered separately from that of nDNA. Several factors and considerations are taken into account for mtDNA variants, which are summarised here (DiMauro & Schon, 2001; Montoya *et al.*, 2009).

- As is the case with nuclear variants, population allele frequency is one of the most important aspects for defining a variant pathogenic. A variant can thus not be present in a patient as well as control groups. A high allele frequency weighs more towards high confidence to be benign. The variants of interest or mutation must be present in the varied mitochondrial genetic background.
- The heteroplasmy and homoplasmy levels should also be reported. Lower heteroplasmy levels should be interpreted according to the sample or tissue on which functional tests were performed.
- A mutation is considered to be of high confidence of pathogenicity if it affects a functionally important domain, especially in highly evolutionary conserved regions. It is important to confirm these effects with single fibre PCR analysis.
- Synonymous variants that do not cause alterations on protein level weigh more towards high confidence to be benign, i.e. a variant must not be a recognised non-pathogenic SNP.
- The variant must be the best mtDNA candidate to be pathogenic.

Secondary structures of tRNA molecules must be considered when evaluating the pathogenicity of detected MT-tRNA variants. A scoring system was designed by McFarland *et al.* (2004a) for evaluation of tRNA variants. It was further refined by Mitchell *et al.* (2006) and Yarham *et al.* (2011):

- Mutations must be heteroplasmic.
- Mutated mtDNA proportion should be higher in affected individuals than in control or unaffected family members.
- Mutated mtDNA proportion should be higher in affected tissue.
- Mutated mtDNA should segregate with a biochemical defect.
- Mutations in highly conserved regions or site in the mitochondrial genome weighs more towards pathogenic.
- Mutations should not be present in control samples, i.e. have a very low allele frequency.

Due to the complexity of MD and mutations involved, there are exceptions to the criteria as listed above. For example, heteroplasmy levels should exceed a certain threshold for the mutation or detected variants to be considered to be disease-causing. The threshold levels for mtDNA deletions are usually 60% and the point mutation threshold is 90% (Rossignol *et al.*, 2003; Shoffner & disease, 1995; Shoffner *et al.*, 1990). It is also important to note that different phenotypic expressions and different affected tissue have different threshold levels (Rossignol *et al.*, 2003).

5.8 CONCLUSION ON BIOINFORMATICS IN SOUTH AFRICA

Robust bioinformatics pipelines are key components of the diagnosis and research of rare genetic diseases such as MD (Graham *et al.*, 2018). Here we describe an offline, flexible and open-source bioinformatics pipeline that annotates variants using VEP and filtering of important disease-causing variants using GEMINI from NGS data. It was developed as a first step towards data processing and offers unique advantages for the detection of multiple genetic alterations, including patients of African descent, for which little information is available on genetic databases. The pipeline can be used on exome as well as amplicon NGS data and was designed using NGS data of a predominantly African cohort with rare disease. Identifying the underlying genetic causes for rare disease is particularly challenging in the South African population, with limited bioinformatics support for researchers and non-bioinformaticians. With the increased demand for diagnosing rare genetic diseases using NGS and genetic screening, and the limited support and resources in developing countries such as SA, it is equally important to develop and provide access to bioinformatics pipelines such as this one. With continued development, pipelines in SA could be further refined and made more user-friendly, making them useful for both researchers and clinicians. These refinements could include, for instance, cloud-based interfaces such as RD-Connect, Sophia Genetics, MitoMap and MitoMaster. With such refinements in place, clinicians would more easily be able to investigate genotype-phenotype correlations in rare diseases such as MD, for African patient populations.

CHAPTER 6:

PATHOGENICITY EVALUATION OF DETECTED

mtDNA AND nDNA VARIANTS

The paediatric cohort for this study is extensively described in Chapter 3. As discussed, stratified clinical evaluations were performed on patients with mitochondrial disease features. Subsequently, mitochondrial RC enzyme analysis was performed on collected muscle tissue samples from the entire cohort of 212. Using the clinical information and biochemical profiles, patients were selected for NGS. An in-house bioinformatics pipeline and a previously reported mtDNA pipeline, as discussed in Chapter 5, were used to filter, mine, and sort through nuclear and mitochondrial NGS data. An overview of the variants identified using these pipelines is presented here, including targeted panel-based, WES, and whole mtDNA NGS data. The candidate variants of interest are also described, and the pathogenicity of each variant is carefully evaluated.

Part of this chapter has been published in a peer-reviewed journal, and addresses Objective 5 (see Appendix A.2):

Schoonen, M., Smuts, I., Louw, R., Elson, J.L., Van Dyk, E., Jonck, L., Rodenburg, R.J.T., Van der Westhuizen, F.H. (2018). Panel-based nuclear and mitochondrial next-generation sequencing outcomes of an ethnically diverse paediatric patient cohort with mitochondrial disease. Accepted for publication in the *Journal for Molecular Diagnostics* (manuscript number: JMDI_2017_395).

6.1 OVERVIEW OF VARIANTS DETECTED

As discussed in Chapter 4, three panels were used as a first step towards identifying disease-causing variants in this cohort. The overarching genetic outcome for Panels 1 and 2 is discussed here. Panel 3 has already been discussed in full by Jonck (2014). The overarching genetic outcome for exome sequencing for a subset of selected African cases is discussed here.

6.1.1 Panel sequencing

6.1.1.1 Panel 1: Agilent HaloPlex — complex I associated genes

In total, 32 patients were sequenced using Panel 1. On average, approximately 740 variants were detected per patient. African patients averaged 783 variants, which is 200 more variants than for their non-African counterparts, for which approximately 677 variants were detected on average (Figure 6.1). The variants detected are classified as either novel or reported variants, of which the reported variants are further classified as either pathogenic, likely pathogenic, benign, likely benign, or VUS under the ACMG guidelines, as described in Section 5.6. The number of novel and reported variants is shown in Figure 6.2, indicating approximately 120 novel variants and approximately 610 reported variants that were detected on average. The number of novel and reported variants detected in Africans averaged at approximately 130 and 650 respectively, and that in non-Africans averaged at approximately 120 and 550, respectively.

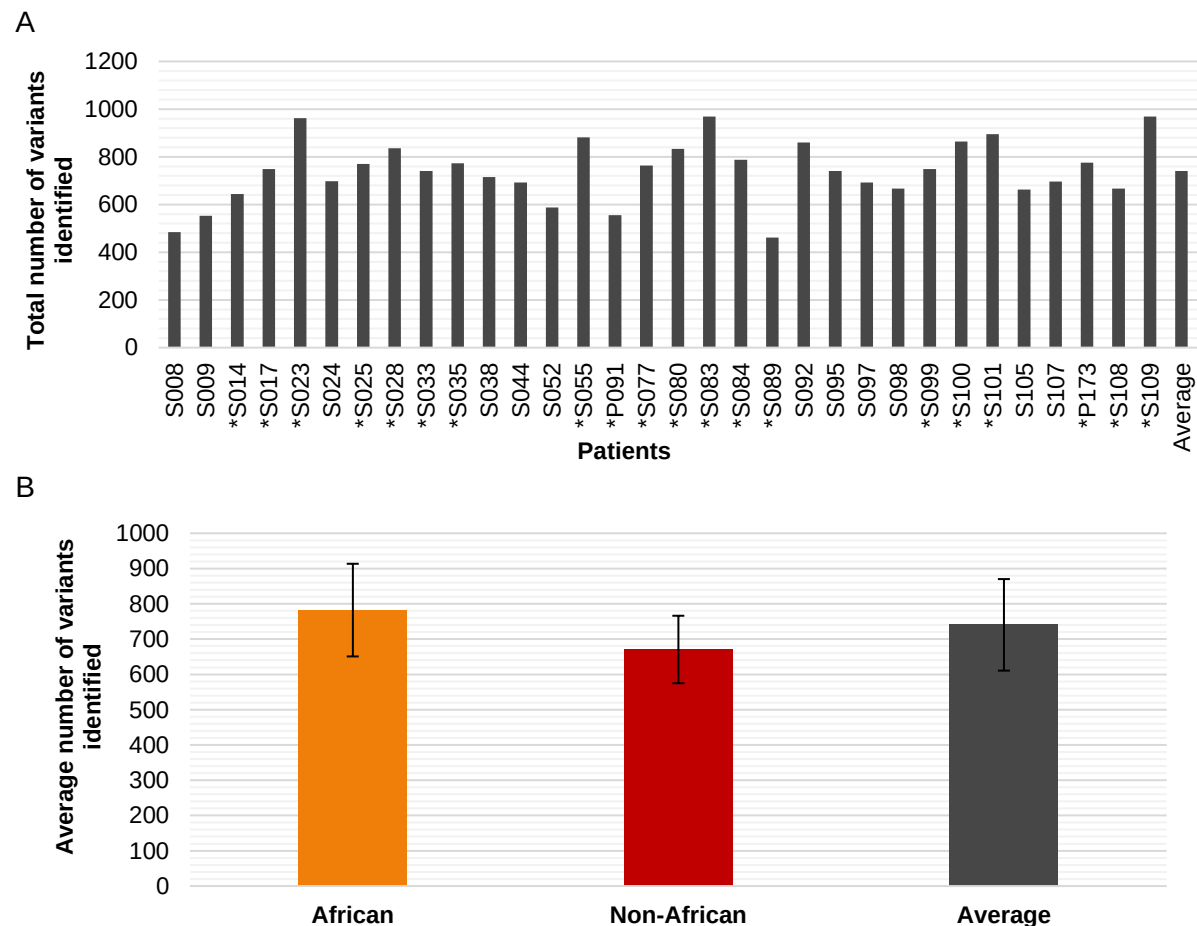
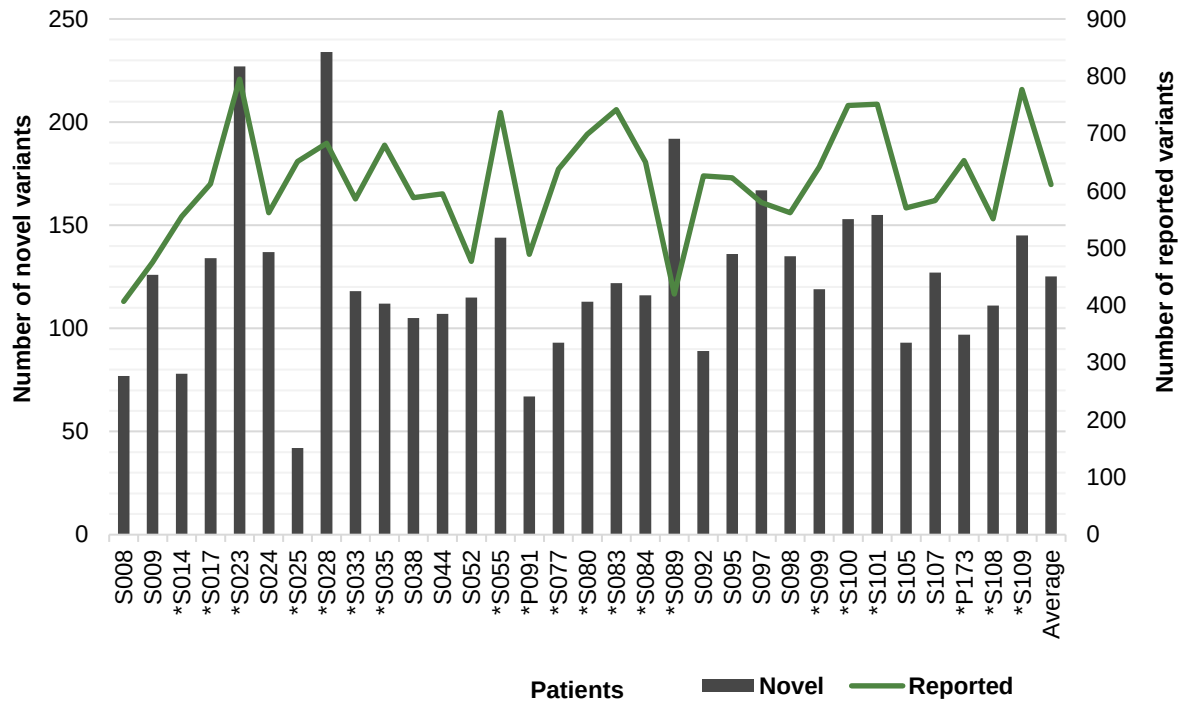


Figure 6.1: Overview of Panel 1 NGS findings for 32 selected patients. Panel A: The total number of variants detected in each of the 32 patients. Africans are marked with an asterisk. Panel B: The average number of variants detected for Africans and non-Africans. The black bar represents the average number of variants detected for both African and non-African patients. The orange bar indicates African patients, and the red bar indicates non-Africans.

A



B

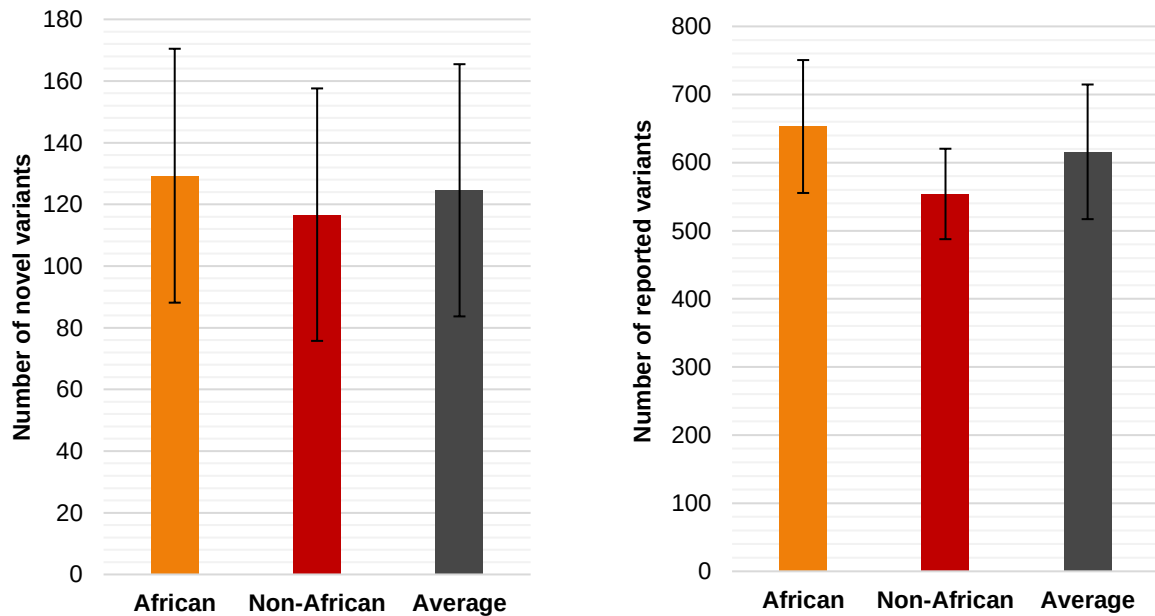


Figure 6.2: Novel and reported NGS variants detected in 32 patients. Panel A: The total number of novel (left vertical axis) and previously reported (right vertical axis) variants detected in each patient. Africans are marked with an asterisk. Panel B: Average number of novel (left) and reported (right) variants detected in African and non-African patients, respectively. The black bars represent the average number of variants detected for both African and non-African patients. The orange bar indicates African patients, and the red bar indicates non-Africans.

6.1.1.2 Panel 2: Ion Ampliseq — complex I–IV associated genes

In total, 48 patients were sequenced using Panel 2. The total number of variants identified in each patient is shown in Figure 6.3A. Approximately 255 variants were detected per patient on average (Figure 6.3B). For African patients approximately 276 variants and for non-Africans approximately 196 variants (Figure 6.3B). were detected on average. Just as it was the case for Panel 1, these detected variants were classified as novel or reported variants, where the latter were evaluated and further classified as either pathogenic, likely pathogenic, benign, likely benign, or VUS. For Africans, approximately 14 novel and 260 reported variants were detected and for non-Africans approximately 11 novel and 240 reported variants were detected on average (Figure 6.4A and Figure 6.4B).

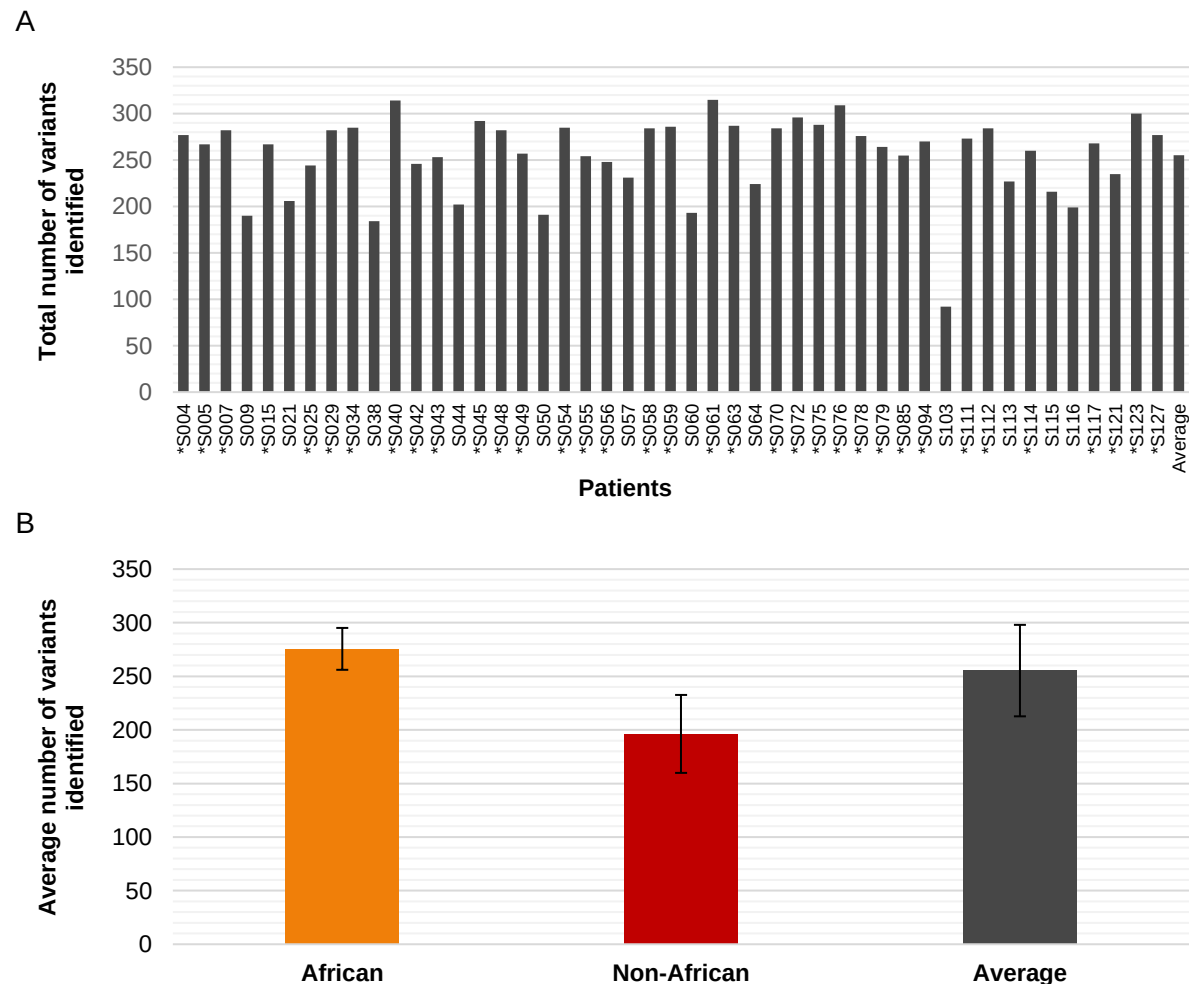


Figure 6.3: Overview of Panel 2 NGS findings for 48 selected patients. Panel A: The total number of variants detected in each of the 48 patients. Africans are marked with an asterisk. Panel B: The average number of variants detected for Africans and non-Africans. The orange bar indicates African patients while the red bar indicates non-Africans. The black bar represents the average number of variants detected for both African and non-African patients.

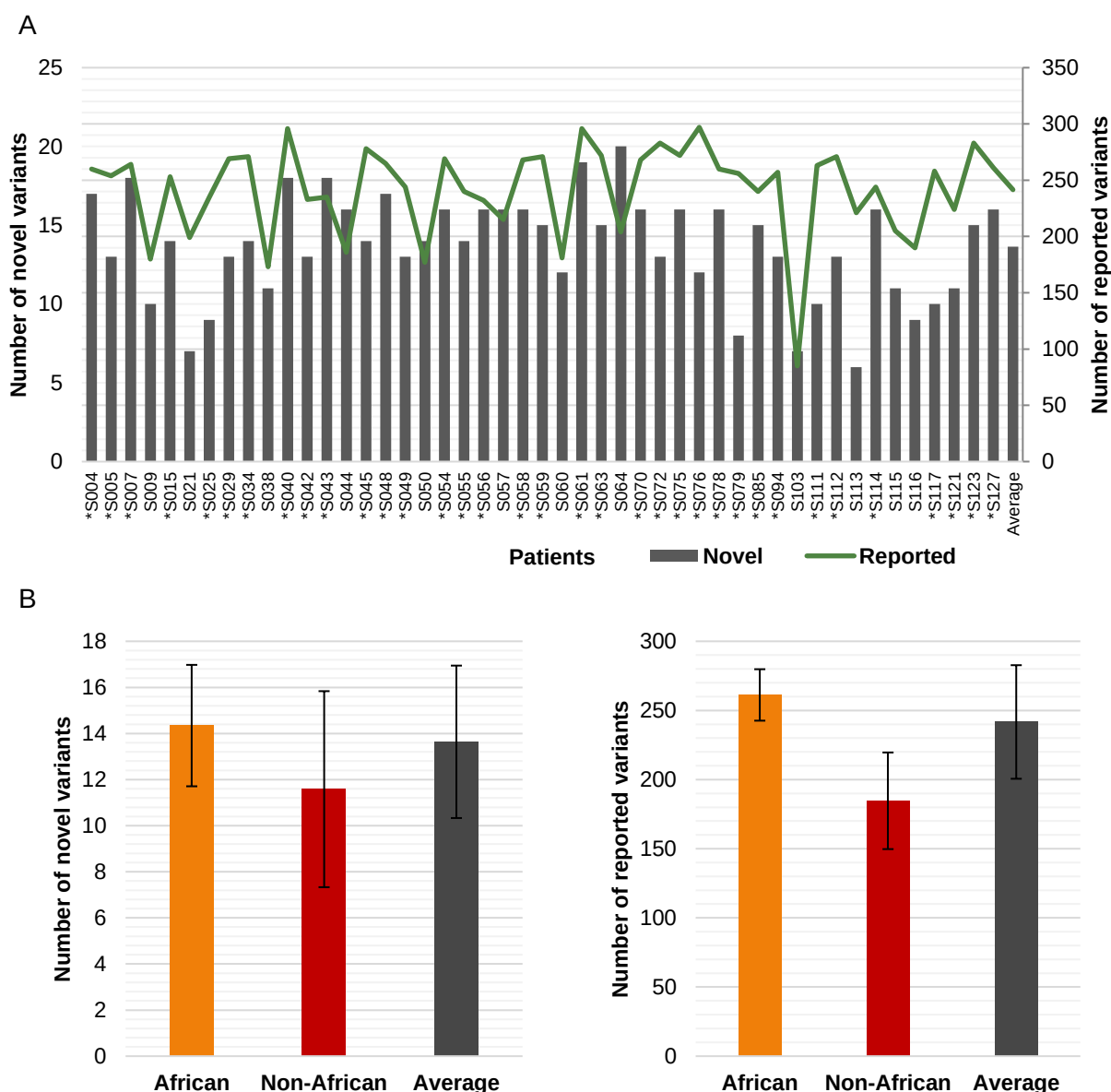


Figure 6.4: Novel and reported NGS variants detected in 48 patients. Panel A: The total number of novel (left vertical axis) and previously reported (right vertical axis) variants detected in each patient. Africans are marked with an asterisk. Panel B: Average number of novel (left) and reported (right) variants detected in African and non-African patients, respectively. The black bars represent the average number of variants detected for both African and non-African patients, the orange bars indicate African patients, and the red bars indicate non-Africans.

6.1.2 Whole exome sequencing

As discussed in Chapter 4, WES was performed on eight randomly selected African cases for which no disease-causing or pathogenic variants could be identified through panel or whole mtDNA sequencing. The total number of variants identified ranged from 46 500–48 500, as seen in Figure 6.5A. The variants detected in these patients are classified into novel or reported

variants, while the latter is subsequently classified as either pathogenic, likely pathogenic, benign, likely benign, or VUS. The number of novel and reported variants is shown in Figure 6.5B.



Figure 6.5: Overview of WES findings for the selected eight African patients. Panel A: Total number of variants detected in each of the eight cases. Panel B: The total number of novel (left vertical axis) and previously reported (right vertical axis) variants detected in each patient.

The variants of interest detected through Panels 1–3 and WES, i.e. novel variants and those previously reported in disease databases (with or without a known disease-causing mechanism) were further investigated and evaluated using the ACMG guidelines set forth by Richards *et al.* (2015). As the majority of MDs are inherited in autosomal recessive patterns, only homozygous variants were investigated. Variants with heterozygous allele calls were not extensively investigated, as these variants are typically not disease-causing and are considered carrier alleles in these patients. Where detected, compound heterozygous variants were, however, evaluated.

6.1.3 Mitochondrial DNA sequencing

The complete mitochondrial genome was sequenced in 123 of the 127 patients in the cohort. Limited sample availability prevented sequencing of four samples: S007, S008, S078, and S107. Whole mtDNA sequencing was successfully applied in 123 cases (79 African and 44 non-African), and a variant of interest was identified in 12 African cases, none of which could be classified as disease-causing. The mtDNA sequencing data for a sub-section of this cohort that presented with a previously reported disease-causing variant ($n = 12$) have been extensively investigated by Van der Walt *et al.* (2012). These variants are listed in Table 6.1.

The majority of the variants identified could not be classified as pathogenic as they fail to meet a number of criteria specific for mtDNA pathogenicity, as discussed in Section 5.7. Notably, a well-known LHON-associated mutation (m.14484T>C at 53% heteroplasmy in muscle) was identified in an African male (S014). Although this patient had clinical features of eye involvement, his clinical phenotype did not match that expected for LHON (such as visual failure and optic atrophy). This inconsistency could be ascribed to varied penetrance of the disease, highlighting the importance of investigating the penetrance of diseases associated with this and other pathogenic variants when detected in understudied populations. An additional 11 mtDNA variants of interest, identified in 10 cases (as listed in Table 6.1) may, however, still prove to have a functional effect on disease initiation or progression. This may still be the case for variants with a comparatively high population allele frequency (above 0.1%) as further investigation may indeed reveal that these variants appear at a higher frequency in these understudied population lineages.

Table 6.1: Summary of mitochondrial DNA variants of importance identified in cohort.

Gene	Variant position ^a	Amino acid change	Mutpred score	ACMG pathogenicity classification ^b	Population frequency, allele frequency ^c	Patient, race, gender	RC deficiency	References ^d
MT-ND6	m.14484T>C	p.Met64Val	-	Likely benign	-	S014, A, M	CI	Brown <i>et al.</i> (1992)
MT-ND2	m.5186A>T	p. Trp239Cys	0.549	VUS	0.0653%, 100%	S097, NA, F	CI	<i>None to date</i>
MT-TA	m.5628T>C	Ala	Y-score: 13	Likely benign	0.15692%, 100%	S059, A, M	CI, CIII, CIV, CII+CIII	Gamba <i>et al.</i> (2012)
MT-TC	m.5814T>C [‡]	Cys	Y-score: 14	VUS	0.33345%, 100%	S058, A, M	CI, CIII, CIV	Santorelli <i>et al.</i> (1997a)
MT-CO3	m.9355A>G	p.Asn50Ser	0.591	VUS	0.06%, 100%	S043, A, F	CIII, CIV	<i>None to date</i>
MT-ND4	m.11087T>C [‡]	p.Phe110Leu	0.788	VUS	0.24%, 100%	S028, A, M	CI, CII+CIII	Gaspar <i>et al.</i> (2015)
MT-ND5	m.13802C>T	p.Thr489Met	0.6	VUS	0.07192%, 100%	S019, NA, F	CI, CII+CIII	Bayona-Bafaluy <i>et al.</i> (2011)
MT-ND6	m.14405A>G	p.Val90Ala	0.597	VUS	0.055%, 100%	S112, A, M	CI, CII, CIV	<i>None to date</i>
MT-ND6	m.14502T>C	p.Ile58Val	0.385	Likely benign	0.39%, 82%	S126, NA, M	CI	Zhao <i>et al.</i> (2009)
MT-CYB	m.14790A>G	p.Asn15Ser	0.704	VUS	0.02288%, 100%	S028, A, M	CI, CII+CIII	Cavadas <i>et al.</i> (2015)
MT-CYB	m.15257G>A	p.Asp171Asn	0.785	VUS	1.30%, 100%	S038NA, F	CI, CIII	Brown <i>et al.</i> (1992)
MT-CYB	m.15272A>G	p.Thr176Ala	0.562	VUS	0.00654%, 100%	S037, NA, F	CIII	Van der Walt <i>et al.</i> (2012)
MT-CYB	m.15735C>T	p.Ala330Val	0.416	Likely benign	0.127497%, 100%	S054, A, F	CII, CIII	Bannwarth <i>et al.</i> (2008)
						S003, A, F	CII+CIII	
					0.127497%, 43%	S048, A, M	CIII, CIV, CII+CIII	
					0.127497%, 100%	S049, A, M	CIII	

^a Variants are listed relative to the rCRS reference sequence (GenBank NC_012920.1).

^b Pathogenicity classification are made for this cohort's patients based on the relative information available.

^c Population frequencies are listed according to the GenBank population database, and allele frequencies are detected in patient samples.

^d References to scientific papers where the variant was first reported on.

Abbreviations: VUS: variant of uncertain significance; A: African; NA: non-African; F: female; M: male; CI–CIV: complexes I–IV.

6.2 NUCLEAR VARIANTS OF IMPORTANCE

A number of variants were considered to be possibly pathogenic or disease-causing in our patients. The findings for panel nDNA and whole exome NGS are summarised in Table 6.2.

Panel nDNA sequencing of 86 cases (61 African and 25 non-African) revealed four variants of interest in two cases. Two pathogenic compound heterozygous *ETFDH* variants were identified in one case (S057), and two possible pathogenic compound heterozygous *SURF1* variants were identified in a second case (S085). WES sequencing of eight African cases revealed nine variants of interest (in six cases), eight of which are considered to be possibly pathogenic (variants in the genes *COQ6*, *RYR1*, *STAC3*, and *ALAS2*) and one in the gene *TRIOBP* as a VUS under the ACMG classification.

Table 6.2: Summary of nuclear DNA variants of interest identified in the cohort.

Gene	Variant position ^a	Amino acid change	refSNP ID	ACMG pathogenicity classification ^b	ExAC frequency ^c , population	Patient, race, gender	RC deficiency	References ^d
DARS2 ¹	g.173808623C>CT,T (+/-)	NA	Novel	VUS	None	S014, A, M	CI	None to date
						S024, NA, M	CI	
						S092, NA, M	CI	
	c.587A>G (+/-)	p.Lys196Arg	rs35515638	Likely benign	0.08256, A	S089, A, M	CI	None to date
				Likely benign	0.0001349, NA	S101, A, F	CI	
	c.1013G>A (+/-)	p.Gly338Glu	rs141298312	Likely benign	0.006536, A	S024, NA, M	CI	None to date
				Likely benign	0.03733, NA	S044, NA, M	CI, CII, CIII, CIV, CII+CIII	
POLG ¹	g.89870461C>T (+/-)	R>Q	Novel	None	None	S023, A, M	CI, CIII, CIV, CII+CIII	None to date
	g.89876634C>T (+/-)	V>M	Novel	None	None			
	c.1636C>T	p.Arg546Cys	rs2307447	Benign	0.01799, A 0, NA	S014, A, M	CI	Cohen <i>et al.</i> (2014)
	c.3708G>T	p.Gln1236His	rs3087374	Benign	0.01377, A 0.0827, NA	S024, NA, M	CI	
						S097, NA, F	CI	
	c.3287G>A (+/-)	p.Arg1096His	rs368435864	VUS	0.00009623, A	S094, A, F	CI	None to date
TAZ ¹	c.515C>T (+/+)	p.Pro172Leu	rs73638278	Benign	0.02808, A	S028, A, M	CI, CII+CIII	None to date
TRMU ¹	c.28G>T (+/+)	p.Ala10Ser	rs11090865	VUS	0.3933, A	S028, A, M	CI, CII+CIII	Guan <i>et al.</i> (2006)
					0.2666, NA	S109, A, F	CI, CIII	
COX4I2 ²	c.412G>A (+/-)	p.Glu138Lys	rs119455950	VUS	0.006936, A	S094, A, F	CI	Shteyer <i>et al.</i> (2009)
NDUFA9 ¹	c.224G>T (+/-)	p.Arg75Leu	rs35263902	VUS	0, A	S049, A, M	CIII	None to date
SDHA ²	c.1523C>T (+/-)	p.Thr508Ile	rs151266052	VUS	0.00821, A	S007, A, M	CII, CIII CII+CIII	Alston <i>et al.</i> (2012)
SDHB ²	c.32G>A (+/-)	p.Arg11His	rs111430410	VUS	0.005455, A	S004, A, F	CIII	Martins <i>et al.</i> (2013)
						S015, A, M	CI, CII, CIII, CIV, CII+CIII	
						S034, A, F	CIII	
SDHD ²	c.34G>A (+/-)	p.Gly12Ser	rs34677591	VUS	0.01018, NA	S050, NA, F	CIII, CIV	Ni <i>et al.</i> (2008)
COQ6 ³	c.41G>A (+/-)	p.Trp14Ter	rs17094161		0.07047, A,)	S002, A, F	CII, CII+CIII	Louw <i>et al.</i> (2018)

	c.859G>T (+/-)	p.Ala49Ser	rs61743884	Likely pathogenic/risk factor	0.03162, A			
	c.278G>A (+/-)	p.Gly93Asp	rs61743864	Benign	0.05479, A	S020, A, M	CII+CIII	<i>None to date</i>
	c.475G>T (+/-)	p.Val163Phe	rs111833521	Benign	0.04316, A	S032, A, F	CIV, CII+CIII	<i>None to date</i>
ETFDH ³	c.1067G>A	p.Gly356Glu	Novel	Pathogenic	None	S057, NA, F	CI, CIII, CIV, CII+CIII	Van der Westhuizen <i>et al.</i> (2017)
	c.1448C>T	p.Pro483Leu	rs377656387	Pathogenic	None			
ACADVL ⁴	194C>T (+/+)	p.Pro65Leu	rs28934585	Likely benign	0.1135, A	S027, A, M	CIII, CIV, CII+CIII	Watanabe <i>et al.</i> (2000)
GLUD2 ⁴	c.1492T>G (+/+)	p.Ser498Ala	rs9697983	Likely benign	0.04979, A	S002, A, F	CIII, CII+CIII	Plaitakis <i>et al.</i> (2010)
SURF1 ²	c.575G>A (+/-)	p.Arg192Gln	rs782021521	Likely pathogenic	0.00001, A	S085, A, F	CIV	Tanigawa <i>et al.</i> (2012)
	c.754_755delAG (+/-)	p.Ser252HisfsTer39	rs782007828	Likely pathogenic	0.00002, A			
RYR1 ⁴	c.8342_8343delTA (+/-)	p.Ile2781ArgfsX49	rs758580075	Likely pathogenic	None, A	S032, A, F	CIV, CII+CIII	Wilmshurst <i>et al.</i> (2010)
	c.11926C>T (+/-)	p.His3976Tyr	rs148772854	Likely pathogenic	0.01287, A			
	c.14524G>A (+/-)	p.Val4842Met	rs193922879	Likely pathogenic	0.001935, A	S033, A, F	CI	Punetha <i>et al.</i> (2016); Wilmshurst <i>et al.</i> (2010)
	c.11193+1G>A (+/-)	-	rs111986316	Likely pathogenic	None			
STAC3 ⁴	c.851G>C (+/+)	p.Trp284Ser	rs140291094	Likely pathogenic	0.0011, A	S011, A, F	CIII	Horstick <i>et al.</i> (2013)
ALAS2 ⁴	c.1757A>T (+/+)	p.Tyr586Phe	rs139596860	Likely pathogenic/VUS	0.009364, A	S117, A, M	CI, CIII, CIV	Balwani <i>et al.</i> (2013); To-Figueras <i>et al.</i> (2011)
TRIOBP ⁴	c.3232C>T (+/+)	p.Arg1078Cys	rs200359708	VUS	0.00030, A	S059, A, M	CI, CIII, CIV, CII+CIII	<i>None to date</i>

¹ Variants identified through Panel sequencing (Panel 1).² Variants identified through Panel sequencing (Panel 2).³ Variants identified through Panel sequencing (Panel 3).⁴ Variants identified through WES.^a Variants are listed relative to the human reference genome sequence, GRCh 37 (hg19).^b Pathogenicity classification are made for this cohort's patients based on the relative information available.^c Population frequencies are listed according to the EXaC population database.^d References to scientific papers where the variant was first reported on.

Abbreviations: VUS: variant of uncertain significance; A: African; NA: non-African; F: female; M: male; CI–CIV: complexes I–IV.

6.2.1 Pathogenic variants

ETFDH

The *ETFDH* (electron-transfer flavoprotein dehydrogenase) compound heterozygous variants (c.1448C>T and c.1067G>A; p.Pro483Leu and p.Gly356Glu; ENST00000307738) were identified in a non-African female (S057) who presented with features of multiple acyl-dehydrogenase deficiency (MADD; OMIM #231680). She had severe muscle weakness, exercise intolerance, chronic fatigue, and hepatomegaly. Metabolic markers of MADD such as dicarboxylic acids, ethylmalonic acid, glutaric acid, as well as acylcarnitines (butyryl-, isovaleryl- and glutaryl carnitine) and acylglycine (hexanoyl- isobutyryl-, isovaleryl and suberylglycine) conjugates were observed in the patient's urine. Furthermore, she had CI, CIII, and CII+CIII RC deficiency, with severely reduced CoQ₁₀ levels in muscle tissue (Louw *et al.*, 2018). These mutations cause structural instability in ETFDH, and were confirmed with functional analysis (Van der Westhuizen *et al.*, 2017). This case, along with two others, is extensively described in Chapter 7. A list of this patient's clinical phenotype and associated MADD phenotypes is given in Table 6.3.

Table 6.3: Cohort phenotypes compared to reported *ETFDH* clinical phenotypes.

Patient, race, gender	Patient phenotype	
S057, NA, F	Developmental delay, migraine, tremor, severe muscle weakness, exercise intolerance, chronic fatigue, hepatomegaly, skeletal involvement	
ETFDH mutation consequences		
Reported manifestations	Clinical phenotype	References
Multiple Acyl Dehydrogenase deficiency (MADD) • Neonatal-onset with congenital anomalies • Neonatal-onset without congenital anomalies • Late-onset form	Neonatal-onset with or without congenital anomalies Nonketotic hypoglycaemia, metabolic acidosis, cardiomyopathy, liver disease, metabolic decompensation, muscle weakness, respiratory failure	Frerman (2001); Fu <i>et al.</i> (2016); Olsen <i>et al.</i> (2007); Van der Westhuizen <i>et al.</i> (2017); Xi <i>et al.</i> (2014)

6.2.2 Likely pathogenic variants

SURF1

An African female (S085), with clinically confirmed Leigh Disease (LD; OMIM #256000) and confirmed mitochondrial CIV deficiency, harboured compound heterozygosity for a missense and frameshift variant (c.575G>A and c.754_755delAG; p.Arg192Gln and p.Ser252HisfsTer39; ENST00000371974,) in exons 6 and 8 in the gene *SURF1* (Surfeit 1). The patient clinically presented early in life with several CNS involvements (nystagmus, extrapyramidal, and pyramidal

symptoms) and muscle involvements (myopathy, hypotonia, and weakness). Furthermore, changes in the basal ganglia/thalami were observed. Metabolic investigation revealed elevated lactic acid. A list of this patient's clinical phenotype and associated LD phenotypes is given in Table 6.4.

Table 6.4: Cohort phenotypes compared to reported SURF1 clinical phenotypes.

Patient, race, gender	Patient phenotype	
S085, A, F	Failure to thrive, developmental delay, developmental regression, central nervous system involvements (nystagmus; extrapyramidal and pyramidal symptoms), myopathy, hypotonia, short stature, radiological changes in the brain (basal ganglia/thalamus), lactic acidosis	
SURF1 mutation consequences		
Reported manifestations	Clinical phenotype	References
Leigh disease due to CIV deficiency	Bilateral lesions in the brainstem, thalamus, basal ganglia, cerebellum, and spinal cord. Necrosis, capillary proliferation in brainstem. High levels of pyruvate and lactate in cerebrospinal fluids Developmental delay, hypotonia, nystagmus, and episodic metabolic acidosis	Danis <i>et al.</i> (2018); Sonam <i>et al.</i> (2017); Tanigawa <i>et al.</i> (2012); Wedatilake <i>et al.</i> (2013); Zhu <i>et al.</i> (1998)

SURF1 is directly involved with cytochrome c oxidase (COX) maintenance and assembly. LoF mutations in *SURF1* cause major structural instability in COX, and are responsible for the phenotype of LD as clinically diagnosed in the case reported here. The frameshift variant has previously been reported as pathogenic for a different (Japanese) patient, with substantial clinical and experimental supporting evidence and a confirmed LoF mechanism (Tanigawa *et al.*, 2012). Consequently, this variant with low African allele frequency is classified as likely pathogenic in this case. The missense variant has not previously been associated with a clinical phenotype, and the extremely low allele frequency suggests moderate likelihood to be pathogenic.

COQ6

An African female (S002) with confirmed primary CoQ₁₀ deficiency was found to be carrying compound heterozygous variants (c.41G>A and c.859G>T; p.Trp14Ter and p.Ala49Ser; ENST00000394026) in the gene *COQ6* (coenzyme Q₁₀ monooxygenase 6). One of the variants identified in a highly conserved region, c.41G>A, results in a premature truncation of the protein with high loss-of-function (LoF) probability. Clinically, she presented at four years and four months of age with a history of severe weakness since birth. Other features were noted upon clinical examination and included macrocephaly, severe hypotonia and head lag, pseudo-hypertrophy of calves and triceps, limitation of extension at knees and elbows, and reduced reflexes. No other candidate *COQ6* gene variants were identified in this case. A list of this patient's clinical phenotype and associated primary CoQ₁₀ deficiency phenotypes is given in Table 6.5.

Table 6.5: Cohort phenotypes compared to reported COQ6 clinical phenotypes.

Patient, race, gender	Patient phenotype	
S002, A, F	Macrocephaly, developmental delay, retinopathy, myopathy, muscle weakness, muscle fibre atrophy, fat infiltration, radiological changes in the brain (white matter atrophy), elevated creatine kinase	
COQ6 mutation consequences		
Reported manifestations	Clinical phenotype	References
Primary CoQ ₁₀	Nephrotic syndrome, sensorineural deafness, seizure, white matter abnormalities, ataxia, and facial dysmorphism	Alcázar-Fabra <i>et al.</i> (2018); Heeringa <i>et al.</i> (2011)

The compound heterozygous COQ6 variants (c.41G>A and c.859G>T) identified in S002 have been extensively investigated elsewhere, where functional and structural analyses showed significantly decreased levels of COQ6 (Louw *et al.*, 2018). This protein is directly involved with CoQ₁₀ biosynthesis, and mutations in this gene result in primary CoQ₁₀ deficiency (Alcázar-Fabra *et al.*, 2018; Heeringa *et al.*, 2011). This correlates well with the clinical and biochemical profiles observed in the case reported here. Based on these experimental findings, the correlation between the observed and reported phenotypes, lead to the classification that these variants are likely to be pathogenic.

RYR1

Two sets of compound heterozygous variants, associated with centronuclear myopathy (CNM) and minicore myopathy and external ophthalmoplegia (MMEO; OMIM #255320), were identified in the gene *RYR1* (ryanodine 1) in two African females (S032 and S033) using WES. For S032, a previously reported pathogenic variant (c.8342_8343delTA; p.Ile2781ArgfsX49; ENST00000359596) (Wilmschurst *et al.*, 2010) with high LoF confidence and a missense variant (c.11926C>T; p.His3976Tyr; ENST00000359596), both in highly conserved regions, were identified. She presented with severe hypotonia, myopathy with myopathic facial features, and initially did not show signs of external ophthalmoplegia. She also had an affected sibling with myopathic facial features and external ophthalmoplegia. A muscle biopsy sample was collected from the sibling, who revealed thickened connective tissue between muscle fibers and evidence of fat infiltration. The female reported here had CIV RC deficiency and her brother had CI+CIII deficiency. Both siblings presented with an additional CII+CIII deficiency.

For S033, a previously reported pathogenic missense variant (c.14524G>A; p.Val482Met; ENST00000359596) (Wilmschurst *et al.*, 2010) and a frameshift variant (c.11193+1G>A; ENST00000359596) were identified. Both variants are in highly conserved regions. Both patients' clinical phenotypes and reported phenotypes is given in Table 6.6.

Table 6.6: Cohort phenotypes compared to reported RYR1 clinical phenotypes.

Patient, race, gender	Patient phenotype	
S032, A, F	Failure to thrive, developmental delay, external ophthalmoplegia, ptosis, hypotonia, myopathy, muscle weakness, elevated lactic acid	
S033, A, F	Failure to thrive, developmental delay, external ophthalmoplegia, hypotonia, myopathy, muscle weakness, renal involvement	
RYR1 mutation consequences		
Reported manifestations	Clinical phenotype	References
Minicore myopathy with external ophthalmoplegia	Congenital muscular dystrophy, external ophthalmoplegia, myopathy, muscle weakness, hypotonia, developmental delay, progressive respiratory failure	Beggs and Agrawal (2013); Monnier <i>et al.</i> (2009); Wilmshurst <i>et al.</i> (2010)

Four previously reported pathogenic *RYR1* variants, two of which are previously reported to be pathogenic, were identified in two compound heterozygotes: S032 (c.8342_8343delTA and c.11926C>T) and S033 (c.14524G>A and c.11193+1G>A). The c.14524G>A and c.8342_8343delTA variants are classified as founder mutations for South African patient populations with centronuclear myopathy (CNM) and MMEO (Horstick *et al.*, 2013). *RYR1* encodes a homotetrameric calcium channel in skeletal muscle and regulates cytosolic Ca²⁺ levels (Blackburn *et al.*, 2017). Dysfunctional *RYR1* disrupts the Ca²⁺ balance, directly affecting different mitochondrial functions such as ATP synthesis regulation and reactive oxygen species generation, consequently contributing to myopathy, external ophthalmoplegia, and ptosis (Görlach *et al.*, 2015; Wilmshurst *et al.*, 2010; Zhou *et al.*, 2007). The two cases reported here had clinical phenotypes consistent with previously reported cases. For S032 and her brother, the specific type of congenital myopathy is still unclear. Both had neurogenic features which are absent in multi-minicore myopathy. The siblings, whose features correlated more strongly with CNM, were similar in their manifestation (albeit with slight, but significant, differences). For example, the female initially did not have external ophthalmoplegia, while the brother did. Importantly, no homozygous variants have been reported in patients with CNM (Abath Neto *et al.*, 2015). S033 had correlating multi-minicore myopathic features as seen in previously reported cases, which included mild myopathic facial features with dense external ophthalmoplegia (Punetha *et al.*, 2016; Wilmshurst *et al.*, 2010). Both sets of compound heterozygous variants are classified as likely pathogenic in these cases due to substantial supporting clinical and experimental evidence confirming the LoF mechanism, the confirmed founder effect for South African populations with a low or absent allele frequency in several population databases, and multiple *in silico* algorithms supporting a deleterious effect on the gene product for the two frameshift variants.

STAC3

A homozygous mutation (c.851G>C; p.Trp284Ser; ENST00000332782) was detected using WES in an African female (S011) in the gene *STAC3* (SH3 and cysteine-rich domains 3), and was first described by Horstick *et al.* (2013) in five families with Native American Myopathy (NAM; OMIM #255995). The phenotype observed in the female described here is similar to that for a case described elsewhere, and is given in Table 6.7 along with reported clinical phenotypes (Horstick *et al.*, 2013). She was born prematurely and had intra-uterine growth restriction. Biochemically she presented with isolated CIII deficiency in muscle.

Table 6.7: Cohort phenotypes compared to reported *STAC3* clinical phenotypes.

Patient, race, gender	Patient phenotype	
S011, A, F	Failure to thrive, developmental delay, relative macrocephaly, ptosis, myopathy, low nasal bridge.	
STAC3 mutation consequences		
Reported manifestations	Clinical phenotype	References
Native American Myopathy	Developmental delay, congenital myopathy, congenital weakness, skeletal anomalies, ptosis, cleft palate, oral hypotonia, poor feeding, progressive scoliosis	Grzybowski <i>et al.</i> (2017); Horstick <i>et al.</i> (2013); Stamm <i>et al.</i> (2008)

STAC3 encodes a putative muscle-specific adaptor protein, which takes part in excitation-contraction in muscle. Disruptions within this protein are thought to cause reduction in mitochondrial Ca^{2+} , either intra- or extra mitochondrial, which have a direct effect on OXPHOS (Blackburn *et al.*, 2017). Clinical features, as a result, include congenital myopathy with facial dysmorphic features including ptosis, which is consistent with the case reported here. This variant was previously sequenced as part of the 1000 Genomes Project, and was undetected in 113 Caucasian controls. This variant is thus classified as likely pathogenic in this case due to its previously confirmed pathogenicity (with substantial supporting evidence) and low African allele frequency.

ALAS2

An African male (S117), with skin and muscle involvement, carried a homozygous gain-of-function variant (c.1757A>T; p.Tyr586Phe; ENST00000330807) in exon 11 of the gene *ALAS2* (5'-aminolevulinic synthase 2) (To-Figueras *et al.*, 2011). This variant was first described in a Spanish patient with erythropoietic porphyria by To-Figueras *et al.* (2011), who states that *ALAS2* acts as a modifier gene in patients with erythropoietic porphyria (X-linked protoporphyria; OMIM #300752). Another manifestation is sideroblastic anaemia type 1 (OMIM #300751), which is also a result of *ALAS2* mutations. However, this patient did not present with symptoms for the latter.

He instead presented with swelling of his face, hands, and feet, and experienced non-specific bodily pain, symptoms which are more similar to X-linked protoporphyria. He had depigmented skin lesions in his face, on the extensor areas of the fore and upper arms, and over his knees and lateral right thigh. Furthermore, he suffered from severe muscle weakness with decreased muscle bulk in all four limbs, with muscle histology revealing atypical dermatomyositis. Biochemically he had confirmed CI, CIII, and CIV deficiency in muscle. A summary of this patient's clinical phenotype and other reported phenotypes is given in Table 6.8.

Table 6.8: Cohort phenotypes compared to reported ALAS2 clinical phenotypes.

Patient, race, gender	Patient phenotype	
S117, A, M	Myopathy, weakness, exercise intolerance, skin involvement, auto-immune disorder, muscle fibre atrophy, inflammatory cell infiltrate	
ALAS2 mutation consequences		
Reported manifestations	Clinical phenotype	References
X-linked Protoporphyria, erythropoietic	Severe photosensitivity, liver disease, anaemia and iron deficiency	Balwani <i>et al.</i> (2013); Brancaleoni <i>et al.</i> (2016); Ducamp <i>et al.</i> (2012); To-Figueras <i>et al.</i> (2011)

The patient described here, S117, did not present with typical X-linked protoporphyria, the primary phenotype associated with *ALAS2* mutations (Balwani *et al.*, 2013; To-Figueras *et al.*, 2011). His primary cause of disease is an interplay between two deficiencies: dermatomyositis and X-linked porphyria. The muscle histology from this patient was suggestive of an inflammatory myopathy, but the inflammatory cell infiltrate was perivascular, and the typical pattern of peripheral muscle cell atrophy in dermatomyositis was not demonstrated. As the patient had definitive skin involvement, which included swelling and redness, it is suggested that the patient has porphyria rather than inflamed myopathy. His mitochondrial dysfunction could arise from impaired electron transport as a direct result of bone marrow haem synthetic dysfunction. The clinical significance is still unclear and therefore this variant is classified as a variant of uncertain significance. A case can be made, however, that this variant should be classified as likely pathogenic since it has a low African allele frequency, and there is substantial supporting evidence for its pathogenicity.

6.2.3 Variants of uncertain significance

TRIOBP

A *TRIOBP* (trio- and filamentous-actin-binding protein) homozygous missense variant (c.3232C>T; p.Arg1078Cys; ENST00000406386) was identified using WES in an African male (S059) who presented with developmental delay, visual impairment, muscle weakness and hypotonia, and clinodactyly. Furthermore, he presented with mild facial dysmorphisms, which

included an epicanthic fold and low set ears. Most notably the patient had hearing impairment, which was confirmed by abnormal auditory brainstem response. Mutations in this gene are known to cause autosomal recessive deafness (OMIM #609823), a feature found in the case presented here. Metabolic profiles revealed significant ketosis associated with dicarboxylic aciduria involving C6–C10 acids (i.e., adipic-, suberic-, and sebacic acid), with a normal amino acid profile. The clinical phenotypes of the patient as well as previously reported phenotypes are listed in Table 6.9.

Table 6.9: Cohort phenotypes compared to reported *TRIOBP* clinical phenotypes.

Patient, race, gender	Patient phenotype	
S059, A, M	Developmental delay, extrapyramidal symptoms, visual impairment, sensorineural deafness, muscle weakness, gastro intestinal involvement, elevated lactic acid	
TRIOBP mutation consequences		
Reported manifestations	Clinical phenotype	References
Autosomal recessive deafness	Non-syndromic hearing loss, deafness	Kitajiri <i>et al.</i> (2010); Riazuddin <i>et al.</i> (2006); Shahin <i>et al.</i> (2006)

S059 had a severe mitochondrial phenotype, where deafness was most notable. The latter is consistent with reported cases harbouring *TRIOBP* mutations; however the former has not yet been associated with mutations in this gene (Riazuddin *et al.*, 2006; Shahin *et al.*, 2006). The underlying genetic cause for this patient's mitochondrial phenotype is still unclear. This variant, c.3232C>T, has not yet been associated with a clinical phenotype and therefore lacks substantial supporting evidence of pathogenicity. Therefore, this variant is classified as a variant of uncertain significance.

6.2.4 Likely benign and benign variants

DARS2

Three *DARS2* variants, two previously reported (p.Lys196Arg, p.Gly338Glu) and one novel frameshift (g.173808623C>CT) were identified in six patients: three African and three non-African. *DARS2* provides instructions for making mitochondrial aspartyl-tRNA synthetase — an enzyme responsible for attaching an aspartic acid to the correct tRNA which helps with assembling of amino acids into proteins. A mutation in this gene causes a severe autosomal recessive mitochondrial phenotype known as Leukoencephalopathy with Brainstem and Spinal Cord Involvement and Lactate Elevation (LBSL; OMIM #611105). The majority of patients with LBSL harbour compound heterozygous *DARS2* mutations, and to date only a few cases with a homozygous mutation survived beyond birth (Miyake *et al.*, 2011; Shimojima *et al.*, 2017). From

the more than 50 disease-causing or pathogenic variants identified, the three variants identified here, all of which were detected as heterozygous, have not previously been associated with a LBSL phenotype or previously described in literature. The clinical phenotypes of the patient cohort, as well as previously reported phenotypes, are given in Table 6.10.

Table 6.10: Cohort phenotypes compared to reported DARS2 clinical phenotypes.

Patient, race, gender	Patient phenotype	
S014, A, M	Developmental regression, pyramidal symptoms, ptosis, hypotonia, muscle weakness, hepatomegaly, basal/ganglia changes, white matter atrophy	
S024, NA, M	Developmental regression, ADHD, epilepsy, absences, ataxia, tremors, peripheral neuropathy, hypotonia, exercise intolerance, radiological changes in the brain, lactic acidosis	
S044, NA, M	Developmental regression, ADHD, depression, epilepsy, myoclonic seizures, absences, tonic/clonic seizures, migraine, pyramidal symptoms, exercise intolerance, hepatic failure	
S089, A, M	Developmental regression, epilepsy, tonic seizures, ataxia, tremor, nystagmus, pyramidal symptoms, cerebellar atrophy, lactic acidosis	
S092, NA, M	Macrocephaly, developmental delay, dysmorphism, ataxia, tremor, visual impairment, hypotonia, hypermobile joints, cerebellar atrophy	
S101, A, F	Hypotonia, myopathy, muscle weakness	
DARS2 mutation consequences		
Reported manifestations	Clinical phenotype	References
Brainstem and spinal Cord involvement and lactate elevation (LBSL)	Pyramidal, cerebellar and dorsal column dysfunction, white matter abnormalities, elevated levels of lactic acid, developmental delay, cognitive impairment, epilepsy and peripheral neuropathy, hypotonia, failure to thrive	Miyake <i>et al.</i> (2011); Rathore <i>et al.</i> (2017); Shimojima <i>et al.</i> (2017)

The non-African case, S024, is the only patient who harboured two *DARS2* variants, the novel frameshift and a previously reported missense variant. He furthermore presented with the most severe mitochondrial phenotype compared to the other cases described here in which a *DARS2* variant was detected. He displays typical LBSL phenotype patterns, suggesting an underlying genetic mutation in *DARS2*. This specific case has not yet been extensively investigated. Three cases, one non-African (S044) and two African (S014 and S089), showed similar and overlapping previously reported LBSL phenotype patterns. The remaining cases, S092 and S101, had no correlating phenotypes. The two missense variants have relatively low allele frequencies but are nevertheless classified as benign and tolerated by various computational algorithms including SIFT and Polyphen-2. The novel variant was not listed in either the ExAC or GnomAD databases.

The variants reported here have not been extensively investigated with functional analysis and are therefore classified as VUS. No other significant *DARS2* variants were detected in any of the six cases. These variants still remain of interest for the African population as phenotypes showed slight correlating LBSL patterns. It is important to note that phenotypes reported for LBSL overlap with clinical phenotypes associated with *POLG* manifestations. Patients with suspected LBSL or

any of the POLG spectrum disorders should therefore be extensively investigated, both clinically and genetically, in order to prevent incorrect diagnosis of either of the manifestations.

POLG

Two *POLG* variants were identified in three patients. Polymerase gamma, or *POLG*, is found in mitochondria and plays a role in mtDNA replication and repair. There are currently over 180 known *POLG* mutations, all of which result in a number of clinical manifestations. The most common *POLG* manifestations (Cohen *et al.*, 2018), with their respective clinical phenotypes, is given in Table 6.11.

The three cases presented here (S014, S024, and S097) had isolated CI deficiency with three different clinical phenotypes observed in each, as seen in Table 6.11. The first African case, S014, showed similar but inconclusive phenotypes associated with AHS and arPEO. The second case, S024, had symptoms similar to ANS, but he did not present with ophthalmoplegia at the time of this investigation. The latter's phenotype fits more with LBSL, as previously described. The last case, S097, showed symptoms correlating with MCHS, but without myopathy.

Table 6.11: Cohort phenotypes compared to reported *POLG* clinical phenotypes.

Patient, race, gender	Patient phenotype	
S014, A, M	Developmental regression, pyramidal symptoms, ptosis, hypotonia, muscle weakness, hepatomegaly, basal ganglia changes, white matter atrophy	
S024, NA, M	Developmental regression, ADHD, epilepsy, absences, ataxia, tremors, neuropathy, hypotonia, exercise intolerance, radiological changes in the brain, lactic acidosis	
S097, NA, F	Failure to thrive, developmental delay, brainstem involvement, lactic acidosis	
POLG mutation consequences		
Reported manifestations	Clinical phenotype	References
Alpers-Huttenlocher syndrome (AHS)	Childhood-onset progressive encephalopathy with epilepsy, hepatic failure	Anagnostou <i>et al.</i> (2016); Cohen <i>et al.</i> (2014); Cohen <i>et al.</i> (2018); Cohen and Naviaux (2010); Davidzon <i>et al.</i> (2006); Naviaux and Nguyen (2004); Nguyen <i>et al.</i> (2006); Saneto <i>et al.</i> (2013); Wong <i>et al.</i> (2008)
Childhood myocerebrohepatopathy spectrum (MCHS)	Developmental delay or dementia, lactic acidosis, myopathy with failure to thrive	
Myoclonic epilepsy myopathy sensory ataxia (MEMSA)	Epilepsy, ataxia, myopathy without ophthalmoplegia	
Ataxia neuropathy spectrum (ANS)	Spectrum: Mitochondrial recessive ataxia syndrome and sensory ataxia neuropathy dysarthria, ophthalmoplegia Symptoms: seizures, ataxia, ophthalmoplegia, myopathy (rare)	
Autosomal recessive progressive external ophthalmoplegia (arPEO)	Progressive weakness of the extraocular eye muscles resulting in ptosis and ophthalmoparesis	
Autosomal dominant progressive external ophthalmoplegia (adPEO)	Myopathy and often variable degrees of sensorineural hearing loss, axonal neuropathy, ataxia, depression, parkinsonism, hypogonadism, cataracts	

The patients described here could fall in the POLG-related spectrum based on similar features observed. However, the variants are classified as benign on several disease databases (ClinVar and OMIM) suggesting a high probability that the variants are benign in these cases as well. These variants still remain of interest to African populations as none of the variants were previously investigated with additional functional analysis.

TAZ

A TAZ variant (c.515C>T, p.Pro172Leu) was detected in an African male (S028) biochemically diagnosed with isolated CI deficiency. The TAZ gene is highly expressed in cardiac muscle as its gene product, tafazzin, plays an important role in the remodelling of immature cardiolipin. Tafazzin is found in mitochondrial membranes and stabilises RC complexes (Gonzalvez *et al.*, 2013). Loss of cardiolipin, due to mutations in the TAZ gene, has a direct effect on CI biogenesis and other respiratory activities. The phenotypes associated with these biochemical instabilities are cardiomyopathy, muscle myopathy, and Barth syndrome (BTHS; also known as 3-methylglutaconic aciduria type II) (Thompson *et al.*, 2016). This patient did not present with a typical TAZ-associated phenotype but with hypotonia, as found in many patients, especially males with BTHS (Borna *et al.*, 2017; Imai-Okazaki *et al.*, 2018). Furthermore, markedly elevated 3-methylglutaconic acid levels were detected in urine. A more detailed clinical profile for S028 is given in Table 6.12 along with phenotypes of previously reported manifestations.

Table 6.12: Cohort phenotypes compared to reported TAZ clinical phenotypes.

Patient, race, gender	Patient phenotype	
S028, A, M	Hypotonia, muscle weakness, gastro-intestinal involvement, renal involvement, short stature, skeletal involvements.	
TAZ mutation consequences		
Reported manifestations	Clinical phenotype	References
3-methylglutaconic aciduria type II (Barth syndrome,)	Dilated cardiomyopathy, neutropenia, skeletal myopathy, abnormal mitochondria, short stature, hypocholesterolemia, cognitive dysfunction, 3-methylglutaconic aciduria	Barth <i>et al.</i> (1983); Borna <i>et al.</i> (2017); Imai-Okazaki <i>et al.</i> (2018); McKenzie <i>et al.</i> (2006); Yen <i>et al.</i> (2008)

There are some conflicting interpretations of pathogenicity observed for the case reported here. The most common feature of BTHS is cardiomyopathy, which was absent in our patient. However, the elevated 3-methylglutaconic acid levels, a metabolomics diagnostic hallmark for BTHS, do suggest a BTHS phenotype for this patient, albeit with different clinical manifestations. The suspicion of BTHS for this patient can only be confirmed with functional analysis on the TAZ protein, which unfortunately could not yet be performed at the time of this investigation.

ACADVL

A homozygous *ACADVL* variant (194C>T, p.Pro65Leu) was identified in an African male, S027. This variant was previously reported as pathogenic (Watanabe *et al.*, 2000), with the clinical phenotype of very long chain acyl-CoA dehydrogenase deficiency. The clinical features of our patient and reported clinical phenotypes are listed in Table 6.13.

Table 6.13: Cohort phenotypes compared to reported *ACADVL* clinical phenotypes.

Patient, race, gender	Patient phenotype	
S027, A, M	Developmental delay, epilepsy, absences, myoclonic seizures, visual impairment, neuropathy, hypotonia, muscle weakness, hepatomegaly, radiological changes in the brain (brain atrophy, dilated ventricle horn, cerebellar atrophy), elevated lactic acid	
ACADVL mutation consequences		
Reported manifestations	Clinical phenotype	References
Very long-chain acyl-CoA dehydrogenase (ACADVL)	Hypoglycemia, lethargy, muscle weakness, liver involvement	Diekman <i>et al.</i> (2015); Evans <i>et al.</i> (2016).

The variant identified here (p.Pro67Leu) is considered a complex mutant with another allele, where lysine is substituted for glutamine at position 247 of the *ACADVL* gene (p.Lys247Gln). The p.Pro67Leu mutation causes an exon 3 skipping, which suggests severe effects on protein levels. Functional analysis, however, showed no reduction in protein levels for the p.Pro65Leu mutation in the case reported by Watanabe *et al.* (2000), but the p.Lys247Gln showed dramatically reduced enzyme activity (Watanabe *et al.*, 2000). Our case had clinical phenotypes inconsistent with other reported cases (Diekman *et al.*, 2015; Evans *et al.*, 2016; Watanabe *et al.*, 2000). Furthermore, a very high African allele frequency of 11% is reported for this specific detected variant. Considering this, it is concluded that the detected variant is benign in this case, even though it has been reported as pathogenic by (Watanabe *et al.*, 2000).

GLUD2

A homozygous *GLUD2* variant was identified in an African female (S002). The variant, p.Ser498Ala, was previously reported as pathogenic by Plaitakis *et al.* (2010), with the clinical phenotype of PD. The clinical features of our patient and the reported clinical phenotypes is given in Table 6.14.

Table 6.14: Cohort phenotypes compared to reported *GLUD2* clinical phenotypes.

Patient, race, gender	Patient phenotype	
S002, A, F	Macrocephaly, developmental delay, retinopathy, myopathy, muscle weakness, muscle fibre atrophy, fat infiltration, radiological changes in the brain (white matter atrophy), elevated creatine kinase	
GLUD2 mutation consequences		
Reported manifestations	Clinical phenotype	References
Parkinson disease	Tremor, bradykinesia, postural instability	Plaitakis <i>et al.</i> (2010); Uitti <i>et al.</i> (2005)

The patient reported here, S002, did not present with any signs or symptoms of PD at the time of this investigation. As previously discussed in Section 6.2.1, she has confirmed CoQ₁₀ deficiency with likely pathogenic variants identified in the gene *COQ6*. The African allele frequency for this specific variant is relatively high at 5%, suggesting that the variant is benign with high confidence in this African case.

6.3 DISCUSSION AND CONCLUSIONS

Three sequencing approaches were followed in this investigation in order to identify previously reported variants, with or without a disease causing mechanism, and novel variants. The three approaches included mtDNA sequencing and two nuclear approaches, targeted gene panel NGS and WES. From the overall results, as discussed in Section 6.1, it was seen throughout the two nuclear approaches that more variants were detected in Africans compared to non-Africans, however not significantly more. This is consistent with recently published results which comment on the genetic variety and diversity found in African populations residing mostly in Sub-Saharan Africa (Campbell & Tishkoff, 2010; Choudhury *et al.*, 2017; Gurdasani *et al.*, 2015; Schlebusch *et al.*, 2012).

Mitochondrial genome sequencing revealed relatively low diagnostic outcome in this cohort. It is known that mtDNA point mutations are more prone to cause a phenotype in adults rather than children (Rahman, 2018; Viscomi & Zeviani, 2018) given that the mutation load has been reached and that the point mutation is situated in a highly conserved region. Large scale deletions or rearrangements can be sporadic and are thought to take place during development of an embryo. None of our cases presented with a large scale deletion or rearrangements. The mtDNA mutations known to cause the most common syndromes, such as MELAS, MERRF, and LHON, were absent. The one case, S014, who did present with the well-known LHON mutation (m.14484T>C), was strictly classified as benign, due to an inconsistent phenotype observed in this case when compared to reported cases (Wallace *et al.*, 1988). The mtDNA novel variants

identified in the cohort still have the potential to be either pathogenic or benign, each requiring functional investigations to determine pathogenicity.

Several mutations in nuclear genes have been associated with a specific clinical MD phenotype, as discussed Section 2.2.3. A large number of genes (more than 250), directly involved with OXPHOS structure and function, are known to cause primary MD. In only three cases, S002, S085 and S057, a primary MD could be identified when considering clinical, biochemical and genetic findings (panel sequencing). A variant for the most frequent childhood manifestation of MD, known as LD, was identified in only one African case, S085, and in only one case, S002, a primary CoQ₁₀ deficiency compound heterozygous variants could be identified. The third case, presented with pathogenic gene mutations (*ETFDH*) which is indirectly involved with OXPHOS but are still considered a primary MD. The remaining five cases (S032, S033, S011, S117 and S059), all presented with secondary MD gene variants, notably identified through WES. These findings show a relatively poor diagnostic outcome for panel sequencing, as only two cases harboured pathogenic and likely pathogenic variants. Initial indications from limited WES data are much more promising as nine likely pathogenic variants were identified in six cases using WES compared to five variants identified in three cases using panel NGS and mtDNA sequencing.

Considering all the cases in this study, in only five cases (S057, S085, S032, S033, and S011) a strong genotype-phenotype correlation could be established and a moderately strong correlation in two cases (S002 and S117). For the remaining two cases, S014 and S059, a non-specific correlation was observed. Observations like these serve as a strong motivation that a “genetics first/only” approach, without supporting clinical and biochemical investigation, is not suitable in such understudied, ethnically diverse populations. Furthermore, when following a genetic approach, it is concluded that panel sequencing may only be an efficient approach in populations where the genotype-phenotype correlations are well-established for specific monogenic diseases. For heterogeneous diseases such as MD, however, even in such populations, WES/WGS performs significantly better compared to a targeted gene-panel approach (Haack *et al.*, 2012; Lieber *et al.*, 2013; Neveling *et al.*, 2013; Wortmann *et al.*, 2015; Wortmann *et al.*, 2017). The results reported are thus in line with proposals that WES should be considered as the primary option for genetic investigations in heterogeneous inherited diseases such as MD and, in fact, may be particularly ideal in understudied, ethnically diverse populations where there is evidence of inconsistencies with documented MD phenotypes. However, in such populations the value of extensive clinical and biochemical (structural and functional) investigations to support molecular genetic data outcomes, should not be neglected, and at this time are, in fact, more crucial than in well studied populations.

CHAPTER 7:

REPORTING OF A NOVEL *ETFDH* MUTATION — A CASE REPORT

Variants of importance were cautiously evaluated before it could be identified as pathogenic or disease-causing (as discussed in Sections 5.6). Two particular variants (compound heterozygotes) of interest were identified through panel sequencing in the gene *ETFDH*, in a paediatric patient (see Section 6.2.1). These variants, one of which was novel, were functionally investigated and showed significant damage being done on protein level. These variants were identified in two other cases — as described in this chapter — one of which was homozygous for the novel variant (Index case). Here, we report on the three patients who harboured these mutations as a case report.

This chapter has been published in a peer-reviewed journal, and addresses Objective 6 (see Appendix A.3):

Van der Westhuizen F.H., Smuts, I., Honey, E., Louw, R., Schoonen, M., Jonck, L., Dercksen, M. (2017). A novel mutation in ETFDH manifesting as severe neonatal-onset multiple acyl-CoA dehydrogenase deficiency. *Journal of the Neurological Sciences*, 384: 121–125 [DOI: 10.1016/j.jns.2017.11.012] — with permission.

7.1 ABSTRACT

Neonatal-onset multiple acyl-CoA dehydrogenase deficiency (MADD) type I is an autosomal recessive disorder of the electron transfer flavoprotein function characterized by a severe clinical and biochemical phenotype, including congenital abnormalities with unresponsiveness to riboflavin treatment as distinguishing features. From a retrospective study, relying mainly on metabolic data, we have identified a novel mutation, c.1067G>A (p.Gly356Glu) in exon 8 of *ETFDH*, in three South African Caucasian MADD patients with the index patient presenting the hallmark features of type I MADD and two patients with compound heterozygous

(c.1067G>A + c.1448C>T) mutations presenting with MADD type III. SDS-PAGE/Western-blot confirmed the significant effect of this mutation on ETFDH structural instability. The identification of this novel mutation in three families originating from the South African Afrikaner population is significant to direct screening and strategies for this disease, which amongst the organic acidemias routinely screened for, is relatively frequently observed in this population group.

7.2 INTRODUCTION

Multiple acyl-CoA dehydrogenase deficiency (MADD, OMIM:231680), also known as glutaric aciduria type II (GAI), is an autosomal recessive metabolic disorder. Mutations in the *ETFA*, *ETFB* and *ETFDH* genes result in the deficient function of alpha and beta subunits of the electron transfer protein (ETF) or ETF-coenzyme Q oxidoreductase, respectively. Dysfunction in either of these two flavoproteins leads to compromised fatty acid and amino acid oxidation as well as choline metabolism. The MADD phenotype can be divided into three clinical classes: two neonatal-onset forms with (type I) or without (type II) congenital anomalies or a mild/delayed late-onset form (type III), which shows a variable degree of response to riboflavin treatment (Frerman, 2001; Grünert, 2014; Liang *et al.*, 2009). The neonatal-onset form has a poor prognosis, presenting within the first week(s) of life with severe hypoglycaemia, mild to severe hyperammonaemia, metabolic acidosis, with or without dysmorphic features, hepatic-, cardiac and renal involvement. The phenotypic presentation of type III is highly variable, characterized by episodic lethargy, vomiting, hypoglycaemia, metabolic acidosis with or without hyperammonaemia and hepatomegaly. Painful myopathy can be progressive and abnormal lipid storage and free carnitine depletion is commonly observed (Frerman, 2001; Grünert, 2014).

The biochemical characterization of MADD includes organic acid and acylcarnitine profiling, which indicates, increased levels of aliphatic mono- and dicarboxylic acids, acylglycine conjugates as well as an increase of C4–C18 acylcarnitines in blood (Frerman, 2001). Although mutations are most frequently encountered in the *ETFDH* gene in the late-onset form of MADD, for the most severe (type I) neonatal form of the disease mutations in *ETFDH*, *ETFA*, and *ETFB* are found more or less in equal frequency (Fu *et al.*, 2016; Grünert, 2014; Liang *et al.*, 2009; Olsen *et al.*, 2007; Xi *et al.*, 2014; Yamada *et al.*, 2016; Yotsumoto *et al.*, 2008).

As mutation screening is not currently offered routinely in our (South African) population, and since biochemical data through routine metabolic screening suggests that MADD may be in the same order of prevalence as other organic acidemias in South African populations (Dercksen *et al.*, 2012; Van der Watt *et al.*, 2010) we've begun to retrospectively investigate the molecular genetics of MADD with metabolic data as the denominator. From the initial explorative investigations in three Caucasian patients, we've identified a novel pathogenic variant present in

all three cases, with the homozygous form in the index case presenting as a (type I) severe neonatal onset form of the disease.

7.3 PATIENTS AND METHODS

Informed consent was obtained as part of current studies with ethical approval numbers 91/98 (University of Pretoria) and NW-00170-13-S1 (North-West University). Table 7.1 summarizes the clinical, biochemical and genetic features of the three patients.

7.3.1 Patient 1 (index patient)

The firstborn male child from non-consanguineous Afrikaner parents presented shortly after birth with severe metabolic decompensation unresponsive to standard metabolic intervention. He was born prematurely via caesarean section and needed admission to the intensive care immediately after delivery. A loud systolic murmur was audible and the chest X-ray showed a prominent right ventricle and absent pulmonary conus. The ECG displayed an extreme right QRS axis and a dysplastic pulmonary valve with a pulmonary outflow gradient measuring 57 mmHG indicated a severe pulmonary stenosis. The foramen ovale was patent with a right to left shunt and there was a large atrial septum aneurism. The renal ultrasound showed dilated calyces in the upper lobes more severely on the left indicating hydronephrosis. After a relative normal 24 h complicated only with hyperlactanaemia and suspected hypothyroidism his condition rapidly deteriorated with persistent hypoglycaemia requiring high doses of intravenous glucose, metabolic acidosis and hyperammonaemia (548 $\mu\text{mol/L}$). He needed intubation and ventilation on day 6 of life but his condition deteriorated despite administration of riboflavin (200 mg/kg/day) and L-carnitine (100 mg/kg/day). He developed a bulging fontanel, sclerema neonatorum and convulsions and died on day 9 of life. GC-MS organic acid profiling performed on urine at day 3 of life showed dicarboxylic aciduria with the presence of various short and medium chain acylglycine conjugates suggestive of MADD. The metabolic excretion as well as early onset was indicative of a severe neonatal phenotype with a very poor prognosis. His family subsequently had a normal male child delivered after prenatal diagnosis performed on chorionic villi showed the foetus was heterozygous for the reported c.1067G>A mutation.

Table 7.1: Summary of clinical, biochemical and genetic features of patients.

	Patient 1 (index)	Patient 2	Patient 3
Sex	Male	Female	Female
Age of onset/death	Week 1/Day 9	Year 2/undetermined (aged at least past 8 years)	Year 4/23
Main clinical features	Intra-uterine growth retardation Multiple congenital anomalies: • Celiac lesions • Hydronephrosis Riboflavin unresponsive	Progressive muscle weakness Hepatomegaly Riboflavin responsiveness undetermined Exercise intolerance Migraine Hypermobility	Progressive muscle weakness Hepatomegaly Migraine Partial responsive to riboflavin
Routine biochemistry	Elevated blood lactate Hypoglycaemia Elevated ammonia No urine ketones Thrombocytopenia Metabolic acidosis	Mildly elevated liver enzymes Normal creatine Elevated lactate:pyruvate No urinary ketones Ammonia?	Metabolic acidosis Elevated ammonia Hypoglycaemia Elevated creatine kinase Elevated transaminase No urinary ketones
Muscle or liver histology	No histology done	No histology done	Liver biopsy indicated macrovascular steatosis
Routine metabolic results	<i>Urine Organic acid profiling:</i> Lactic acid briefly elevated but normalized ↑↑↑ethylmalonic acid ↑↑2-hydroxyglutaric acid ↑↑↑dicarboxylic acids including glutaric-, suberic acid, sebacic acid, decenoic acid ↑↑Krebs cycle intermediates ↑↑↑glycine conjugates: isovalerylglycine, isobutyrylglycine, hexanoylglycine, suberylglycine <i>Urine acylcarnitine profiling:</i> ↑↑↑isovalerylcarnitine, ↑↑butyrylcarnitine, ↑↑glutaryl carnitine, low free carnitine <i>Urine amino acid profiling:</i> General amino aciduria including elevated sarcosine and glycine	<i>Urine organic acid profiling:</i> ↑ethylmalonic acid ↑2-hydroxyglutaric acid ↑↑↑dicarboxylic acids including glutaric acid, suberic acid, sebacic acid, decenedioic acid ↑↑↑glycine conjugates: hexanoylglycine, suberylglycine <i>Urine carnitine profiling:</i> ↑↑Butyrylcarnitine, ↑↑glutaryl carnitine, ↑ isovalerylcarnitine, Low free carnitine <i>Urine amino acids profiling:</i> ↑↑Sarcosine, ↑↑glycine	<i>Urine Organic acid profiling:</i> Normal lactic acid ↑ethylmalonic acid ↑↑dicarboxylic acids including adipic acid, suberic acid ↑glycine conjugates: hexanoylglycine, sovalerylglycine, suberylglycine <i>Urine acylcarnitine profiling:</i> ↑↑↑Butyrylcarnitine, ↑↑↑glutaryl carnitine, ↑ octanoylcarnitine, ↑ isovalerylcarnitine, Normal free carnitine <i>Urine amino acid profiling:</i> ↑Sarcosine, ↑↑glycine
Specialized investigations	None	Muscle respiratory chain enzymes deficient (CI, CIII, CII+III, CIV; Muscle CoQ ₁₀ levels reduced.	Oxidation of C16 and C14 fatty acids both reduced in fibroblasts (20% vs. controls mean; C16:C14 = 0.5) and leukocytes (5% vs. controls mean; C16:C14 = 0.96)
ETFDH variant/s (dbSNP)	c.1067G>A +/-; p.Gly356Glu (novel)	c.1067G>A +/-; p.Gly356Glu c.1448C>T +/-; p.Pro483Leu (377656387)	c.1067G>A +/-; p.Gly356Glu c.1448C>T +/-; p.Pro483Leu (377656387)

Abbreviations: C4: butyryl-/isobutyrylcarnitine; C5: isovalerylcarnitine; C5-DC: glutaryl carnitine; C8: octanoylcarnitine; CBC: complete blood count; AST: aspartate aminotransferase; ALT: alanine aminotransferase; GGT: gamma-glutamyltransferase.

7.3.2 Patient 2

An eight-year-old female patient presented with the complaint of clumsiness, chronic fatigue, exercise intolerance, muscle weakness and severe migrainous-like headaches. She was born at term from nonconsanguineous parents and had three normal siblings. Her early gross motor development was normal and she was able to walk before the age of one year. At the age of 17 months the parents observed that she had poor head control related to muscle weakness. Since the age of three years the weakness slowly progressed. Her body mass index for her age was 13.3 and fell on Z-score — 1.8 at the age of eight years. She had a waddling gait and mild proximal muscle weakness, but she did not have a Gower sign. Polyminimyoclonus resembling spinal muscular atrophy was observed. The deep tendon reflexes were suppressed, but present. She had no other cerebellar signs. Hypermobility of all her joints was observed. Her general level of functioning was on the level of 6.5 years.

She had hepatomegaly with mildly elevated liver enzymes. Renal functions were normal, creatine kinase was normal and the lactate:pyruvate ratio was 16.3. The routine metabolic investigations showed general dicarboxylic aciduria including acylglycine and acylcarnitine conjugates. Spinal muscular atrophy was excluded: the patient did not have a homozygous deletion in the telomeric region of the survival motor neuron on exon seven. A muscle biopsy was performed for the determination of respiratory enzyme analyses, after which the patient and family was not responsive to any further follow up.

7.3.3 Patient 3

This patient was born at term from non-consanguineous parents with one healthy sibling and no previous family history of neuromuscular disease. She initially presented with hypotonia and only started to walk at the age of 20 months. The progressive myopathic phenotype of unknown cause was confirmed at the age of two years after a comprehensive clinical evaluation. Metabolic decompensation was observed at the age of three years when she was admitted with coma, severe hyperammonemia (394 $\mu\text{mol/L}$ which decreased to 90 $\mu\text{mol/L}$ after hospitalization), hepatomegaly and hypoglycaemia without urinary ketones. No myoglobinuria were noted, but prerenal failure was suspected. Her liver enzymes were elevated, together with elevated creatine kinase. A liver biopsy indicated macrovascular steatosis with no necrosis. The metabolic investigations showed increased levels of urine mono- and dicarboxylic acid, acylglycine as well as short and long chain acylcarnitine conjugation. The latter in combination with general clinical presentation of myopathy and liver involvement was suggestive of MADD. She received riboflavin and L-carnitine treatment together with a low protein diet and improved partially, waking up from

the coma and showing stronger motor function. Progressive deterioration of muscular function did however occur and she suffered from episodic migraines. Regular follow-up metabolic investigations were done to assess her metabolic status, which showed moderate to limited biochemical improvement after implementation of therapeutic intervention. She did not present with cognitive impairment and became a fully functional adult. She unfortunately died from a stroke-like episode at age 23 before the genetic diagnosis was confirmed.

7.3.4 Metabolic and biochemical investigations

Metabolic investigations were performed essentially as described before (Reinecke *et al.*, 2012). Respiratory chain enzyme activities and CoQ₁₀ levels were analysed in muscle homogenates (600 × g supernatants) for Patient 2 as described before (Itkonen *et al.*, 2013; Smuts *et al.*, 2010) and normalized to citrate synthase (UCS). Fatty acid oxidation of [9,10-³H] palmitic (C16) and myristic (C14) acids were performed in fibroblasts and leukocytes for Patient 3 as described before (Brivet *et al.*, 1995).

7.3.5 Mutation and protein analyses

The outcomes reported here are the results from a retrospective investigation of selected cases, using next-generation sequencing (NGS, Ion Torrent) of candidate genes. Primary NGS data analysis and secondary data mining were conducted using an in-house bioinformatics pipeline and web-based annotation tools such as Ensembl online VEP runner (McLaren *et al.*, 2016) and data mining tools, including GEMINI (v0.18.0) (Paila *et al.*, 2013), OMIM, ClinVar, and available allele frequency data from the extensive Exome Aggregation Consortium (ExAC) database. The potential pathogenic variants identified and allelic distribution in all of these cases were confirmed using Sanger sequencing and, as the c.1067G>A variant is part of a homopolymeric region, also restriction length fragment polymorphism (RFLP) analyses Figure 7.1.

SDS-PAGE/Western-blot analysis was used to confirm the steady state levels of muscle ETFDH and, due to the limited material available, could only be done in Patient 2. Immunodetection using the antibodies against ETFDH (Abcam, ab131376) and β-actin (Abcam ab8227) were conducted and expression quantified against total protein load using the procedures described for Bio-Rad's V3 workflow (Bio-Rad Laboratories).

7.4 RESULTS

Table 7.1 summarizes the clinical, biochemical and other specialized investigations for the three cases. The following key findings for metabolic, biochemical and genetic analyses are reported.

7.4.1 Metabolic and biochemical investigations

Routine and follow up metabolic investigations in urine throughout the history of all three cases showed some of the hallmark metabolites associated with MADD, notably dicarboxylic acids, ethylmalonic acid, glutaric acid as well as acylcarnitines- (butyryl-, isovaleryl- and glutaryl carnitine) and acylglycine (hexanoyl-, isobutyryl-, isovaleryl- and suberylglycine) conjugates. Patient 1 had not undergone any biopsies at time of his death. Based on her clinical features, Patient 2 was early on selected to undergo a muscle biopsy for respiratory chain enzyme analyses, which revealed combined complex I, III, II+III and IV deficiencies. The follow up investigation of CoQ₁₀ levels in 600 x g supernatants showed significantly reduced levels at 0.86 nmol/UCS (controls 2.680–8.470 nmol/UCS, n = 37). For Patient 3, with an early suspicion of MADD, fatty acid oxidation confirmed the functional deficiency with palmitic (C16) and myristic (C14) acid oxidations almost equally reduced in fibroblasts and leukocytes.

7.4.2 Mutation and protein analysis

A c.1067G>A (p.Gly356Glu) variant on the *ETFDH* gene was detected in homozygous form in Patient 1 (Figure 7.1). This variant does not appear on any clinical or genetic databases, nor in literature consulted and was shown to be heterozygous in both parents as well as in an unaffected sibling. This novel variant is predicted to be deleterious by the SIFT prediction software and disease-causing by MutationTaster prediction software. For both Patients 2 and 3, a previously documented c.1448C>T (p.Pro483Leu) mutation and the putative deleterious c.1067G>A variant was identified in heterozygous forms. Although the patient and family of Patient 2 were lost to follow up, the carrier status for the c.1067G>A variant was confirmed in mother and sibling for Patient 3. SDS-PAGE/Western-blot analysis (Figure 7.2) of the steady state levels for ETFDH in the only available tissue material (muscle from Patient 2) showed a significant (83%) reduction compared to controls.

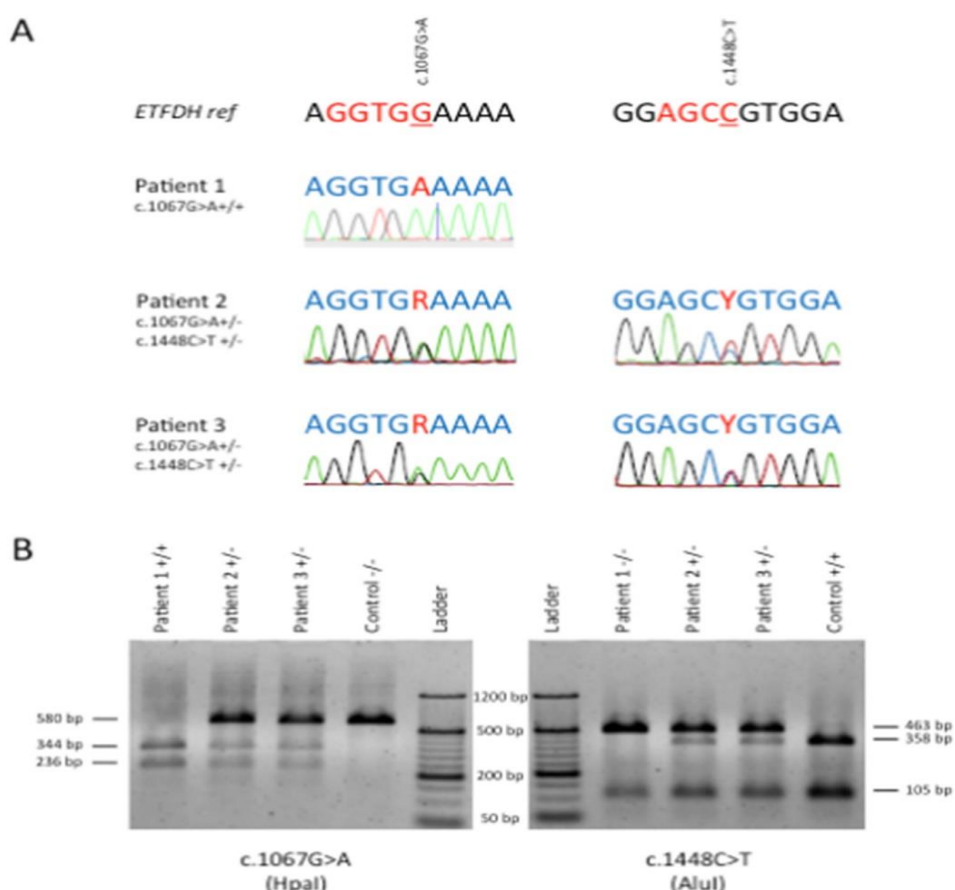


Figure 7.1: Partial sequence alignment of sequence data (A) and RFLP analysis (B) of patients. In Panel A the sequence validation data for the three patients is shown in alignment with the reference sequence for the *ETFDH* gene (ENS00000171503) at the positions for the c.1067G>A and C.1448C>T mutations. R indicates a heterozygous G to A nucleotide base change and Y a heterozygous C to T nucleotide base change. In Panel B the results from restriction fragment length polymorphism (RFLP) analyses for the patients and appropriate controls are shown. For this, DNA was first amplified using PCR and selected primers for a 580 bp region covering the c.1067>A mutation (fwd: GCACATAGTGCTCCAAATAC; rev: CATGCCTGGCTAATCTTTCC) and a 568 bp region covering the c.1448>T mutation (fwd: CACACATTTGGGCAGTTTCG; rev: AAAGTATCTGTCCATCGGG), respectively. The resulting fragments were digested with *HpaI* (GGTGA(N8)) and *AluI* (AGCT), as indicated in red for the two mutations in Panel A. In both cases the enzymes only cut at the positions when the mutations are present. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

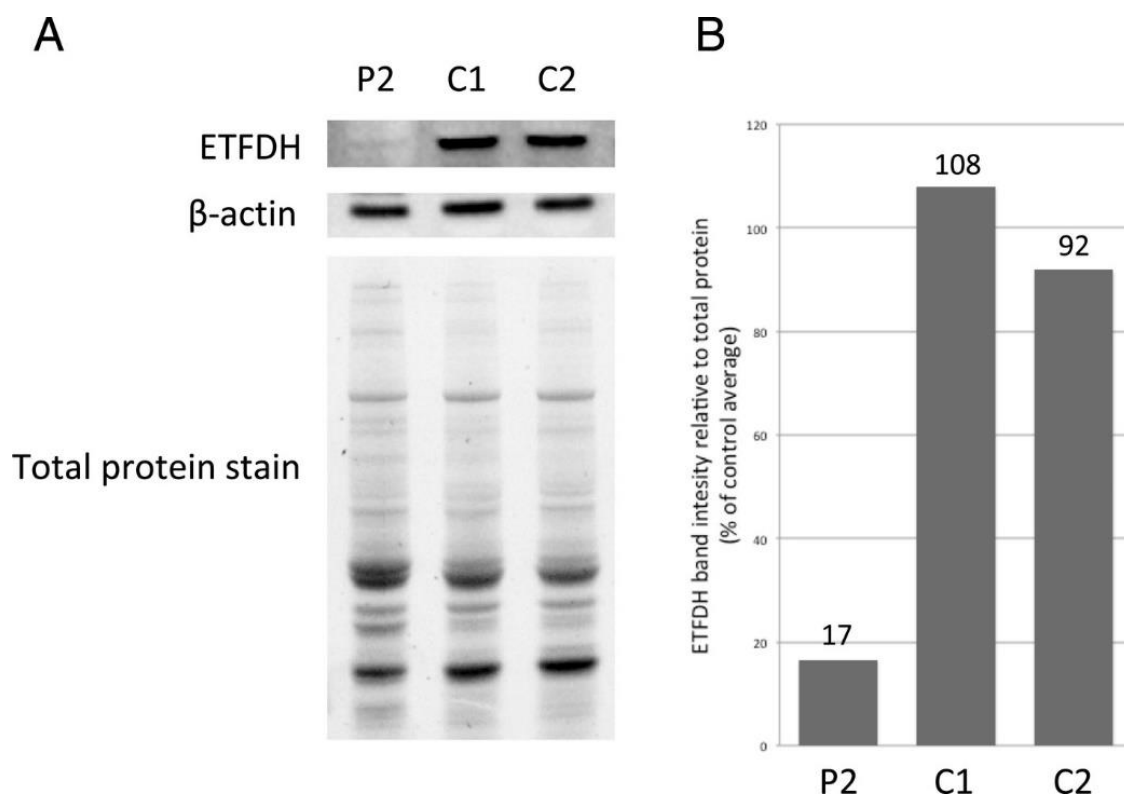


Figure 7.2: Structural instability of muscle ETFDH in Patient 2. An enriched mitochondrial preparation from muscle of Patient 2 (P2, with c.1067G>A+ c.1448C>T variants) and two healthy controls (C1, C2) were separated on SDS-PAGE and immune-stained for ETFDH and β-actin (top, Panel A). The band intensities for ETFDH were normalized to total protein content in each lane (bottom, Panel A), of which the quantified results are shown in Panel B.

7.5 DISCUSSION AND CONCLUSION

Three patients presenting with clinical and metabolic features of MADD were retrospectively investigated to identify the possible genetic cause for the disease. Targeting genes possibly involved, we identified a novel pathogenic variant (c.1067G>A) in *ETFDH* in all three patients. This variant was homozygous in the index patient (Patient 1), who presented with a MADD type I phenotype, and compound heterozygous with the well-documented pathogenic variant c.1448C>T in two other (Patients 2 and 3). Although the trajectories of clinical and further investigations were different, all three patients presented with clear clinical and metabolic features indicative of MADD type I (Patient 1), or type III (Patient 2 and 3). Patient 1, homozygous for the novel c.1067G>A variant, presented in first days of life with the most severe (type I) neonatal onset clinical and metabolic features of MADD, notably with congenital features and being unresponsive to riboflavin treatment. Patient 2, presenting with a multi-systemic neuromuscular features early in life, was initially diagnosed with a combined mitochondrial respiratory chain disorder in her ninth year. A deeper investigation into CoQ₁₀ involvement and metabolomics resulted in the final MADD diagnosis and the genetic outcome presented here. Patient 3 was

diagnosed with MADD early in life via metabolic and functional fatty acid oxidation tests, and were able to follow riboflavin treatment. In this case, having a variant associated with a riboflavin responsive (c.1448C>T) and unresponsive form (c.1067G>A), it is not surprising that this patient showed a partial responsiveness to riboflavin. A closer look at this previously undocumented c.1067G>A variant, indicate a substitution of a highly conserved Glu for Gly at position 356, putatively affecting the FAD domain of ETFDH (Watmough & Frerman, 2010; Zhang *et al.*, 2006). Comparing it to other type I mutations also do not highlight any evident structural similarities that can inform on any molecular basis for its pathogenicity. It is also notably the first associated with this severe phenotype identified in exon 8 of *ETFDH*. We had the limitation of having biological material for only one of the patients for further structural ETFDH analysis (Patient 2, with the compound heterozygous c.1067G>A + c.1448C>T variants). Considering that the homozygous c.1448C>T variant is known to result in significantly reduced ETFDH activity ($\pm$ 30% of WT) and steady state stability (<10% of WT) under low riboflavin conditions (Cornelius *et al.*, 2012), it can be concluded from the structural analysis (Figure 7.2) that the presence of the c.1067G>A variant on the alternative allele to the c. c.1448C>T variant should have a similar impact on the instability of the native ETFDH protein. We propose that this result, in addition to the predictive parameters of the variant and the genetics in these families, support the suggestion that this novel variant can be added to the list of mutations causing a type I MADD phenotype. Finally, in conclusion, from the limited family histories and genetics that could be gathered from these three cases, no relation between these families could as yet be established. Nevertheless, we believe it is reasonable to suspect that a common founder for the c.1067G>A mutation may have originated within this population group. The reasons for this is, firstly, that all three of these patients originate from Caucasian Afrikaner families, which is a southern African population group having descended from Europe since the 17th century; and, secondly, that this mutation has not yet been reported in the better screened European population groups from which the Afrikaner population originated. Considering the possibility that this mutation, in addition to the milder riboflavin-responsive c.1448C>T mutation, may have a relatively high frequency in MADD patients of this population group, we propose that this mutation should be included in screening strategies for families with a history of MADD in this and other related population groups.

CHAPTER 8:

MUTATION SCREENING OF TWO PUTATIVE POPULATION VARIANTS IN THE *GCDH* AND *MPV17* GENES

8.1 INTRODUCTION

In Chapter 2, information regarding the mitochondrion is outlined, including its functional and structural properties as well as the genetic aspects of mitochondrial genes, both mtDNA and nDNA. Chapter 2 also discusses nDNA mutations that cause disruptions in the OXPHOS system, which is primarily responsible for ATP production. Several genes outside the OXPHOS system can contribute to deficiencies within OXPHOS. However, this is only one of numerous direct contributors to MD since gene mutations in other systems, such as glycolysis, the TCA cycle and β -oxidation may also cause MD. Two of the genes involved with MD which lie outside OXPHOS, namely *GCDH* and *MPV17*, are described in the current chapter. Mutations in these genes result in rare diseases, namely glutaric acidemia type I (GA-1) and mtDNA depletion syndrome (MDDS) respectively, for which the diagnosis is often difficult. While GA-1 has a unique set of clinical features and a well-defined metabolic signature, MDDS can present in many different clinical features. The latter is made even more complicated by the heterogeneity of the clinical features associated with MDDS, which often results in the misdiagnosis of patients. Nonetheless, it is crucial to rapidly diagnose patients with either of these diseases in order to start treatment and alleviate symptoms. To this end, major advances have been made with regards to rapid diagnosis via genetic screening. A founder mutation in the gene *CGDH* was first identified by Van der Watt *et al.* (2010) in 11 black South African patients clinically diagnosed with GA-1. Additionally, a founder mutation in the gene *MPV17* was identified by Meldau *et al.* (2018) in 24 black South African patients harbouring MDDS. Based on these two discoveries, South African patients with suspected clinical features of either disease, mainly from the Western Cape and Eastern Cape provinces, are routinely screened for GA-1 and/or MDDS at the NHLS at the *Red Cross Children's Hospital* in Cape Town, South Africa.

In Chapter 3, the cohort of paediatric patients with MD used in this investigation is discussed. As mentioned, the cohort originates from the Northern provinces of South Africa — a part of the country for which no screening for these founder mutations has been performed to date. Considering this, it is important to not only establish genotype-phenotype correlations in patients originating from these provinces, but also to initiate routine screening initiatives for these founder mutations at several other facilities in Gauteng, Kwa-Zulu Natal and the North-West Province.

Here, two cases who harboured a homozygous *GCDH* mutation (c.877G>A, p.A293T), and two cases who harboured the same heterozygous mutation, are described. None of the patients from this cohort harboured a homozygous *MPV17* mutation. Furthermore, the screening processes for both mutations are described here.

8.2 BACKGROUND ON FOUNDER MUTATION ON THE GENE *GCDH*, KNOWN TO BE INVOLVED WITH GLUTARIC ACIDEMIA TYPE I

8.2.1 Glutaric acidemia type I

The MM enzyme, glutaryl-CoA dehydrogenase (EC1.3.99.7, *GCDH*), is mainly responsible for catalysing the oxidative decarboxylation of glutaryl-CoA to crotonyl-CoA in the catabolic pathway of the ketogenic amino acids L-lysine, L-hydroxylysine, and L-tryptophan (Boy *et al.*, 2017), as illustrated in Figure 8.1. The gene *GCDH*, roughly 45kDa in size, uses electron transfer flavoprotein as its electron acceptor and is ubiquitously expressed in the liver. However, recent publications demonstrate that the kidneys might also be involved in long term disease (Kölker *et al.*, 2015). Mutations in this gene result in GA-1 (OMIM 231670), an autosomal recessive inherited neurometabolic disorder. GA-1 has been observed in over 200 cases worldwide with a global prevalence of 1 in 110 000 newborn children (Boy *et al.*, 2017; Hedlund *et al.*, 2006; Lindner *et al.*, 2004; Strauss & Morton, 2003; Strauss *et al.*, 2003).

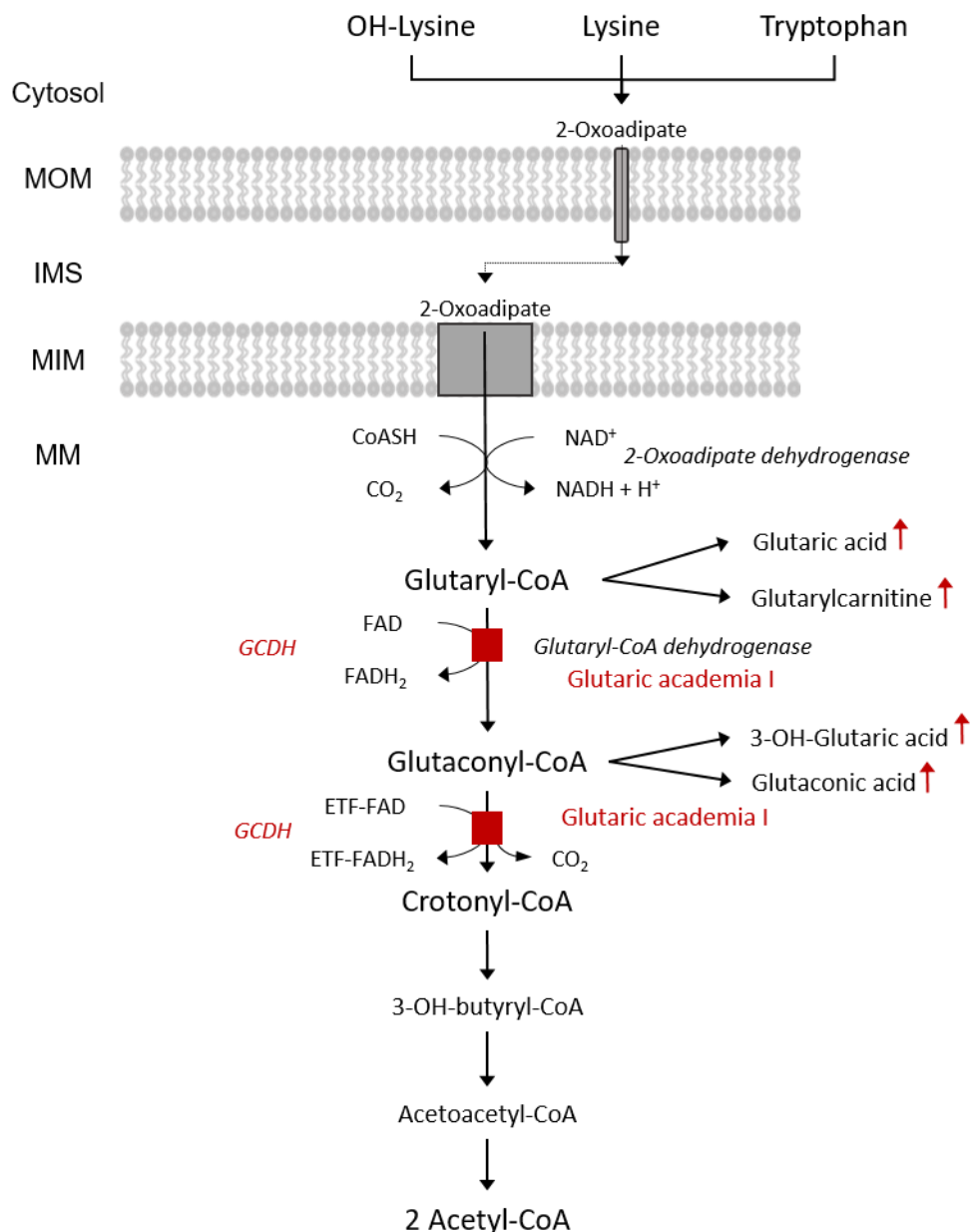


Figure 8.1: Simplified illustration of the L-lysine and L-tryptophan catabolic pathway. Lysine and tryptophan enter into the mitochondrial matrix from the cytosol, and is catabolised by glutaryl-CoA dehydrogenase to ultimately form two acetyl-CoA molecules. Mutations in the gene *GCDH* directly impact the functioning of glutaryl-CoA dehydrogenase, thereby preventing glutaryl-CoA catabolism to crotonyl-CoA and subsequently results in elevated concentrations of glutaric acid, glutarylcarnitine, 3-OH-glutaric acid, and glutaconic acid (indicated by red arrows) in urine. Abbreviations: MOM: mitochondrial outer membrane; IMS: intermembrane space; MIM: mitochondrial inner membrane; MM: mitochondrial matrix; NAD⁺: nicotinamide adenine dinucleotide — oxidised; NADH: nicotinamide adenine dinucleotide — reduced; FAD: flavin adenine dinucleotide; FADH₂: flavin adenine dinucleotide — reduced; ETF-FAD: electron transfer flavoprotein-flavin adenine dinucleotide; ETF-FADH: electron transfer flavoprotein-flavin adenine dinucleotide — reduced.

Biochemically, GA-1 is usually detected with routine newborn screening, genetic screening, and metabolomics analyses of urine, plasma, and/or cerebrospinal fluids using GC-MS, and/or electrospray-ionisation tandem mass spectrometry, but can easily be misdiagnosed for similar clinical phenotypes (Kölker *et al.*, 2011; Lindner *et al.*, 2004). During metabolomics analysis, elevated GA, 3-hydroxyglutaric acid (3-OH-GA), and glutaconic acid is observed in all three bodily fluids previously mentioned (Boy *et al.*, 2017; Goodman *et al.*, 1998). These elevated metabolites which are a characteristic of GA-1, often accumulate and become toxic, resulting in brain damage. Furthermore, these mutations cause macrocephaly, muscular hypotonia and often developmental delay in neonates and infants older than three months (Hedlund *et al.*, 2006). If detected early enough, GA-1 can be treated. However, this is very difficult as signs and symptoms often occur only after an encephalopathic crisis that causes irreversible neurological damage (Thomas *et al.*, 2018). Decreasing dietary protein greatly decreases the production of GA and was first suggested by Goodman *et al.* (1975). Recent studies showed effective treatment by following a low protein diet as well as L-carnitine supplementation (Thomas *et al.*, 2018). Monitoring of patients with GA-1 is, however, of the essence and should not be neglected.

In South Africa, the diagnosis of rare diseases such as GA-1 amongst children, especially those living in rural areas, is challenging due to limited diagnostic centres and infrastructure (Meldau *et al.*, 2016; Van der Westhuizen *et al.*, 2015). However, in recent years significant progress has been made in diagnosing GA-1, especially in African populations by Van der Watt *et al.* (2010) and Thomas *et al.* (2018), who detected a population-specific mutation in African populations of the Western Cape province in South Africa. In total, 23 cases with GA-1 presented with the same disease-causing mutation in the *GCDH* gene, where an alanine substitutes for threonine at position 293 (c.877G>A, p.A293T), and which has to date only been observed in African populations. According to Van der Watt *et al.* (2010), the carrier rate for this mutation is 1 in 36, with a prediction score of 1 in 5184 newborn children in African populations.

A cohort of predominantly African paediatric patients originating from the northern parts of South Africa were retrospectively screened for the same mutations as reported by Van der Watt *et al.* (2010) and Thomas *et al.* (2018), and are described here.

8.2.2 Methods

8.2.2.1 Patient selection and samples

At the time of screening, the cohort consisted of 218 patients, all of whom were clinically referred to *Steve Biko Academic Hospital* for clinical evaluations. Mutational screening analysis was performed on the gDNA of 116 patients with a confirmed MD after clinical evaluation and a

positive biochemical analysis (enzymology of RC complexes — see Section 3.4 for more information). These patients are numbered with the prefix “S” followed by the number, for example S001 (as described in Section 4.7). Limited sample availability prevented screening of 11 patients with an MD. Those patients who did not present with a biochemical RC deficiency are numbered with the prefix “P” followed by the number, for example P001.

8.2.2.2 Metabolic and biochemical investigation

A number of patients (n = 145) in our cohort underwent metabolic screening (organic acid and/or aminoacid), as described by Reinecke *et al.* (2012). Of the 145 patients, 50 had normal metabolic urine profiles and 95 had abnormalities in detected urine. No metabolic data were available for 73 patients due to insufficient sample quantity or quality. Biochemical enzyme analysis of CI–CIV was performed using 600 x g muscle supernatant, and normalised against CS, as previously discussed in Section 3.4.

8.2.2.3 Genetic screening analysis

The p.A293T mutation was detected by PCR amplification followed by restriction fragment length polymorphism (RFLP). Briefly, amplification was done using a mismatch primer designed to introduce an acyl restriction enzyme cutting site in the absence of the mutation. DNA concentrations ranged from 2–200 ng/μL and were stored in low TE Buffer [10 mM Tris-HCl (pH 8.0)]. The forward and reverse primers’ sequences are listed in Table 8.1, and result in a 104 bp PCR product.

Table 8.1: Forward and reverse primers used for PCR amplification for the gene GCDH.

Primer name	Primer sequence	PCR size
<i>GCDH</i> 5754-Forward	5'-CCTTCGGCTGCCTGAACGAC-3'	104 bp
<i>GCDH</i> 5858-Reverse	5'-ACACCTGTCGAGGGCGTACTG-3'	

DNA was amplified using SuperTherm Gold polymerase from Separation Scientific (STG, CAT #JMR-851, Johannesburg, South Africa) in 50 μL reactions. The optimal annealing temperature for the primers was at 62°C and a total of 35 cycles of 30 sec each was used. The PCR products were stored at 4°C prior to RFLP. RFLP was done using the restriction enzyme *BsaH1* (Cutting site: GRCGYC, New England Biolabs, CAT #R0556S, MA, USA). For digestion, 15 μL PCR product was used, 0.5 μL *BsaH1* (1 unit), and 2.5 μL 10x NEBuffer (1 unit) in 20 μL reactions. Incubation was done at 37°C for three hours using a thermal cycler.

Digestion of samples without mutations (wildtype) results in two fragments of 19 bp and 85 bp respectively, while in the presence of a mutation the recognition site is lost and results in an undigested fragment size of 104 bp (as indicated in Figure 8.2). The products of restriction enzyme digestion were separated on a 4% (w/v) agarose gel in TBE buffer, stained with 0.5 µg/mL EtBr, followed by visual inspection under UV-light. An O'RangeRuler, 100 bp ladder was used (CAT #SM0623, ThermoFisher, MA, USA). Digestion of samples was repeated and interpreted as: 1) a prominent 104 bp band indicating a homozygous mutation; 2) three prominent bands at bp positions 104, 85 and 19 indicating a heterozygous mutation; or 3) faint or no bands indicating unsuccessful restriction enzyme digestion. Samples that presented with either a homozygous or heterozygous mutation were confirmed by Sanger sequencing (Inqaba Biotech, Pretoria, South Africa). The Sanger sequencing data were aligned against the *GCHD* reference (NG_009292.1 and NC_000019.8) using CLC Genomics workbench (v11, Qaigen CA, USA).

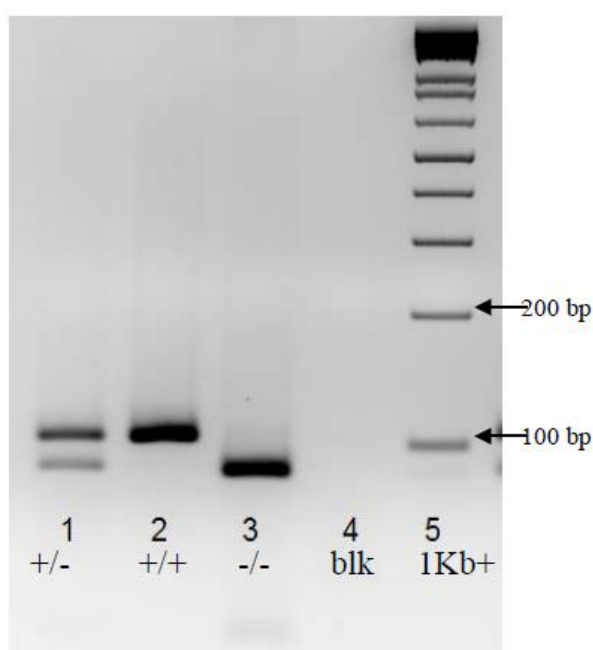


Figure 8.2: Agarose RFLP results for *BsaH1* digestion. Examples of a heterozygous (+/-) mutation in Lane 1, homozygous (+/+) mutation in Lane 2 and the wildtype (-/-) in Lane 3, using a 100 bp ladder. Three bands are visible for heterozygous mutations at 104 bp, 85 bp and 19 bp (the latter of which is not visible in this example). Only one band of 104 bp is visible for homozygous mutations since the recognition site is lost. The wildtype samples will result in two bands at 85 bp and 19 bp (the latter of which is not visible in this example). Abbreviations: blk: Blank; 1Kb+: 10 000bp ladder; bp: base pair.

8.2.3 Results

8.2.3.1 Mutation analysis

Restriction length fragment polymorphism analysis was done for 116 patients (See Table 4.6). The digestion agarose gel electrophoresis results are given in Figure 8.3–Figure 8.7. Digestion with the enzyme was repeated for the four patients that presented with either a homozygous or heterozygous mutation to confirm the initial digestion results (see Figure 8.7), and Sanger sequencing was done to confirm the mutations detected in the four patients (see Figure 8.8).

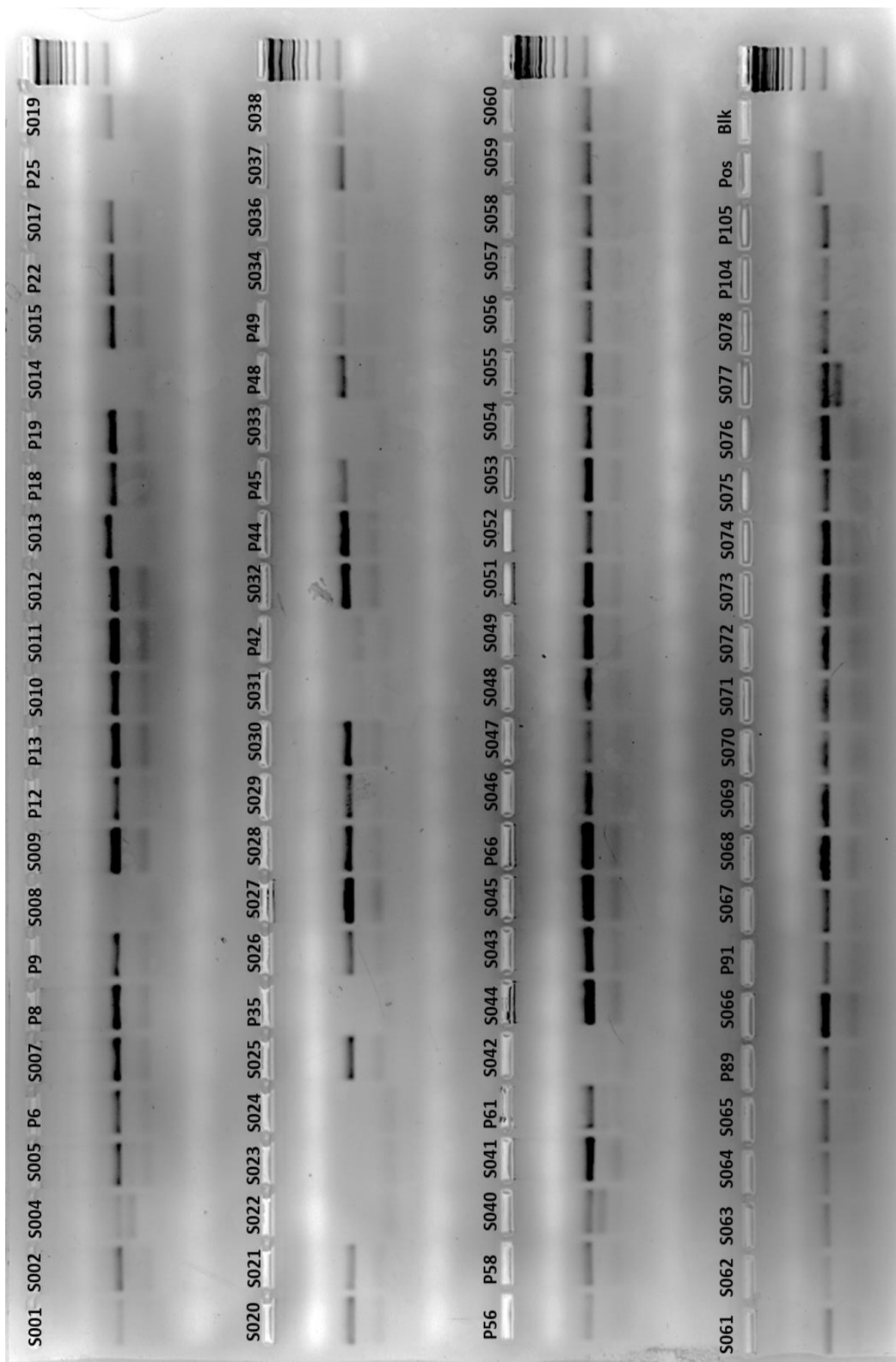


Figure 8.3: Image of agarose gel with restriction enzyme digested in samples P006–P105 and S001–S078. A 100 bp DNA ladder was used.
Abbreviations: Pos: positive control; Blk: blank.

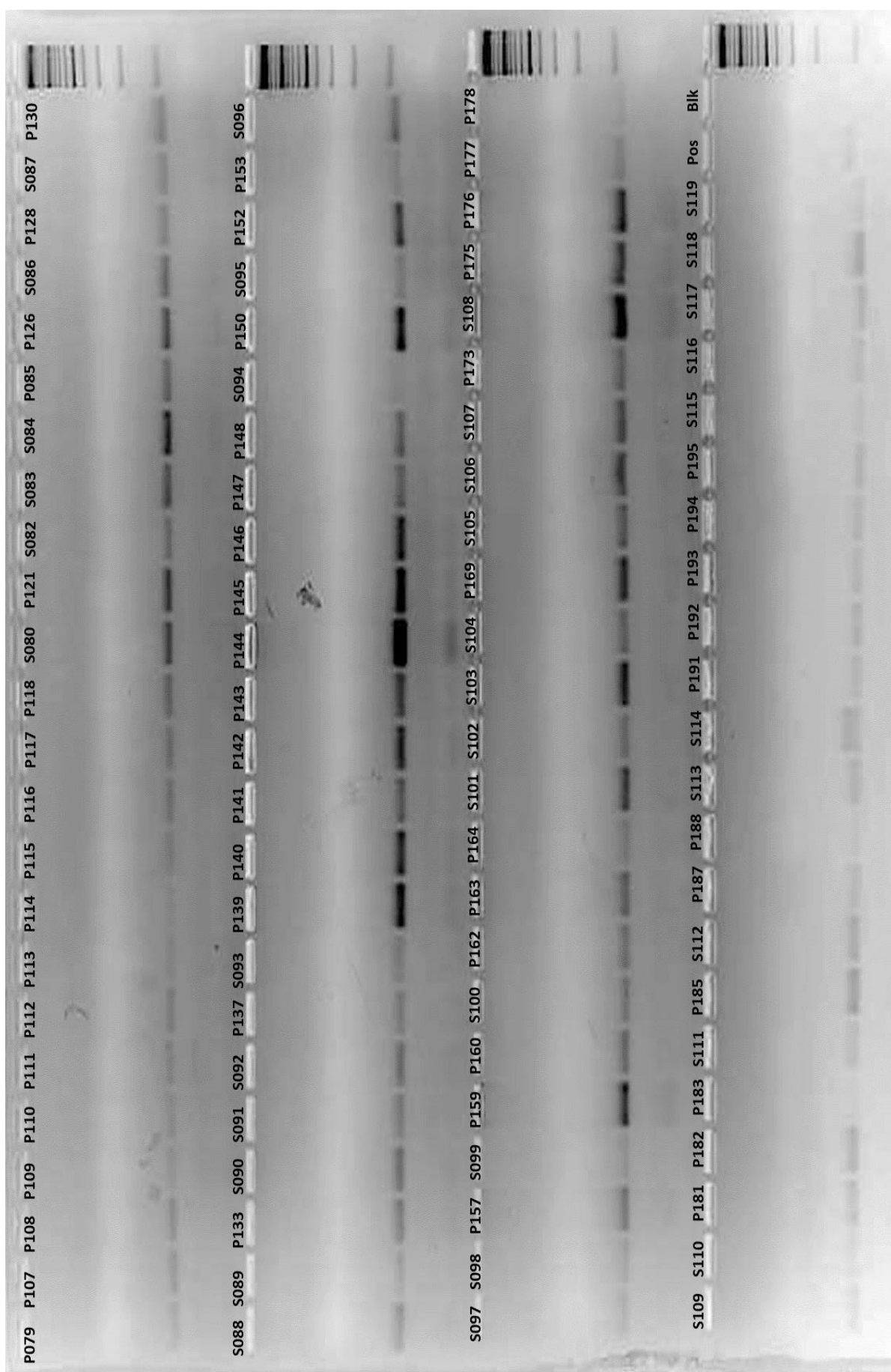


Figure 8.4: Image of agarose gel with restriction enzyme digested samples P079–P195 and S080–S119. A 100 bp DNA ladder was used. Abbreviations: Pos: positive control; Blk: blank.

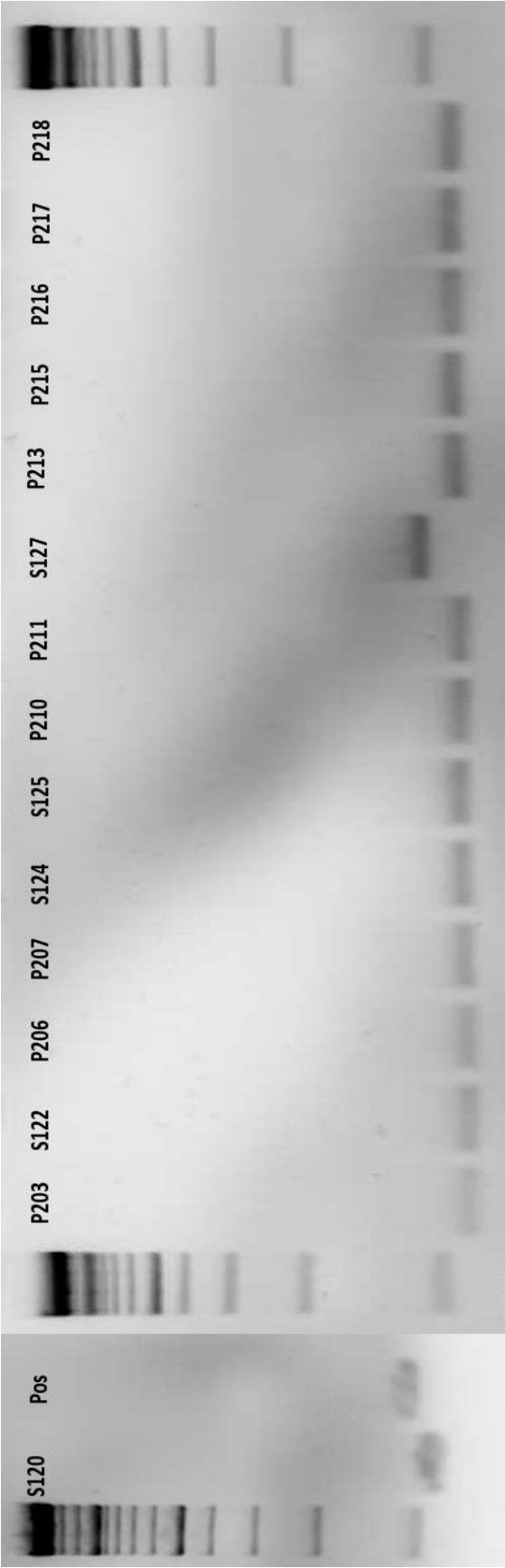


Figure 8.5: Image of agarose gel with restriction digested DNA samples P203–P218 and S120–S127. A 100 bp DNA ladder was used.
Abbreviations: Pos: positive control.

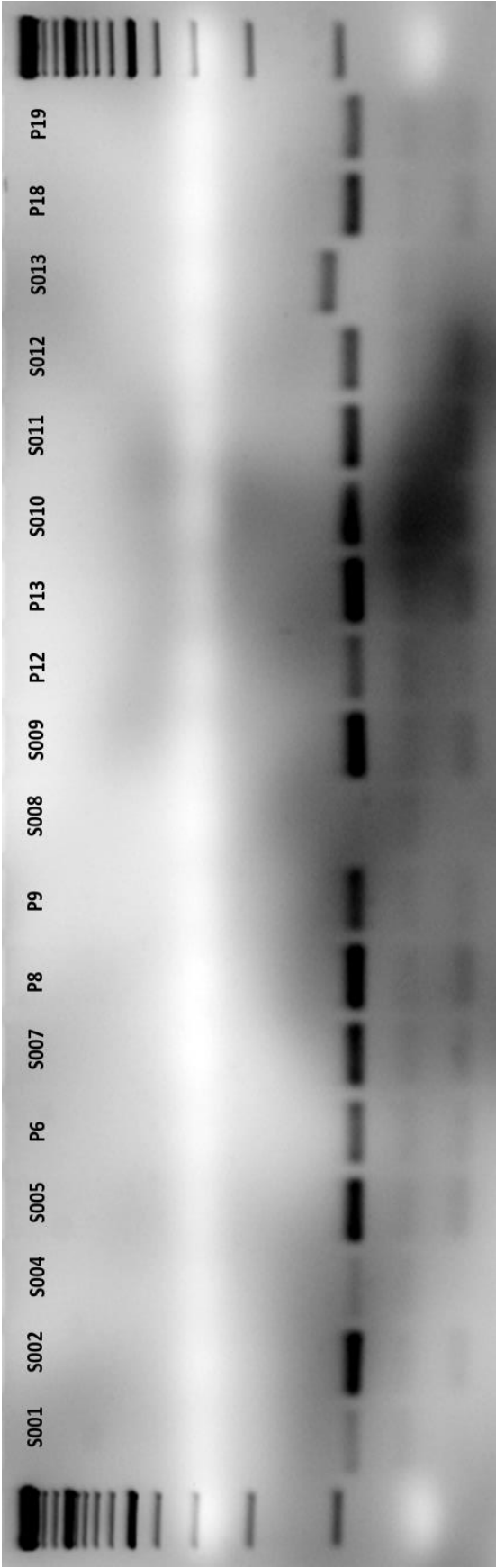


Figure 8.6: Image of agarose gel with restriction digested DNA samples that were repeated. The samples were P6–P19 and S001–S013. A 100 bp DNA ladder was used.

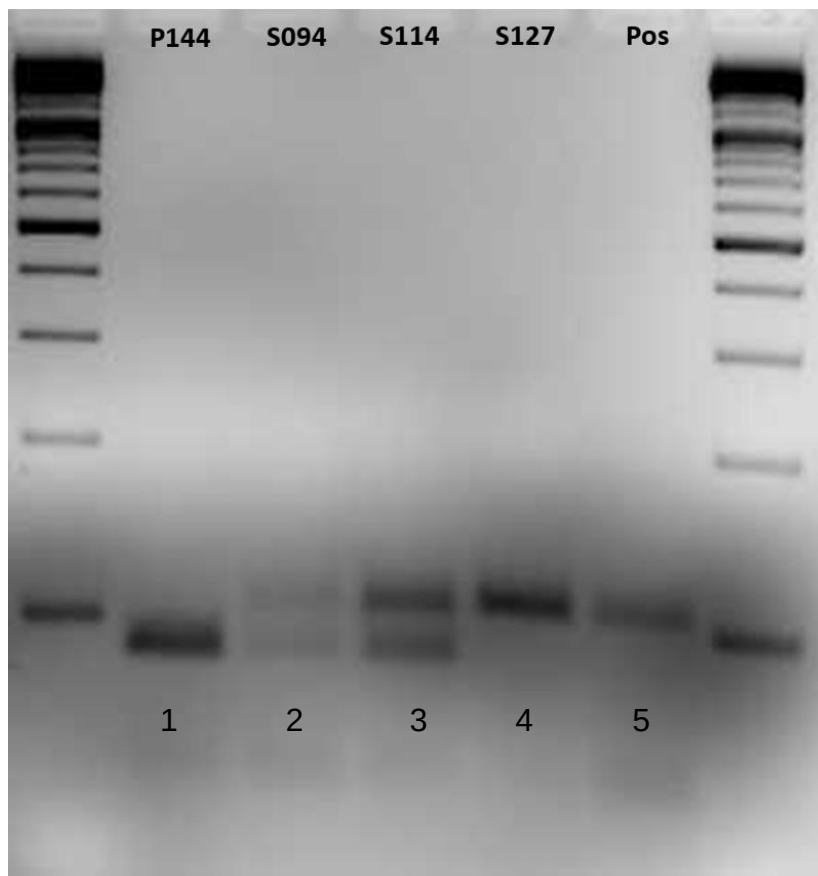


Figure 8.7: Image of agarose gel with digested DNA samples using a 100 bp ladder. Lane 1 is a repeat of sample P144 with no clear mutation. Lanes 2 and 3 are heterozygotes for the mutation (S094 and S114). Lane 4 (S127) is a homozygote for the mutation. Lane 5 is the positive control sample. Abbreviations: Pos: positive control.

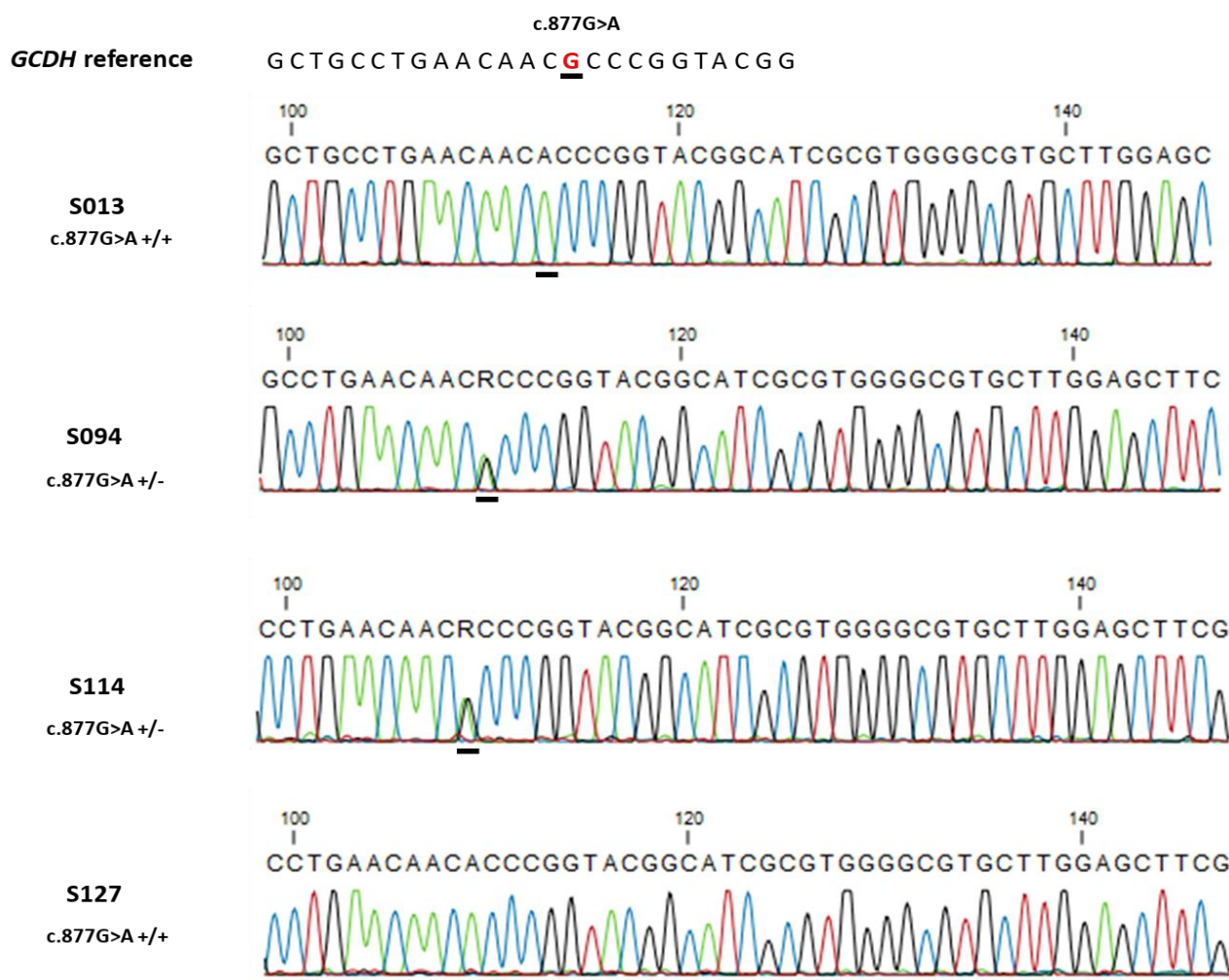


Figure 8.8: Partial Sanger sequence validation data for the four patients is shown in alignment with the reference sequence for the gene *GCDH* (ENSG00000105607) at the position for the c.877G>A mutation. Patients S013 and S127 presented with homozygous mutations while S094 and S114 presented with heterozygous mutations.

8.2.4 Case descriptions

The p.A293T (c.877G>A) disease-causing mutation for GA-1 was detected as homozygous in both alleles of two non-related African females (S013 and S127), and as heterozygous in another two non-related African females (S094 and S114). Table 8.2 summarises the clinical, metabolic, biochemical and genetic investigations for these four patients.

Table 8.2: Summary of the clinical, metabolic and biochemical information of patients.

	S013	S094	S114	S127
Age, race, gender and haplogroup				
	<1, A, F, L3e3b1	1–2, A, F, L2a1b1	6–10, A, F, L0d2a1a	1–2, A, F, L2a1b1a
Main clinical manifestations				
	DR, CNS, MI, RB, macrocephaly	DD, CNS	DD, MI	DD, CNS, MI, RB
Metabolic analysis profiles				
Organic acid	Abnormal GA ↑ 3-OH-GA ↑	Normal	Normal	Abnormal GA ↑ 3-OH-GA ↑
Carnitine	Abnormal glutaryl carnitine ↑↑	Abnormal free carnitines ↓↓↓	Normal	No information available
Amino acid	Normal	Normal	Abnormal amino acids	No information available
Carbohydrate	Abnormal trace amounts of glucose and lactose	Normal	Normal	No information available
RC enzyme analysis				
	CIII, CII+CIII	CI	CI, CIV	CI
Genetic analysis				
	c.877G>A, p.A293T +/+	c.877G>A, p.A293T +/-	c.877G>A, p.A293T +/-	c.877G>A, p.A293T +/+

Abbreviations: DD: developmental delay; DR: developmental regression; CNS: central nervous system; MI: muscle involvement; RB: radiological changes in the brain; A: African; F: female; GA: glutaric acid; 3-OH-GA: 3-hydroxyglutaric acid; RC: respiratory chain; CI–CIV: complex I–IV; ↑: elevated; ↑↑: highly elevated; ↓↓↓: severely decreased; +/+ : homozygous; +/-: heterozygous.

8.2.4.1 Case 1: Patient S013

At the age of 18 months, Patient S013 presented with a concern about neuro-regression. Although she was able to sit and crawl at eight months, she experienced flu-like symptoms and lost the ability to sit and crawl. Furthermore, she had macrocephaly (see Figure 8.9), hypotonia, muscle weakness, bilateral ankle clonus, no head control, and dystonic posturing of the upper limbs. Her progress over time was marked by progressive weakness and hypotonia. She did not have any episodes of decompensation and her renal function and muscle histology was normal. A MRI of the brain showed bilateral increased signal in the putamina, caudate nuclei and subcortical white matter in T2. Metabolic profiles revealed elevated GA, 3-OH-GA and severely elevated glutaryl carnitine:GA ratios in urine. Biochemically she presented with CIII and CII+CIII RC deficiency. As this patient presented clinically, biochemically, and genetically with GA-1, it was concluded that the homozygous mutation identified in this patient was indeed the cause for this

patient's phenotype. Segregation analysis can be done to confirm the pathogenicity, but is not necessary.

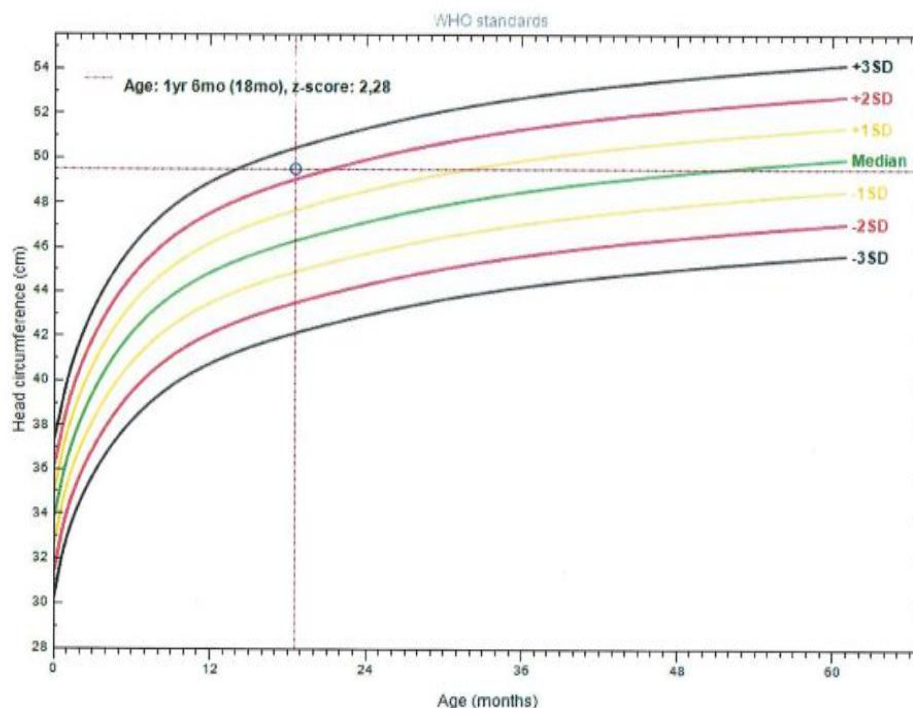


Figure 8.9: World health organization head circumference-for-age. Patient S013 presents with a Z-score of 2.28 and is in the ~95th percentile. This indicates macrocephaly for this patient at her age (World Health Organization, 2018).

8.2.4.2 Case 2: Patient S127

At the age of two years, Patient S127 presented with right focal dystonia of mostly the arm, central hypotonia and developmental delay. Her reflexes were difficult to elicit but presented with hypermobility in the joints. Furthermore, her dystonia increased over time with the right leg becoming more involved. The tongue also became involved and speech was affected. No episodes of decompensations were observed. She later gained milestones that allowed her to attend crèche, where she started to draw and also started to walk with a splint on her right leg. She still had head lag, but gross motor function classification system of level II. She had a positive Babinski on the left side and striatal toe on the right side. Her head circumference was within the normal range, indicating an absence of macrocephaly (see Figure 8.10). Leigh disease is suspected as she presented with asymptomatic GA-1 features. Her metabolic profile, however, is indicative of GA-1, as she demonstrated elevated urinary levels of GA, and 3-OH-GA. Her genetic profile revealed a homozygous p.A293T mutation, also indicating GA-1 in this patient. At the time of this investigation, her clinical profile, however, does not correlate completely. She might present

with late-onset GA-1, and should therefore be monitored frequently. She is on a low lysine diet as recommended (Boy *et al* 2017).

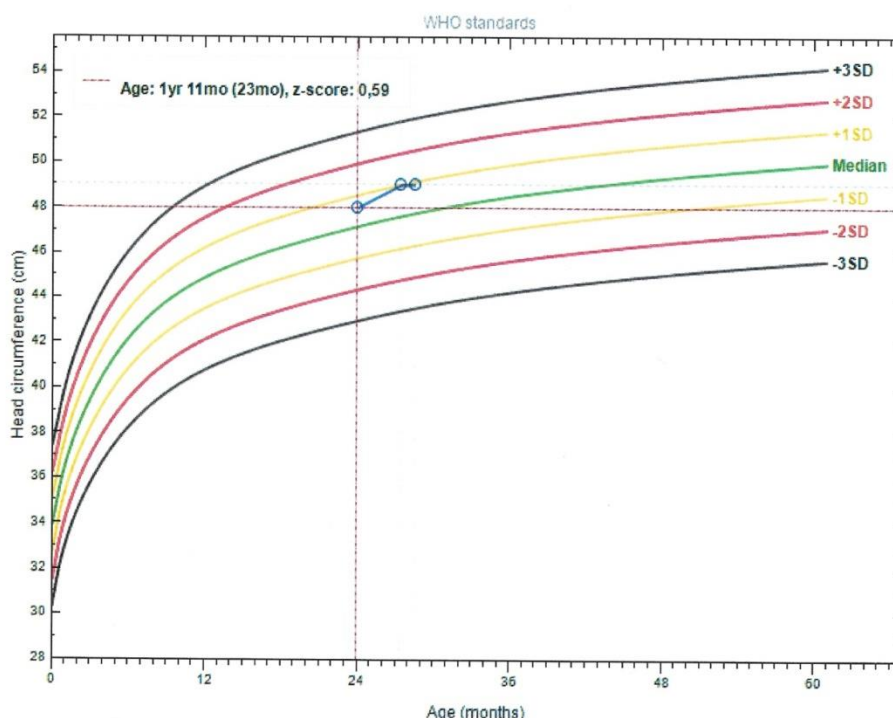


Figure 8.10: World health organization head circumference-for-age. Patient S127 presents with a Z-score of 0.59 indicating a normal head circumference for her age (World Health Organization, 2018).

8.2.4.3 Case 3: Patient S094

Patient S094 had a clinical phenotype typical of patients with GA-1, which included signs of developmental delay as well as CNS involvement with ataxia and cerebellar symptoms and extrapyramidal/pyramidal symptoms, with the absence of macrocephaly. Furthermore, an atypical metabolic profile was observed since organic acid analysis, using GC-MS, revealed that GA and/or 3-OH-GA were within normal ranges. Carnitine analysis revealed severely decreased free carnitines, which are necessary for detoxification, and could lead to decreased energy production. It is often recommended that L-carnitine therapy is administered to patients with low free carnitine levels in urine. The fatty acid oxidation of this patient should be investigated as carnitines are responsible for the transfer of long-chain fatty acids from the cytoplasm into the MM for oxidation (Longo *et al.*, 2006). As this patient presented with a heterozygous mutation, it was concluded that the mutation identified here is not the cause of disease. Segregation analysis should confirm which of the patient's parents carries the mutant allele.

8.2.4.4 Case 4: Patient S114

Patient S114 presented with myopathy and weakness in muscles, however no signs of macrocephaly were seen, which contradicts typical GA-1 phenotypes. At the time of investigation, this patient did not have metabolic data available since not all patients evaluated at the *Steve Biko Academic Hospital* undergo metabolic analysis. Although she presented with a heterozygous mutation, it was concluded that it is not a cause for disease due to its autosomal recessive inheritance pattern. However, expanded genetic analysis and metabolic investigations should be considered for this patient in order to find other disease-causing variants.

8.3 BACKGROUND ON FOUNDER MUTATION ON THE GENE MPV17, KNOWN TO BE INVOLVED WITH MITOCHONDRIAL DEPLETION SYNDROME

8.3.1 Mitochondrial DNA depletion syndrome

Mitochondrial DNA depletion syndrome (MDDS) is an autosomal recessive disease with diverse clinical manifestations, caused by mutations in nuclear genes usually involved with mtDNA maintenance or replication (e.g. *C10orf2*, *DGUOK*, *MPV17*, *RRM2B*, *SUCLA2*, *SUCGL1*, *SLC25A4*, *OPA1*, *TK2*, *TYMP*, *POLG*, *POLG2*, and *PEO1*) (Copeland, 2012; Löllgen & Weiher, 2015). It can manifest early in life and often leads to death in infants and children since multiple organs can be affected due to its heterogeneous nature. Some of the manifestations or syndromes include hepatocerebral MDDS, Alpers-Huttenlocher syndrome, myopathic MDDS, encephalomyopathic MDDS, and mitochondrial neurogastrointestinal encephalomyopathy. The first manifestation, hepatocerebral MDDS, affects the liver and its function, but can also result in other clinical manifestations, which include psychomotor delay, hypotonia, and neurological dysfunction. Mutations in the genes *POLG*, *C12orf2*, *DGUOK* and *MPV17* are known causes for hepatocerebral MDDS (Ferrari *et al.*, 2005). Alpers-Huttenlocher syndrome causes hepatic and neurological dysfunction and has been linked to the gene *POLG*, with 54 mutations previously reported to be pathogenic (Naviaux & Nguyen, 2004). Patients with myopathic MDDS, which has been associated with mutations in *POLG*, *RRM2B* and *TK2* (Bourdon *et al.*, 2007), typically presents with hypotonia, muscle weakness and failure to thrive (Buchaklian *et al.*, 2012; Saada *et al.*, 2001). Encephalomyopathic MDDS causes hypotonia, muscle weakness, psychomotor delay, sensorineural hearing loss, lactic acidosis and neurological dysfunctions due to mutations in the genes *RRM2B*, *TK2*, and *SUCLA2* (Lesko *et al.*, 2010). Lastly, mitochondrial neurogastrointestinal encephalomyopathy causes gastrointestinal dysmotility, ptosis and neurological dysfunction (Nishino *et al.*, 1999; Shaibani *et al.*, 2009).

Meldau *et al.* (2018) reported on a population-specific mutation in the gene *MPV17* that causes mitochondrial neurohepatopathy in African children. The Mpv17 protein is situated on the MOM and acts as a non-selective channel. It is responsible for regulating trans-membrane potential in order to maintain mitochondrial homeostasis as well the production of ROS (Antonenkova *et al.*, 2015; Meldau *et al.*, 2018). Under normal physiological conditions, the channel opens partially, alleviating excess potential which limits excessive ROS production. However, if this important modulator is compromised due to mutations, it results in excessive ROS production and the accumulation thereof. This leads to mtDNA instability and depletion, a pathophysiological condition known as MDDS (as illustrated in Figure 8.11). Mutations in this gene cause severe clinical manifestations that include failure to thrive, hypotonia, hepatosplenomegaly, and jaundice. A novel homozygous nonsense mutation (c.106C>T, p.Gln36Ter) in the gene *MPV17* was identified in three Black-African patients by Meldau *et al.* (2018), who also reported an estimated carrier frequency for this population at 1 in 68 and a newborn incidence of 1 in 18 496.

A cohort of predominantly African paediatric patients originating from the northern parts of South Africa were retrospectively screened for the same mutations as reported by Meldau *et al.* (2018), and is described here.

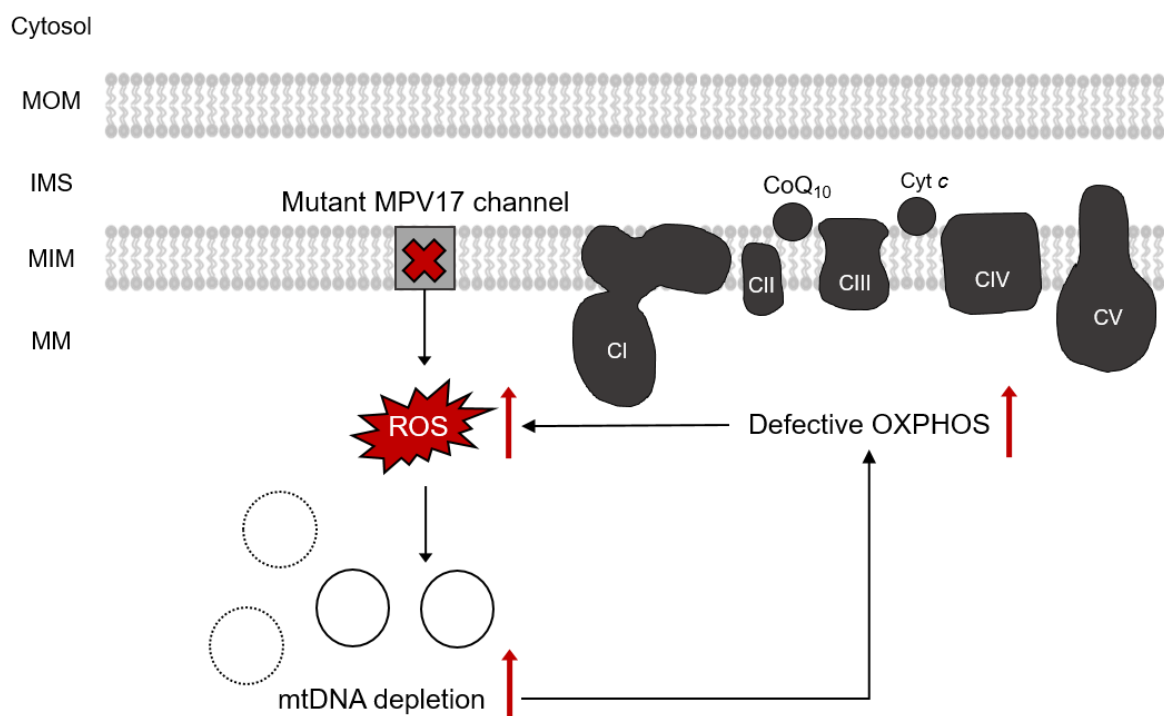


Figure 8.11: Distributed consequences when Mpv17 is compromised due to a mutation in the gene *MPV17*. Mutations in the modulator responsible for reactive oxygen species (ROS) production cause accumulation of ROS that alters mtDNA content. The depleted amount of mtDNA affects the oxidative phosphorylation system directly, which in turn produces more ROS. Abbreviations: CI–CV: complexes I–V; CoQ₁₀: coenzyme Q₁₀; Cyt c: cytochrome C; ROS: reactive oxygen species; mtDNA: mitochondrial DNA; OXPHOS: oxidative phosphorylation; MOM: mitochondrial outer membrane; IMS: intermembrane space; MIM: mitochondrial inner membrane; MM: mitochondrial matrix.

8.3.2 Methods

8.3.2.1 Patient selection and samples

The same approach was followed for patient selection as described in Section 8.2.2.

8.3.2.2 Metabolic and biochemical investigation

The same approach for metabolic and biochemical investigations was followed as described in Section 8.2.2.

8.3.2.3 Genetic screening analysis

The reported c.106C>T variant, as described by Meldau *et al.* (2018), was detected by PCR amplification, followed by RFLP. Briefly, two primer sets (listed in Table 8.3) were amplified using SuperTherm Gold polymerase from Separation Scientific (STG, CAT #JMR-851, Johannesburg, South Africa) in 50 µL reactions. The gDNA input ranged from 2–200 ng/µL.

Table 8.3: Forward and reverse primers used for PCR amplification for the gene MPV17.

Primer name	Primer Sequence	PCR size
MPV17-forward	5'-CTCCCTCCTTGGAATGGCAG-3'	336
MPV17-reverse	5'-TCCCTGCCCAAGTATCAGA-3'	

The optimal annealing temperature for the primers was at 58°C and a total of 35 cycles of 30 sec each was used. The PCR products were stored at 4°C prior to RFLP. RFLP was done using the restriction enzyme *PvuII* (Cutting site: CAGCTG, New England Biolabs, CAT R0151S, MA, USA). For digestion, 15 µL PCR product was used, 0.5 µL *PvuII* (1 unit), and 2.5 µL 10x NEBuffer (1 unit) in 20 µL reactions. Incubation was done at 37°C for two and a half hours using a thermal cycler.

Digestion of samples without mutations (wildtype) results in two fragments of 244 bp and 96 bp respectively, while in the presence of a mutation the recognition site is lost and results in an undigested fragment size of 336 bp (Figure 8.12). The products of restriction enzyme digestion were separated on a 3% (w/v) agarose gel in TBE buffer stained with 0.5 µg/mL EtBr, followed by visual inspection under UV-light. An O'RangeRuler, 100 bp ladder was used from ThermoFisher (CAT #SM0623, ThermoFisher, MA, USA). Digestion of samples was repeated

and interpreted as: 1) a prominent 336 bp band indicating a homozygous mutation; 2) three prominent bands at bp positions 336, 244 and 96 indicating a heterozygous mutation; or 3) faint or no bands indicating unsuccessful restriction enzyme digestion.

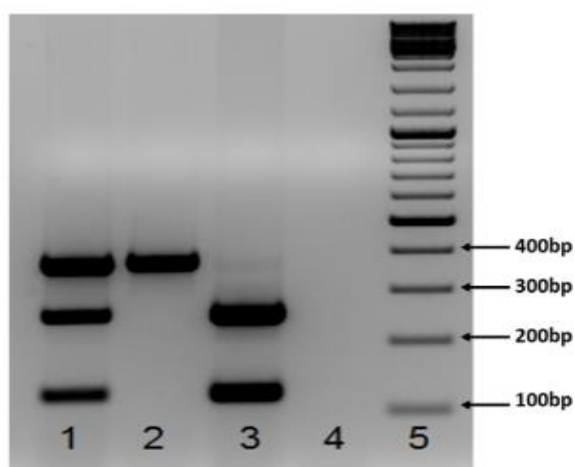


Figure 8.12: Agarose RFLP results for *PvuII* digestion. Example for heterozygous (+/-) mutation in Lane 1, homozygous (+/+) mutation in Lane 2 and the wildtype (-/-) in Lane 3, using a 100 bp ladder. Three bands are visible for heterozygous mutations at 336 bp, 244 bp and 96 bp. Only one band of 336 bp is visible for homozygous mutations, since the recognition site is lost. The wildtype samples will result in two bands at 244 bp and 96 bp. Abbreviations: bp: base pair.

8.3.3 Results

The p.Gln36Ter (c.106C>T) homozygous mutation for mitochondrial neurohepatopathy MDDS was undetected in all 116 patients.

8.3.3.1 Mutation analysis

Restriction length fragment polymorphism analysis was done for 116 patients (patients (See Table 4.6). The digestion agarose gel electrophoresis results are given in Figure 8.13–Figure 8.15. Digestion with the enzyme was repeated for 12 patients that presented with either a homozygous or heterozygous mutation, or those that presented with faint bands, to confirm the initial digestion results (see Figure 8.7). Only one patient (S028) initially presented with a homozygous mutation, but the repeat analysis confirmed this as a wild-type and not a mutation (Figure 8.16). Furthermore, throughout the initial digestion, faint bands can be seen at roughly 50bp in Figure 8.13–Figure 8.14, which represent primer dimers that should not be confused with DNA bands indicating a heterozygous mutation.

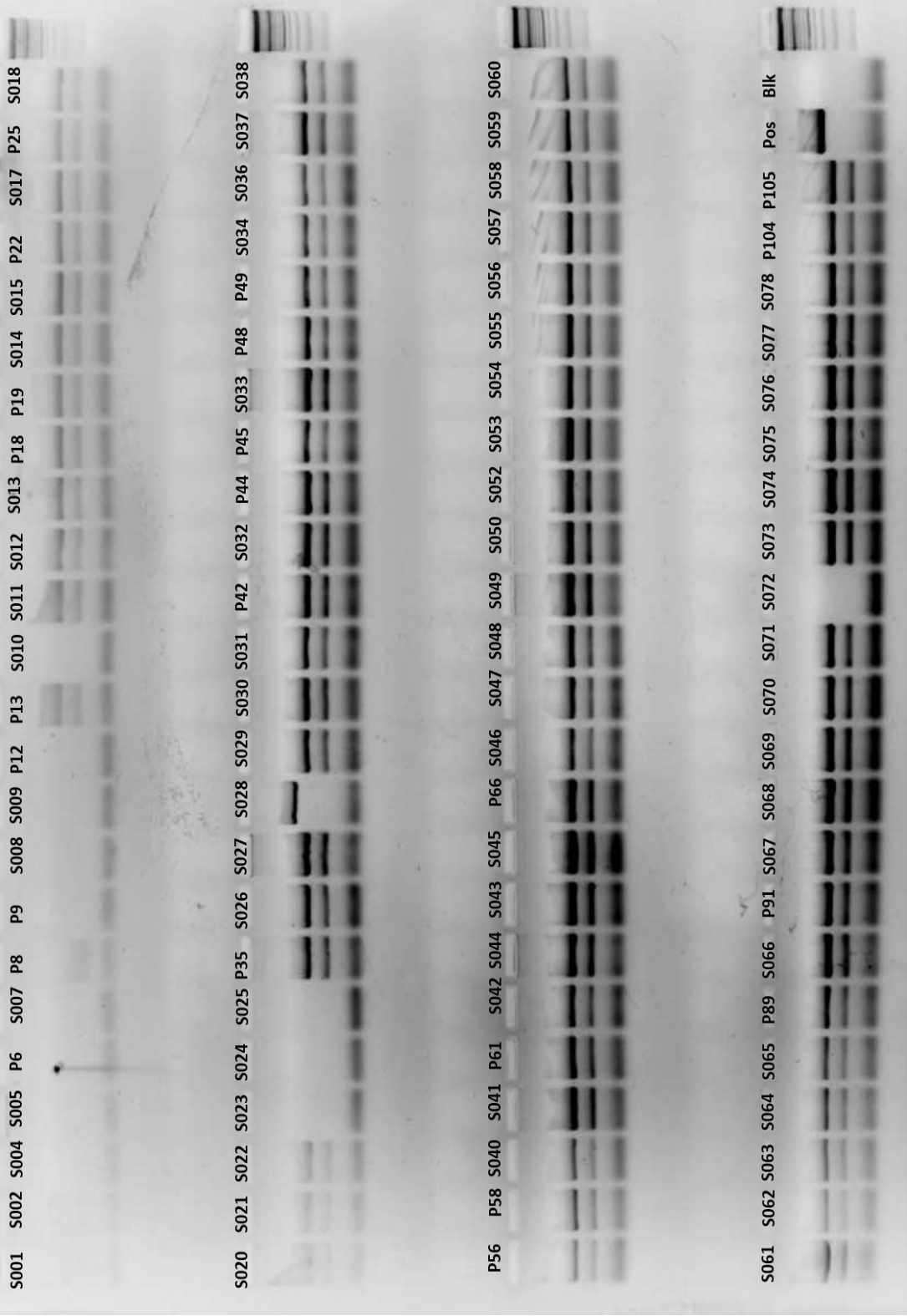


Figure 8.13: Image of agarose gel with restriction enzyme digested samples P006–P105 and S001–S078. A 100 bp DNA ladder was used. Abbreviations: Pos: positive control; Blk: blank.

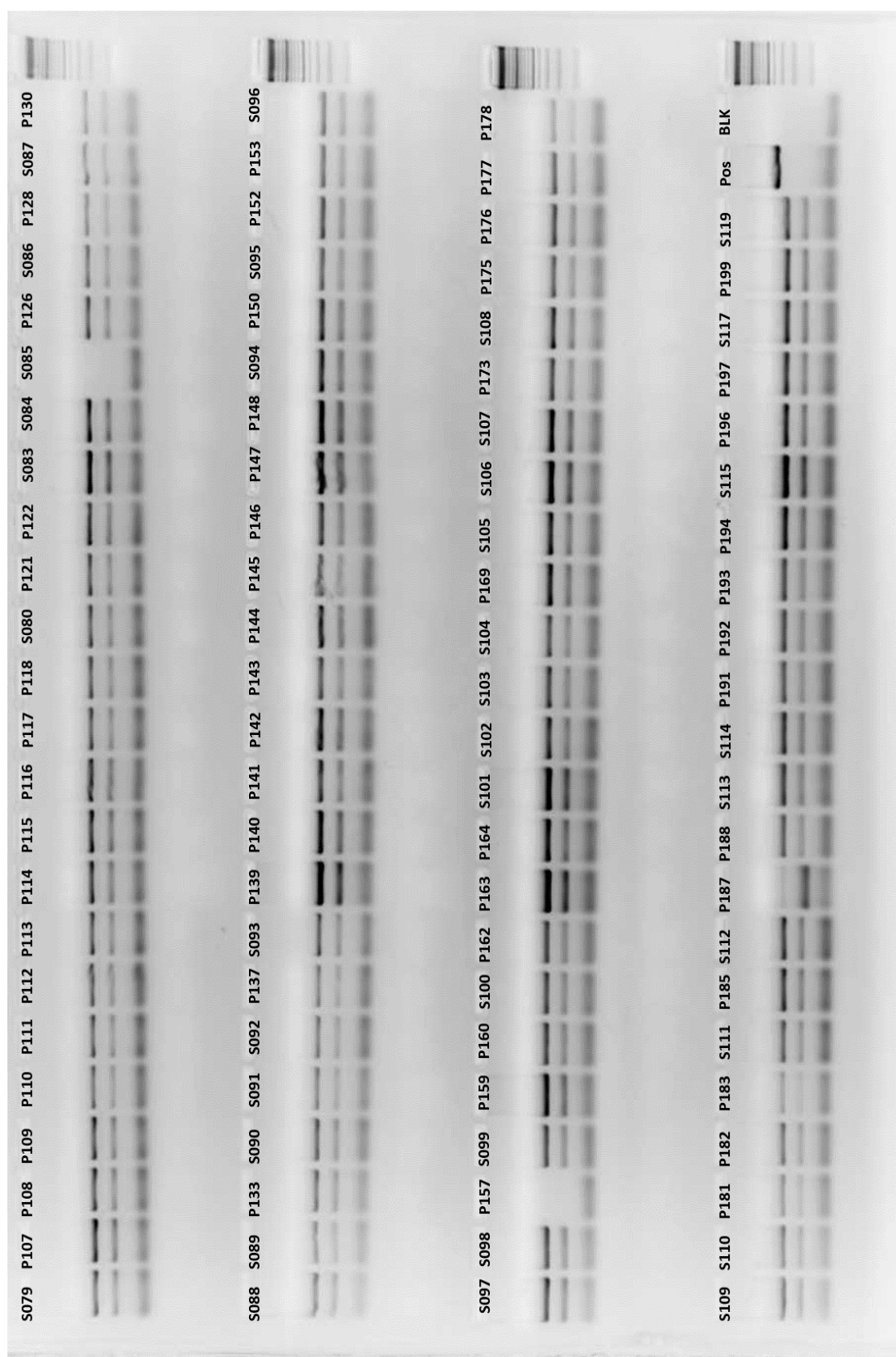


Figure 8.14: Image of agarose gel with restriction enzyme digested samples P079–P195 and S080–S119. A 100 bp DNA ladder was used.
 Abbreviations: Pos: positive control; BLK: blank.

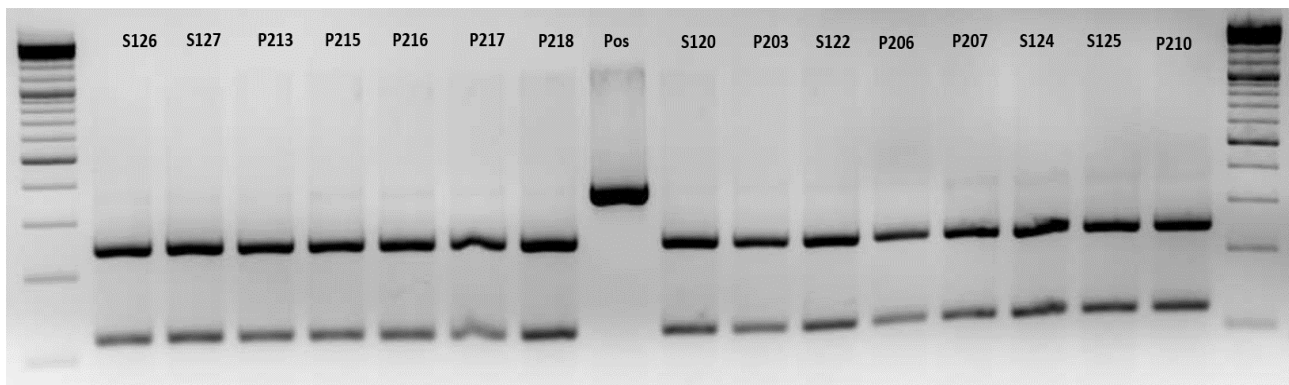


Figure 8.15: Image of agarose gel with restriction digested DNA samples P203–P218 and S120–S127. A 100 bp DNA ladder was used. Abbreviations: Pos: positive control.

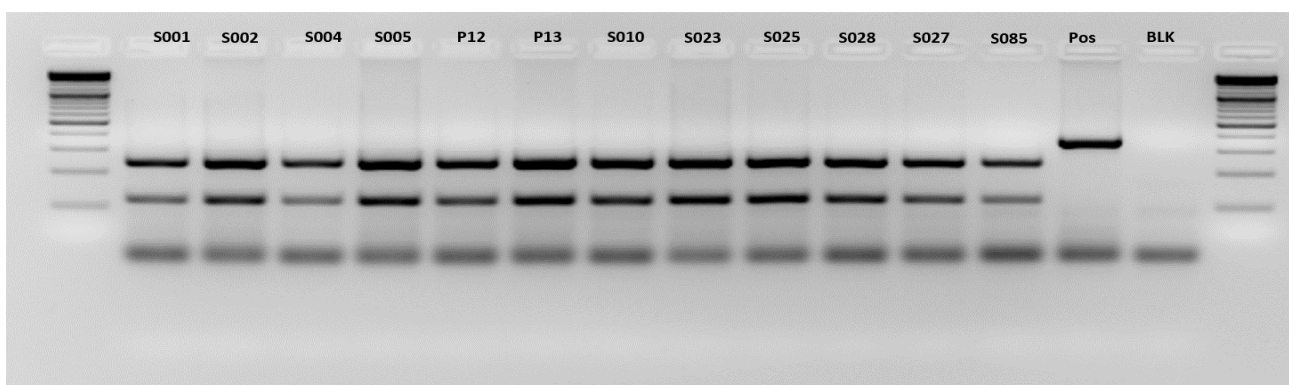


Figure 8.16: Image of agarose gel with restriction digested DNA samples. Lanes 1–12 are samples that underwent RFLP a second time (repeated samples). All samples are wild-type and no heterozygous or homozygous mutations were detected. A 100 bp DNA ladder was used. Abbreviations: Pos: positive control; BLK: blank.

8.4 DISCUSSION

8.4.1 Glutaric acidemia type I

Glutaric acidemia type I is a rare autosomal inherited neurometabolic disorder that occurs during infancy or early childhood. Recent studies identified several patients that presented with late-onset neurological symptoms associated with GA-1 (Külkens *et al.*, 2005). Mutations in the gene *GCDH* have been reported in more than 200 GA-1 cases and are the leading cause for this disease (Goodman *et al.*, 1998; Zschocke *et al.*, 2000). Van der Watt *et al.* (2010) identified a population-specific mutation (c.877G>A, p.A293T) in African patients with GA-1, which was found to affect 1 in 36 newborns (n = 750), giving a prevalence rate of 1 in 5184 in the South African population. The cohort of patients described by Van der Watt *et al.* (2010) was mainly from the Western Cape regions with only a few residing in Kwa-Zulu Natal and Central South Africa. Here,

we present four unrelated Black-African cases (two homozygotes and two heterozygotes) with this same mutation in the gene *GCDH* (c.877G>A, p.A293T), all originating from the Northern Provinces of South Africa.

One patient, S013, presented with an early-onset GA-1 phenotype while the second patient, S127, had an asymptomatic GA-1 phenotype at the time of this investigation. However, it is still possible for S127 to present with GA-1 symptoms later in life. Furthermore, biochemically both patients with the homozygous p.A293T mutation had elevated concentrations of GA and 3-OH-GA in their urine. The two patients with a heterozygous p.A293T mutation (S094 and S114) had asymptomatic clinical phenotypes as well as non-correlating biochemical profiles. Although only a small number of patients was screened, a carrier frequency of 1 in 50 patients is found in our cohort. A more precise newborn prevalence number can only be identified when more patients from the Northern parts of South African are screened for this mutation. As newborn screening is the most effective method for identifying this mutation, clinicians are urged to screen Black-African newborn babies for GA-1.

8.4.2 Mitochondrial DNA depletion syndrome

MPV17-related MDs such as hepatocerebral depletion syndrome, Alpers-Huttenlocher syndrome, myopathic depletion syndrome, encephalomyopathic depletion syndrome, neurogastrointestinal encephalomyopathy and neurohepatopathy MDDS are considered to be rare genetic diseases worldwide. In South Africa, a newborn prevalence is estimated at 1 in 18 496 (Meldau *et al.*, 2018) for neurohepatopathy MDDS. Recent studies on *MPV17*-related MD, specifically neurohepatopathy, identified a South African population-specific novel nonsense variant, c.106C>T (p.Gln36Ter) in black African patients originating from Western Cape Province using exome sequencing. As this is a founder mutation for MDDS, we screened our cohort of patients for the same mutation. This specific variant was, however, not detected in any of our patients, originating from the Northern provinces of South Africa. Recent studies showed that pathogenic *MPV17* variants can present later in life and are associated with adult-onset MDDS as well as focal COX deficiency and other multisystemic disorders (Blakely *et al.*, 2012; Garone *et al.*, 2012). Furthermore, it is possible that newborns, juvenile or adult patients are overlooked or misdiagnosed due to clinical selection bias.

CHAPTER 9:

SUMMARY, CONCLUSIONS AND FUTURE PROSPECTS

In this chapter the rationale, experimental approach, and key findings of this study are summarised and critically evaluated, and final conclusions are drawn.

9.1 RATIONALE BEHIND AN INVESTIGATION INTO MITOCHONDRIAL DISEASE AETIOLOGY IN SOUTH AFRICAN POPULATIONS

The aim of this study was to investigate underlying genetic causes, for both mtDNA and nDNA genes known to be involved with MD, in a clinically and biochemically defined paediatric cohort. As discussed in detail in Chapter 2, MDs are a complex, heterogeneous group of inherited disorders, causing severe phenotypes which place considerable physical, social and economic burden on patients, family members, and medical personnel. The clinical symptoms of MD can arise in childhood or later in life and can affect multiple organs, with a predominance of neurological and myopathic features (Alston *et al.*, 2017). The current range of literature mostly originates from non-African, mostly homogenous populations, and highlights the complexity of MDs in these populations. Even though a significant number of MD phenotypes have a well-defined genotype correlation, some of the aetiology is still unclear. This is the case for both well-defined and understudied patient populations, which consequently makes diagnosis of MDs challenging. Understanding MD aetiology, with well-documented clinical, biochemical and genetic information, contributes greatly to fast and effective routine diagnosis - and subsequent treatment actions for patients. This scenario is seen at specialist diagnostic and research facilities worldwide where a diversity of expertise is combined to perform services from different fields of specialization to establish a clear and precise diagnosis. Notably also at these specialist facilities, diagnostic services are inevitably accompanied by a strong emphasis on research. Examples of such established institutions where the contributors to this study have made personal observations include *The Wellcome Trust Centre for Mitochondrial Research* and the *MRC Centre*

for *Neuromuscular Disease* (both in the UK); the *Radboud Center for Mitochondrial Medicine*, (Nijmegen, The Netherlands); and the *Murdoch Children's Research Institute* (Melbourne Australia). Such specialised institutions, however, do not exist in many developing countries such as South Africa which tend to have understudied, ethnically diverse populations. These limitations have, for some time, been recognized by clinicians, specialists and scientists working in these populations, as well as those working in ethnically diverse cities in developed countries. However, more recently, international organisations involved with rare diseases, such as the International Rare Diseases Research Consortium (IRDiRC, <http://www.irdirc.org/about-us/vision-goals/>), have set specific goals in recognition of, and with the aim to address these global disparities.

In South Africa, basic clinical diagnostic services at selected centres are provided for patients with IMDs, including limited services for MD. Considering MD, recent reports (Meldau *et al.*, 2016; Van der Westhuizen *et al.*, 2015) highlight the limited capacity offered for patients as there are only a limited number of both clinicians who specialise in rare neuromuscular diseases such as MD, and supporting diagnostic and research entities in the country. These specialists are considered key in the diagnosis and management of children and adults with MD or suspected MD (Wortmann *et al.*, 2017). This limited capacity is a major contributing factor to the lack of information on MD aetiology in South Africa. Nevertheless, progress has been made in selected cases or small cohort studies over the past decade.

An example is the research investigation set forth by Van der Walt *et al.* (2012), who performed an initial study to evaluate mtDNA involvement in a selected number of paediatric patients from a cohort originating from the Northern parts of South Africa. This cohort has been assembled by Prof. I. Smuts (paediatric neurologist) and colleagues at the *Steve Biko Academic Hospital*, Pretoria, since 1998. This was the first extensive genetic investigation on the mitochondrial genome, which also served to compare the genetic and biochemical data with the phenotypical features observed in these patients. Although these genetic investigations were performed on a small-scale, a large number of novel variants of interest was identified - as was expected for African patients (Campbell & Tishkoff, 2010; Gurdasani *et al.*, 2015; Sherman *et al.*, 2018). A small number of candidate pathogenic variants were finally identified, confirming what was reported anecdotally in this population for paediatric patients (Van der Westhuizen *et al.*, 2015). The aetiology was therefore still unresolved in this unique cohort of paediatric patients from Prof. I. Smuts, thus forming the rationale for this investigation.

In response to these initial results, as highlighted by Smuts *et al.* (2010), Meldau *et al.* (2016), and Van der Westhuizen *et al.* (2015), a more extensive molecular genetic investigation was devised to include the nuclear genome. This was considered justified and timely as nuclear gene involvement is expected to be much more prevalent in paediatric patients with MD compared to mitochondrial genome involvement, which is more prevalent in adults (Craven *et al.*, 2017;

DiMauro & Schon, 2003; Shoubbridge, 2001). The following aims and objectives were therefore formulated to address this problem.

9.2 RESEARCH AIMS AND OBJECTIVES

The aim of this study was to investigate the aetiology of MD in a South African paediatric patient cohort originating from the northern provinces of South Africa. The following objectives were therefore formulated to address this aim.

9.2.1 Objective 1: Nuclear gene selection for high throughput next-generation sequencing

The first objective of this study was to design a strategy whereby nuclear genes associated with complexes I, II, III, and IV, as well as those associated with CoQ₁₀, could be sequenced in a high throughput manner, using targeted selection technology and NGS.

In recent years, WES has become a frequently used diagnostic option for patients with clinically suspected MD. A number of advantages and disadvantages of WES have been described in recent literature. The advantages include higher diagnostic yield compared to panel sequencing and fewer highly invasive procedures, such as muscle biopsy collection which are performed on patients. A major disadvantage is the number of incidental findings detected and that different types of disease-causing gene variants in specific region types of interest could be missed as there are limitations in capturing the whole exome (Lionel *et al.*, 2018; Meienberg *et al.*, 2016). Even though WES is a relatively fast diagnostic procedure with accurate genetic results, at the time of this study, as far as the author could determine, it has not yet been considered as a “first approach” option in developing countries for MD due to a number of reasons including limited infrastructure, finance, and sequencing data processing capability, and storage. Most crucially, ethical issues associated with the generation of expansive genetic data for diagnostics in understudied populations has been a highly contentious issue (De Vries & Pepper, 2012b). Ethical issues regarding WES include possible exploitation of these invaluable African sequencing data due to the richness in diversity thereof and incidental findings that might surface in patients, which could lead to legal consequences (De Vries *et al.*, 2012). Considering these circumstances, panel/targeted NGS was considered a prudent option for this study. Initially, it was decided that the first step was to investigate only the genes known to be involved with MDs at the time, even though the genes were identified as disease-causing primarily in non-Africans. The three panels were designed and consist of approximately 140 genes, all known to be involved with MDs. Genes were selected based on previously reported publications and in-house data from *Radboud Centre*

for *Mitochondrial Medicine* (Nijmegen, The Netherlands). In the end, three gene panels were designed and used for NGS, addressing Objective 1.

9.2.2 Objective 2: Patient selection for next-generation sequencing

The second objective of this study was to identify patients to be included for both nuclear DNA and whole mitochondrial genome sequencing, and to isolate DNA from muscle and/or blood samples from the selected patients.

Paediatric patients were selected for nDNA panel sequencing based on certain criteria, as there were limitations to both the number of genes (due to the reagent kits used) and samples (due to funding limitations) which could be sequenced. One important criterion was muscle RC enzyme activities. The patients with the lowest enzyme activities compared to the lowest control value were included. Clinical and other biochemical data available were also considered and discussed with the clinician before final inclusion. An important aspect of this study for panel sequencing was to include patients who were clinically evaluated as well as biochemically diagnosed with an RC enzyme deficiency. These criteria were also used for whole mitochondrial genome sequencing, with one exception: there was no limit in the number of patients to be sequenced, as long as each patient was clinically and biochemically diagnosed with an MD. In the end, 85 patients, comprising 61 Africans and 24 non-Africans, met the criteria and were included for panel NGS and 123 patients, comprising 79 African and 44 non-African, were included for whole mtDNA sequencing. Limited sample availability prevented whole mtDNA sequencing in four cases. DNA was isolated for the entire cohort of 127 patients.

9.2.3 Objective 3: Generate high quality next-generation sequencing data

The third objective of this study was to generate high quality NGS data from selected nuclear genes, as well as high quality data from whole mitochondrial genome sequencing.

Of the 212 patients in the South African mitochondrial cohort, all originating from the Northern Provinces of South Africa, 127 cases, 81 Africans and 46 non-Africans, were clinically and biochemically diagnosed with an MD. DNA was isolated from the entire cohort of 127 patients, which was used in library preparation for NGS. In the end, nuclear DNA sequences were produced and reported for 85 patients, while whole mitochondrial DNA sequences were produced and reported for 123 patients. Complete whole mitochondrial DNA sequences will shortly be made available to both GenBank and Phylotree to contribute to the current global set of African population mtDNA sequence data. The third objective of this study was thus successfully achieved.

9.2.4 Objective 4: Design a bioinformatics pipeline with the focus on nuclear DNA sequencing data

The fourth objective was to design and write a bioinformatics pipeline to efficiently identify novel, reported and disease-causing/pathogenic variants from NGS data, specifically nuclear DNA sequencing data.

This objective was addressed in Chapter 5, which is presented in this thesis as an accepted paper at a suitable peer-reviewed scientific journal. Robust bioinformatics pipelines are key components for diagnosis of rare diseases, specifically MD, which is considered a heterogeneous group of disorders for which diagnosis is often a daunting task. An offline pipeline was developed, i.e. where no internet connection is required to process data. This was considered an important aspect in developing countries where there is often limited/unstable internet access. It was developed as a first step towards NGS data processing for detection of multiple genetic alterations, including patients of African descent for which little information is available. The pipeline is developed to process smaller data sets, i.e. panel sequencing data as well as larger datasets, i.e., exome data. Identifying underlying genetic causes for rare diseases are particularly challenging in ethnically diverse populations, with limited bioinformatics support for researchers and non-bioinformaticians. It is therefore important to develop and continually provide access to bioinformatics pipelines, such as this one. This pipeline is publically available (aimed at clinicians and researchers), and can be found on GitHub (https://github.com/aseyffert/hetero_annotate). With continued development this pipeline could be further refined and made more user-friendly for both clinicians and researchers — most notably this pipeline could be expanded to annotate and filter GRCh38 aligned NGS data.

9.2.5 Objective 5: Pathogenicity evaluation of identified variants

The fifth objective was to assess the pathogenicity of novel, reported and previously reported disease-causing nuclear variants identified in patients.

The objective was addressed in Chapters 6, particularly Section 6.2. Chapter 6 is presented in this thesis as a submitted paper at a suitable scientific journal. The bioinformatics pipeline, as described in Chapter 5, was used to process NGS data for 85 patients.

Mitochondrial genome sequencing revealed, not unexpectedly, relatively low diagnostic outcome in this cohort, as the most common mtDNA pathogenic variants (for example variants causing LHON, MELAS, and MERRF) were absent, while a number of uncommon previously reported pathogenic mitochondrial variants were identified in this cohort. However, substantial

pathogenicity evidence was mostly insufficient for these previously reported variants and consequently they could not be classified as pathogenic in these patients. A large number of novel mtDNA variants was detected and identified; however, none have been extensively investigated and they still need to be followed up. Targeted gene panel sequencing was the first prudent step for investigating nuclear genes known to be involved with MD. Given the overarching results from panel sequencing, as described in Chapter 6, it was evident that this approach (where only a number of mitochondrial-associated genes were targeted) was mostly ineffective in identifying disease-causing or pathogenic variants in a large number of patients. Through panel sequencing, pathogenic and likely pathogenic variants could be identified in only two compound heterozygote cases. A number of reported pathogenic variants identified in African cases, could not be classified as pathogenic due to a general lack of substantial evidence of pathogenicity, which included but was not limited to inconsistent genotype-phenotype correlations and higher than expected African allele frequencies found in population databases.

Although targeted/gene panel NGS is considered a prudent NGS approach in many diagnostic settings, whilst being aware that this is a heterogeneous, understudied patient population and that an expanded gene panel may not necessarily increase diagnostic yield (Plutino *et al.*, 2018), genetic investigations were expanded to include WES to probe its outcome on a small subset of eight African patients where no initial NGS results were obtained. Initial indications from limited WES data were much more promising; nine likely pathogenic variants were identified in six cases compared to five variants identified in three cases using panel NGS and mtDNA sequencing.

Considering all the cases in this study, in only five cases (S057, S085, S032, S033 and S011) a strong genotype-phenotype correlation could be established and a moderately strong correlation in two cases (S002 and S117). For the remaining cases ($n = 78$), a non-specific correlation was observed. Observations like these serve as strong motivation that a “genetics first/only” approach, without supporting clinical and biochemical investigation, is not suitable in such understudied, ethnically diverse populations. Furthermore, when following a genetic approach, it was concluded that panel sequencing could be an efficient approach in populations where the genotype-phenotype correlations are well-established for specific monogenic diseases. For heterogeneous diseases such as MD, however, even in such populations, WES/WGS performs significantly better compared to a targeted gene-panel approach (Dillon *et al.*, 2018; Lionel *et al.*, 2018; Taylor *et al.*, 2014).

The results reported here are therefore in line with proposals that WES should be considered as the primary option for genetic investigations in heterogeneous inherited diseases such as MD, and, in fact, may be particularly ideal in understudied, ethnically diverse populations where there is evidence of inconsistencies with documented MD phenotypes. However, in such populations the value of extensive clinical and biochemical (structural and functional) investigations to support

molecular genetic data outcomes, should not be neglected, and at this time are, more crucial than in well studied populations.

9.2.6 Objective 6: Follow up on a pathogenic mutation in a selected case

The sixth objective was to report on a selected case for which extensive functional analyses were performed.

This objective was addressed in Chapter 7, which is presented in this thesis as a published peer-reviewed paper (Van der Westhuizen *et al.*, 2017). Here, three patients — one index case — were extensively described. One patient from the cohort described here presented with both a novel and previously reported pathogenic *ETFDH* variation. These variants were detected using panel sequencing, more specifically targeted genes involved with CoQ₁₀ biosynthesis and functioning, including genes known to be involved with primary as well as secondary CoQ₁₀ deficiency. This patient was evaluated at the *Steve Biko Academic Hospital* by Prof. I. Smuts. The two other cases were retrospectively investigated and initially evaluated by Dr. E. Honey (University of Pretoria) and Drs. M. Lippert and M. Engelbrecht (private paediatricians, Pretoria).

The three cases described by Van der Westhuizen *et al.* (2015) had a wide age range spanning over three decades; the youngest patient (index case) showed signs and symptoms of a severe metabolic defect in his first week of life and passed away on day nine while the oldest patient presented with MADD features at age four and passed away at age 23. The three cases showed similar urine metabolic features typical of MADD but varied clinical features, and were only later classified as MADD after extensive genetic and biochemical investigations were performed on the one case from this cohort described here, where a novel disease-causing *ETFDH* variant was confirmed. MADD patients are often responsive to riboflavin treatment, which can alleviate severe signs and symptoms when detected early in life. It is therefore of great importance to be able to identify metabolic defects early in life either through newborn screening, sequencing (both panel and WES), or routine metabolomics. The capacity to perform such routine procedures is still lacking to great extend in developing countries, as can be seen from these three cases where immediate follow-up investigations could not be performed.

9.2.7 Objective 7: Population specific mutation screening

The seventh and last objective was to screen for two putative founder mutations known to be involved with mitochondrial dysfunction in the South African population.

Patients from the Cape provinces of South Africa are routinely being screened for several metabolic and other rare diseases at the NHLS and at the University of Cape Town. Recent publications by Van der Watt *et al.* (2010) and Meldau *et al.* (2018) identified two founder mutations in the genes *GCDH* and *MPV17*. These mutations are only found in patients from African descent. As a result, routine screening is performed in paediatric patients to identify the gene mutations which cause GA-1 and MDDS at the NHLS under the supervision of S. Meldau. This research, conducted and performed by Van der Watt *et al.* (2010) and Meldau *et al.* (2018), highlights the genotype and phenotype differences found in African patients compared to non-Africans, especially seen in patients with GA-1. There is furthermore a difference observed in the clinical features for GA-1 in patients from Cape provinces (so-called Coloured people) compared to patients from Northern Provinces (mostly Black-Africans). The latter was seen in this cohort where the two Black-African females with the reported *GCDH* mutation (c.877G>A, p.A293T) presented with different clinical phenotypes — which was inconsistent with reported cases originating from the Cape Provinces (Van der Watt *et al.*, 2010). Furthermore, the Black-African clinical phenotype did not only differ from Coloured cases, but also differed among Black-African cases in our cohort, as one of the two females with this mutation had severe GA-1 while the other had almost no typical GA-1 signs or symptoms. From this anomaly it can be concluded that African patients may have unique genotype-phenotype correlations for some pathogenic variants and screening for disease-causing mutations as a first option may therefore be largely ineffective in ethnic diverse cohorts. In the end, 116 patients from our cohort of predominantly Black-African patients were screened for the same founder mutations in the genes *GCDH* and *MPV17*, which had been previously identified. A homozygous *GCDH* mutation was identified in two Black Africans. The *MPV17* mutation was absent from the entire cohort. Objective 7 was thus successfully addressed.

9.3 FINAL CONCLUSION AND FUTURE PROSPECTS

To conclude, this study was the first in size and scope to investigate the underlying genetic cause of MD in an understudied, ethnically diverse cohort predominantly of African descent, all clinically diagnosed with an MD. The genetic techniques used include whole mitochondrial genome sequencing, targeted gene panel sequencing, and (as a pilot study) WES on a subset of African cases.

In short, from this study the following conclusions were made and presented in various forms (for degree qualification purposes it was important to disseminate this expanded knowledge in suitable forms as publications and presentations, as listed in Appendix A):

- 1) The limited domestic capacity for clinical assessment of patients with neurological features, and, most notably, the limited number of clinical experts, as well as scientists and diagnostic facilities in South Africa, contributes greatly to the general lack on MD aetiology information for South African populations.
- 2) Increased rare disease bioinformatics capacity in South Africa is crucial for the analysis of African-specific NGS data. At this point in time there is bioinformatics support for non-communicable disease research for South African populations, but comparatively little with a specific focus on rare disease research and diagnosis.
- 3) Targeted gene panel sequencing may be suitable and sufficient for identifying pathogenic variants for well-understood diseases in well-studied patient populations. For these populations well-defined genotype-phenotype correlations allow the construction of highly targeted gene panels for sequencing. For populations which aren't well-studied, however, it is much less effective since the correct identification of relevant target genes relies heavily on a good understanding of the genotype-phenotype correlation for the particular disease.
- 4) WES, which is fast becoming the diagnostic tool of choice for heterogeneous diseases with unknown or only partially understood aetiology, such as MD, is more appropriate when genotype-phenotype correlation information is lacking. Its untargeted nature improves diagnostic yield compared to targeted sequencing (panel sequencing). The limited results from this study support this conclusion.
- 5) The genotype-phenotype correlations found in well-studied non-African patient populations do not always hold for under-studied African patient populations. This is evident in this cohort where variants previously reported as pathogenic were identified in several cases, but with none of the correlated phenotypical features present.
- 6) Screening for variants as a first option is mostly insufficient in identifying pathogenic variants in understudied ethnically diverse patient populations. It can, however, be sufficient in detection of known pathogenic variants in well-established patient populations.
- 7) A WES-first/only diagnostic approach, which is often favoured since it eliminates potentially invasive biochemical analyses, is not sufficient for cases where a heterogeneous disease (such as MD) is suspected in a patient who is part of an under-studied population (such as the African population). It is therefore still important to not neglect the value of clinical and biochemical data when investigating heterogeneous diseases (such as MD) in understudied patient populations. In these cases, biochemical and other structural/functional molecular investigations are particularly important since they provide a means of confirming the structural and functional effects of genetic variation.

The results of this study not only give better insight into the aetiology of this particular cohort of MD patients, but also provide new knowledge thereof. A number of recommendations for future developments can thus be made.

Meldau *et al.* (2018) considered the diagnostic outcomes for SA patients to be relatively low, as there has been a genetic confirmation in fewer than 70 cases to date, even when combining the two centres (situated in the Western Cape Province and Gauteng Province). In a country with a population of ~52 million in 2011, this translated to obtaining a genetically confirmed diagnosis in ~1:800 000 people. This ratio would not have changed significantly since then, even when adding the limited outcomes of this much more extensive study. When comparing the estimated prevalence of MD at 1:5000 worldwide (Alston *et al.*, 2017; Chinnery, 2014; Gorman *et al.*, 2016; Gorman *et al.*, 2015), bearing in mind that an epidemiological study has yet to be conducted, it is still clearly evident that the number of undiagnosed cases in SA is staggeringly high. This clearly reflects on the limited access of patients and limited diagnostic capacity in the country. This limited capacity, coupled with the evidence presented here that more expansive genetics and supporting clinical and biochemical investigations are required, prompts the obvious recommendation that the number and efficiency of clinical and specialised diagnostic/laboratory and research facilities need to increase significantly. This is crucial not only to enable diagnosis and improve patient prospects, but also to enable the extended research into the aetiology of MDs that is required. With health funding directed toward more pressing health issues worldwide, a sensible intermediary step would be to collaborate with experts and established institutions in other countries (in both developed and developing countries), and to form national and international networks. This would allow the sharing of capacities, data, and expertise, and the opportunity to combine efforts to acquire funding. Indeed, such efforts have been employed in the past, as is evident from the affiliations of collaborators to this study. More recently, the formation of the *International Centre for Genomic Research in Neuromuscular Disease*, a collaboration between five lower- and middle-income countries (including SA), with leading scientists in the United Kingdom (UK), has been funded by the Medical Research Council (UK) and will commence in 2019. Although the objectives of this large study are aimed at clinical training and genomics of the wider set of neuromuscular diseases, the efforts of this collaboration promise to greatly contribute to new knowledge of the genetics of MD in this population, and can thus form the basis of significantly improved future genetic diagnoses.

As discussed in Chapter 4 and 6, the genetic investigations in this study mainly focussed on targeted gene panel sequencing and whole mtDNA NGS. Both applications proved relatively ineffective, as the diagnostic yield was lower than expected. Future research and investigations could continue to explore genetic approaches, in particularly WES, which is known to have improved diagnostic yield. To this end it would be advantageous to perform WES on the entire cohort described here. Furthermore, performing WES on a larger number of patients would not

only expand the knowledge on underlying genetic causation of MD, but would also have significant benefits and contributions towards the global effort to gather genetic information of all world populations — in particular since there is a great lack of genetic information of African populations south of the Sahara (Choudhury *et al.*, 2017; De Vries & Pepper, 2012b; De Vries *et al.*, 2012; Gurdasani *et al.*, 2015). As genetic diagnoses increasingly become the norm, it will be important that future investigations include research on genetic counselling and all of its aspects (increase trained genetic counsellors as well as the number of facilities where counselling is provided). Counselling would help patients and their parents understand the complexity of genetic diseases, in particular understanding what role the genetic component of MD plays in their prognosis. Considering the lacking information on MD aetiology (due to lacking information regarding the genotype-phenotype correlation) it will be of great value to perform a wider range of laboratory investigations in selected samples as required, such as cellular respiration on cell lines (e.g., cybrid cell line investigations to identify mutant mtDNA), histology on skeletal muscle and other tissues, and metabolic investigations. Another approach to consider in order to inform on the effect of novel variants in cell line models would be the use of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) in conjunction with Cas9 (CRISPR-associated genes). However, one should exercise some caution as there are several ethical concerns regarding gene editing using CRISPR/Cas9 in clinical and diagnostic settings.

As discussed in Chapter 5, bioinformatics plays a crucial part in identifying disease-causing variants from large datasets. There are currently a number of platforms and consortiums available online where both phenotypic and genotypic data can be analysed and shared (RD-Connect, <https://rd-connect.eu>, SOPHiA Genetics, <https://www.sophiagenetics.com>). Individual rare diseases can be searched for in, and investigated through these available databases, which could ultimately aid in diagnoses. For future research in bioinformatics on African rare disease data it would be useful to develop such databases where data can be shared between clinicians and researchers. It would also be favourable to collaborate with existing consortiums that use these platforms to improve diagnosis of MD in Southern Africa. There are, however, several ethical issues to take into account regarding African genetic data, and one should exercise some caution on how data are being shared and with whom it is shared (De Vries & Pepper, 2012a).

Finally, a new reference genome should be considered for African patient populations, as was recently highlighted by Sherman *et al.* (2018), who found that African genomes contain approximately 10% more DNA (effecting 315 protein-coding genes), than is found in the current (mainly Caucasian) human reference genome (GRCh38). Such a new reference genome would allow better characterisation of rare disease genotypes in African patient populations, and in turn facilitate the development of a better understanding of the genotype-phenotype correlation for MDs.

REFERENCES

- Abath Neto, O., Martins, C.d.A., Carvalho, M., Chadi, G., Seitz, K.W., Oliveira, A.S.B., Reed, U.C., Laporte, J. & Zanoteli, E. 2015. DNM2 mutations in a cohort of sporadic patients with centronuclear myopathy. *Genetics and molecular biology*, 38(2):147-151.
- Abu-Libdeh, B., Douiev, L., Amro, S., Shahrour, M., Ta-Shma, A., Miller, C., Elpeleg, O. & Saada, A. 2017. Mutation in the COX4I1 gene is associated with short stature, poor weight gain and increased chromosomal breaks, simulating Fanconi anemia. *European Journal of Human Genetics*, 25(10):1142.
- Acevedo, F., Huerta, E., Burgeff, C., Koleff, P. & Sarukhán, J. 2011. Integrative genomics viewer. *NATURE Biotechnology* 29(1).
- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. & Sunyaev, S.R. 2010. A method and server for predicting damaging missense mutations. *Nature methods*, 7(4):248.
- Alcázar-Fabra, M., Trevisson, E. & Brea-Calvo, G. 2018. Clinical syndromes associated with Coenzyme Q10 deficiency. *Essays in biochemistry*, 62(3):377-398.
- Alston, C.L., Berti, C.C., Blakely, E.L., Oláhová, M., He, L., McMahon, C.J., Olpin, S.E., Hargreaves, I.P., Nolli, C. & McFarland, R. 2015. A recessive homozygous p. Asp92Gly SDHD. *Human genetics*, 134(8):869-879.
- Alston, C.L., Davison, J.E., Meloni, F., van der Westhuizen, F.H., He, L., Hornig-Do, H.-T., Peet, A.C., Gissen, P., Goffrini, P. & Ferrero, I. 2012. Recessive germline SDHA and SDHB mutations causing leukodystrophy and isolated mitochondrial complex II deficiency. *Journal of medical genetics*, 49(9):569-577.
- Alston, C.L., Rocha, M.C., Lax, N.Z., Turnbull, D.M. & Taylor, R.W. 2017. The genetics and pathology of mitochondrial disease. *J Pathol*, 241(2):236-250.
- Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F. & Hamosh, A. 2014. OMIM. org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic acids research*, 43(D1):D789-D798.
- Anagnostou, M.E., Ng, Y.S., Taylor, R.W. & McFarland, R. 2016. Epilepsy due to mutations in the mitochondrial polymerase gamma (POLG) gene: A clinical and molecular genetic review. *Epilepsia*, 57(10):1531-1545.
- Anderson, S., Bankier, A.T., Barrell, B.G., De Bruijn, M., Coulson, A.R., Drouin, J., Eperon, I., Nierlich, D., Roe, B.A. & Sanger, F. 1981. Sequence and organization of the human mitochondrial genome. *Nature*, 290(5806):457-465.
- Andrews, B., Carroll, J., Ding, S., Fearnley, I.M. & Walker, J.E. 2013. Assembly factors for the membrane arm of human complex I. *Proceedings of the National Academy of Sciences*, 110(47):18934-18939.
- Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. & Howell, N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature genetics*, 23(2):147.
- Angebault, C., Charif, M., Guegen, N., Piro-Megy, C., Mousson de Camaret, B., Procaccio, V., Guichet, P.-O., Hebrard, M., Manes, G. & Leboucq, N. 2015. Mutation in NDUFA13/GRIM19 leads to early onset

hypotonia, dyskinesia and sensorial deficiencies, and mitochondrial complex I instability. *Human molecular genetics*, 24(14):3948-3955.

Antonenkova, V.D., Isomursu, A., Mennerich, D., Vapola, M.H., Weiher, H., Kietzmann, T. & Hiltunen, J.K. 2015. The human mtDNA depletion syndrome gene MPV17 encodes a non-selective channel that modulates membrane potential. *Journal of Biological Chemistry*:jbc. M114. 608083.

Antonicka, H., Leary, S.C., Guercin, G.-H., Agar, J.N., Horvath, R., Kennaway, N.G., Harding, C.O., Jaksch, M. & Shoubridge, E.A. 2003a. Mutations in COX10 result in a defect in mitochondrial heme A biosynthesis and account for multiple, early-onset clinical phenotypes associated with isolated COX deficiency. *Human molecular genetics*, 12(20):2693-2702.

Antonicka, H., Mattman, A., Carlson, C.G., Glerum, D.M., Hoffbuhr, K.C., Leary, S.C., Kennaway, N.G. & Shoubridge, E.A. 2003b. Mutations in COX15 produce a defect in the mitochondrial heme biosynthetic pathway, causing early-onset fatal hypertrophic cardiomyopathy. *The American Journal of Human Genetics*, 72(1):101-114.

Applegarth, D.A. & Toone, J.R. 2000. Incidence of inborn errors of metabolism in British Columbia, 1969–1996. *Pediatrics*, 105(1):e10-e10.

Artuch, R., Brea-Calvo, G., Briones, P., Aracil, A., Galván, M., Espinós, C., Corral, J., Volpini, V., Ribes, A. & Andreu, A.L. 2006. Cerebellar ataxia with coenzyme Q10 deficiency: diagnosis and follow-up after coenzyme Q10 supplementation. *Journal of the neurological sciences*, 246(1):153-158.

Astuti, D., Latif, F., Dallol, A., Dahia, P.L., Douglas, F., George, E., Sköldbberg, F., Husebye, E.S., Eng, C. & Maher, E.R. 2001. Gene mutations in the succinate dehydrogenase subunit SDHB cause susceptibility to familial pheochromocytoma and to familial paraganglioma. *The American Journal of Human Genetics*, 69(1):49-54.

Baertling, F., Al-Murshedi, F., Sánchez-Caballero, L., Al-Senaidi, K., Joshi, N.P., Venselaar, H., den Brand, M.A., Nijtmans, L.G. & Rodenburg, R.J. 2017. Mutation in mitochondrial complex IV subunit COX5A causes pulmonary arterial hypertension, lactic acidemia, and failure to thrive. *Human mutation*, 38(6):692-703.

Baertling, F., van den Brand, M., Hertecant, J.L., Al-Shamsi, A., van den Heuvel, L., Distelmaier, F., Mayatepek, E., Smeitink, J.A., Nijtmans, L.G. & Rodenburg, R.J. 2015. Mutations in COA6 cause cytochrome c oxidase deficiency and neonatal hypertrophic cardiomyopathy. *Human mutation*, 36(1):34-38.

Balsa, E., Marco, R., Perales-Clemente, E., Szklarczyk, R., Calvo, E., Landázuri, M.O. & Enríquez, J.A. 2012. NDUFA4 is a subunit of complex IV of the mammalian electron transport chain. *Cell metabolism*, 16(3):378-386.

Balwani, M., Bloomer, J., Desnick, R., of the NIH-Sponsored, P.C. & Network, R.D.C.R. 2013. X-linked protoporphyria.

Bannwarth, S., Procaccio, V., Rouzier, C., Fragaki, K., Poole, J., Chabrol, B., Desnuelle, C., Pouget, J., Azulay, J. & Attarian, S. 2008. Rapid identification of mitochondrial DNA (mtDNA) mutations in neuromuscular disorders by using surveyor strategy. *Mitochondrion*, 8(2):136-145.

Barel, O., Shorer, Z., Flusser, H., Ofir, R., Narkis, G., Finer, G., Shalev, H., Nasasra, A., Saada, A. & Birk, O.S. 2008. Mitochondrial complex III deficiency associated with a homozygous mutation in UQCRC. *The American Journal of Human Genetics*, 82(5):1211-1216.

Barth, P., Scholte, H., Berden, J., Van der Klei-Van Moorsel, J., Luyt-Houwen, I., Veer-Korthof, E.T.V.T., Van der Harten, J. & Sobotka-Plojhar, M. 1983. An X-linked mitochondrial disease affecting cardiac muscle, skeletal muscle and neutrophil leucocytes. *Journal of the neurological sciences*, 62(1-3):327-355.

Bayona-Bafaluy, M.P., López-Gallardo, E., Montoya, J. & Ruiz-Pesini, E. 2011. Maternally inherited susceptibility to cancer. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1807(6):643-649.

- Baysal, B.E., Ferrell, R.E., Willett-Brozick, J.E., Lawrence, E.C., Myssiorek, D., Bosch, A., van der Mey, A., Taschner, P.E., Rubinstein, W.S. & Myers, E.N. 2000. Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. *Science*, 287(5454):848-851.
- Beard, S., Spector, E., Seltzer, W., Frerman, F. & Goodman, S. 1993. MUTATIONS IN ELECTRON-TRANSFER FLAVOPROTEIN-UBIQUINONE OXIDOREDUCTASE (ETF-QO) IN GLUTARIC ACIDEMIA TYPE-II (GA2). (In. CLINICAL RESEARCH organised by: SLACK INC 6900 GROVE RD, THOROFARE, NJ 08086. p. A271-A271).
- Beggs, A.H. & Agrawal, P.B. 2013. Multiminicore disease.
- Benit, P., Slama, A., Cartault, F., Giurgea, I., Chretien, D., Lebon, S., Marsac, C., Munnich, A., Rötig, A. & Rustin, P. 2004. Mutant NDUFS3 subunit of mitochondrial complex I causes Leigh syndrome. *Journal of medical genetics*, 41(1):14-17.
- Berger, I., Herschkovitz, E., Shaag, A., Edvardson, S., Saada, A. & Elpeleg, O. 2008. Mitochondrial complex I deficiency caused by a deleterious NDUFA11 mutation. *Annals of neurology*, 63(3):405-408.
- Bezawork-Geleta, A., Rohlena, J., Dong, L., Pacak, K. & Neuzil, J. 2017. Mitochondrial Complex II: At the Crossroads. *Trends in Biochemical Sciences*.
- Bianciardi, L., Imperatore, V., Fernandez-Vizarra, E., Lopomo, A., Falabella, M., Furini, S., Galluzzi, P., Grosso, S., Zeviani, M. & Renieri, A. 2016. Exome sequencing coupled with mRNA analysis identifies NDUFAF6 as a Leigh gene. *Molecular genetics and metabolism*, 119(3):214-222.
- Blachly-Dyson, E. & Forte, M. 2001. VDAC channels. *IUBMB life*, 52(3-5):113-118.
- Blackburn, P.R., Selcen, D., Gass, J.M., Jackson, J.L., Macklin, S., Cousin, M.A., Boczek, N.J., Klee, E.W., Dimberg, E.L. & Kennelly, K.D. 2017. Whole exome sequencing of a patient with suspected mitochondrial myopathy reveals novel compound heterozygous variants in RYR1. *Molecular genetics & genomic medicine*, 5(3):295-302.
- Blakely, E.L., Butterworth, A., Hadden, R.D., Bodi, I., He, L., McFarland, R. & Taylor, R.W. 2012. MPV17 mutation causes neuropathy and leukoencephalopathy with multiple mtDNA deletions in muscle. *Neuromuscular Disorders*, 22(7):587-591.
- Borna, N.N., Kishita, Y., Ishikawa, K., Nakada, K., Hayashi, J.I., Tokuzawa, Y., Kohda, M., Nyuzuki, H., Yamashita-Sugahara, Y., Nasu, T., Takeda, A., Murayama, K., Ohtake, A. & Okazaki, Y. 2017. A novel mutation in TAZ causes mitochondrial respiratory chain disorder without cardiomyopathy. *J Hum Genet*, 62(5):539-547.
- Bourdon, A., Minai, L., Serre, V., Jais, J.-P., Sarzi, E., Aubert, S., Chrétien, D., de Lonlay, P., Paquis-Flucklinger, V. & Arakawa, H. 2007. Mutation of RRM2B, encoding p53-controlled ribonucleotide reductase (p53R2), causes severe mitochondrial DNA depletion. *Nature genetics*, 39(6):776.
- Bourgeron, T., Rustin, P., Chretien, D., Birch-Machin, M., Bourgeois, M., Viegas-Péquignot, E., Munnich, A. & Rötig, A. 1995. Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency. *Nature genetics*, 11(2):144.
- Boy, N., Mühlhausen, C., Maier, E.M., Heringer, J., Assmann, B., Burgard, P., Dixon, M., Fleissner, S., Greenberg, C.R. & Harting, I. 2017. Proposed recommendations for diagnosing and managing individuals with glutaric aciduria type I: second revision. *Journal of inherited metabolic disease*, 40(1):75-101.
- Boyer, P.D. 1997. The ATP synthase—a splendid molecular machine. *Annual review of biochemistry*, 66(1):717-749.
- Brancaleoni, V., Balwani, M., Granata, F., Graziadei, G., Missineo, P., Fiorentino, V., Fustinoni, S., Cappellini, M., Naik, H. & Desnick, R.J.C.g. 2016. X-chromosomal inactivation directly influences the phenotypic manifestation of X-linked protoporphyria. 89(1):20-26.

Brandon, M.C., Lott, M.T., Nguyen, K.C., Spolim, S., Navathe, S.B., Baldi, P. & Wallace, D.C. 2005. MITOMAP: a human mitochondrial genome database—2004 update. *Nucleic acids research*, 33(suppl_1):D611-D613.

Brivet, M., Slama, A., Saudubray, J.-M., Legrand, A. & Lemonnier, A. 1995. Rapid diagnosis of long chain and medium chain fatty acid oxidation disorders using lymphocytes. *Annals of clinical biochemistry*, 32(2):154-159.

Brown, M., Voljavec, A., Lott, M., MacDonald, I. & Wallace, D. 1992. Leber's hereditary optic neuropathy: a model for mitochondrial neurodegenerative diseases. *The FASEB Journal*, 6(10):2791-2799.

Buchaklian, A.H., Helbling, D., Ware, S.M. & Dimmock, D.P. 2012. Recessive deoxyguanosine kinase deficiency causes juvenile onset mitochondrial myopathy. *Molecular genetics and metabolism*, 107(1):92-94.

Calvo, S.E., Compton, A.G., Hershman, S.G., Lim, S.C., Lieber, D.S., Tucker, E.J., Laskowski, A., Garone, C., Liu, S., Jaffe, D.B., Christodoulou, J., Fletcher, J.M., Bruno, D.L., Goldblatt, J., Dimauro, S., Thorburn, D.R. & Mootha, V.K. 2012. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci Transl Med*, 4(118):118ra110.

Calvo, S.E., Tucker, E.J., Compton, A.G., Kirby, D.M., Crawford, G., Burt, N.P., Rivas, M., Guiducci, C., Bruno, D.L. & Goldberger, O.A. 2010. High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nature genetics*, 42(10):851.

Campbell, M.C. & Tishkoff, S.A.J.C.b. 2010. The evolution of human genetic and phenotypic variation in Africa. 20(4):R166-R173.

Carelli, V. & La Morgia, C. 2018. Clinical syndromes associated with mtDNA mutations: where we stand after 30 years. *Essays in biochemistry*, 62(3):235-254.

Carrozzo, R., Murray, J., Santorelli, F. & Capaldi, R. 2000. The T9176G mutation of human mtDNA gives a fully assembled but inactive ATP synthase when modeled in Escherichia coli. *FEBS letters*, 486(3):297-299.

Casali, C., Santorelli, F.M., Damati, G., Bernucci, P., Debiase, L. & DiMauro, S. 1995. A novel mtDNA point mutation in maternally inherited cardiomyopathy. *Biochemical and biophysical research communications*, 213(2):588-593.

Cavadas, B., Soares, P., Camacho, R., Brandão, A., Costa, M.D., Fernandes, V., Pereira, J.B., Rito, T., Samuels, D.C. & Pereira, L. 2015. Fine time scaling of purifying selection on human nonsynonymous mtDNA mutations based on the worldwide population tree and mother-child pairs. *Human mutation*, 36(11):1100-1111.

Chen, L., Cui, Y., Jiang, D., Ma, C.Y., Tse, H.f., Hwu, W.L. & Lian, Q. 2017. Management of Leigh Syndrome: current status and new insights. *Clinical Genetics*.

Chinnery, P.F. 2014. Mitochondrial disorders overview (2014.). University of Washington, Seattle: GeneReviews.

Chinnery, P.F. & Hudson, G. 2013. Mitochondrial genetics. *British medical bulletin*, 106(1):135-159.

Choudhury, A., Ramsay, M., Hazelhurst, S., Aron, S., Bardien, S., Botha, G., Chimusa, E.R., Christoffels, A., Gamielien, J. & Sefid-Dashti, M.J.J.N.c. 2017. Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans. 8(1):2062.

Christian, B.E. & Spremulli, L.L. 2012. Mechanism of protein biosynthesis in mammalian mitochondria. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1819(9):1035-1054.

Čížková, A., Stránecký, V., Mayr, J.A., Tesařová, M., Havlíčková, V., Paul, J., Ivánek, R., Kuss, A.W., Hansíková, H. & Kaplanová, V. 2008. TMEM70 mutations cause isolated ATP synthase deficiency and neonatal mitochondrial encephalocardiomyopathy. *Nature genetics*, 40(11):1288.

- Cohen, B.H., Chinnery, P.F. & Copeland, W.C. 2014. POLG-related disorders.
- Cohen, B.H., Chinnery, P.F. & Copeland, W.C. 2018. GeneReviews®[Internet]. University of Washington, Seattle.
- Cohen, B.H. & Naviaux, R.K. 2010. The clinical diagnosis of POLG disease and other mitochondrial DNA depletion disorders. *Methods*, 51(4):364-373.
- Copeland, W.C. 2012. Defects in mitochondrial DNA replication and human disease. *Critical reviews in biochemistry and molecular biology*, 47(1):64-74.
- Cornelius, N., Frerman, F.E., Corydon, T.J., Palmfeldt, J., Bross, P., Gregersen, N. & Olsen, R.K. 2012. Molecular mechanisms of riboflavin responsiveness in patients with ETF-QO variations and multiple acyl-CoA dehydrogenation deficiency. *Human molecular genetics*, 21(15):3435-3448.
- Craven, L., Alston, C.L., Taylor, R.W. & Turnbull, D.M. 2017. Recent Advances in Mitochondrial Disease. *Annu Rev Genomics Hum Genet*, 18:257-275.
- Crimi, M., Papadimitriou, A., Galbiati, S., Palamidou, P., Fortunato, F., Bordini, A., Papandreou, U., Papadimitriou, D., Hadjigeorgiou, G.M. & Drogari, E. 2004. A new mitochondrial DNA mutation in ND3 gene causing severe Leigh syndrome with early lethality. *Pediatric research*, 55(5):842.
- Crofts, A.R., Holland, J.T., Victoria, D., Kolling, D.R., Dikanov, S.A., Gilbreth, R., Lhee, S., Kuras, R. & Kuras, M.G. 2008. The Q-cycle reviewed: How well does a monomeric mechanism of the bc₁ complex account for the function of a dimeric complex? *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1777(7):1001-1019.
- Damore, M.E., Speiser, P.W., Slonim, A.E., New, M.I., Shanske, S., Xia, W., Santorelli, F.M. & DiMauro, S. 1999. Early onset of diabetes mellitus associated with the mitochondrial DNA T14709C point mutation: patient report and literature review. *Journal of Pediatric Endocrinology and Metabolism*, 12(2):207-214.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T. & Sherry, S.T. 2011. The variant call format and VCFtools. *Bioinformatics*, 27(15):2156-2158.
- Danis, D., Brennerova, K., Skopkova, M., Kurdiova, T., Ukropec, J., Stanik, J., Kolnikova, M. & Gasperikova, D. 2018. Mutations in SURF1 are important genetic causes of Leigh syndrome in Slovak patients. *Endocrine regulations*, 52(2):110-118.
- Darin, N., Oldfors, A., Moslemi, A.R., Holme, E. & Tulinius, M. 2001. The incidence of mitochondrial encephalomyopathies in childhood: clinical features and morphological, biochemical, and DNA abnormalities. *Annals of neurology*, 49(3):377-383.
- Davidzon, G., Greene, P., Mancuso, M., Klos, K.J., Ahlskog, J.E., Hirano, M. & DiMauro, S. 2006. Early-onset familial parkinsonism due to POLG mutations. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 59(5):859-862.
- Dawkins, H.J., Draghia-Akli, R., Lasko, P., Lau, L.P., Jonker, A.H., Cutillo, C.M., Rath, A., Boycott, K.M., Baynam, G. & Lochmüller, H. 2017. Progress in rare diseases research 2010–2016: an IRDiRC perspective. *Clinical and translational science*.
- De Lonlay, P., Valnot, I., Barrientos, A., Gorbatyuk, M., Tzagoloff, A., Taanman, J.-W., Benayoun, E., Chrétien, D., Kadhon, N. & Lombès, A. 2001. A mutant mitochondrial respiratory chain assembly protein causes complex III deficiency in patients with tubulopathy, encephalopathy and liver failure. *Nature genetics*, 29(1):57.
- De Meirleir, L., Seneca, S., Lissens, W., De Clercq, I., Eyskens, F., Gerlo, E., Smet, J. & Van Coster, R. 2004. Respiratory chain complex V deficiency due to a mutation in the assembly gene ATP12. *Journal of medical genetics*, 41(2):120-124.
- De Vries, D.D., Van Engelen, B.G., Gabreëls, F.J., Ruitenbeek, W. & Van Oost, B.A. 1993. A second missense mutation in the mitochondrial ATPase 6 gene in Leigh's syndrome. *Annals of neurology*, 34(3):410-412.

- De Vries, J. & Pepper, M. 2012a. Genomic sovereignty and the African promise: mining the African genome for the benefit of Africa. *Journal of Medical Ethics*:medethics-2011-100448.
- De Vries, J. & Pepper, M. 2012b. Genomic sovereignty and the African promise: mining the African genome for the benefit of Africa. *Journal of Medical Ethics*: medethics-2011-100448.
- De Vries, J., Slabbert, M. & Pepper, M.S. 2012. Ethical, legal and social issues in the context of the planning stages of the Southern African Human Genome Programme. *Medicine and Law*, 31:119.
- De Wit, E., Delport, W., Rugamika, C.E., Meintjes, A., Möller, M., Van Helden, P.D., Seoighe, C. & Hoal, E.G.J.H.g. 2010. Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. 128(2):145-153.
- Debray, F.-G., Morin, C., Janvier, A., Villeneuve, J., Maranda, B., Laframboise, R., Lacroix, J., Decarie, J.-C., Robitaille, Y. & Lambert, M. 2011. LRPPRC mutations cause a phenotypically distinct form of Leigh syndrome with cytochrome c oxidase deficiency. *Journal of medical genetics*, 48(3):183-189.
- Dercksen, M., Duran, M., IJlst, L., Mienie, L.J., Reinecke, C.J., Ruiter, J., Waterham, H.R. & Wanders, R.J. 2012. Clinical variability of isovaleric acidemia in a genetically homogeneous population. *Journal of inherited metabolic disease*, 35(6):1021-1029.
- Deschauer, M., Bamberg, C., Claus, D., Zierz, S., Turnbull, D. & Taylor, R. 2003. Late-onset encephalopathy associated with a C11777A mutation of mitochondrial DNA. *Neurology*, 60(8):1357-1359.
- Diekman, E.F., Ferdinandusse, S., Van Der Pol, L., Waterham, H.R., Ruiter, J.P., IJlst, L., Wanders, R.J., Houten, S.M., Wijburg, F.A. & Blank, A.C.J.G.i.M. 2015. Fatty acid oxidation flux predicts the clinical severity of VLCAD deficiency. 17(12):989.
- Dillon, O.J., Lunke, S., Stark, Z., Yeung, A., Thorne, N., Gaff, C., White, S.M. & Tan, T.Y.J.E.J.o.H.G. 2018. Exome sequencing has higher diagnostic yield compared to simulated disease-specific panels in children with suspected monogenic disorders. 26(5):644.
- DiMauro, S. & Moraes, C.T. 1993. Mitochondrial encephalomyopathies. *Archives of Neurology*, 50(11):1197-1208.
- DiMauro, S. & Schon, E.A. 2001. Mitochondrial DNA mutations in human disease. *American Journal of Medical Genetics Part A*, 106(1):18-26.
- DiMauro, S. & Schon, E.A. 2003. Mitochondrial respiratory-chain diseases. *New England Journal of Medicine*, 348(26):2656-2668.
- Distelmaier, F., Koopman, W.J., van den Heuvel, L.P., Rodenburg, R.J., Mayatepek, E., Willems, P.H. & Smeitink, J.A. 2009. Mitochondrial complex I deficiency: from organelle dysfunction to clinical disease. *Brain*, 132(4):833-842.
- Doimo, M., Desbats, M.A., Cerqua, C., Cassina, M., Trevisson, E. & Salviati, L. 2014. Genetics of coenzyme q10 deficiency. *Mol Syndromol*, 5(3-4):156-162.
- Dolezal, P., Likic, V., Tachezy, J. & Lithgow, T. 2006. Evolution of the molecular machines for protein import into mitochondria. *Science*, 313(5785):314-318.
- Ducamp, S., Schneider-Yin, X., de Rooij, F., Clayton, J., Fratz, E.J., Rudd, A., Ostapowicz, G., Varigos, G., Lefebvre, T. & Deybach, J.-C.J.H.m.g. 2012. Molecular and functional analysis of the C-terminal region of human erythroid-specific 5-aminolevulinic synthase associated with X-linked dominant protoporphyria (XLDPP). 22(7):1280-1288.
- Duchen, M.R. 2004. Mitochondria in health and disease: perspectives on a new mitochondrial biology. *Molecular aspects of medicine*, 25(4):365-451.
- Duncan, A.J., Bitner-Glindzicz, M., Meunier, B., Costello, H., Hargreaves, I.P., López, L.C., Hirano, M., Quinzii, C.M., Sadowski, M.I. & Hardy, J. 2009. A nonsense mutation in COQ9 causes autosomal-

recessive neonatal-onset primary coenzyme Q10 deficiency: a potentially treatable form of mitochondrial disease. *The American Journal of Human Genetics*, 84(5):558-566.

Dunning, C.R., McKenzie, M., Sugiana, C., Lazarou, M., Silke, J., Connelly, A., Fletcher, J., Kirby, D., Thorburn, D. & Ryan, M. 2007. Human CIA30 is involved in the early assembly of mitochondrial complex I and mutations in its gene cause disease. *The EMBO journal*, 26(13):3227-3237.

Enriquez, J.A. & Lenaz, G. 2014. Coenzyme q and the respiratory chain: coenzyme q pool and mitochondrial supercomplexes. *Molecular syndromology*, 5(3-4):119-140.

Evans, M., Andresen, B.S., Nation, J., Boneh, A.J.M.g. & metabolism. 2016. VLCAD deficiency: follow-up and outcome of patients diagnosed through newborn screening in Victoria. 118(4):282-287.

Fassone, E. & Rahman, S. 2012. Complex I deficiency: clinical features, biochemistry and molecular genetics. *Journal of medical genetics*, 49(9):578-590.

Fernández-Vizarra, E. & Zeviani, M. 2015. Nuclear gene mutations as the cause of mitochondrial complex III deficiency. *Frontiers in genetics*, 6.

Ferrari, G., Lamantea, E., Donati, A., Filosto, M., Briem, E., Carrara, F., Parini, R., Simonati, A., Santer, R. & Zeviani, M. 2005. Infantile hepatocerebral syndromes associated with mutations in the mitochondrial DNA polymerase- γ A. *Brain*, 128(4):723-731.

Frerman, F. 2001. Defects of electron transfer flavoprotein and electron transfer flavoprotein-ubiquinone oxidoreductase Glutaric aciduria type II. *The metabolic and molecular basis of inherited disease*:2357-2365.

Freyer, C., Stranneheim, H., Naess, K., Mourier, A., Felser, A., Maffezzini, C., Lesko, N., Bruhn, H., Engvall, M. & Wibom, R. 2015. Rescue of primary ubiquinone deficiency due to a novel COQ7 defect using 2, 4-dihydroxybenzoic acid. *Journal of medical genetics*, 52(11):779-783.

Friederich, M.W., Coughlin, C.R., O'Rourke, C., Lovell, M.A., Gowan, K. & Van Hove, J.L. 2015. Mutations in NDUFB10 result in isolated complex I deficiency due to incomplete assembly of complex I holoenzyme. *Mitochondrion*, 24:S26.

Friedman, J.R. & Nunnari, J. 2014. Mitochondrial form and function. *Nature*, 505(7483):335-343.

Fu, H.-x., Liu, X.-y., Wang, Z.-q., Jin, M., Wang, D.-n., He, J.-j., Lin, M.-t. & Wang, N. 2016. Significant clinical heterogeneity with similar ETFDH genotype in three Chinese patients with late-onset multiple acyl-CoA dehydrogenase deficiency. *Neurological Sciences*, 37(7):1099-1105.

Gamba, J., Kiyomoto, B.H., de Oliveira, A.S.B., Gabbai, A.A., Schmidt, B. & Tengan, C.H. 2012. The mutations m. 5628T> C and m. 8348A> G in single muscle fibers of a patient with chronic progressive external ophthalmoplegia. *Journal of the neurological sciences*, 320(1-2):131-135.

Garone, C., Rubio, J.C., Calvo, S.E., Naini, A., Tanji, K., DiMauro, S., Mootha, V.K. & Hirano, M. 2012. MPV17 mutations causing adult-onset multisystemic disorder with multiple mitochondrial DNA deletions. *Archives of neurology*, 69(12):1648-1651.

Genome Mining. 2017. GEMINI: a flexible framework for exploring genome variation. [<https://gemini.readthedocs.io/en/latest/>] Date of access: 30 Oct 2018.

Gerards, M., Sluiter, W., van den Bosch, B.J., de Wit, L., Calis, C.M., Frentzen, M., Akbari, H., Schoonderwoerd, K., Scholte, H. & Jongbloed, R.J. 2010. Defective complex I assembly due to C20orf7 mutations as a new cause of Leigh syndrome. *Journal of medical genetics*, 47(8):507-512.

Ghezzi, D., Goffrini, P., Uziel, G., Horvath, R., Klopstock, T., Lochmüller, H., D'Adamo, P., Gasparini, P., Strom, T.M. & Prokisch, H. 2009. SDHAF1, encoding a LYR complex-II specific assembly factor, is mutated in SDH-defective infantile leukoencephalopathy. *Nature genetics*, 41(6):654.

Ghezzi, D., Saada, A., D'Adamo, P., Fernandez-Vizarra, E., Gasparini, P., Tiranti, V., Elpeleg, O. & Zeviani, M. 2008. FASTKD2 nonsense mutation in an infantile mitochondrial encephalomyopathy associated with cytochrome c oxidase deficiency. *The American Journal of Human Genetics*, 83(3):415-423.

Giachin, G., Bouverot, R., Acajjaoui, S., Pantalone, S. & Soler-López, M. 2016. Dynamics of human mitochondrial complex I assembly: implications for neurodegenerative diseases. *Frontiers in molecular biosciences*, 3.

Giles, R.E., Blanc, H., Cann, H.M. & Wallace, D.C. 1980. Maternal inheritance of human mitochondrial DNA. *Proceedings of the National academy of Sciences*, 77(11):6715-6719.

Gonzalez, F., D'Aurelio, M., Boutant, M., Moustapha, A., Puech, J.-P., Landes, T., Arnauné-Pelloquin, L., Vial, G., Taleux, N. & Slomianny, C.J.B.e.B.A.-M.B.o.D. 2013. Barth syndrome: cellular compensation of mitochondrial dysfunction and apoptosis inhibition due to changes in cardiolipin remodeling linked to tafazzin (TAZ) gene mutation. 1832(8):1194-1206.

Goodman, S.I., Markey, S.P., Moe, P.G., Miles, B.S. & Teng, C.C. 1975. Glutaric aciduria; a "new" disorder of amino acid metabolism. *Biochemical medicine*, 12(1):12-21.

Goodman, S.I., Stein, D.E., Schlesinger, S., Christensen, E., Schwartz, M., Greenberg, C.R. & Elpeleg, O.N. 1998. Glutaryl-CoA dehydrogenase mutations in glutaric acidemia (type I): review and report of thirty novel mutations. *Human mutation*, 12(3):141.

Görlach, A., Bertram, K., Hudecova, S. & Krizanová, O. 2015. Calcium and ROS: a mutual interplay. *Redox biology*, 6:260-271.

Gorman, G.S., Chinnery, P.F., DiMauro, S., Hirano, M., Koga, Y., McFarland, R., Suomalainen, A., Thorburn, D.R., Zeviani, M. & Turnbull, D.M. 2016. Mitochondrial diseases. *Nat Rev Dis Primers*, 2:16080.

Gorman, G.S., Schaefer, A.M., Ng, Y., Gomez, N., Blakely, E.L., Alston, C.L., Feeney, C., Horvath, R., Yu-Wai-Man, P. & Chinnery, P.F. 2015. Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Annals of neurology*, 77(5):753-759.

Graham, E., Lee, J., Price, M., Tarailo-Graovac, M., Matthews, A., Engelke, U., Tang, J., Kluijtmans, L.A., Wevers, R.A. & Wasserman, W.W.J.J.o.i.m.d. 2018. Integration of genomics and metabolomics for prioritization of rare disease variants: a 2018 literature review. 41(3):435-445.

Grünert, S.C. 2014. Clinical and genetical heterogeneity of late-onset multiple acyl-coenzyme A dehydrogenase deficiency. *Orphanet journal of rare diseases*, 9(1):117.

Grzybowski, M., Schänzer, A., Pepler, A., Heller, C., Neubauer, B.A. & Hahn, A.J.N. 2017. Novel STAC3 Mutations in the first non-Amerindian patient with Native American myopathy. 48(06):451-455.

Guan, M.-X., Yan, Q., Li, X., Bykhovskaya, Y., Gallo-Teran, J., Hajek, P., Umeda, N., Zhao, H., Garrido, G. & Mengesha, E. 2006. Mutation in TRMU related to transfer RNA modification modulates the phenotypic expression of the deafness-associated mitochondrial 12S ribosomal RNA mutations. *The American Journal of Human Genetics*, 79(2):291-302.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O. & Choudhury, A.J.N. 2015. The African genome variation project shapes medical genetics in Africa. 517(7534):327.

Haack, T.B., Danhauser, K., Haberberger, B., Hoser, J., Strecker, V., Boehm, D., Uziel, G., Lamantea, E., Invernizzi, F. & Poulton, J. 2010. Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nature genetics*, 42(12):1131.

Haack, T.B., Haberberger, B., Frisch, E.-M., Wieland, T., Iuso, A., Gorza, M., Strecker, V., Graf, E., Mayr, J.A. & Herberg, U. 2012. Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. *Journal of medical genetics*, 49(4):277-283.

- Hallmann, K., Kudin, A.P., Zsurka, G., Kornblum, C., Reimann, J., Stüve, B., Waltz, S., Hattingen, E., Thiele, H. & Nürnberg, P. 2015. Loss of the smallest subunit of cytochrome c oxidase, COX8A, causes Leigh-like syndrome and epilepsy. *Brain*, 139(2):338-345.
- Ham, M., Han, J., Osann, K., Smith, M. & Kimonis, V.J.M. 2018. Meta-analysis of genotype-phenotype analysis of OPA1 mutations in autosomal dominant optic atrophy.
- Hao, H.-X., Khalimonchuk, O., Schraders, M., Dephoure, N., Bayley, J.-P., Kunst, H., Devilee, P., Cremers, C.W., Schiffman, J.D. & Bentz, B.G. 2009. SDH5, a gene required for flavination of succinate dehydrogenase, is mutated in paraganglioma. *Science*, 325(5944):1139-1142.
- Harding, A., Holt, I., Sweeney, M., Brockington, M. & Davis, M. 1992. Prenatal diagnosis of mitochondrial DNA8993 T----G disease. *American journal of human genetics*, 50(3):629.
- Hatefi, Y. 1985. The mitochondrial electron transport and oxidative phosphorylation system. *Annual review of biochemistry*, 54(1):1015-1069.
- Haut, S., Brivet, M., Touati, G., Rustin, P., Lebon, S., Garcia-Cazorla, A., Saudubray, J.M., Boutron, A., Legrand, A. & Slama, A. 2003. A deletion in the human QP-C gene causes a complex III deficiency resulting in hypoglycaemia and lactic acidosis. *Human genetics*, 113(2):118-122.
- Hayashi, J.-I., Ohta, S., Takai, D., Miyabayashi, S., Sakuta, R., Goto, Y.-i. & Nonaka, I. 1993. Accumulation of mtDNA with a mutation at position 3271 in tRNA^{Leu} (UUR) gene introduced from a MELAS patient to HeLa cells lacking mtDNA results in progressive inhibition of mitochondrial respiratory function. *Biochemical and biophysical research communications*, 197(3):1049-1055.
- Hedlund, G.L., Longo, N. & Pasquali, M. 2006. Glutaric acidemia type 1. (In: American Journal of Medical Genetics Part C: Seminars in Medical Genetics organised by: Wiley Online Library. p. 86-94).
- Heeringa, S.F., Chernin, G., Chaki, M., Zhou, W., Sloan, A.J., Ji, Z., Xie, L.X., Salvati, L., Hurd, T.W. & Vega-Warner, V. 2011. COQ6 mutations in human patients produce nephrotic syndrome with sensorineural deafness. *The Journal of clinical investigation*, 121(5):2013-2024.
- Hejzlarová, K., Mracek, T., Vrbacký, M., Kaplanova, V., Karbanova, V., Nuskova, H., Pecina, P. & Houstek, J. 2014. Nuclear genetic defects of mitochondrial ATP synthase. *Physiological research*, 63:S57.
- Herbert, M. & Turnbull, D.J.N.R.M.C.B. 2018. Progress in mitochondrial replacement therapies. 19(2):71-72.
- Herrmann, J.M., Bihlmaier, K. & Mesecke, N. 2007. The Enzymes. Elsevier. p. 345-366).
- Hinson, J.T., Fantin, V.R., Schönberger, J., Breivik, N., Siem, G., McDonough, B., Sharma, P., Keogh, I., Godinho, R. & Santos, F. 2007. Missense mutations in the BCS1L gene as a cause of the Björnstad syndrome. *New England Journal of Medicine*, 356(8):809-819.
- Hoefs, S.J., Dieteren, C.E., Distelmaier, F., Janssen, R.J., Epplen, A., Swarts, H.G., Forkink, M., Rodenburg, R.J., Nijtmans, L.G. & Willems, P.H. 2008. NDUFA2 complex I mutation leads to Leigh disease. *The American Journal of Human Genetics*, 82(6):1306-1315.
- Hoefs, S.J., van Spronsen, F.J., Lenssen, E.W., Nijtmans, L.G., Rodenburg, R.J., Smeitink, J.A. & van den Heuvel, L.P. 2011. NDUFA10 mutations cause complex I deficiency in a patient with Leigh disease. *European Journal of Human Genetics*, 19(3):270.
- Hopper, R.K., Carroll, S., Aponte, A.M., Johnson, D.T., French, S., Shen, R.-F., Witzmann, F.A., Harris, R.A. & Balaban, R.S. 2006. Mitochondrial matrix phosphoproteome: effect of extra mitochondrial calcium. *Biochemistry*, 45(8):2524-2536.
- Horstick, E.J., Linsley, J.W., Dowling, J.J., Hauser, M.A., McDonald, K.K., Ashley-Koch, A., Saint-Amant, L., Satish, A., Cui, W.W. & Zhou, W. 2013. Stac3 is a component of the excitation-contraction coupling machinery and mutated in Native American myopathy. *Nature communications*, 4:1952.

Horvath, R., Schneiderat, P., Schoser, B.G., Gempel, K., Neuen-Jacob, E., Plöger, H., Müller-Höcker, J., Pongratz, D., Naini, A. & DiMauro, S. 2006. Coenzyme Q10 deficiency and isolated myopathy. *Neurology*, 66(2):253-255.

Howell, N., Bindoff, L., McCullough, D., Kubacka, I., Poulton, J., Mackey, D., Taylor, L. & Turnbull, D. 1991. Leber hereditary optic neuropathy: identification of the same mitochondrial ND1 mutation in six pedigrees. *American journal of human genetics*, 49(5):939.

Huigsloot, M., Nijtmans, L.G., Szklarczyk, R., Baars, M.J., van den Brand, M.A., HendriksFranssen, M.G., van den Heuvel, L.P., Smeitink, J.A., Huynen, M.A. & Rodenburg, R.J. 2011. A mutation in C2orf64 causes impaired cytochrome c oxidase assembly and mitochondrial cardiomyopathy. *The American Journal of Human Genetics*, 88(4):488-493.

Imai-Okazaki, A., Kishita, Y., Kohda, M., Yatsuka, Y., Hirata, T., Mizuno, Y., Harashima, H., Hirono, K., Ichida, F. & Noguchi, A.J.T.J.o.p. 2018. Barth Syndrome: Different Approaches to Diagnosis. 193:256-260.

Indrieri, A., van Rahden, V.A., Tiranti, V., Morleo, M., Iaconis, D., Tammaro, R., D'Amato, I., Conte, I., Maystadt, I. & Demuth, S. 2012. Mutations in COX7B cause microphthalmia with linear skin lesions, an unconventional mitochondrial disease. *The American Journal of Human Genetics*, 91(5):942-949.

Isohanni, P., Carroll, C.J., Jackson, C.B., Pohjanpelto, M., Lönnqvist, T. & Suomalainen, A. 2018. Defective mitochondrial ATPase due to rare mtDNA m. 8969G> A mutation—causing lactic acidosis, intellectual disability, and poor growth. *neurogenetics*:1-5.

Itkonen, O., Suomalainen, A. & Turpeinen, U. 2013. Mitochondrial coenzyme Q10 determination by isotope-dilution liquid chromatography–tandem mass spectrometry. *Clinical chemistry*, 59(8):1260-1267.

Jain-Ghai, S., Cameron, J.M., Al Maawali, A., Blaser, S., MacKay, N., Robinson, B. & Raiman, J. 2013. Complex II deficiency—a case report and review of the literature. *American journal of medical genetics Part A*, 161(2):285-294.

Jaksch, M., Klopstock, T., Kurlermann, G., Dörner, M., Hofmann, S., Kleinle, S., Hegemann, S., Weissert, M., Müller-Höcker, J. & Pongratz, D. 1998. Progressive myoclonus epilepsy and mitochondrial myopathy associated with mutations in the tRNASer (UCN) gene. *Annals of neurology*, 44(4):635-640.

Janer, A., Antonicka, H., Lalonde, E., Nishimura, T., Sasarman, F., Brown, G.K., Brown, R.M., Majewski, J. & Shoubridge, E.A. 2012. An RMND1 Mutation causes encephalopathy associated with multiple oxidative phosphorylation complex deficiencies and a mitochondrial translation defect. *The American Journal of Human Genetics*, 91(4):737-743.

Janssen, A.J., Trijbels, F.J., Sengers, R.C., Smeitink, J.A., Van den Heuvel, L.P., Wintjes, L.T., Stoltenberg-Hogenkamp, B.J. & Rodenburg, R.J. 2007. Spectrophotometric assay for complex I of the respiratory chain in tissue samples and cultured fibroblasts. *Clinical chemistry*, 53(4):729-734.

Jonassen, T. & Clarke, C.F. 2000. Isolation and functional expression of human COQ3, a gene encoding a methyltransferase required for ubiquinone biosynthesis. *Journal of Biological Chemistry*, 275(17):12381-12387.

Jonck, L.-M. 2014. CoenzymeQ10-associated gene mutations in South African patients with respiratory chain deficiencies. North-West University. (Dissertation).

Jonckheere, A.I., Renkema, G.H., Bras, M., van den Heuvel, L.P., Hoischen, A., Gilissen, C., Nabuurs, S.B., Huynen, M.A., de Vries, M.C. & Smeitink, J.A. 2013. A complex V ATP5A1 defect causes fatal neonatal mitochondrial encephalopathy. *Brain*, 136(5):1544-1554.

Jonckheere, A.I., Smeitink, J.A. & Rodenburg, R.J. 2012. Mitochondrial ATP synthase: architecture, function and pathology. *Journal of inherited metabolic disease*, 35(2):211-225.

Karczewski, K.J., Weisburd, B., Thomas, B., Solomonson, M., Ruderfer, D.M., Kavanagh, D., Hamamsy, T., Lek, M., Samocha, K.E. & Cummings, B.B. 2016. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic acids research*, 45(D1):D840-D845.

- Karki, R., Pandya, D., Elston, R.C. & Ferlini, C.J.B.m.g. 2015. Defining “mutation” and “polymorphism” in the era of personal genomics. 8(1):37.
- Keogh, M., Steele, H. & Chinnery, P.F. 2018. (In Chinnery, P.F. & Keogh, M., eds. Clinical Mitochondrial Medicine Cambridge: Cambridge University Press. p. pp. 69-83).
- Khan, I., Crouch, J.D., Bharti, S.K., Sommers, J.A., Carney, S.M., Yakubovskaya, E., Garcia-Diaz, M., Trakselis, M.A. & Brosh, R.M. 2016. Biochemical Characterization of the Human Mitochondrial Replicative Twinkle Helicase SUBSTRATE SPECIFICITY, DNA BRANCH MIGRATION, AND ABILITY TO OVERCOME BLOCKADES TO DNA UNWINDING. *Journal of Biological Chemistry*, 291(27):14324-14339.
- Kircher, M., Witten, D.M., Jain, P., O’roak, B.J., Cooper, G.M. & Shendure, J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics*, 46(3):310.
- Kitajiri, S.-i., Sakamoto, T., Belyantseva, I.A., Goodyear, R.J., Stepanyan, R., Fujiwara, I., Bird, J.E., Riazuddin, S., Riazuddin, S. & Ahmed, Z.M.J.C. 2010. Actin-bundling protein TRIOBP forms resilient rootlets of hair cell stereocilia essential for hearing. 141(5):786-798.
- Kmita, K. & Zickermann, V. 2013. Accessory subunits of mitochondrial complex I: Portland Press Limited.
- Kölker, S., Christensen, E., Leonard, J.V., Greenberg, C.R., Boneh, A., Burlina, A.B., Burlina, A.P., Dixon, M., Duran, M. & Cazorla, A.G. 2011. Diagnosis and management of glutaric aciduria type I—revised recommendations. *Journal of inherited metabolic disease*, 34(3):677.
- Kölker, S., Valayannopoulos, V., Burlina, A.B., Sykut-Cegielska, J., Wijburg, F.A., Teles, E.L., Zeman, J., Dionisi-Vici, C., Barić, I. & Karall, D. 2015. The phenotypic spectrum of organic acidurias and urea cycle disorders. Part 2: the evolving clinical phenotype. *Journal of inherited metabolic disease*, 38(6):1059-1074.
- Komaki, H., Akanuma, J., Iwata, H., Takahashi, T., Mashima, Y., Nonaka, I. & Goto, Y.-i. 2003. A novel mtDNA C11777A mutation in Leigh syndrome. *Mitochondrion*, 2(4):293-304.
- Koopman, W.J., Beyrath, J., Fung, C.W., Koene, S., Rodenburg, R.J., Willems, P.H. & Smeitink, J.A. 2016. Mitochondrial disorders in children: toward development of small-molecule treatment strategies. *EMBO molecular medicine*, 8(4):311-327.
- Kotlar, A.V., Trevino, C.E., Zwick, M.E., Cutler, D.J. & Wingo, T.S. 2018. Bystro: rapid online variant annotation and natural-language filtering at whole-genome scale. *Genome biology*, 19(1):14.
- Külkens, S., Harting, I., Sauer, S., Zschocke, J., Hoffmann, G., Gruber, S., Bodamer, O. & Kölker, S. 2005. Late-onset neurologic disease in glutaryl-CoA dehydrogenase deficiency. *Neurology*, 64(12):2142-2144.
- Lagier-Tourenne, C., Tazir, M., López, L.C., Quinzii, C.M., Assoum, M., Drouot, N., Busso, C., Makri, S., Ali-Pacha, L. & Benhassine, T. 2008. ADCK3, an ancestral kinase, is mutated in a form of recessive ataxia associated with coenzyme Q10 deficiency. *The American Journal of Human Genetics*, 82(3):661-672.
- Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M. & Maglott, D.R. 2013. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic acids research*, 42(D1):D980-D985.
- Lane, A., Soodyall, H., Arndt, S., Ratshikhopha, M., Jonker, E., Freeman, C., Young, L., Morar, B. & Toffie, L.J.A.J.o.P.A.T.O.P.o.t.A.A.o.P.A. 2002. Genetic substructure in South African Bantu-speakers: Evidence from autosomal DNA and Y-chromosome studies. 119(2):175-185.
- Laugel, V., This-Bernd, V., Cormier-Daire, V., Speeg-Schatz, C., de Saint-Martin, A. & Fischbach, M. 2007. Early-onset ophthalmoplegia in Leigh-like syndrome due to NDUFV1 mutations. *Pediatric neurology*, 36(1):54-57.
- Lesko, N., Naess, K., Wibom, R., Solaroli, N., Nennesmo, I., von Döbeln, U., Karlsson, A. & Larsson, N.-G. 2010. Two novel mutations in thymidine kinase-2 cause early onset fatal encephalomyopathy and severe mtDNA depletion. *Neuromuscular Disorders*, 20(3):198-203.

- Letts, J.A., Fiedorczuk, K. & Sazanov, L.A. 2016. The architecture of respiratory supercomplexes. *Nature*, 537(7622):644.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078-2079.
- Liang, W.-C., Ohkuma, A., Hayashi, Y.K., López, L.C., Hirano, M., Nonaka, I., Noguchi, S., Chen, L.-H., Jong, Y.-J. & Nishino, I. 2009. ETFDH mutations, CoQ10 levels, and respiratory chain activities in patients with riboflavin-responsive multiple acyl-CoA dehydrogenase deficiency. *Neuromuscular Disorders*, 19(3):212-216.
- Lieber, D.S., Calvo, S.E., Shanahan, K., Slate, N.G., Liu, S., Hershman, S.G., Gold, N.B., Chapman, B.A., Thorburn, D.R. & Berry, G.T. 2013. Targeted exome sequencing of suspected mitochondrial disorders. *Neurology*, 80(19):1762-1770.
- Lindner, M., Kölker, S., Schulze, A., Christensen, E., Greenberg, C. & Hoffmann, G. 2004. Neonatal screening for glutaryl-CoA dehydrogenase deficiency. *Journal of inherited metabolic disease*, 27(6):851-859.
- Lionel, A.C., Costain, G., Monfared, N., Walker, S., Reuter, M.S., Hosseini, S.M., Thiruvahindrapuram, B., Merico, D., Jobling, R. & Nalpathamkalam, T.J.G.i.M. 2018. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. 20(4):435.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. & Law, M. 2012. Comparison of next-generation sequencing systems. *BioMed Research International*, 2012.
- Loeffen, J., Elpeleg, O., Smeitink, J., Smeets, R., Stöckler-Ipsiroglu, S., Mandel, H., Sengers, R., Trijbels, F. & Van Den Heuvel, L. 2001. Mutations in the complex I NDUFS2 gene of patients with cardiomyopathy and encephalomyopathy. *Annals of neurology*, 49(2):195-201.
- Loeffen, J., Smeitink, J., Triepels, R., Smeets, R., Schuelke, M., Sengers, R., Trijbels, F., Hamel, B., Mullaart, R. & van den Heuvel, L. 1998. The first nuclear-encoded complex I mutation in a patient with Leigh syndrome. *The American Journal of Human Genetics*, 63(6):1598-1608.
- Löllgen, S. & Weiher, H. 2015. The role of the Mpv17 protein mutations of which cause mitochondrial DNA depletion syndrome (MDDS): lessons from homologs in different species. *Biological chemistry*, 396(1):13-25.
- Longo, N., Amat di San Filippo, C. & Pasquali, M. 2006. Disorders of carnitine transport and the carnitine cycle. (In: American Journal of Medical Genetics Part C: Seminars in Medical Genetics organised by: Wiley Online Library. p. 77-85).
- López, L.C., Schuelke, M., Quinzii, C.M., Kanki, T., Rodenburg, R.J., Naini, A., DiMauro, S. & Hirano, M. 2006. Leigh syndrome with nephropathy and CoQ10 deficiency due to decaprenyl diphosphate synthase subunit 2 (PDSS2) mutations. *The American Journal of Human Genetics*, 79(6):1125-1129.
- Lott, M.T., Leipzig, J.N., Derbeneva, O., Xie, H.M., Chalkia, D., Sarmady, M., Procaccio, V. & Wallace, D.C. 2013. mtDNA variation and analysis using MITOMAP and MITOMASTER. *Current protocols in bioinformatics*:1.23. 21-21.23. 26.
- Louw, R., Smuts, I., Wilsenach, K.-L., Jonck, L.-M., Schoonen, M. & Van der Westhuizen, F.H. 2018. The dilemma of diagnosing coenzyme Q10 deficiency in muscle. *Molecular genetics and metabolism*.
- Luo, C., Long, J. & Liu, J. 2008. An improved spectrophotometric method for a more specific and accurate assay of mitochondrial complex III activity. *Clinica Chimica Acta*, 395(1):38-41.
- Luo, S., Valencia, C.A., Zhang, J., Lee, N.-C., Slone, J., Gui, B., Wang, X., Li, Z., Dell, S., Brown, J., Chen, S.M., Chien, Y.-H., Hwu, W.-L., Fan, P.-C., Wong, L.-J., Atwal, P.S. & Huang, T. 2018. Biparental Inheritance of Mitochondrial DNA in Humans.

- Malicdan, M.C.V., Vilboux, T., Ben-Zeev, B., Guo, J., Eliyahu, A., Pode-Shakked, B., Dori, A., Kakani, S., Chandrasekharappa, S.C. & Ferreira, C.R. 2018. A novel inborn error of the coenzyme Q10 biosynthesis pathway: cerebellar ataxia and static encephalomyopathy due to COQ5 C-methyltransferase deficiency. *Human mutation*, 39(1):69-79.
- Mancuso, M., Ferraris, S., Nishigaki, Y., Azan, G., Mauro, A., Sammarco, P., Krishna, S., Tay, S.K., Bonilla, E. & Romansky, S.G. 2005. Congenital or late-onset myopathy in patients with the T14709C mtDNA mutation. *Journal of the neurological sciences*, 228(1):93-97.
- Mancuso, M., Orsucci, D., Volpi, L., Calsolaro, V. & Siciliano, G. 2010. Coenzyme Q10 in neuromuscular and neurodegenerative disorders. *Current drug targets*, 11(1):111-121.
- Mannella, C.A. 2006. Structure and dynamics of the mitochondrial inner membrane cristae. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1763(5-6):542-548.
- Manouvrier, S., Rötig, A., Hannebique, G., Gheerbrandt, J., Royer-Legrain, G., Munnich, A., Parent, M., Grünfeld, J., Largilliere, C. & Lombs, A. 1995. Point mutation of the mitochondrial tRNA (Leu) gene (A 3243 G) in maternally inherited hypertrophic cardiomyopathy, diabetes mellitus, renal failure, and sensorineural deafness. *Journal of medical genetics*, 32(8):654-656.
- Martín, M.A., Blázquez, A., Gutierrez-Solana, L.G., Fernández-Moreira, D., Briones, P., Andreu, A.L., Garesse, R., Campos, Y. & Arenas, J. 2005. Leigh syndrome associated with mitochondrial complex I deficiency due to a novel mutation in the NDUFS1 gene. *Archives of neurology*, 62(4):659-661.
- Martins, R.G., Nunes, J.B., Máximo, V., Soares, P., Peixoto, J., Catarino, T., Rito, T., Soares, P., Pereira, L. & Sobrinho-Simões, M. 2013. A founder SDHB mutation in Portuguese paraganglioma patients. *Endocrine-related cancer*, 20(6):L23-L26.
- Massa, V., Fernandez-Vizarra, E., Alshahwan, S., Bakhsh, E., Goffrini, P., Ferrero, I., Mereghetti, P., D'Adamo, P., Gasparini, P. & Zeviani, M. 2008. Severe infantile encephalomyopathy caused by a mutation in COX6B1, a nucleus-encoded subunit of cytochrome c oxidase. *The American Journal of Human Genetics*, 82(6):1281-1289.
- Masucci, J.P., Davidson, M., Koga, Y., Schon, E.A. & King, M.P. 1995. In vitro analysis of mutations causing myoclonus epilepsy with ragged-red fibers in the mitochondrial tRNA (Lys) gene: two genotypes produce similar phenotypes. *Molecular and cellular biology*, 15(5):2872-2881.
- Mayr, J.A., Havlíčková, V., Zimmermann, F., Magler, I., Kaplanová, V., Ješina, P., Pecinová, A., Nůšková, H., Koch, J. & Sperl, W. 2010. Mitochondrial ATP synthase deficiency due to a mutation in the ATP5E gene for the F1 ϵ subunit. *Human molecular genetics*, 19(17):3430-3439.
- McFarland, R., Clark, K.M., Morris, A.A., Taylor, R.W., Macphail, S., Lightowlers, R.N. & Turnbull, D.M. 2002. Multiple neonatal deaths due to a homoplasmic mitochondrial DNA mutation. *Nature genetics*, 30(2):145.
- McFarland, R., Elson, J.L., Taylor, R.W., Howell, N. & Turnbull, D.M. 2004a. Assigning pathogenicity to mitochondrial tRNA mutations: when 'definitely maybe' is not good enough. *TRENDS in Genetics*, 20(12):591-596.
- McFarland, R., Schaefer, A.M., Gardner, J.L., Lynn, S., Hayes, C.M., Barron, M.J., Walker, M., Chinnery, P.F., Taylor, R.W. & Turnbull, D.M. 2004b. Familial myopathy: new insights into the T14709C mitochondrial tRNA mutation. *Annals of neurology*, 55(4):478-484.
- McKenzie, M., Lazarou, M., Thorburn, D.R. & Ryan, M.T. 2006. Mitochondrial respiratory chain supercomplexes are destabilized in Barth Syndrome patients. *J Mol Biol*, 361(3):462-469.
- McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P. & Cunningham, F. 2016. The ensembl variant effect predictor. *Genome biology*, 17(1):122.
- Meienberg, J., Bruggmann, R., Oexle, K. & Matyas, G.J.H.g. 2016. Clinical sequencing: is WGS the better WES? , 135(3):359-362.

Meldau, S., De Lacy, R., Riordan, G., Goddard, E., Pillay, K., Fieggen, K., Marais, D. & Van der Watt, G. 2018. Identification of a single MPV17 nonsense-associated altered splice variant in 24 South African infants with mitochondrial neurohepatopathy. *Clinical genetics*.

Meldau, S., Riordan, G., Van der Westhuizen, F., Elson, J.L., Smuts, I., Pepper, M.S. & Soodyall, H. 2016. Could we offer mitochondrial donation or similar assisted reproductive technology to South African patients with mitochondrial DNA disease? *SAMJ: South African Medical Journal*, 106(3):234-236.

Miller, C., Wang, L., Ostergaard, E., Dan, P. & Saada, A. 2011. The interplay between SUCLA2, SUCLG2, and mitochondrial DNA depletion. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1812(5):625-629.

Mitchell, A.L., Elson, J.L., Howell, N., Taylor, R.W. & Turnbull, D.M. 2006. Sequence variation in mitochondrial complex I genes: mutation or polymorphism? *Journal of medical genetics*, 43(2):175-179.

Miyake, N., Yamashita, S., Kurosawa, K., Miyatake, S., Tsurusaki, Y., Doi, H., Saito, H. & Matsumoto, N. 2011. A novel homozygous mutation of DARS2 may cause a severe LBSL variant. *Clinical genetics*, 80(3):293-296.

Mollet, J., Delahodde, A., Serre, V., Chretien, D., Schlemmer, D., Lombes, A., Boddaert, N., Desguerre, I., De Lonlay, P. & De Baulny, H.O. 2008. CABP1 gene mutations cause ubiquinone deficiency with cerebellar ataxia and seizures. *The American Journal of Human Genetics*, 82(3):623-630.

Mollet, J., Giurgea, I., Schlemmer, D., Dallner, G., Chretien, D., Delahodde, A., Bacq, D., de Lonlay, P., Munnich, A. & Rötig, A. 2007. Prenyldiphosphate synthase, subunit 1 (PDSS1) and OH-benzoate polyprenyltransferase (COQ2) mutations in ubiquinone deficiency and oxidative phosphorylation disorders. *The Journal of clinical investigation*, 117(3):765-772.

Monnier, N., Laquerrière, A., Marret, S., Goldenberg, A., Marty, I., Nivoche, Y. & Lunardi, J.J.N.D. 2009. First genomic rearrangement of the RYR1 gene associated with an atypical presentation of lethal neonatal hypotonia. 19(10):680-684.

Montoya, J., López-Gallardo, E., Díez-Sánchez, C., López-Pérez, M.J. & Ruiz-Pesini, E.J.B.e.B.A.-B. 2009. 20 years of human mtDNA pathologic point mutations: carefully reading the pathogenicity criteria. 1787(5):476-483.

Moraes, C.T., Ciacci, F., Silvestri, G., Shanske, S., Sciacco, M., Hirano, M., Schon, E.A., Bonilla, E. & DiMauro, S. 1993. Atypical clinical presentations associated with the MELAS mutation at position 3243 of human mitochondrial DNA. *Neuromuscular Disorders*, 3(1):43-50.

Mulder, N.J., Christoffels, A., De Oliveira, T., Gamielien, J., Hazelhurst, S., Joubert, F., Kumuthini, J., Pillay, C.S., Snoep, J.L. & Bishop, Ö.T. 2016. The development of computational biology in South Africa: successes achieved and lessons learnt. *PLoS computational biology*, 12(2):e1004395.

Naviaux, R.K. & Nguyen, K.V. 2004. POLG mutations associated with Alpers' syndrome and mitochondrial DNA depletion. *Annals of neurology*, 55(5):706-712.

Neveling, K., Feenstra, I., Gilissen, C., Hoefsloot, L.H., Kamsteeg, E.J., Mensenkamp, A.R., Rodenburg, R.J., Yntema, H.G., Spruijt, L., Vermeer, S., Rinne, T., van Gassen, K.L., Bodmer, D., Lugtenberg, D., de Reuver, R., Buijsman, W., Derks, R.C., Wieskamp, N., van den Heuvel, B., Ligtenberg, M.J., Kremer, H., Koolen, D.A., van de Warrenburg, B.P., Cremers, F.P., Marcelis, C.L., Smeitink, J.A., Wortmann, S.B., van Zelst-Stams, W.A., Veltman, J.A., Brunner, H.G., Scheffer, H. & Nelen, M.R. 2013. A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum Mutat*, 34(12):1721-1726.

Nguyen, K.V., Sharief, F.S., Chan, S.S., Copeland, W.C. & Naviaux, R.K. 2006. Molecular diagnosis of Alpers syndrome. *Journal of hepatology*, 45(1):108-116.

Ni, Y., Zbuk, K.M., Sadler, T., Patocs, A., Lobo, G., Edelman, E., Platzer, P., Orloff, M.S., Waite, K.A. & Eng, C. 2008. Germline mutations and variants in the succinate dehydrogenase genes in Cowden and Cowden-like syndromes. *The American Journal of Human Genetics*, 83(2):261-268.

- Niemann, S. & Müller, U. 2000. Mutations in SDHC cause autosomal dominant paraganglioma, type 3. *Nature genetics*, 26(3):268.
- Nijtmans, L.G., Ugalde, C., van den Heuvel, L.P. & Smeitink, J.A. 2004. Mitochondrial Function and Biogenesis. Springer. p. 149-176).
- Nishino, I., Spinazzola, A. & Hirano, M. 1999. Thymidine phosphorylase gene mutations in MNGIE, a human mitochondrial disorder. *Science*, 283(5402):689-692.
- Niyazov, D.M., Kahler, S.G. & Frye, R.E. 2016. Primary mitochondrial disease and secondary mitochondrial dysfunction: importance of distinction for diagnosis and treatment. *Molecular syndromology*, 7(3):122-137.
- Nunnari, J. & Suomalainen, A. 2012. Mitochondria: in sickness and in health. *Cell*, 148(6):1145-1159.
- Ogasahara, S., Engel, A.G., Frens, D. & Mack, D. 1989. Muscle coenzyme Q deficiency in familial mitochondrial encephalomyopathy. *Proceedings of the National Academy of Sciences*, 86(7):2379-2382.
- Ogilvie, I., Kennaway, N.G. & Shoubridge, E.A. 2005. A molecular chaperone for mitochondrial complex I assembly is mutated in a progressive encephalopathy. *The Journal of clinical investigation*, 115(10):2784-2792.
- Olsen, R.K., Olpin, S.E., Andresen, B.S., Miedzybrodzka, Z.H., Pourfarzam, M., Merinero, B., Frerman, F.E., Beresford, M.W., Dean, J.C. & Cornelius, N. 2007. ETFDH mutations as a major cause of riboflavin-responsive multiple acyl-CoA dehydrogenation deficiency. *Brain*, 130(8):2045-2054.
- Onat, O.E., Gulsuner, S., Bilguvar, K., Basak, A.N., Topaloglu, H., Tan, M., Tan, U., Gunel, M. & Ozcelik, T. 2013. Missense mutation in the ATPase, aminophospholipid transporter protein ATP8A2 is associated with cerebellar atrophy and quadrupedal locomotion. *European Journal of Human Genetics*, 21(3):281.
- Oquendo, C., Antonicka, H., Shoubridge, E., Reardon, W. & Brown, G. 2004. Functional and genetic studies demonstrate that mutation in the COX15 gene can cause Leigh syndrome. *Journal of medical genetics*, 41(7):540-544.
- Ostergaard, E., Rodenburg, R.J., van den Brand, M., Thomsen, L.L., Duno, M., Batbayli, M., Wibrand, F. & Nijtmans, L. 2011. Respiratory chain complex I deficiency due to NDUFA12 mutations as a new cause of Leigh syndrome. *Journal of medical genetics*, 48(11):737-740.
- Ostergaard, E., Weraarpachai, W., Ravn, K., Born, A.P., Jønson, L., Duno, M., Wibrand, F., Shoubridge, E.A. & Vissing, J. 2015. Mutations in COA3 cause isolated complex IV deficiency associated with neuropathy, exercise intolerance, obesity, and short stature. *Journal of medical genetics*:jmedgenet-2014-102914.
- Paila, U., Chapman, B.A., Kirchner, R. & Quinlan, A.R. 2013. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS computational biology*, 9(7):e1003153.
- Papadopoulou, L.C., Sue, C.M., Davidson, M.M., Tanji, K., Nishino, I., Sadlock, J.E., Krishna, S., Walker, W., Selby, J. & Glerum, D.M. 1999. Fatal infantile cardioencephalomyopathy with COX deficiency and mutations in SCO2, a COX assembly gene. *Nature genetics*, 23(3):333.
- Petersen, D.C., Libiger, O., Tindall, E.A., Hardie, R.-A., Hannick, L.I., Glashoff, R.H., Mukerji, M., Fernandez, P., Haacke, W. & Schork, N.J.J.P.g. 2013. Complex patterns of genomic admixture within southern Africa. 9(3):e1003309.
- Petruzzella, V., Vergari, R., Puzifferri, I., Boffoli, D., Lamantea, E., Zeviani, M. & Papa, S. 2001. A nonsense mutation in the NDUFS4 gene encoding the 18 kDa (AQDQ) subunit of complex I abolishes assembly and activity of the complex in a patient with Leigh-like syndrome. *Human molecular genetics*, 10(5):529-536.
- Plaitakis, A., Latsoudis, H., Kanavouras, K., Ritz, B., Bronstein, J.M., Skoula, I., Mastorodemos, V., Papapetropoulos, S., Borompokas, N. & Zaganas, I. 2010. Gain-of-function variant in GLUD2 glutamate dehydrogenase modifies Parkinson's disease onset. *European Journal of Human Genetics*, 18(3):336.

- Plutino, M., Chaussenot, A., Rouzier, C., Ait-El-Mkadem, S., Fragaki, K., Paquis-Flucklinger, V. & Bannwarth, S. 2018. Targeted next generation sequencing with an extended gene panel does not impact variant detection in mitochondrial diseases. *BMC medical genetics*, 19(1):57.
- Punetha, J., Kesari, A., Uapinyoying, P., Giri, M., Clarke, N.F., Waddell, L.B., North, K.N., Ghaoui, R., O'Grady, G.L. & Oates, E.C. 2016. Targeted re-sequencing emulsion PCR panel for myopathies: results in 94 cases. *Journal of neuromuscular diseases*, 3(2):209-225.
- Quinzii, C., Kattah, A., Naini, A., Akman, H., Mootha, V., DiMauro, S. & Hirano, M. 2005. Coenzyme Q deficiency and cerebellar ataxia associated with an aprataxin mutation. *Neurology*, 64(3):539-541.
- Quinzii, C., Naini, A., DiMauro, S. & Hirano, M. 2008. Human CoQ₁₀ deficiencies. *Biofactors*, 32(1-4):113-118.
- Quinzii, C., Naini, A., Salviati, L., Trevisson, E., Navas, P., DiMauro, S. & Hirano, M. 2006. A mutation in para-hydroxybenzoate-polyprenyl transferase (COQ2) causes primary coenzyme Q10 deficiency. *The American Journal of Human Genetics*, 78(2):345-349.
- Rahman, S. 2018. (In Chinnery, P.F. & Keogh, M., eds. *Clinical Mitochondrial Medicine* Cambridge: Cambridge University p. pp. 10-21).
- Rahman, S., Blok, R., Dahl, H.H., Danks, D., Kirby, D., Chow, C., Christodoulou, J. & Thorburn, D. 1996. Leigh syndrome: clinical features and biochemical and DNA abnormalities. *Annals of neurology*, 39(3):343-351.
- Rak, M., Bénit, P., Chrétien, D., Bouchereau, J., Schiff, M., El-Khoury, R., Tzagoloff, A. & Rustin, P. 2016. Mitochondrial cytochrome c oxidase deficiency. *Clinical Science*, 130(6):393-407.
- Ramey, C. 2011. The Bourne-Again SHell. *The Architecture of Open Source Applications*.
- Rathore, G., Star, L., Larsen, P. & Rizzo, W. 2017. Novel mutation of DARS2 gene leads to a rare and aggressive presentation of leukoencephalopathy with brainstem and spinal cord involvement and lactate elevation (LBSL)(P4. 174). *Neurology*, 88(16 Supplement):P4. 174.
- Reid, F.M., Vernham, G.A. & Jacobs, H.T. 1994. A novel mitochondrial point mutation in a maternal pedigree with sensorineural deafness. *Human mutation*, 3(3):243-247.
- Reinecke, C.J., Koekemoer, G., Van der Westhuizen, F.H., Louw, R., Lindeque, J.Z., Mienie, L.J. & Smuts, I. 2012. Metabolomics of urinary organic acids in respiratory chain deficiencies in children. *Metabolomics*, 8(2):264-283.
- Rendón, O.Z., Antonicka, H., Horvath, R. & Shoubridge, E.A. 2016. A Mutation in the Flavin Adenine Dinucleotide-Dependent Oxidoreductase FOXRED1 Results in Cell-Type-Specific Assembly Defects in Oxidative Phosphorylation Complexes I and II. *Molecular and cellular biology*, 36(16):2132-2140.
- Riazuddin, S., Khan, S.N., Ahmed, Z.M., Ghosh, M., Caution, K., Nazli, S., Kabra, M., Zafar, A.U., Chen, K. & Naz, S. 2006. Mutations in TRIOBP, which encodes a putative cytoskeletal-organizing protein, are associated with nonsyndromic recessive deafness. *The American Journal of Human Genetics*, 78(1):137-143.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E. & Spector, E. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine*, 17(5):405.
- Rodenburg, R.J. 2011. Biochemical diagnosis of mitochondrial disorders. *Journal of inherited metabolic disease*, 34(2):283-292.
- Rorbach, J., Richter, R., Wessels, H.J., Wydro, M., Pekalski, M., Farhoud, M., Kühl, I., Gaisne, M., Bonnefoy, N. & Smeitink, J.A. 2008. The human mitochondrial ribosome recycling factor is essential for cell viability. *Nucleic acids research*, 36(18):5787-5799.

- Rossignol, R., Faustin, B., Rocher, C., Malgat, M., Mazat, J.-P. & Letellier, T.J.B.J. 2003. Mitochondrial threshold effects. *370(3):751-762*.
- Saada, A., Edvardson, S., Rapoport, M., Shaag, A., Amry, K., Miller, C., Lorberboum-Galski, H. & Elpeleg, O. 2008. C6ORF66 is an assembly factor of mitochondrial complex I. *The American Journal of Human Genetics*, 82(1):32-38.
- Saada, A., Shaag, A., Mandel, H., Nevo, Y., Eriksson, S. & Elpeleg, O. 2001. Mutant mitochondrial thymidine kinase in mitochondrial DNA depletion myopathy. *Nature genetics*, 29(3):342.
- Saada, A., Vogel, R.O., Hoefs, S.J., van den Brand, M.A., Wessels, H.J., Willems, P.H., Venselaar, H., Shaag, A., Barghuti, F. & Reish, O. 2009. Mutations in NDUFAF3 (C3ORF60), encoding an NDUFAF4 (C6ORF66)-interacting complex I assembly protein, cause fatal neonatal mitochondrial disease. *The American Journal of Human Genetics*, 84(6):718-727.
- Sacconi, S., Salviati, L., Nishigaki, Y., Walker, W.F., Hernandez-Rosa, E., Trevisson, E., Delplace, S., Desnuelle, C., Shanske, S. & Hirano, M. 2008. A functionally dominant mitochondrial DNA mutation. *Human molecular genetics*, 17(12):1814-1820.
- Sacconi, S., Salviati, L., Sue, C.M., Shanske, S., Davidson, M.M., Bonilla, E., Naini, A.B., De Vivo, D.C. & DiMauro, S. 2003. Mutation screening in patients with isolated cytochrome c oxidase deficiency. *Pediatr Res*, 53(2):224-230.
- Salatino, S. & Ramraj, V. 2016. BrowseVCF: a web-based application and workflow to quickly prioritize disease-causative variants in VCF files. *Briefings in bioinformatics*, 18(5):774-779.
- Salviati, L., Sacconi, S., Murer, L., Zacchello, G., Franceschini, L., Laverda, A., Basso, G., Quinzii, C., Angelini, C. & Hirano, M. 2005. Infantile encephalomyopathy and nephropathy with CoQ10 deficiency: a CoQ10-responsive condition. *Neurology*, 65(4):606-608.
- Salviati, L., Trevisson, E., Hernandez, M.A.R., Casarin, A., Pertegato, V., Doimo, M., Cassina, M., Agosto, C., Desbats, M.A. & Sartori, G. 2012. Haploinsufficiency of COQ4 causes coenzyme Q10 deficiency. *Journal of medical genetics*, 49(3):187-191.
- Sánchez-Caballero, L., Guerrero-Castillo, S. & Nijtmans, L. 2016. Unraveling the complexity of mitochondrial complex I assembly: a dynamic process. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1857(7):980-990.
- Saneto, R.P., Cohen, B.H., Copeland, W.C. & Naviaux, R.K. 2013. Alpers-Huttenlocher syndrome. *Pediatric neurology*, 48(3):167-178.
- Santorelli, F.M., Siciliano, G., Casali, C., Basirico, M.G., Carrozzo, R., Calvosa, F., Sartucci, F., Bonfiglio, L., Murri, L. & DiMauro, S. 1997a. Mitochondrial tRNACys gene mutation (A5814G): a second family with mitochondrial encephalopathy. *Neuromuscular Disorders*, 7(3):156-159.
- Santorelli, F.M., Tanji, K., Kulikova, R., Shanske, S., Vilarinho, L., Hays, A.P. & DiMauro, S. 1997b. Identification of a novel mutation in the mtDNA ND5 gene associated with MELAS. *Biochemical and biophysical research communications*, 238(2):326-328.
- Sarzi, E., Brown, M.D., Lebon, S., Chretien, D., Munnich, A., Rotig, A. & Procaccio, V. 2007. A novel recurrent mitochondrial DNA mutation in ND3 gene is associated with isolated complex I deficiency causing Leigh syndrome and dystonia. *American journal of medical genetics Part A*, 143(1):33-41.
- Schlebusch, C.M., Skoglund, P., Sjödin, P., Gattepaille, L.M., Hernandez, D., Jay, F., Li, S., De Jongh, M., Singleton, A. & Blum, M.G.J.S. 2012. Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *1227721*.
- Schrier, S.A. & Falk, M.J. 2011. Mitochondrial disorders and the eye. *Current opinion in ophthalmology*, 22(5):325.
- Schwarz, J.M., Cooper, D.N., Schuelke, M. & Seelow, D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature methods*, 11(4):361.

- Seyffert, A.S. 2018. A set of BASH scripts for annotating and filtering Next-Generation Sequencing data using Ensembl's offline VEP runner GEMINI. [https://github.com/aseyffert/hetero_annotate] Date of access: 30 Oct 2018.
- Shahin, H., Walsh, T., Sobe, T., Rayan, A.A., Lynch, E.D., Lee, M.K., Avraham, K.B., King, M.-C. & Kanaan, M. 2006. Mutations in a novel isoform of TRIOBP that encodes a filamentous-actin binding protein are responsible for DFNB28 recessive nonsyndromic hearing loss. *The American Journal of Human Genetics*, 78(1):144-152.
- Shaibani, A., Shchelochkov, O.A., Zhang, S., Katsonis, P., Lichtarge, O., Wong, L.-J. & Shinawi, M. 2009. Mitochondrial neurogastrointestinal encephalopathy due to mutations in RRM2B. *Archives of neurology*, 66(8):1028-1032.
- Shearer, A.E., Eppsteiner, R.W., Booth, K.T., Ephraim, S.S., Gurrola II, J., Simpson, A., Black-Ziegelbein, E.A., Joshi, S., Ravi, H. & Giuffre, A.C. 2014. Utilizing ethnic-specific differences in minor allele frequency to recategorize reported pathogenic deafness variants. *The American Journal of Human Genetics*, 95(4):445-453.
- Shepherd, D. & Garland, P. 1969. The kinetic properties of citrate synthase from rat liver mitochondria. *Biochemical Journal*, 114(3):597-610.
- Sherman, R.M., Forman, J., Antonescu, V., Puiu, D., Daya, M., Rafaels, N., Boorgula, M.P., Chavan, S., Vergara, C. & Ortega, V.E.J.N.G. 2018. Assembly of a pan-genome from deep sequencing of 910 humans of African descent.1.
- Shimajima, K., Higashiguchi, T., Kishimoto, K., Miyatake, S., Miyake, N., Takanashi, J.-i., Matsumoto, N. & Yamamoto, T. 2017. A novel DARS2 mutation in a Japanese patient with leukoencephalopathy with brainstem and spinal cord involvement but no lactate elevation. *Human genome variation*, 4:17051.
- Shoffner, J.J.T.m. & disease, m.b.o.i. 1995. Oxidative phosphorylation disease.
- Shoffner, J.M., Lott, M.T., Lezza, A.M., Seibel, P., Ballinger, S.W. & Wallace, D.C. 1990. Myoclonic epilepsy and ragged-red fiber disease (MERRF) is associated with a mitochondrial DNA tRNA^{Lys} mutation. *Cell*, 61(6):931-937.
- Shoubridge, E.A.J.H.m.g. 2001. Nuclear genetic defects of oxidative phosphorylation. 10(20):2277-2284.
- Shteyer, E., Saada, A., Shaag, A., Kidess, R., Revel-Vilk, S. & Elpeleg, O. 2009. Exocrine pancreatic insufficiency, dyserythropoietic anemia, and calvarial hyperostosis are caused by a mutation in the COX4I2 gene. *The American Journal of Human Genetics*, 84(3):412-417.
- Sjöstrand, F.S. 2018. The evolution of membrane models. *Structure and Properties of Cell Membrane Structure and Properties of Cell Membranes*, 1.
- Skladal, D., Halliday, J. & Thorburn, D.R. 2003. Minimum birth prevalence of mitochondrial respiratory chain disorders in children. *Brain*, 126(8):1905-1912.
- Smeitink, J., Sengers, R., Trijbels, F. & van den Heuvel, L. 2001. Human NADH: ubiquinone oxidoreductase. *Journal of bioenergetics and biomembranes*, 33(3):259-266.
- Smeitink, J. & van den Heuvel, L. 1999. Human mitochondrial complex I in health and disease. *American journal of human genetics*, 64(6):1505.
- Smith, P.K., Krohn, R.I., Hermanson, G., Mallia, A., Gartner, F., Provenzano, M., Fujimoto, E., Goeke, N., Olson, B. & Klenk, D. 1985. Measurement of protein using bicinchoninic acid. *Analytical biochemistry*, 150(1):76-85.
- Smuts, I., Louw, R., Du Toit, H., Klopper, B., Mienie, L.J. & van der Westhuizen, F.H. 2010. An overview of a cohort of South African patients with mitochondrial disorders. *Journal of inherited metabolic disease*, 33(3):95-104.

- Smuts, I., Van der Westhuizen, F.H., Louw, R., Mienie, L.J., Engelke, U.F., Wevers, R.A., Mason, S., Koekemoer, G. & Reinecke, C.J. 2013. Disclosure of a putative biosignature for respiratory chain disorders through a metabolomics approach. *Metabolomics*, 9(2):379-391.
- Sonam, K., Bindu, P.S., Srinivas Bharath, M.M., Govindaraj, P., Gayathri, N., Arvinda, H.R., Chiplunkar, S., Nagappa, M., Sinha, S., Khan, N.A., Nunia, V., Paramasivam, A., Thangaraj, K. & Taly, A.B. 2017. Mitochondrial oxidative phosphorylation disorders in children: Phenotypic, genotypic and biochemical correlations in 85 patients from South India. *Mitochondrion*, 32:42-49.
- Soto, I.C., Fontanesi, F., Liu, J. & Barrientos, A. 2012. Biogenesis and assembly of eukaryotic cytochrome c oxidase catalytic core. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1817(6):883-897.
- Spiegel, R., Shaag, A., Mandel, H., Reich, D., Penyakov, M., Hujeirat, Y., Saada, A., Elpeleg, O. & Shalev, S.A. 2009. Mutated NDUFS6 is the cause of fatal neonatal lactic acidemia in Caucasus Jews. *European Journal of Human Genetics*, 17(9):1200.
- Stamm, D.S., Aylsworth, A.S., Stajich, J.M., Kahler, S.G., Thorne, L.B., Speer, M.C. & Powell, C.M.J.A.J.o.M.G.P.A. 2008. Native American myopathy: congenital myopathy with cleft palate, skeletal anomalies, and susceptibility to malignant hyperthermia. 146(14):1832-1841.
- Stevens, W.R. & Rago, S.A. 2013. Advanced programming in the UNIX environment: Addison-Wesley.
- Stock, D., Leslie, A.G. & Walker, J.E. 1999. Molecular architecture of the rotary motor in ATP synthase. *Science*, 286(5445):1700-1705.
- Strauss, K.A. & Morton, D.H. 2003. Type I glutaric aciduria, part 2: a model of acute striatal necrosis. (*In*. American Journal of Medical Genetics Part C: Seminars in Medical Genetics organised by: Wiley Online Library. p. 53-70).
- Strauss, K.A., Puffenberger, E.G., Robinson, D.L. & Morton, D.H. 2003. Type I glutaric aciduria, part 1: natural history of 77 patients. (*In*. American Journal of Medical Genetics Part C: Seminars in Medical Genetics organised by: Wiley Online Library. p. 38-52).
- Sugiana, C., Pagliarini, D.J., McKenzie, M., Kirby, D.M., Salemi, R., Abu-Amero, K.K., Dahl, H.-H.M., Hutchison, W.M., Vascotto, K.A. & Smith, S.M. 2008. Mutation of C20orf7 disrupts complex I assembly and causes lethal neonatal mitochondrial disease. *The American Journal of Human Genetics*, 83(4):468-478.
- Sun, F., Huo, X., Zhai, Y., Wang, A., Xu, J., Su, D., Bartlam, M. & Rao, Z. 2005. Crystal structure of mitochondrial respiratory membrane protein complex II. *Cell*, 121(7):1043-1057.
- Szklarczyk, R., Wanschers, B.F., Nijtmans, L.G., Rodenburg, R.J., Zschocke, J., Dikow, N., van den Brand, M.A., Hendriks-Franssen, M.G., Gilissen, C. & Veltman, J.A. 2012. A mutation in the FAM36A gene, the human ortholog of COX20, impairs cytochrome c oxidase assembly and is associated with ataxia and muscle hypotonia. *Human molecular genetics*, 22(4):656-667.
- Tamiya, G., Makino, S., Hayashi, M., Abe, A., Numakura, C., Ueki, M., Tanaka, A., Ito, C., Toshimori, K. & Ogawa, N. 2014. A mutation of COX6A1 causes a recessive axonal or mixed form of Charcot-Marie-Tooth disease. *The American Journal of Human Genetics*, 95(3):294-300.
- Tanigawa, J., Kaneko, K., Honda, M., Harashima, H., Murayama, K., Wada, T., Takano, K., Iai, M., Yamashita, S., Shimbo, H., Aida, N., Ohtake, A. & Osaka, H. 2012. Two Japanese patients with Leigh syndrome caused by novel SURF1 mutations. *Brain Dev*, 34(10):861-865.
- Taylor, R.W., Pyle, A., Griffin, H., Blakely, E.L., Duff, J., He, L., Smertenko, T., Alston, C.L., Neeve, V.C. & Best, A.J.J. 2014. Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies. 312(1):68-77.
- Taylor, R.W., Schaefer, A.M., Barron, M.J., McFarland, R. & Turnbull, D.M. 2004. The diagnosis of mitochondrial muscle disease. *Neuromuscular Disorders*, 14(4):237-245.

- Taylor, R.W., Singh-Kler, R., Hayes, C.M., Smith, P.E. & Turnbull, D.M. 2001. Progressive mitochondrial disease resulting from a novel missense mutation in the mitochondrial DNA ND3 gene. *Annals of neurology*, 50(1):104-107.
- Taylor, R.W. & Thorburn, D. 2018. (In P. Chinnery & Keogh, M., eds. *Clinical Mitochondrial Medicine* Cambridge: Cambridge University Press. p. pp. 39-51).
- The Ensembl genome database project - NCBI - NIH. 2018. Variant Effect Predictor - VEP script. [<https://www.ensembl.org/info/docs/tools/vep/index.html>] Date of access: 30 Oct 2018.
- Thomas, A., Dobbels, E.F., Springer, P.E., Ackermann, C., Cotton, M.F. & Laughton, B. 2018. Favourable outcome in a child with symptomatic diagnosis of Glutaric aciduria type 1 despite vertical HIV infection and minor head trauma. *Metabolic brain disease*, 33(2):537-544.
- Thompson, R., Johnston, L., Taruscio, D., Monaco, L., Bérout, C., Gut, I.G., Hansson, M.G., Peter-Bram, A., Patrinos, G.P. & Dawkins, H. 2014. RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. *Journal of general internal medicine*, 29(3):780-787.
- Thompson, W.R., DeCroes, B., McClellan, R., Rubens, J., Vaz, F.M., Kristaponis, K., Avramopoulos, D. & Vernon, H.J.J.G.i.M. 2016. New targets for monitoring and therapy in Barth syndrome. 18(10):1001.
- Thyagarajan, D., Shanske, S., Vazquez-Memije, M., Devivo, D. & Dimauro, S. 1995. A novel mitochondrial ATPase 6 point mutation in familial bilateral striatal necrosis. *Annals of neurology*, 38(3):468-472.
- Timmers, H.J., Gimenez-Roqueplo, A.-P., Mannelli, M. & Pacak, K. 2009. Clinical aspects of SDHx-related pheochromocytoma and paraganglioma. *Endocrine-related cancer*, 16(2):391-400.
- To-Figueras, J., Ducamp, S., Clayton, J., Badenas, C., Delaby, C., Ged, C., Lyoumi, S., Gouya, L., de Verneuil, H. & Beaumont, C. 2011. ALAS2 acts as a modifier gene in patients with congenital erythropoietic porphyria. *Blood*, 118(6):1443-1451.
- Truscott, K.N., Brandner, K. & Pfanner, N. 2003. Mechanisms of protein import into mitochondria. *Current Biology*, 13(8):R326-R337.
- Tsukihara, T., Aoyama, H., Yamashita, E. & Tomizaki, T. 1995. Structures of metal sites of oxidized bovine heart cytochrome c oxidase at 2.8 angstroms. *Science*, 269(5227):1069.
- Tucker, E.J., Wanschers, B.F., Szklarczyk, R., Mountford, H.S., Wijeyeratne, X.W., van den Brand, M.A., Leenders, A.M., Rodenburg, R.J., Reljić, B. & Compton, A.G. 2013. Mutations in the UQCC1-interacting protein, UQCC2, cause human complex III deficiency associated with perturbed cytochrome b protein expression. *PLoS genetics*, 9(12):e1004034.
- Tuppen, H.A., Blakely, E.L., Turnbull, D.M. & Taylor, R.W. 2010. Mitochondrial DNA mutations and human disease. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1797(2):113-128.
- Uitti, R.J., Baba, Y., Wszolek, Z.K., Putzke, D.J.J.P. & disorders, r. 2005. Defining the Parkinson's disease phenotype: initial symptoms and baseline characteristics in a clinical cohort. 11(3):139-145.
- Usami, S.i., Abe, S., Kasai, M., Shinkawa, H., Moeller, B., Kenyon, J.B. & Kimberling, W.J. 1997. Genetic and clinical features of sensorineural hearing loss associated with the 1555 mitochondrial mutation. *The Laryngoscope*, 107(4):483-490.
- Van den Bosch, B., Gerards, M., Sluiter, W., Stegmann, A., Jongen, E., Hellebrekers, D., Oegema, R., Lambrichts, E., Prokisch, H. & Danhauser, K. 2012. Defective NDUFA9 as a novel cause of neonatally fatal complex I disease. *Journal of medical genetics*, 49(1):10-15.
- Van der Walt, E.M., Smuts, I., Taylor, R.W., Elson, J.L., Turnbull, D.M., Louw, R. & van Der Westhuizen, F.H. 2012. Characterization of mtDNA variation in a cohort of South African paediatric patients with mitochondrial disease. *European Journal of Human Genetics*, 20(6):650-656.

- Van der Watt, G., Owen, E.P., Berman, P., Meldau, S., Watermeyer, N., Olpin, S.E., Manning, N.J., Baumgarten, I., Leisegang, F. & Henderson, H. 2010. Glutaric aciduria type 1 in South Africa-high incidence of glutaryl-CoA dehydrogenase deficiency in black South Africans. *Mol Genet Metab*, 101(2-3):178-182.
- Van der Westhuizen, F.H., Sinxadi, P.Z., Dandara, C., Smuts, I., Riordan, G., Meldau, S., Malik, A.N., Sweeney, M.G., Tsai, Y. & Towers, G.W. 2015. Understanding the implications of mitochondrial DNA variation in the health of black southern African populations: The 2014 Workshop. *Human mutation*, 36(5):569-571.
- Van der Westhuizen, F.H., Smuts, I., Honey, E., Louw, R., Schoonen, M., Jonck, L.-M. & Dercksen, M. 2017. A novel mutation in ETFDH manifesting as severe neonatal-onset multiple acyl-CoA dehydrogenase deficiency. *Journal of the Neurological Sciences*.
- Van Rahden, V.A., Fernandez-Vizarra, E., Alawi, M., Brand, K., Fellmann, F., Horn, D., Zeviani, M. & Kutsche, K. 2015. Mutations in NDUFB11, encoding a complex I component of the mitochondrial respiratory chain, cause microphthalmia with linear skin defects syndrome. *The American Journal of Human Genetics*, 96(4):640-650.
- Venter, L., Lindeque, Z., van Rensburg, P.J., Van der Westhuizen, F., Smuts, I. & Louw, R. 2015. Untargeted urine metabolomics reveals a biosignature for muscle respiratory chain deficiencies. *Metabolomics*, 11(1):111-121.
- Visapää, I., Fellman, V., Vesa, J., Dasvarma, A., Hutton, J.L., Kumar, V., Payne, G.S., Makarow, M., Van Coster, R. & Taylor, R.W. 2002. GRACILE syndrome, a lethal metabolic disorder with iron overload, is caused by a point mutation in BCS1L. *The American Journal of Human Genetics*, 71(4):863-876.
- Viscomi, C. & Zeviani, M. 2018. (In Chinnery, P.F. & Keogh, M., eds. *Clinical Mitochondrial Medicine* Cambridge: Cambridge University Press. p. pp. 1-9).
- Wallace, D.C., Singh, G., Lott, M.T., Hodge, J.A., Schurr, T.G., Lezza, A., Elsas, L.J. & Nikoskelainen, E.K. 1988. Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science*, 242(4884):1427-1430.
- Wang, K., Li, M. & Hakonarson, H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, 38(16):e164-e164.
- Wanschers, B.F., Szklarczyk, R., van den Brand, M.A., Jonckheere, A., Suijskens, J., Smeets, R., Rodenburg, R.J., Stephan, K., Helland, I.B. & Elkamil, A. 2014. A mutation in the human CBP4 ortholog UQCC3 impairs complex III assembly, activity and cytochrome b stability. *Human molecular genetics*, 23(23):6356-6365.
- Watanabe, H., Oritani, K.E., Fukao, T., Xiang-Qian, S., Aoyama, T., IJlst, L., Ruiten, J., Wanders, R.J. & Kondo, N. 2000. Molecular basis of very long chain acyl-CoA dehydrogenase deficiency in three Israeli patients: identification of a complex mutant allele with P65L and K247Q mutations, the former being an exonic mutation causing exon 3 skipping. *Human mutation*, 15(5):430.
- Watmough, N.J. & Frerman, F.E. 2010. The electron transfer flavoprotein: ubiquinone oxidoreductases. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1797(12):1910-1916.
- Wedatilake, Y., Brown, R.M., McFarland, R., Yapfite-Lee, J., Morris, A.A., Champion, M., Jardine, P.E., Clarke, A., Thorburn, D.R. & Taylor, R.W. 2013. SURF1 deficiency: a multi-centre natural history study. *Orphanet journal of rare diseases*, 8(1):96.
- Weissensteiner, H., Forer, L., Fuchsberger, C., Schopf, B., Kloss-Brandstatter, A., Specht, G., Kronenberg, F. & Schonherr, S. 2016. mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Res*, 44(W1):W64-69.
- Weraarpachai, W., Antonicka, H., Sasarman, F., Seeger, J., Schrank, B., Kolesar, J.E., Lochmüller, H., Chevrette, M., Kaufman, B.A. & Horvath, R. 2009. Mutation in TACO1, encoding a translational activator of COX I, results in cytochrome c oxidase deficiency and late-onset Leigh syndrome. *Nature genetics*, 41(7):833.

- Weraarpachai, W., Sasarman, F., Nishimura, T., Antonicka, H., Auré, K., Rötig, A., Lombès, A. & Shoubbridge, E.A. 2012. Mutations in C12orf62, a factor that couples COX I synthesis with cytochrome c oxidase assembly, cause fatal neonatal lactic acidosis. *The American Journal of Human Genetics*, 90(1):142-151.
- Wilmschurst, J., Lillis, S., Zhou, H., Pillay, K., Henderson, H., Kress, W., Müller, C., Ndondo, A., Cloke, V. & Cullup, T. 2010. RYR1 mutations are a common cause of congenital myopathies with central nuclei. *Annals of neurology*, 68(5):717-726.
- Wirth, C., Brandt, U., Hunte, C. & Zickermann, V. 2016. Structure and function of mitochondrial complex I. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1857(7):902-914.
- Wohlrab, H. 2009. Transport proteins (carriers) of mitochondria. *IUBMB life*, 61(1):40-46.
- Wolf, N.I. & Smeitink, J.A. 2002. Mitochondrial disorders A proposal for consensus diagnostic criteria in infants and children. *Neurology*, 59(9):1402-1405.
- Wong, L.-J.C. 2013. Mitochondrial disorders caused by nuclear genes: Springer.
- Wong, L.J.C., Naviaux, R.K., Brunetti-Pierri, N., Zhang, Q., Schmitt, E.S., Truong, C., Milone, M., Cohen, B.H., Wical, B. & Ganesh, J. 2008. Molecular and clinical genetics of mitochondrial diseases due to POLG mutations. *Human mutation*, 29(9):E150-E172.
- World Health Organization. 2018. The WHO Child Growth Standards: Head circumference-for-age. http://www.who.int/childgrowth/standards/hc_for_age/en/ Date of access: November 2018.
- Wortmann, S.B., Koolen, D.A., Smeitink, J.A., van den Heuvel, L. & Rodenburg, R.J. 2015. Whole exome sequencing of suspected mitochondrial patients in clinical practice. *J Inherit Metab Dis*, 38(3):437-443.
- Wortmann, S.B., Mayr, J.A., Nuoffer, J.M., Prokisch, H. & Sperl, W. 2017. A Guideline for the Diagnosis of Pediatric Mitochondrial Disease: The Value of Muscle and Skin Biopsies in the Genetics Era. *Neuropediatrics*, 48(4):309-314.
- Xi, J., Wen, B., Lin, J., Zhu, W., Luo, S., Zhao, C., Li, D., Lin, P., Lu, J. & Yan, C. 2014. Clinical features and ETFDH mutation spectrum in a cohort of 90 Chinese patients with late-onset multiple acyl-CoA dehydrogenase deficiency. *Journal of inherited metabolic disease*, 37(3):399-404.
- Yaffe, M.P. 1999. The machinery of mitochondrial inheritance and behavior. *Science*, 283(5407):1493-1497.
- Yamada, K., Kobayashi, H., Bo, R., Takahashi, T., Purevsuren, J., Hasegawa, Y., Taketani, T., Fukuda, S., Ohkubo, T. & Yokota, T. 2016. Clinical, biochemical and molecular investigation of adult-onset glutaric acidemia type II: Characteristics in comparison with pediatric cases. *Brain and Development*, 38(3):293-301.
- Yarham, J.W., Al-Dosary, M., Blakely, E.L., Alston, C.L., Taylor, R.W., Elson, J.L. & McFarland, R. 2011. A comparative analysis approach to determining the pathogenicity of mitochondrial tRNA mutations. *Human mutation*, 32(11):1319-1325.
- Yen, T.Y., Hwu, W.L., Chien, Y.H., Wu, M.H., Lin, M.T., Tsao, L.Y., Hsieh, W.S. & Lee, N.C. 2008. Acute metabolic decompensation and sudden death in Barth syndrome: report of a family and a literature review. *Eur J Pediatr*, 167(8):941-944.
- Yotsumoto, Y., Hasegawa, Y., Fukuda, S., Kobayashi, H., Endo, M., Fukao, T. & Yamaguchi, S. 2008. Clinical and molecular investigations of Japanese cases of glutaric acidemia type 2. *Molecular genetics and metabolism*, 94(1):61-67.
- Zhang, J., Frerman, F.E. & Kim, J.-J.P. 2006. Structure of electron transfer flavoprotein-ubiquinone oxidoreductase and electron transfer to the mitochondrial ubiquinone pool. *Proceedings of the National Academy of Sciences*, 103(44):16212-16217.

- Zhao, F., Guan, M., Zhou, X., Yuan, M., Liang, M., Liu, Q., Liu, Y., Zhang, Y., Yang, L. & Tong, Y. 2009. Leber's hereditary optic neuropathy is associated with mitochondrial ND6 T14502C mutation. *Biochemical and biophysical research communications*, 389(3):466-472.
- Zhou, H., Jungbluth, H., Sewry, C.A., Feng, L., Bertini, E., Bushby, K., Straub, V., Roper, H., Rose, M.R. & Brockington, M. 2007. Molecular mechanisms and phenotypic variation in RYR1-related congenital myopathies. *Brain*, 130(8):2024-2036.
- Zhu, Z., Yao, J., Johns, T., Fu, K., De Bie, I., Macmillan, C., Cuthbert, A.P., Newbold, R.F., Wang, J., Chevrette, M., Brown, G.K., Brown, R.M. & Shoubridge, E.A. 1998. SURF1, encoding a factor involved in the biogenesis of cytochrome c oxidase, is mutated in Leigh syndrome. *Nat Genet*, 20(4):337-343.
- Zschocke, J., Quak, E., Guldberg, P. & Hoffmann, G.F. 2000. Mutation analysis in glutaric aciduria type I. *Journal of medical genetics*, 37(3):177-181.

APPENDIX A: SCIENTIFIC OUTPUTS FROM THIS INVESTIGATION

Outputs directly emanating from this investigation:

Appendix A.1 179

Schoonen, M., Seyffert, A.S., Van der Westhuizen, F.H. & Smuts, I. (2019). A bioinformatics pipeline for rare genetic diseases in South African patients. *South African Journal of Science*, 2019;115(3/4), Art. #4876. [DOI: <https://doi.org/10.17159/sajs.2019/4876>]

Appendix A.2 192

Schoonen, M., Smuts, I., Louw, R., Elson, J.L., Van Dyk, E., Jonck, L., Rodenburg, R.J.T. & Van der Westhuizen, F.H. (2019). Panel-based nuclear and mitochondrial next-generation sequencing outcomes of an ethnically diverse paediatric patient cohort with mitochondrial disease. *Journal for Molecular Diagnostics*. [DOI: <https://doi.org/10.1016/j.jmoldx.2019.02.002>]

Appendix A.3 245

Van der Westhuizen F.H., Smuts, I., Honey, E., Louw, R., Schoonen, M., Jonck, L. & Dercksen, M. (2017). A novel mutation in ETFDH manifesting as severe neonatal-onset multiple acyl-CoA dehydrogenase deficiency. *Journal of the Neurological Sciences*, 384: 121–125. [DOI: [10.1016/j.jns.2017.11.012](https://doi.org/10.1016/j.jns.2017.11.012)].

Outputs indirectly emanating from this investigation:

Appendix A.4.....250

Louw, R., Smuts, I., Wilsenach, K., Jonck, L., Schoonen, M. & Van der Westhuizen, F.H. (2018). The dilemma of diagnosing coenzyme Q₁₀ deficiency in muscle. *Molecular Genetics and Metabolism*, 125(1–2): 38–43 [DOI: <https://doi.org/10.1016/j.ymgme.2018.02.015>].

Conference outputs emanating from this investigation:

Schoonen, M., Smuts, I., Louw, R., Elson, J.L., Van Dyk, E., Jonck, L., Rodenburg, R.J.T. & Van der Westhuizen, F.H. (2018). Mitochondrial disease in South Africa: Clinical, biochemical and genetic outcome. *RareX, Rare disease conference, Johannesburg, South Africa, September 2018. Oral presentation.*

Schoonen, M., Smuts, I., Louw, R., Elson, J.L., Van Dyk, E., Jonck, L., Rodenburg, R. & Van der Westhuizen, F.H. (2018). Mitochondrial disease in South Africa: Clinical, biochemical and genetic outcome. *Joint conference of the Federation of African and South Africa Society for Biochemistry and Molecular Biology (FASBMB-SASBMB), Potchefstroom, South Africa, 8–12 Jul 2018. Oral presentation.*

Mienie, L., Louw, R., Smuts, I., Wilsenach, K., Jonck, L., Schoonen, M. & Van der Westhuizen, F.H. (2018). The dilemma of diagnosing coenzyme Q₁₀ deficiency in muscle. *Joint conference of the Federation of African and South Africa Society for Biochemistry and Molecular Biology (FASBMB-SASBMB), Potchefstroom, South Africa, 8–12 Jul 2018. Poster presentation.*

Schoonen, M., Seyffert, A.S., Van der Westhuizen, F.H. & Smuts, I. (2018). A bioinformatics pipeline for rare genetic diseases in South African patients. *Joint conference of the Federation of African and South Africa Society for Biochemistry and Molecular Biology (FASBMB-SASBMB), Potchefstroom, South Africa, 8–12 Jul 2018. Poster presentation.*

Mereis, M., Van der Westhuizen F.H., Smuts, I., Honey, E., Louw, R., Schoonen, M., Jonck, L. & Dercksen, M. (2018). A novel mutation in ETFDH manifesting as severe neonatal-onset multiple acyl-CoA dehydrogenase deficiency. *Joint conference of the Federation of African and South Africa Society for Biochemistry and Molecular Biology (FASBMB-SASBMB), Potchefstroom, South Africa, 8–12 Jul 2018. Poster presentation.*

Van der Westhuizen, F.H., Meldau, S., Schoonen, M., Elson, J.L., Louw, R. & Smuts, I. (2017). Reviewing current outcomes of genetic testing for mitochondrial disease in South Africa. *Euromit, International meeting on mitochondrial pathology, Germany, 11–15 June 2017. Poster presentation.*

Van der Westhuizen, F.H., Smuts, I., Honey E., Lippert M.M., Engelbrecht, M., Schoonen, M., Jonck L., Louw, R. & Dercksen, M. (2017). Multiple acyl-CoA dehydrogenase deficiency due to a novel homozygous and compound heterozygous mutation in the ETFDH gene in three South African patients. *INFORM 2017 - Symposium of the International Network for Fatty Acid Oxidation Research and Management. 3–4 Sep, Rio de Janeiro, Brazil. Poster presentation.*

Van der Westhuizen, F.H., Smuts, I., Honey E., Lippert M.M., Engelbrecht, M., Schoonen, M., Jonck L., Louw, R. & Dercksen, M. (2017). Multiple acyl-CoA dehydrogenase deficiency due to a novel homozygous and compound heterozygous mutation in the ETFDH gene in three South African patients. *ICIEM 2017 – International Congress of Inborn Errors of Metabolism. 5–8 Sep, Rio de Janeiro, Brazil. Poster presentation.*

Schoonen, M., Smuts, I., Louw, R., Van Dyk, E. & Van der Westhuizen, F.H. (2017). Mitochondrial research using Ion Torrent technology. *Ion World Tour, Johannesburg, South Africa. Oral presentation.*

Schoonen, M., Smuts, I., Louw, R., Van Dyk, E. & Van der Westhuizen, F.H. (2016). Detecting and identifying mitochondrial disease causing variants using a custom Ion AmpliSeq panel. *Ion World Tour, South Africa. Oral presentation.*

Schoonen, M., Smuts, I., Rodenburg R., Van Dyk, E., Van der Sluis, R., Louw, R. & Van der Westhuizen, F.H. (2015). Nuclear gene variants associated with complex I deficiency in Southern African children diagnosed with muscle RCD. *FEBS/EMBO, Mitochondria in life, death, and disease, Crete, Greece, 12-16 October 2015. Poster presentation.*

Schoonen, M., Smuts, I., Rodenburg R., Van Dyk, E., Van der Sluis, R., Louw, R. & Van der Westhuizen, F.H. (2015). Nuclear gene variants associated with complex I deficiency in Southern African children diagnosed with muscle RCD. *7th Mitochondrial Physiology School, MiPSchool, Cape Town South Africa, 24 - 27 March 2015. Oral presentation.*

Schoonen M., Smuts I., Rodenburg R., Van Dyk E., Louw R. & Van der Westhuizen F.H. (2015) Nuclear gene variants associated with complex I of the respiratory chain in Southern African children diagnosed with RCDs. *Genomics Research Institute (GRI) Symposium, Pretoria, South Africa. Poster presentation.*

Schoonen M., Smuts I., Rodenburg R., Van Dyk E., Louw R. & Van der Westhuizen F.H. (2014) Nuclear gene variants associated with complex I deficiency in Southern African children diagnosed with muscle respiratory chain deficiency. *9th Euromit Conference, Tampere, Finland, 16-19 June 2014. Poster presentation.*

Van der Westhuizen, F.H., Schoeman, E.M., Schoonen, M., Elson, J.L., Louw, R., & Smuts, I. (2013). Mitochondrial disorders in South African patients: a next generation sequencing approach to unravel the aetiology. *15th Southern African Society of Human Genetics Congress 6-9 October 2013. Johannesburg, South Africa. Oral Presentation.*

Non-specialised community contribution emanating from this investigation:

Bioinformatics are considered a specialised field where data are being processed on unique ways to answer biological questions. It was there for important to learn bioinformatics skills on a deeper level. This knowledge obtained through various courses (online and local) was shared with other students and researchers in the form of data and software carpentry workshops. I formed part of the bioinformatics community in Potchefstroom where I became a qualified data and software carpentry instructor (<https://datacarpentry.org/> and <https://software-carpentry.org/>) Workshops where I was involved with were held at several universities including North-West University (2016, 2017 and 2018), University of Namibia (2016) and University of Cape Town (2017).

**AUTHORS:**Maryke Schoonen¹ Albertus S. Seyffert² Francois H. van der Westhuizen¹ Izelle Smuts³ **AFFILIATIONS:**¹Human Metabolomics, North-West University, Potchefstroom, South Africa²Centre for Space Research, North-West University, Potchefstroom, South Africa³Department of Paediatrics and Child Health, Steve Biko Academic Hospital, University of Pretoria, Pretoria, South Africa**CORRESPONDENCE TO:**

Maryke Schoonen

EMAIL:

mschoonen28@gmail.com

DATES:

Received: 18 Apr. 2018

Revised: 01 Nov. 2018

Accepted: 03 Dec. 2018

Published: 27 Mar. 2019

HOW TO CITE:

Schoonen M, Seyffert AS, Van der Westhuizen FH, Smuts I. A bioinformatics pipeline for rare genetic diseases in South African patients. *S Afr J Sci.* 2019;115(3/4), Art. #4876, 3 pages. <https://doi.org/10.17159/sajs.2019/4876>

ARTICLE INCLUDES:☒ Peer review☒ [Supplementary material](#)**DATA AVAILABILITY:**☐ Open data set☐ All data included☒ On request from authors☐ Not available☐ Not applicable**EDITORS:**

Pascal Bessong

Marco Weinberg

KEYWORDS:

computational tools; African cohort; next-generation sequencing; rare disease

FUNDING:

South African Medical Research Council

A bioinformatics pipeline for rare genetic diseases in South African patients

The research fields of bioinformatics and computational biology are growing rapidly in South Africa. Bioinformatics pipelines play an integral part in handling sequencing data, which are used to investigate the aetiology of common and rare diseases. Bioinformatics platforms for common disease aetiology are well supported and continuously being developed in South Africa. However, the same is not the case for rare diseases aetiology research. Investigations into the latter rely on international cloud-based tools for data analyses and ultimately confirmation of a genetic disease. However, these tools are not necessarily optimised for ethnically diverse population groups. We present an in-house developed bioinformatics pipeline to enable researchers to annotate and filter variants in either exome or amplicon next-generation sequencing data. This pipeline was developed using next-generation sequencing data of a predominantly African cohort of patients diagnosed with rare disease.

Significance:

- We demonstrate the feasibility of in-country development of ethnicity-sensitive, automated bioinformatics pipelines using free software in a South African context.
- We provide a roadmap for development of similarly ethnicity-sensitive bioinformatics pipelines.

Introduction

The research fields of bioinformatics and computational biology are both growing rapidly in South Africa, with an ever-increasing number of both small and large laboratories having access to next-generation sequencing (NGS) technologies. This increase in sequencing capability continues to stimulate the development of a variety of platforms, databases and initiatives, such as H3Africa (<https://www.h3africa.org>), the South African Human Genome Programme (<http://sahgp.sanbi.ac.za>), and the South African Bioinformatics Institute (<https://www.sanbi.ac.za>), to support sequencing data analysis. However, to date, the majority of these developments have been focused on common disease related research, because diseases like HIV, fibromyalgia, tuberculosis and malaria are major health challenges in southern Africa.¹

Southern African researchers doing rare-disease related research (diseases affecting 6–8% of the global population)² still rely on flexible, international cloud-based tools, such as Bystro³, BrowseVCF⁴, and RD-Connect⁵, as their analysis needs have not yet been fully met locally. For European populations, these tools allow researchers to leverage well-established genotype–phenotype correlations to guide investigations into rare disease aetiology. However, these correlations do not always hold for African populations, thus significantly reducing the power of and degree to which these online tools can be relied on in the South African research context. A niche therefore exists for a tool that is both sensitive to population heterogeneity and general enough in nature to enable effective domestic rare disease research.

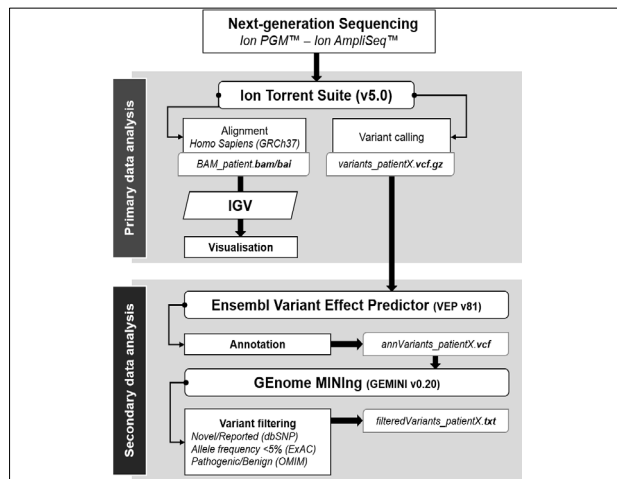
African population groups are heterogeneous with large genetic variety and limited information on genotype–phenotype correlations.⁶ Even though data are processed using the same reference genome, some aspects, such as allele frequency for disease-causing variants and genotype–phenotype correlation, could differ between ethnically diverse population groups⁷ and must be taken into account.

In this paper, we present a bioinformatics pipeline developed in-house to address these limitations. This pipeline processes NGS data of ethnically diverse population groups without any strong prior assumptions regarding genotype–phenotype correlations. This offline pipeline, suitable for the analysis of both exome and amplicon sequencing data, is written in Bash (or the Bourne-Again SHell) and uses only open-source software. To allow other researchers to benefit from this work, the pipeline has been made available on GitHub under the GNU GPLv3 software licence.⁸ The Ensembl Variant Effect Predictor (VEP)⁹ offline script is used for variant annotation, and the Genome Mining (GEMINI)¹⁰ command line database management tool for variant filtering. The pipeline is easily adjustable with regard to what annotations are made, and how they are filtered, which is especially useful when working with NGS data from ethnically diverse patients. The bioinformatics pipeline for variant annotation and filtering of amplicon and exome sequencing data presented here has been successfully used in research forming part of the project ‘Investigating the aetiology of South African paediatric patients diagnosed with mitochondrial disorders’.^{11–14}

Bioinformatics pipeline

Workflows leveraging NGS technology mainly consist of two parts: primary and secondary analysis. Figure 1 shows what this workflow looks like when our newly developed pipeline is incorporated. During primary analysis, patient samples are prepared and sequenced on a specialised platform such as Ion Torrent (used in our research case) or Illumina. These platforms further perform the requisite signalcalling, basecalling, reference sequence alignment and variant calling. The output from this primary analysis, for a single sample, is a Variant Call Format (VCF) file listing all the variants for that sample (*variants_patientX.vcf.gz* in Figure 1).¹⁵ Secondary analysis, often with the identification of disease-causing variants in mind, entails variant annotation and filtering. Variant annotation is typically done using purpose-built tools, such as ANNOVAR¹⁶ and VEP runner (used in our research case) that

annotate variants with relevant metadata such as variant type, variant allele frequency and predicted impact. Variant filtering, in which variants of interest are identified, can be done using tools like BrowseVCF, RD-Connect and GEMINI (used in our research case). These tools filter variants based on the metadata associated with each, which allows a researcher to find the variants most relevant to their investigation.



PGM, Personal Genome Machine; BAM, binary alignment/map; IGV, Integrative Genomics Viewer; VCF, Variant Call Format; dbSNP, database for single nucleotide polymorphisms; ExAC, Exome Aggregation Consortium; OMIM, Online Mendelian Inheritance in Man.

Figure 1: Next-generation sequencing bioinformatics pipeline followed for identifying disease-causing variants from Ion Torrent sequencing data. The first component of this illustration is primary data analysis, which is a semi-automated process done using the Ion Torrent Software (Torrent Suite). The second component illustrates software used for secondary data analysis.

The pipeline presented in this paper is focused exclusively on secondary analysis because the problematic potential population biases, which stem from variant annotations made based on European-population-centric research results, influence only the secondary analysis results. The pipeline utilises the offline VEP script's flexibility to annotate variants sufficiently comprehensively so that population-sensitive variant filtering can be achieved using GEMINI's high-specificity querying capabilities. Our pipeline is intended to form part of a larger NGS data analysis workflow, and is only responsible for variant annotation and filtering.

Secondary data analysis

Bash is the default command shell on most modern Unix-like systems (and Unix-like tools for Windows), and incorporates useful features from the Korn shell and the C shell.^{17,18} The ability for Bash users to write powerful scripts to automate analysis has made it a popular choice for research implementations, typically in the form of command line tools and analysis pipelines. Many of these command line tools are available as free software, giving researchers who use the Bash shell access to a plethora of bioinformatics tools with which to do their research.

The main advantage of using Bash is that it allows powerful automation of established tools in a natural way while minimising both the number of introduced software dependencies and the need for more advanced analysis infrastructure. These are important advantages considering the rare-disease focus of this pipeline. For researchers who focus exclusively on rare diseases, whose expertise and research interests often lie outside bioinformatics, a pipeline based on Bash allows flexible analyses that remain easily repeatable over long periods of time, while necessitating minimal skills development and infrastructure investment.

Here, we present the bioinformatics pipeline developed using these tools during our research. Scripts were run on Slackware v14.12 with GNU Bash, version 4.4.12.

As primary analysis, using the Ion Torrent system, delivers sequencing results in batches, it seemed prudent to write the secondary analysis scripts to operate in a batch fashion. This approach, coupled with the built-in automation that comes with scripting, ensures consistency across the samples in each batch and across different batches.

In the secondary analysis pipeline, the scripts implement two steps: variant annotation and data mining. Variant annotation is done using the VEP script, which can be downloaded from Ensembl's website.⁹ Data mining is done using the command line tool GEMINI, which uses the SQLite relational database management system to enable effective sequencing data mining.¹⁰ Each of these steps has a Bash script dedicated to it: **vep_single.sh** and **gemini_single.sh**, respectively (the '.sh' extension indicates that these text files are Bash scripts). These scripts are embedded into the **vep_batch.sh** and **gemini_batch.sh** scripts, respectively, which run them for each .vcf.gz file in a given directory. The **hetero_annotate.sh** script binds these two batch scripts into a full pipeline, calling each in sequence and managing their inputs and outputs. See Appendices 1–6 in the supplementary material for the full contents of these scripts. A brief description of what each script does is given below.

The **vep_single.sh** script (Supplementary appendix 1) takes an Ion Torrent-generated input .vcf.gz file as its first argument, and runs the VEP for it. It takes an output file as its second argument (a .vcf file), to which the annotated output of the VEP is written. Alongside these two arguments, a number of additional arguments are passed to the VEP script, with the most notable being the **--fields [list]** argument, which allows for the specification of the required annotation fields included in the annotated output .vcf file. This argument (as well as the specification of the VEP arguments) is handled in an extensible way by the **vep_single.sh** script, and the list of desired fields can be built up over multiple lines, or in multiple groups. A list of all possible fields and a list of all possible additional arguments can be found in the VEP's documentation.^{9,19} The VEP also generates a 'statistic run report' (.html file) when run, containing general statistics that give information on, among other things, the number of variants processed, number of overlapping genes, and number of novel/reported variants.

The **gemini_single.sh** (Supplementary appendix 2) script takes a VEP-annotated input .vcf file as its first argument and loads it into a SQLite database (a .db file). This file is then mined for relevant variants using user-defined SQLite database query specifications. Each line in the queries_spec.txt (Supplementary appendix 3) auxiliary text file contains one such query specification, and consists of two comma-separated fields: the query's name and the relevant SQLite query snippet. Queries can easily be added to or removed from this file, or a different such file can be specified in the **gemini_single.sh** script. An example of such a query, which filters based on variants' allele frequency in African populations, is: **rareAFR,aaf_1kg_afr <= 0.01**.

What information these queries should return for each variant is controlled by the **cols** variable, which is used to build the full SQLite query. Some examples of columns that can be included in a query are **is_coding** (which is true if the variant is in a coding region), **rs_ids** (which lists the rsIDs associated with each variant), and **aaf_1kg_afr** (which stores the allele frequency of the variant in the AFR population as reported in the 1000 Genomes Project). For a list of all possible columns, see GEMINI's documentation.²⁰ The **cols** variable is handled similarly to the **--fields [list]** argument to the VEP, and is similarly extensible. The output of each query is a list of variants, with information from the database columns specified in the **cols** variable, stored in a .txt file that has the query's name as filename. These files constitute the main output of the pipeline.

Once all the queries have been performed, a **meta_info.txt** file is generated that summarises the lengths of the lists contained in the query output files. This file, along with the generated query outputs, is saved in the folder specified as the second argument to the **gemini_single.sh** script.

The **vep_batch.sh** and **gemini_batch.sh** scripts (Supplementary appendices 4 and 5, respectively) each run their counterpart script, as discussed above, for all the .vcf.gz files in a given directory. Both

of these scripts take the directory containing the relevant input files as their first argument, with the second argument being the directory to which output should be written.

Finally, the **hetero_annotate.sh** script (Supplementary appendix 6) administers the operation of the two batch scripts described above. The first argument of this script is the directory containing the Ion Torrent output **.vcf.gz** files (the first argument for the **vep_batch.sh** and **vep_single.sh** scripts), and the second argument is the directory to which the GEMINI output should be written (the second argument to the **gemini_batch.sh** and **gemini_single.sh** scripts).

From here the user can further prioritise and filter the variants according to the criteria of their choice. For our African data set, variants were first filtered based on the novelty of these variants. Second, variants were filtered based on African allele frequencies, as reported in Exome aggregation consortium (ExAC²¹). In addition, for mostly non-African population groups, known disease-causing variants can be identified from the data set using databases such as the Online Mendelian Inheritance in Man (OMIM²²) and ClinVar²³.

Conclusion

Robust bioinformatics pipelines are key components for diagnosis and research of rare genetic diseases. Here we describe an offline, flexible and open-source bioinformatics pipeline that annotates variants using VEP and filtering of important disease-causing variants using GEMINI from NGS data. It was developed as a first prudent step towards data processing and offers unique advantages for the detection of multiple genetic alterations, including in patients of African descent for whom little information is available. The pipeline can be used on exome as well as amplicon NGS data and was designed using NGS data of a predominantly African cohort with rare disease. Identifying underlying genetic causes for rare disease is particularly challenging in the South African population, with limited bioinformatics support for researchers and non-bioinformaticians. With the increased burden to diagnose rare genetic diseases using NGS and genetic screening, and the limited support and resources in developing countries such as South Africa, it is equally important to develop and provide access to bioinformatics pipelines, such as this one. With continued development, pipelines in South Africa could be further refined and made more user-friendly, making them useful for both researchers and clinicians. These refinements could include, for instance, cloud-based interfaces like those used in developed countries. With such refinements in place, clinicians would more easily be able to investigate genotype–phenotype correlation in rare diseases for African population groups.

Acknowledgements

We acknowledge the Medical Research Council of South Africa for financial support in funding the overarching research project titled ‘Investigating the aetiology of South African paediatric patients diagnosed with mitochondrial disorders’.

Authors' contributions

M.S. and A.S.S. were responsible for: conceptualisation, methodology and writing – the initial draft. F.H.v.d.W. was responsible for: writing – revisions, student supervision and project leadership. I.S. was responsible for: writing – revisions, student supervision and funding acquisition.

References

- Mulder NJ, Christoffels A, De Oliveira T, Gamielien J, Hazelhurst S, Joubert F, et al. The development of computational biology in South Africa: Successes achieved and lessons learnt. *PLoS Comput Biol*. 2016;12(2): e1004395, 15 pages. <https://doi.org/10.1371/journal.pcbi.1004395>
- Dawkins HJ, Draghia-Akli R, Lasko P, Lau LP, Jonker AH, Cuttillo CM, et al. Progress in rare diseases research 2010–2016: An IRDiRC perspective. *Clin Transl Sci*. 2018;11(1):11–20. <https://doi.org/10.1111/cts.12501>
- Kotlar AV, Trevino CE, Zwick ME, Cutler DJ, Wingo TS. Bystro: rapid online variant annotation and natural-language filtering at whole-genome scale. *Genome Biol*. 2018;19(1), Art. #14, 11 pages. <https://doi.org/10.1186/s13059-018-1387-3>
- Salatino S, Ramraj V. BrowseVCF: A web-based application and workflow to quickly prioritize disease-causative variants in VCF files. *Brief Bioinform*. 2016;18(5):774–779. <https://doi.org/10.1093/bib/bbw054>
- Thompson R, Johnston L, Taruscio D, Monaco L, Bérout C, Gut IG, et al. RD-Connect: An integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. *J Gen Intern Med*. 2014;29(3):780–787. <https://doi.org/10.1007/s11606-014-2908-8>
- Van der Westhuizen FH, Sinxadi PZ, Dandara C, Smuts I, Riordan G, Meldau S, et al. Understanding the implications of mitochondrial DNA variation in the health of black southern African populations: The 2014 Workshop. *Hum Mutat*. 2015;36(5):569–571. <https://doi.org/10.1002/humu.22789>
- Shearer AE, Eppsteiner RW, Booth KT, Ephraim SS, Gurrola II J, Simpson A, et al. Utilizing ethnic-specific differences in minor allele frequency to recategorize reported pathogenic deafness variants. *Am J Hum Genet*. 2014;95(4):445–453. <https://doi.org/10.1016/j.ajhg.2014.09.001>
- Seyffert AS. A set of BASH scripts for annotating and filtering next-generation sequencing data using Ensembl's offline VEP runner GEMINI [software]. c2018 [cited 2018 Oct 01]. Available from: https://github.com/aseyffert/hetero_annotate
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl variant effect predictor. *Genome Biol*. 2016;17(1), Art. #122, 14 pages. <https://doi.org/10.1101/042374>
- Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: Integrative exploration of genetic variation and genome annotations. *PLoS Comput Biol*. 2013;9(7), e1003153, 8 pages. <https://doi.org/10.1371/journal.pcbi.1003153>
- Van der Westhuizen FH, Smuts I, Honey E, Louw R, Schoonen M, Jonck L-M, et al. A novel mutation in ETFDH manifesting as severe neonatal-onset multiple acyl-CoA dehydrogenase deficiency. *J Neurol Sci*. 2017;384:121–125. <https://doi.org/10.1016/j.jns.2017.11.012>
- Louw R, Smuts I, Wilsenach K-L, Jonck L-M, Schoonen M, Van der Westhuizen FH. The dilemma of diagnosing coenzyme Q10 deficiency in muscle. *Mol Genet Metab*. 2018;125(1–2):38–48. <https://doi.org/10.1016/j.ymgme.2018.02.015>
- Van der Walt EM, Smuts I, Taylor RW, Elson JL, Turnbull DM, Louw R, et al. Characterization of mtDNA variation in a cohort of South African paediatric patients with mitochondrial disease. *Eur J Hum Genet*. 2012;20(6):650–656. <https://doi.org/10.1038/ejhg.2011.262>
- Smuts I, Louw R, Du Toit H, Klopper B, Mienie LJ, Van der Westhuizen FH. An overview of a cohort of South African patients with mitochondrial disorders. *J Inher Metab Dis*. 2010;33(3):95–104. <http://dx.doi.org/10.1007/s10545-009-9031-8>
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16), e164, 7 pages. <https://doi.org/10.1093/nar/gkq603>
- Ramey C, editor. Bash, the Bourne–Again Shell. In: Proceedings of The Romanian Open Systems Conference & Exhibition (ROSE 1994), The Romanian UNIX User's Group (GURU); 1994 November 3–5; Bucharest, Romania.
- Stevens WR, Rago SA. Advanced programming in the UNIX environment. Ann Arbor, MI: Addison-Wesley; 2013.
- The Ensembl Genome Database Project, NCBI, NIH. Variant Effect Predictor [software]. c2018 [cited 2018 Oct 01]. Available from: <https://www.ensembl.org/info/docs/tools/vep/index.html>
- Genome Mining. GEMINI: A flexible framework for exploring genome variation [software]. c2017 [cited 2018 Oct 01]. Available from: <https://gemini.readthedocs.io/en/latest/>
- 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*. 2016;45(D1):D840–D845. <https://doi.org/10.1093/nar/gkw971>
- Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res*. 2014;43(D1):D789–D798. <https://doi.org/10.1093/nar/gku1205>
- Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2013;42(D1):D980–D985. <https://doi.org/10.1093/nar/gkt1113>

Appendix A.1

Supplementary material to: Schoonen et al. S Afr J Sci. 2019;115(3/4), Art. #4876, 3 pages

How to cite:

Schoonen M, Seyffert AS, Van der Westhuizen FH, Smuts I. A bioinformatics pipeline for rare genetic diseases in South African patients [supplementary material]. S Afr J Sci. 2019;115(3/4), Art. #4876, 10 pages. <https://doi.org/10.17159/sajs.2019/4876/suppl>

Appendix 1: The **vep_single.sh** script

Script that takes an Ion Torrent Variant Caller Format file and annotates it using Ensembl's offline VEP runner

```
#!/usr/bin/env bash

# This script takes an Ion Torrent VCF file and annotates it using Ensembl VEP
# Example: bash vep_single.sh /path/to/vcf/input.vcf.gz /path/to/vcf/input.vcf

# NOTE: Filenames starting with "tmp_" indicate temporary files that are deleted

# ==== Preparation ====

num_forks=4

# Directory where VEP is installed. [This may differ for your installation]
vep_dir="${HOME}/scripts/ensembl-vep"

# ---- VCF annotation fields ----

# Specify the .vcf annotation fields. They can be added in groups as below.
vep_fields="Consequence,Codons,Amino_acids,Gene,SYMBOL,Feature,EXON"
vep_fields="${vep_fields},PolyPhen,SIFT,check_alleles"

# ---- VEP arguments ----

# Specify the arguments to pass to VEP. They can be added in groups as below.
# NOTE: There is a space after "${vep_args}" in each additional group.
vep_args="--cache --offline --format vcf --vcf --buffer_size 25000"
```

Appendix A.1

```
vep_args="${vep_args} --sift b --polyphen b --symbol --numbers"
vep_args="${vep_args} --af_1kg --pubmed"
vep_args="${vep_args} --fields ${vep_fields}"

# ---- Handle arguments ----

in_file=$1                                # Path to input .vcf.gz file
out_file=$2                               # Path to output .vcf file

# ===== Run VEP =====

cp ${in_file} ${vep_dir}/tmp_input.vcf.gz
cd ${vep_dir}

echo "... running Ensembl VEP..."

./vep -i tmp_input.vcf.gz -o tmp_output.vcf -fork ${num_forks} ${vep_args}

# ---- Cleanup ----

rm tmp_input.vcf.gz
cd -

mv ${vep_dir}/tmp_output.vcf ${out_file}
mv ${vep_dir}/tmp_output.vcf_summary.html ${out_file}_summary.html
```

Appendix A.1

Appendix 2: The **gemini_single.sh** script

The annotated VEP file generated from Ensembl VEP runner is loaded into a SQLite database (GEMINI) for querying.

```
#!/usr/bin/env bash

# This script takes a VEP-annotated VCF file and loads it into a SQLite
# database for querying. It produces a set of variant lists, as well as a
# meta_info.txt file in the output directory.
# Example: bash gemini_single.sh /path/to/vcf/input.vcf /path/to/output/dir/

# ==== Preparation ====
# Specify the number of CPU cores GEMINI can use, and the queries_spec file
num_cores=4
qfile="../../queries_spec.txt"

# ---- GEMINI query columns ----
# Define columns of interest. They can be added in groups as below.
cols="chrom, start, end, gene, exon, ref, alt, qual, type, cyto_band"
cols="${cols}, is_coding, codon_change, aa_change"
cols="${cols}, num_hom_ref, num_hom_alt, num_het"
cols="${cols}, impact, impact_severity, is_lof, is_conserved"
cols="${cols}, depth, qual_depth, rs_ids, in_omim, pfam_domain"
cols="${cols}, clinvar_sig, clinvar_origin, clinvar_disease_name"
cols="${cols}, polyphen_pred, polyphen_score, sift_pred, sift_score"
cols="${cols}, in_hm3, in_esp, in_1kg, aaf_1kg_all, aaf_1kg_eur, aaf_1kg_afr"

# ---- GEMINI load arguments ----
gemini_args="--skip-gerp-bp"

# ---- Handle arguments ----
in_file=$1          # Annotated .vcf input file
```

Appendix A.1

```
out_dir=$2          # Output directory for query results files

# ==== Run GEMINI ====

# ---- Load into SQLite database ----

cp ${in_file} tmp_vcf.vcf
mkdir tmp_out_dir

gemini load -v tmp_vcf.vcf -t VEP --cores ${num_cores} ${gemini_args} tmp_gS.db

# ---- Query database ----

echo "Querying..."
while read line
do
    qname=$(echo ${line} | cut -d "," -f 1)
    qsnip=$(echo ${line} | cut -d "," -f 2)
    query="select ${cols} from variants where ${qsnip}"

    printf "... ${qname}...\n"
    echo gemini query -q "${query}" tmp_gS.db > tmp_out_dir/${qname}.txt
    printf "done\n"
done < ${qfile}

# ---- Generate meta_info.txt ----

for in_file in tmp_out_dir/*.txt; do
    wc -l ${in_file} >> tmp_out_dir/meta_info.txt
done

# ---- Move files to out_dir ----

if [ -d ${out_dir}/queries ]; then
    echo "Clobbering contents of queries folder..."
    rm -r ${out_dir}/queries/*
else
```

Appendix A.1

```
echo "Creating queries folder"
mkdir ${out_dir}/queries
fi

# NOTE: This works because meta_info.txt gets moved _before_ the rest...
mv ./tmp_out_dir/meta_info.txt ${output_dir}
mv ./tmp_out_dir/* ${out_dir}/queries/

# ---- Cleanup ----
rm ./tmp_gS.db ./tmp_vcf.vcf
rmdir tmp_out_dir
```

Appendix A.1

Appendix 3: The **queries_spec.txt** text file

Text file listing the queries GEMINI should make; called in the **gemini_single.sh** script. Queries can easily be removed or added.

```
LoF,is_lof = 1
RareNovelLoF,is_lof = 1 and (in_dbsnp = 0 or aaf <= 0.01)
omimClinivar,in_omim = 1 or clinvar_disease_name is not NULL
siftPolyphen,sift_pred = 'deleterious' or polyphen_pred = 'probably_damaging'
allVariants,
coding,is_coding = 1
nonCoding,is_coding = 0
codingSynonymous,is_coding = 1 and impact LIKE 'synonymous_coding'
codingNonSynonymous,is_coding = 1 and impact LIKE 'non_syn_coding'
codingStopgain,is_coding = 1 and impact LIKE 'stop_gain'
codingStoploss,is_coding = 1 and impact LIKE 'stop_loss'
codingFrameshift,is_coding = 1 and impact LIKE 'frame_shift'
reported,in_dbsnp = 1
reportedLoF,is_lof = 1 and in_dbsnp = 1
reportedNonLoF,is_lof = 0 and in_dbsnp = 1
reportedCoding,in_dbsnp = 1 and is_coding = 1
reportedCodingLoF,is_lof = 1 and in_dbsnp = 1 and is_coding = 1
reportedCodingNonLoF,is_lof = 0 and in_dbsnp = 1 and is_coding = 1
novel,in_dbsnp = 0
novelLoF,is_lof = 1 and in_dbsnp = 0
novelNonLoF,is_lof = 0 and in_dbsnp = 0
novelCoding,in_dbsnp = 0 and is_coding = 1
novelCodingLoF,is_lof = 1 and in_dbsnp = 0 and is_coding = 1
novelCodingNonLoF,is_lof = 0 and in_dbsnp = 0 and is_coding = 1
```

Appendix A.1

Appendix 4: The **vep_batch.sh** script

Files are annotated in a batch matter, where each file (.vcf.gz) in a folder is annotated.

```
#!/usr/bin/env bash

# This script is a wrapper for vep_single.sh which calls it for each .vcf.gz
# file in the input directory. It outputs a VEP-annotated VCF file for each
# input file in the output directory.

# Example: bash vep_batch.sh /path/to/input/dir/ /path/to/output/dir/
# NOTE: Absolute paths are safest since they're unambiguous

# ---- Handle arguments ----

in_dir=$1          # Path to input directory
out_dir=$2         # Path to output directory

# ---- Run VEP for batch ----

for in_file in ${in_dir}/*.vcf.gz
do
    echo "Processing ${in_file} [VEP]..."
    out_file_name="$(basename ${in_file} .gz)"
    bash vep_single.sh ${in_file} ${out_dir}/${out_file_name}
done
```

Appendix A.1

Appendix 5: The **gemini_batch.sh** script

Files are queried in a batch matter, where each file (.vcf) in a folder is queried.

```
#!/usr/bin/env bash

# This script is a wrapper for gemini_single.sh which calls it for each .vcf
# file in the input directory. See that script for output details.
# Example: bash gemini_batch.sh /path/to/input/dir/ path/to/output/dir/
# NOTE: Absolute paths are safest since they're unambiguous

in_dir=$1          # Path to input directory
out_dir=$2         # Path to output directory

# ---- Run GEMINI for batch ----
for in_file in ${in_dir}/*.vcf
do
    echo "Processing ${in_file} [GEMINI]..."

    out_subdir="$(basename ${in_file} .vcf)"

    # Create output subdirectory (each file gets its own)
    out_dir_full=${out_dir}/${out_subdir}
    mkdir ${out_dir_full}

    bash gemini_single.sh ${in_file} ${out_dir_full}
done
```

Appendix A.1

Appendix 6: The **hetero_annotate.sh** script

This script uses the `vep_single.sh` and `gemini_single.sh` scripts (via their batch wrappers) to build a batch-capable variant annotation and filtering pipeline.

```
#!/usr/bin/env bash

# This script uses the vep_single.sh and gemini_single.sh scripts (via their
# batch wrappers) to build a batch-capable variant annotation and filtering
# pipeline. The underlying tools are the offline Ensembl VEP runner and GEMINI
# This script should ideally be used on patient folders in a batch fashion.
# Example: bash hetero_annotate.sh /path/to/input/dir/ /path/to/output/dir/
# NOTE: Absolute paths are safest since they're unambiguous

# ---- Handle arguments ----
in_dir=$1          # Path to input directory
out_dir=$2         # Path to output directory

# ---- Prepare out_dir ----
# Check whether out_dir exists (1) and is empty (2)
if ! [ -d ${out_dir} ]; then
    echo "Creating main output directory: ${out_dir}..."
    mkdir ${out_dir}
else
    if ! [ -z "$(ls -A ${out_dir})" ]; then
        echo "Empty main output directory first..."
        exit 1
    fi
fi

mkdir ${out_dir}/vep_out

# ---- Run VEP ----
cd src/vep_scripts
```

Appendix A.1

```
bash vep_batch.sh ${in_dir} ${out_dir}/vep_out
```

```
cd -
```

```
# ---- Run GEMINI ----
```

```
cd src/gemini_scripts
```

```
bash gemini_batch.sh ${out_dir}/vep_out ${out_dir}
```

```
cd -
```

Accepted Manuscript

Panel-Based Nuclear and Mitochondrial Next-Generation Sequencing Outcomes of an Ethnically Diverse Pediatric Patient Cohort with Mitochondrial Disease

Maryke Schoonen, Izelle Smuts, Roan Louw, Joanna L. Elson, Etresia van Dyk, Lindi-Maryn Jonck, Richard J.T. Rodenburg, Francois H. van der Westhuizen



PII: S1525-1578(17)30603-7

DOI: <https://doi.org/10.1016/j.jmoldx.2019.02.002>

Reference: JMDI 778

To appear in: *The Journal of Molecular Diagnostics*

Accepted Date: 6 February 2019

Please cite this article as: Schoonen M, Smuts I, Louw R, Elson JL, van Dyk E, Jonck L-M, Rodenburg RJT, van der Westhuizen FH, Panel-Based Nuclear and Mitochondrial Next-Generation Sequencing Outcomes of an Ethnically Diverse Pediatric Patient Cohort with Mitochondrial Disease, *The Journal of Molecular Diagnostics* (2019), doi: <https://doi.org/10.1016/j.jmoldx.2019.02.002>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Panel-based nuclear and mitochondrial next-generation sequencing outcomes of an ethnically diverse paediatric patient cohort with mitochondrial disease

Maryke Schoonen,* Izelle Smuts,† Roan Louw,* Joanna L Elson,*‡ Etresia van Dyk,* Lindi-Maryn Jonck,* Richard JT Rodenburg,§ and Francois H van der Westhuizen*

From Human Metabolomics,* North-West University, Potchefstroom, South Africa; the Department of Paediatrics and Child Health,† Steve Biko Academic Hospital, University of Pretoria, Pretoria, South Africa; the Institute of Genetic Medicine,‡ Newcastle University, Newcastle upon Tyne, United Kingdom; and the Department of Paediatrics,§ Radboud Centre for Mitochondrial Medicine, Radboud University Medical Centre, Nijmegen, The Netherlands.

Running head: NGS of paediatric patients with MD.

Funding: Supported in part by the Medical Research Council of South Africa under the project title: “Investigating the aetiology of South African paediatric patients diagnosed with mitochondrial disorders”.

Corresponding Author:

Francois H van der Westhuizen

Human Metabolomics

North-West University

Potchefstroom, South Africa

E-mail: francois.vanderwesthuizen@nwu.ac.za

Disclosures: None declared.

Footnote: Current address of E.v.D.: ThermoFisher.

Abstract

Mitochondrial disease (MD) is a group of rare inherited disorders with clinical heterogeneous phenotypes. Recent advances in next-generation sequencing (NGS) allows for rapid genetic diagnostics in patients who suffer from MD, resulting in significant strides into determining the aetiology of the disease. This, however, has not been the case in many patient populations. We report on a molecular diagnostic study using mitochondrial DNA (mtDNA) and targeted nuclear DNA (nDNA) NGS of an extensive cohort of predominantly sub-Saharan African paediatric patients with clinical and biochemically defined MD. Patients in this novel cohort presented mostly with muscle involvement (73%). Of the original 212 patients in this cohort, a muscle respiratory chain deficiency was identified in 127 cases. Genetic analyses were conducted for these 127 cases based on biochemical deficiencies, for both mitochondrial (n=123), and nDNA using panel-based NGS (n=86). As a pilot investigation, whole exome sequencing was performed in a subset of African patients (n=8). These analyses resulted in the identification of a previously reported pathogenic mtDNA variant, and seven pathogenic or likely pathogenic nDNA variants (*ETFDH*, *SURF1*, *COQ6*, *RYR1*, *STAC3*, *ALAS2*, *TRIOBP*); most of which identified via whole exome sequencing. This study contributes to the lacking knowledge on MD aetiology in an understudied, ethnically diverse population, highlights inconsistencies in genotype-phenotype correlations, and proposes future directions of diagnostic approaches in such patient populations.

Introduction

Mitochondria are ubiquitous in the human body and serve mainly as the energy producing organelle via oxidative phosphorylation (OXPHOS). This metabolic pathway comprises five protein complexes (CI-V), consisting of a total of 92 structural subunits encoded by both mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) genes.^{1, 2} Underlying genetic mutations create disruptions within this system which manifest clinically, often affecting multiple highly energy demanding organs simultaneously. The heterogeneous class of clinical phenotypes associated with such mutations is referred to as mitochondrial disease (MD).³ A large number of genes (at least 289), extending beyond the structural OXPHOS genes, have been identified as being involved in MD.⁴

Traditionally, MDs are diagnosed through extensive clinical evaluation including biochemical tissue analysis, followed by genetic screening for selected mutations. The current suite of next-generation sequencing (NGS) options available for MD diagnosis, and for heterogeneous disease diagnosis in general, includes targeted panel sequencing, unbiased whole-exome sequencing (WES), and whole genome sequencing (WGS); each with particular advantages, disadvantages and considerations. Recent publications advocate for a “genetics first” diagnostic approach, with the promise of eliminating the need for functional and biochemical analyses in the majority of diagnoses.^{5, 6} This raises some concerns for understudied ethnically diverse populations in developing countries where relatively little progress has been made towards understanding the genetic aetiology of MD, and where the genotype-phenotype correlations are poorly understood and are inconsistent with those for, e.g., non-African populations. To address these limitations, we and others have undertaken various clinical, biochemical, and genetic studies on MDs in the South African population — one of the few developing countries to do so (Supplemental Figure S1).⁷⁻¹¹

To date, a traditional diagnostic trajectory of extensive clinical evaluation and functional biochemical diagnosis of referred patients, followed by screening for known, common mutations, has been followed.¹² Currently, published patient data, and public genetic and disease databases from

predominantly non-African populations, are used as reference due to the limited (specific) information available on African MD aetiology.

To address these diagnostic challenges for MD in an understudied ethnically diverse population, we report on the outcome of an NGS approach when targeting reported nuclear and mtDNA-encoded genes involved in MD. This approach was investigated in a predominantly African cohort of 212 South African paediatric patients selected based on clinical and muscle respiratory chain (RC) enzymology data; an approach which would be considered as prudent in a diagnostic setting. We highlight the contrasting outcome of a WES approach in a small subset of this patient cohort and, with specific consideration of the genotype-phenotype correlation suggested by selected cases, propose future diagnostic directions which should be considered for similar understudied population groups.

Materials and methods

Patient cohort

Since 2006, more than 6000 patients with neurological symptoms have been referred to the Steve Biko Academic Hospital (a state funded institution in Pretoria, South Africa), and clinically evaluated according to a MD criteria (MDC) scoring system first set forth by Wolf and Smeitink (2002)¹³ and subsequently refined by Smuts et al.⁷ Currently this cohort consists of 212 paediatric patients who manifested clinically with MD signs or symptoms from as early as the neonatal period. This cohort originated from the northern provinces of South Africa and is predominantly African (64%), with an equal number of males and females. Urine and muscle (*vastus lateralis*) samples were collected from the entire cohort for subsequent biochemical and molecular genetic investigations. Limited availability of samples from parents and patients prevented segregation analysis. This study was approved by the Ethics Committees of the University of Pretoria (No. 91/98 and amendments) and the North-West University (NWU-00170-13-A1).

Biochemical analyses

Muscle RC enzyme analyses for complexes I-IV (CI-IV; EC 1.6.5.3, EC 1.3.5.1, EC 1.10.2.2, and EC 1.9.3.1, respectively), and CII+CIII, were performed and normalized against citrate synthase (CS, EC 2.3.3.1) activity for all 212 patients. Enzymes were analyzed in 600 x g homogenates prepared from frozen muscle samples, as previously described.^{7, 14} Other biochemical analyses performed included urine metabolic analysis^{7, 9, 15} and muscle CoQ₁₀ analysis¹⁶ in muscle samples. In total, 127 patients were identified to have an RC deficiency.

Genetic analyses***mtDNA sequencing***

Genomic DNA (gDNA) preparation from muscle homogenate was performed using standard protocol, as previously described.⁸ The Qubit 2.0 Fluorometer (ThermoFisher Scientific, MA) was used for quantification of gDNA. The complete mitochondrial genome was sequenced in 123 patients, all of whom had a known combined clinical and biochemical MD profile. Two chemistries were used: 71 patient samples were sequenced using the 454 GS-FLX platform (Roche, CA), and 52 patient samples were sequenced using the Ion PGM platform (ThermoFisher Scientific, MA). The 454 GS-FLX sequencing, including, library preparation and enrichment, was done according to procedures described by van der Walt et al.⁸ The Ion PGM sequencing, including library preparation and enrichment, were done according to the manufacturer's protocol for Ion Torrent platform. Four samples could not be sequenced due to insufficient sample quantity and quality.

Nuclear panel sequencing

For nuclear gene investigations, Ion Torrent™ amplicon panel sequencing was performed on 86 patients, all of whom had a known combined clinical and biochemical MD profile. Patients were included for sequencing based on their MDC scores for clinical presentations, and the severity of their biochemical deficiencies, ie, patients who had enzyme activity lower than the reference range

when expressed against at least two of three enzyme markers (CS, CII, or CIV), were considered.⁷

Genes to be included in sequencing panels were selected based on their mitochondrial RC involvement, either direct or indirect (Supplemental Table S1, S2, and S3). Three custom gene panels consisting of 136 genes in total were designed and are briefly described. Panel 1 consisted of 78 targeted genes associated with CI deficiency (HaloPlex Target Enrichment System, Agilent Technologies, CA, Supplemental Table S1). This customized panel had a target region size of 360 091 kbp and 99.6% targeted coverage. In total, 30 patients were sequenced using Panel 1.

Panel 2, from Ion AmpliSeq Custom DNA Panels (Thermo Fisher Scientific), had 78 targeted genes associated with CI–IV deficiency (CI=38, CII=6, CIII=10, CIV=24), with a total target region size of 157 834 kbp and 98.2% targeted coverage (Supplemental Table S2). In total, 48 patients were sequenced using Panel 2, of which five patients overlapped with Panel 1 and 10 patients overlapped with Panel 3.

Panel 3, from Ion AmpliSeq, targeted 18 genes known to be involved with primary and secondary CoQ₁₀ deficiency (Supplemental Table S3). The design size was 61 kbp with a targeted coverage of 98%. In total, 26 patients were sequenced in Panel 3 (six patients overlapping with Panel 1). The entire coding region of each gene, including flanking regions of intron-exons, were sequenced using the Ion PGM platform as per manufacturers' protocol (HaloPlex ref G9912C and Ampliseq ref 4480441). The selected genes and panels were not African-population-specific, as the underlying genetic cause for MD is mostly unclear in African populations.

Whole exome sequencing

As an initial comparison on the outcomes of a panel vs WES approach, WES was performed on a subset of eight randomly selected African (Haplogroup L) cases where no strong candidate disease-causing variants had been identified by initial mtDNA and/or nuclear panel sequencing. WES was performed at the Central Analytical Facilities, Stellenbosch University, South Africa, using the Ion Proton sequencer (ThermoFisher Scientific, MA) according to the manufacturer's protocol for the Ion

Torrent platform™. An approximate 95% on-target coverage was achieved, with an average depth coverage of approximately 140x.

Bioinformatics for mtDNA variants

Mitochondrial DNA sequences were aligned against the human mitochondrial revised Cambridge Reference Sequence (rCRS [NC_012920 gi:251831106]). Haplogroup assignment, and variant identification and annotation, was performed using mtDNA-Server (v1.20.0),¹⁷ MitoMap, and MitoMaster.¹⁸ Homoplasmic and heteroplasmic (levels above 30%) non-synonymous variants were further evaluated based on their allele frequency reported in Genbank and appearance on Phylotree^{19, 20}, and those with a population allele frequency below 0.1% were considered significant. Variant pathogenicity was evaluated according to the criteria proposed for mtDNA variants.²¹⁻²³ For example, the MutPred scoring system was used to classify non-synonymous variants in structural subunits of OXPHOS (<http://mutpred.mutdb.org/help.html>, last accessed April 2018).²⁴ A MutPred score above 0.5 suggests a probable damaging impact on protein function, with scores between 0.75 and 1.0 indicating such functional damage on a protein/amino acid with high confidence. Mitochondrial-tRNA variants were individually evaluated using MitoTIP²⁵ and classified according to a scoring system first set forth by McFarland et al.²² and subsequently refined by Yarham et al.²⁶ A low Yarham score (below 10) weighs more towards benign or neutral classification, whereas a score above 10, with substantial evidence from functional tests, weighs more toward a pathogenic classification. Variants were also evaluated using the guidelines put forth by The American College of Medical Genetics and Genomics (ACMG), where possible (notably for mtDNA variants).²⁷

Bioinformatics for nDNA variants

Raw sequencing files, obtained from the Ion PGM, were analyzed with the *Torrent Suite* (v5.0.2). The sequence files were aligned against Genome Reference Consortium Human Build 37 (GRCh37, hg19) followed by coverage analysis and variant calling using the *coverageAnalysis* and *variantCaller*

plugins (v.5.0) from the Torrent Suite, respectively. The Variant Calling Format (VCF) files were further annotated using the offline *Variant Effect Predictor* from Ensembl (<https://uswest.ensembl.org/info/docs/tools/vep/index.html>, v81; last accessed July 2018),²⁸ followed by variant mining using *GEMINI* (v.20).²⁹ The output text files generated using GEMINI contained information on both novel and reported variants. Most notably, they detailed whether or not a variant had previously been reported as pathogenic. Further variant filtering was done using population databases such as Exome Aggregation Consortium (ExAC) and *gnomAD*³⁰ (specifically for African allele frequencies), disease-specific databases such as ClinVar and Online Mendelian Inheritance in Man (OMIM), and sequence databases such as NCBI Genome and RefSeqGene. As supporting evidence, the missense variants of interest were cautiously evaluated using various *in silico* predictive algorithms (*SIFT*, *Polyphen-2*, and *CADD*).³¹⁻³³ These algorithms, however, have been shown to have low sensitivity, specificity, and accuracy.³⁴ Candidate variants of interest were evaluated using ACMG guidelines, and are classified as pathogenic, likely pathogenic, variants of uncertain significance, likely benign, or benign.

Results

Clinical profiles

From the more than 6,000 neurological patients referred for clinical assessment, 212 patients (113 males and 99 females) presented with mitochondrial phenotypes, forming the cohort that is described here. These clinically defined MD patients were comprehensively evaluated and were predominantly of African ancestry (n=130, Figure 1A). Age of onset was as early as the first year of life, including the neonatal period (n=139, Figure 1B). The most common clinical finding was muscle involvement, which manifested in 73% (n=155) of the cohort. Cardiac involvement and deafness was the least observed at 5% (n=10) and 8% (n=17), respectively (Figure 1C).

Biochemical profiles

An RC enzyme complex deficiency was identified in 127 cases (64 males and 63 females); 64 cases had isolated deficiency and the remaining 63 had a combined deficiency. Complex I deficiency, either isolated (n=43) or combined (n=32), was most prevalent, as is frequently reported,^{35, 36} followed by CIII deficiency (n=54; isolated=15 and combined=39; Figure 2B).

Genetic analyses

The findings for whole mtDNA, panel nuclear DNA, and whole exome NGS are summarized in Table 1.^{16,37-44} Whole mtDNA sequencing was successfully applied in 123 cases (79 African and 44 non-African), and a variant of interest was identified in each of 12 African cases, none of which could be classified as disease-causing.⁸ An example of such a variant of interest is the well-known Leber's Hereditary Optic Neuropathy (LHON; OMIM #535000)³⁷ mutation, m.14484T>C, identified in an African male (S014) who did not present with typical LHON symptoms, and was later clinically diagnosed with fucosidosis. An additional 11 mtDNA variants of interest identified in 10 cases may, however, still prove to have a functional effect on disease initiation or progression (Supplemental Table S4^{8,37,45-50}). Notably, this may still be the case for variants with a comparatively high population allele frequency (above 0.1%) since further investigation may indeed reveal that these variants appear at a higher frequency in these understudied population lineages.

Panel nuclear DNA sequencing of 86 cases (61 African and 25 non-African) revealed four variants of interest in two cases. Two pathogenic compound heterozygous *ETFDH* variants were identified in one case (S057), and two possible pathogenic compound heterozygous *SURF1* variants were identified in a second case (S085).

WES sequencing of eight African cases revealed nine variants of interest (in six cases), eight of which (variants in the genes *COQ6*, *RYR1*, *STAC3*, and *ALAS2*) are considered possibly pathogenic and one (in the gene *TRIOBP*) as a variant of uncertain significance under the ACMG classification.

Pathogenic variants

The electron-transfer flavoprotein dehydrogenase (*ETFDH*) compound heterozygous variants (c.1448C>T and c.1067G>A; p.Pro483Leu and p.Gly356Glu; ENST00000307738) were identified in a non-African female (S057) who presented with features of multiple acyl-dehydrogenase deficiency (MADD; OMIM #231680). She had severe muscle weakness, exercise intolerance, chronic fatigue, and hepatomegaly. Metabolic markers of MADD such as dicarboxylic acids, ethylmalonic acid, glutaric acid, as well as acylcarnitines (butyryl-, isovaleryl- and glutarylcarnitine), acylglycine (hexanoyl- isobutyryl-, isovaleryl and suberylglycine), and conjugates were observed in the patient's urine. Furthermore, she had CI, CIII, and CII+CIII RC deficiency, with severely reduced CoQ₁₀ levels in muscle.¹⁶ These mutations cause structural instability in *ETFDH* and were confirmed with functional analysis.³⁸

Likely pathogenic variants

An African female (S085), with clinically confirmed Leigh Disease (LD; OMIM #256000) and confirmed mitochondrial CIV deficiency, harbored compound heterozygosity for a missense and frameshift variant (c.575G>A and c.754_755delAG; p.Arg192Gln and p.Ser252HisfsTer39; ENST00000371974,) in exons 6 and 8 in the gene *SURF1* (Surfeit 1). The patient clinically presented early in life with several CNS involvements (nystagmus; extrapyramidal and pyramidal symptoms) and muscle involvements (myopathy; hypotonia and weakness). Furthermore, changes in the basal ganglia/thalami were observed. Metabolic investigation revealed elevated lactic acid.

An African female (S002) with confirmed primary CoQ₁₀ deficiency¹⁶ was found to be carrying compound heterozygous variants (c.41G>A and c.859G>T; p.Trp14Ter and p.Ala49Ser; ENST00000394026) in the gene *COQ6* (coenzyme Q₁₀ monooxygenase 6). One of the variants identified in a highly conserved region, c.41G>A, results in a premature truncation of the protein with high loss-of-function (LoF) probability. Clinically she presented at four years and four months of age with a history of severe weakness since birth. Other features were noted upon clinical examination and included macrocephaly, severe hypotonia and head lag, pseudo-hypertrophy of

calves and triceps, limitation of extension at knees and elbows, and reduced reflexes. No other candidate *COQ6* gene variants were identified in this case.

Two sets of compound heterozygous variants, associated with centronuclear myopathy (CNM) and minicore myopathy and external ophthalmoplegia (MMEO; OMIM #255320), were identified in the gene *RYR1* (ryanodine 1) in two African females (S032 and S033) using WES. For S032, a previously reported pathogenic variant (c.8342_8343delTA; p.Ile2781ArgfsX49; ENST00000359596)⁴⁰ with high LoF confidence and a missense variant (c.11926C>T; p.His3976Tyr; ENST00000359596), both in highly conserved regions, were identified. She presented with severe hypotonia, myopathy with myopathic facial features, and initially did not show signs of external ophthalmoplegia. She also had an affected sibling with myopathic facial features and external ophthalmoplegia. A muscle biopsy was collected from the sibling who revealed thickened connective tissue between muscle fibers and evidence of fat infiltration. The female reported here had CIV RC deficiency and her brother had CI+CIII deficiency. Both siblings presented with an additional CII+CIII deficiency.

For S033, a previously reported pathogenic missense variant (c.14524G>A; p.Val482Met; ENST00000359596)⁴⁰ and a frameshift variant (c.11193+1G>A; ENST00000359596) were identified. Both variants are in highly conserved regions. She presented with severe hypotonia, mild myopathic facial features, and dense external ophthalmoplegia.

A homozygous mutation (c.851G>C; p.Trp284Ser; ENST00000332782), was detected by WES in an African female (S011) in the gene *STAC3* (SH3 and cysteine-rich domains 3), and was first described by Horstick et al⁴² in five families with Native American Myopathy (NAM; OMIM #255995). The phenotype observed in the female described here is similar to that for a case described elsewhere,⁴² and includes severe myopathy, failure to thrive, developmental delay, relative macrocephaly, and ptosis with no external ophthalmoplegia. She had minor dysmorphic features, including a low nasal bridge. She was born prematurely and had intra uterine growth restriction. Biochemically she presented with isolated CIII deficiency in muscle.

An African male (S117), with skin and muscle involvement, carried a homozygous gain-of-function variant (c.1757A>T; p.Tyr586Phe; ENST00000330807) in exon 11 of the gene *ALAS2* (5'-aminolevulinate synthase 2). This variant was first described in a Spanish patient with erythropoietic porphyria by To-Figueras et al,⁴³ who states that *ALAS2* acts as a modifier gene in patients with erythropoietic porphyria (X-linked protoporphyria; OMIM #300752). Another manifestation is sideroblastic anaemia type 1 (OMIM #300751), which is also as a result of *ALAS2* mutations. However, our patient did not present with symptoms for the latter. He instead presented with swelling of his face, hands, and feet, and experienced non-specific body pain, symptoms which are more similar with X-linked protoporphyria. He had depigmented skin lesions in his face and on the extensor areas of the fore and upper arms, and over the knees and lateral right thigh. Furthermore, he suffered from severe muscle weakness with decreased muscle bulk in all four limbs, with muscle histology revealing atypical dermatomyositis. Biochemically he had confirmed CI, CIII, and CIV deficiency in muscle.

Variants of uncertain significance

A trio- and filamentous-actin-binding protein (*TRIOBP*) homozygous missense variant (c.3232C>T; p.Arg1078Cys; ENST00000406386) was identified using WES in an African male (S059) who presented with developmental delay, visual impairment, muscle weakness and hypotonia, and clinodactyly. Furthermore, he presented with mild facial dysmorphisms, which included an epicanthic fold and low set ears. Most notably the patient had hearing impairment, which was confirmed by abnormal auditory brainstem response. Mutations in this gene are known to cause autosomal recessive deafness (OMIM #609823), a feature found in the case presented here. Metabolic profiles revealed significant ketosis associated with dicarboxylic aciduria involving C6–C10 acids (ie, adipic-, suberic-, and sebacic acid), with a normal amino acid profile.

A number of cases (n=13) presented with variants of uncertain significance, and are listed in Supplemental Table S5.⁵¹⁻⁵⁷ These nuclear variants are classified as likely-pathogenic or pathogenic

according to ClinVar but did not adhere to the ACMG standards and guidelines. For example, previously reported pathogenic variants identified in the genes *NDUFA9*, *SDHA*, *SDHB*, and *POLG* were all detected as heterozygous, whereas homozygous variants in the genes *TRMU*, *ACADVL*, and *GLUD2* had high African allele frequencies. These variants, however, remain of interest for South African population groups.

Discussion

The genetic diagnosis of MD, and identification of the number of genes involved, has rapidly improved since the first mutations were reported in the late 1980s.^{58, 59} Although the use of clinical scoring systems such as MDC, in addition to biochemical evaluation of RC/OXPHOS function in tissue, is still the hallmark of MD diagnostics, it has been increasingly complimented by NGS in recent years, notably exome sequencing.⁶⁰ The main advantage of using NGS in MD diagnoses is that it allows for the discrimination between primary (having a direct genetic aetiology) and secondary MD (caused by non-genetic factors such as environmental toxins).⁶¹ A retrospective investigation into genetic causes of MD was conducted for 127 patients with clinically suspected and biochemically confirmed RC deficiency in an understudied population (predominantly African). Here, we report on two high-throughput NGS techniques (whole mtDNA sequencing and panel sequencing) that were used to find common, previously reported pathogenic or likely pathogenic variants in reported genes involved in MD in this understudied ethnically diverse cohort with mostly unknown MD genotype-phenotype correlations.

The mtDNA sequencing data for a sub-section of this cohort have been extensively investigated elsewhere.⁸ The majority of the variants found could not be classified as pathogenic as they fail to meet a number of mtDNA criteria. Most notably among the observed variants was a well-known LHON-associated mutation (m.14484T>C at 53% heteroplasmy in muscle), identified in an African male (S014). Although this patient had clinical features of eye involvement, his clinical phenotype did not match that expected for LHON (visual failure and optic atrophy). This inconsistency could be

ascribed to varied penetrance of the disease, highlighting the importance of investigating the penetrance of diseases associated with this and other pathogenic variants when detected in understudied populations.

Panel sequencing revealed pathogenic and likely pathogenic variants in two cases: S057 and S085. For S057, the compound heterozygous variants that were identified in *ETFDH* (c.1448T>C and c.1067G>A) have previously been investigated extensively, leading to the classification of these variants as pathogenic with substantial clinical, biochemical, and *in vivo* supporting experimental evidence (from structural and functional analyses).³⁸ For S085, compound heterozygous variants were identified in *SURF1* (c.754_755delAG and c.575G>A). *SURF1* is directly involved with cytochrome c oxidase (COX) maintenance and assembly. LoF mutations in *SURF1* cause major structural instability in COX, and are responsible for the phenotype of LD as clinically diagnosed in the case reported here. The frameshift variant has previously been reported as pathogenic for a different (Japanese) patient, with substantial clinical and experimental supporting evidence and a confirmed LoF mechanism.³⁹ Consequently, this variant, which has a low African allele frequency, was classified as likely pathogenic in this case. The missense variant has not previously been associated with a clinical phenotype, and the extremely low allele frequency suggests moderate likelihood to be pathogenic.

Since the panel sequencing of genes known to be involved in MD revealed a pathogenic or likely pathogenic variant in only two cases (S057 and S85) of the 86 patients investigated, it was evident that this is not an effective approach to follow in this patient population. Although targeted/gene panel NGS is considered a prudent NGS approach in many diagnostic settings, whilst being aware that this is a heterogeneous, understudied patient population and that an expanded gene panel may not necessarily increase diagnostic yield,⁶² genetic investigations were expanded to include WES to probe its outcome on a small subset of eight African patients where no initial NGS results were

obtained. A variant of interest in the genes *COQ6*, *RYR1*, *STAC3*, *ALAS2*, and *TRIOBP* was identified in across six of the eight cases.

The compound heterozygous *COQ6* variants (c.41G>A and c.859G>T) identified in S002 have been extensively investigated elsewhere,¹⁶ where functional and structural analyses showed significantly decreased levels of COQ6. This protein is directly involved with CoQ₁₀ biosynthesis and mutations in this gene result in primary CoQ₁₀ deficiency.^{63, 64} This correlates well with the clinical and biochemical profiles observed in the case reported here. Based on these experimental findings, the correlation between the observed and reported phenotypes, alongside their allele frequencies, these variants are classified as likely pathogenic.

Two previously reported pathogenic and two reported *RYR1* variants were identified in two compound heterozygotes: S032 (c.8342_8343delTA and c.11926C>T) and S033 (c.14524G>A and c.11193+1G>A). The c.14524G>A and c.8342_8343delTA variants are classified as founder mutations for South African patient populations with centronuclear myopathy (CNM) and MMEO.⁴³ *RYR1* encodes a homotetrameric calcium channel in skeletal muscle and regulates cytosolic Ca²⁺ levels.⁶⁵ Dysfunctional *RYR1* disrupts the Ca²⁺ balance, directly affecting different mitochondrial functions such as ATP synthesis regulation and reactive oxygen species generation,^{40, 66, 67} consequently contributing to myopathy, external ophthalmoplegia, and ptosis. The two cases reported here had clinical phenotypes consistent with reported cases. For S032 and her brother, the specific type of congenital myopathy is still unclear. Both had neurogenic features which are absent in multi-minicore myopathy. The siblings, whose features correlated more strongly with CNM, were similar in their manifestation (albeit with slight, but significant, differences). For example, the female initially did not have external ophthalmoplegia, whereas the brother did. Importantly, no homozygous variants have been reported in patients with CNM.⁶⁸ S033 had correlating multi-minicore myopathic features as seen in previously reported cases,^{40, 41} which included mild myopathic facial features with dense external ophthalmoplegia. Both sets of compound heterozygous variants are classified as

likely pathogenic in these cases due to substantial supporting clinical and experimental evidence confirming the LoF mechanism,⁴⁰ the confirmed founder effect for South African populations with a low or absent allele frequency in several population databases, and multiple *in silico* algorithms supporting a deleterious effect on the gene product for the two frameshift variants.

The previously reported pathogenic *STAC3* variant (c.851G>C)⁴² was identified as homozygous in S011. *STAC3* encodes a putative muscle-specific adaptor protein, which takes part in excitation-contraction in muscle. Disruptions within this protein are thought to cause reduction in mitochondrial Ca^{2+} , either intra- or extra mitochondrial, which have a direct effect on OXPHOS.⁶⁵ Clinical features, as a result, include congenital myopathy with facial dysmorphic features including ptosis, which is consistent with the case reported here. This variant was previously sequenced as part of the 1000 Genomes Project, and was undetected in 113 Caucasian controls. Consequently, we classify this variant as likely pathogenic in this case due to its previously confirmed pathogenicity (with substantial supporting evidence) and low African allele frequency.

The homozygous variants identified in *ALAS2* (c.1757A>T) and *TRIOBP* (c.3232C>T) in S117 and S059, respectively, are classified as variants of uncertain significance, and benign or likely benign, respectively, according to several disease databases. The two cases reported here, S117 and S059, had overlapping, but also slightly inconstant phenotypical manifestations compared to reported cases.

Specifically, S117 did not present with typical X-linked protoporphyria,^{43, 44} the primary phenotype associated with *ALAS2* mutations. His primary cause of disease is interplay between two deficiencies: dermatomyositis and X-linked porphyria. The muscle histology from this patient was suggestive of an inflammatory myopathy, but the inflammatory cell infiltrate was perivascular, and the typical pattern of peripheral muscle cell atrophy in dermatomyositis was not demonstrated. As the patient had definitive skin involvement, which included swelling and redness, it is suggested that the patient has porphyria rather than inflamed myopathy. His mitochondrial dysfunction could arise from

impaired electron transport as a direct result of bone marrow haem synthetic dysfunction. The clinical significance is still unclear and therefore this variant is classified as a variant of uncertain significance. A case can be made, however, that this variant should be classified as likely pathogenic since it has a low African allele frequency, and there is substantial supporting evidence for its pathogenicity.

S059 had a severe mitochondrial phenotype, where deafness was most notable. The latter is consistent with reported cases harboring *TRIOBP* mutations,^{69, 70} however the former has not yet been associated with mutations in this gene. The underlying genetic cause for this patient's mitochondrial phenotype is still unclear. This variant, c.3232C>T, has not yet been associated with a clinical phenotype and therefore lacks substantial supporting pathogenicity evidence. Consequently, it was classified as a variant of uncertain significance.

To conclude, in this unique and predominantly African cohort, looking first at nuclear and mitochondrial genes known to be involved in MD (which is in line with current diagnostic practices) resulted in a poor diagnostic outcome, with only a relatively small number of pathogenic or likely pathogenic variants being successfully identified. Initial indications from limited WES data are much more promising as nine likely pathogenic variants were identified in six cases using WES compared to five variants identified in three cases using panel NGS and mtDNA sequencing.

Considering all the cases in this study, in only five cases (S057, S085, S032, S033 and S011) a strong genotype-phenotype correlation could be established and a moderately strong correlation in two cases (S002 and S117). For the remaining two cases S014 and S059, a non-specific correlation was observed. Observations like these serve as a strong motivation that a "genetics first/only" approach, without supporting clinical and biochemical investigation, is not suitable in such understudied, ethnically diverse populations. Furthermore, when following a genetic approach, we conclude that panel sequencing could be an efficient approach in populations where the genotype-phenotype correlations are well-established for specific monogenic diseases. For heterogeneous diseases such

as MD, however, even in such populations, WES/WGS performs significantly better compared to a targeted gene-panel approach.^{4, 5, 71-73} Our results are thus in line with proposals that WES should be considered as the primary option for genetic investigations in heterogeneous inherited diseases such as MD and, in fact, may be particularly ideal in understudied, ethnically diverse populations where there is evidence of inconsistencies with documented MD phenotypes. However, in such populations the value of extensive clinical and biochemical (structural and functional) investigations to support molecular genetic data outcomes, should not be neglected, and at this time are, in fact, more crucial than in well studied populations.

References

1. Chinnery PF, Hudson G: Mitochondrial genetics. *British medical bulletin* 2013, 106:135-159.
2. Anderson S, Bankier AT, Barrell BG, De Bruijn M, Coulson AR, Drouin J, Eperon I, Nierlich D, Roe BA, Sanger F: Sequence and organization of the human mitochondrial genome. *Nature* 1981, 290:457-465.
3. Gorman GS, Chinnery PF, DiMauro S, Hirano M, Koga Y, McFarland R, Suomalainen A, Thorburn DR, Zeviani M, Turnbull DM: Mitochondrial diseases. *Nat Rev Dis Primers* 2016, 2:16080.
4. Wortmann SB, Mayr JA, Nuoffer JM, Prokisch H, Sperl W: A Guideline for the Diagnosis of Pediatric Mitochondrial Disease: The Value of Muscle and Skin Biopsies in the Genetics Era. *Neuropediatrics* 2017, 48:309-314.
5. Wortmann SB, Koolen DA, Smeitink JA, van den Heuvel L, Rodenburg RJ: Whole exome sequencing of suspected mitochondrial patients in clinical practice. *J Inherit Metab Dis* 2015, 38:437-443.
6. Craven L, Alston CL, Taylor RW, Turnbull DM: Recent Advances in Mitochondrial Disease. *Annu Rev Genomics Hum Genet* 2017, 18:257-275.
7. Smuts I, Louw R, Du Toit H, Kloppe B, Mienie LJ, van der Westhuizen FH: An overview of a cohort of South African patients with mitochondrial disorders. *Journal of inherited metabolic disease* 2010, 33:95-104.
8. Van der Walt EM, Smuts I, Taylor RW, Elson JL, Turnbull DM, Louw R, van Der Westhuizen FH: Characterization of mtDNA variation in a cohort of South African paediatric patients with mitochondrial disease. *European Journal of Human Genetics* 2012, 20:650-656.
9. Reinecke CJ, Koekemoer G, Van der Westhuizen FH, Louw R, Lindeque JZ, Mienie LJ, Smuts I: Metabolomics of urinary organic acids in respiratory chain deficiencies in children. *Metabolomics* 2012, 8:264-283.

10. Meldau S, De Lacy R, Riordan G, Goddard E, Pillay K, Fieggen K, Marais D, Van der Watt G: Identification of a single MPV17 nonsense-associated altered splice variant in 24 South African infants with mitochondrial neurohepatopathy. *Clinical genetics* 2018.
11. Van der Watt G, Owen EP, Berman P, Meldau S, Watermeyer N, Olpin SE, Manning NJ, Baumgarten I, Leisegang F, Henderson H: Glutaric aciduria type 1 in South Africa-high incidence of glutaryl-CoA dehydrogenase deficiency in black South Africans. *Mol Genet Metab* 2010, 101:178-182.
12. Van der Westhuizen FH, Sinxadi PZ, Dandara C, Smuts I, Riordan G, Meldau S, Malik AN, Sweeney MG, Tsai Y, Towers GW: Understanding the implications of mitochondrial DNA variation in the health of black southern African populations: The 2014 Workshop. *Human mutation* 2015, 36:569-571.
13. Wolf NI, Smeitink JA: Mitochondrial disorders A proposal for consensus diagnostic criteria in infants and children. *Neurology* 2002, 59:1402-1405.
14. Smith Pe, Krohn RI, Hermanson G, Mallia A, Gartner F, Provenzano M, Fujimoto E, Goeke N, Olson B, Klenk D: Measurement of protein using bicinchoninic acid. *Analytical biochemistry* 1985, 150:76-85.
15. Venter L, Lindeque Z, van Rensburg PJ, Van der Westhuizen F, Smuts I, Louw R: Untargeted urine metabolomics reveals a biosignature for muscle respiratory chain deficiencies. *Metabolomics* 2015, 11:111-121.
16. Louw R, Smuts I, Wilsenach K-L, Jonck L-M, Schoonen M, Van der Westhuizen FH: The dilemma of diagnosing coenzyme Q10 deficiency in muscle. *Molecular genetics and metabolism* 2018.
17. Weissensteiner H, Forer L, Fuchsberger C, Schopf B, Kloss-Brandstatter A, Specht G, Kronenberg F, Schonherr S: mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Res* 2016, 44:W64-69.

18. Lott MT, Leipzig JN, Derbeneva O, Xie HM, Chalkia D, Sarmady M, Procaccio V, Wallace DC: mtDNA variation and analysis using MITOMAP and MITOMASTER. *Current protocols in bioinformatics* 2013;1.23. 21-21.23. 26.
19. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW: GenBank. *Nucleic acids research* 2012, 41:D36-D42.
20. Van Oven M: PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Science International: Genetics Supplement Series* 2015, 5:e392-e394.
21. DiMauro S, Schon EAJ: Mitochondrial DNA mutations in human disease. 2001, 106:18-26.
22. McFarland R, Elson JL, Taylor RW, Howell N, Turnbull DM: Assigning pathogenicity to mitochondrial tRNA mutations: when 'definitely maybe' is not good enough. *TRENDS in Genetics* 2004, 20:591-596.
23. Mitchell AL, Elson JL, Howell N, Taylor RW, Turnbull DM: Sequence variation in mitochondrial complex I genes: mutation or polymorphism? *J Med Genet* 2006, 43:175-179.
24. Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P: Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* 2009, 25:2744-2750.
25. Sonney S, Leipzig J, Lott MT, Zhang S, Procaccio V, Wallace DC, Sondheimer N: Predicting the pathogenicity of novel variants in mitochondrial tRNA with MitoTIP. *PLoS computational biology* 2017, 13:e1005867.
26. Yarham JW, Al-Dosary M, Blakely EL, Alston CL, Taylor RW, Elson JL, McFarland R: A comparative analysis approach to determining the pathogenicity of mitochondrial tRNA mutations. *Human mutation* 2011, 32:1319-1325.
27. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E: Standards and guidelines for the interpretation of sequence variants: a joint

- consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine* 2015, 17:405.
28. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F: The ensembl variant effect predictor. *Genome biology* 2016, 17:122.
 29. Paila U, Chapman BA, Kirchner R, Quinlan AR: GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS computational biology* 2013, 9:e1003153.
 30. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, Hamamsy T, Lek M, Samocha KE, Cummings BB: The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic acids research* 2016, 45:D840-D845.
 31. Ng PC, Henikoff SJ: SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research* 2003, 31:3812-3814.
 32. Adzhubei I, Jordan DM, Sunyaev SR: PolyPhen-2: Predicting functional effect of human missense mutations using PolyPhen-2. *bioinformatics* 2013, 29:21-27.
 33. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J: A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics* 2014, 46:310.
 34. Ernst C, Hahnen E, Engel C, Nothnagel M, Weber J, Schmutzler RK, Hauke J: Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC medical genomics* 2018, 11:35.
 35. McFarland R, Kirby DM, Fowler KJ, Ohtake A, Ryan MT, Amor DJ, Fletcher JM, Dixon JW, Collins FA, Turnbull DM: De novo mutations in the mitochondrial ND3 gene as a cause of infantile mitochondrial encephalopathy and complex I deficiency. *Annals of neurology* 2004, 55:58-64.
 36. Kirby DM, Salemi R, Sugiana C, Ohtake A, Parry L, Bell KM, Kirk EP, Boneh A, Taylor RW, Dahl H-HM: NDUFS6 mutations are a novel cause of lethal neonatal mitochondrial complex I deficiency. *The Journal of clinical investigation* 2004, 114:837-845.

37. Brown M, Voljavec A, Lott M, MacDonald I, Wallace D: Leber's hereditary optic neuropathy: a model for mitochondrial neurodegenerative diseases. *The FASEB Journal* 1992, 6:2791-2799.
38. Van der Westhuizen FH, Smuts I, Honey E, Louw R, Schoonen M, Jonck L-M, Dercksen M: A novel mutation in ETFDH manifesting as severe neonatal-onset multiple acyl-CoA dehydrogenase deficiency. *Journal of the Neurological Sciences* 2017.
39. Tanigawa J, Kaneko K, Honda M, Harashima H, Murayama K, Wada T, Takano K, Iai M, Yamashita S, Shimbo H, Aida N, Ohtake A, Osaka H: Two Japanese patients with Leigh syndrome caused by novel SURF1 mutations. *Brain Dev* 2012, 34:861-865.
40. Wilmshurst J, Lillis S, Zhou H, Pillay K, Henderson H, Kress W, Müller C, Ndondo A, Cloke V, Cullup T: RYR1 mutations are a common cause of congenital myopathies with central nuclei. *Annals of neurology* 2010, 68:717-726.
41. Punetha J, Kesari A, Uapinyoying P, Giri M, Clarke NF, Waddell LB, North KN, Ghaoui R, O'Grady GL, Oates EC: Targeted re-sequencing emulsion PCR panel for myopathies: results in 94 cases. *Journal of neuromuscular diseases* 2016, 3:209-225.
42. Horstick EJ, Linsley JW, Dowling JJ, Hauser MA, McDonald KK, Ashley-Koch A, Saint-Amant L, Satish A, Cui WW, Zhou W: Stac3 is a component of the excitation–contraction coupling machinery and mutated in Native American myopathy. *Nature communications* 2013, 4:1952.
43. To-Figueras J, Ducamp S, Clayton J, Badenas C, Delaby C, Ged C, Lyoumi S, Gouya L, de Verneuil H, Beaumont C: ALAS2 acts as a modifier gene in patients with congenital erythropoietic porphyria. *Blood* 2011, 118:1443-1451.
44. Balwani M, Bloomer J, Desnick R, of the NIH-Sponsored PC, Network RDCR: X-linked protoporphyria. 2013.

45. Gamba J, Kiyomoto BH, de Oliveira ASB, Gabbai AA, Schmidt B, Tengan CH: The mutations m. 5628T> C and m. 8348A> G in single muscle fibers of a patient with chronic progressive external ophthalmoplegia. *Journal of the neurological sciences* 2012, 320:131-135.
46. Santorelli FM, Siciliano G, Casali C, Basirico MG, Carrozzo R, Calvosa F, Sartucci F, Bonfiglio L, Murri L, DiMauro S: Mitochondrial tRNACys gene mutation (A5814G): a second family with mitochondrial encephalopathy. *Neuromuscular Disorders* 1997, 7:156-159.
47. Gaspar R, Santana I, Mendes C, Fernandes AS, Duro D, Simões M, Luís D, Santos MJ, Grazina M: Genetic Variation of MT-ND Genes in Frontotemporal Lobar Degeneration: Biochemical Phenotype-Genotype Correlation. *Neurodegenerative Diseases* 2015, 15:70-80.
48. Bayona-Bafaluy MP, López-Gallardo E, Montoya J, Ruiz-Pesini E: Maternally inherited susceptibility to cancer. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 2011, 1807:643-649.
49. Zhao F, Guan M, Zhou X, Yuan M, Liang M, Liu Q, Liu Y, Zhang Y, Yang L, Tong Y: Leber's hereditary optic neuropathy is associated with mitochondrial ND6 T14502C mutation. *Biochemical and biophysical research communications* 2009, 389:466-472.
50. Cavadas B, Soares P, Camacho R, Brandão A, Costa MD, Fernandes V, Pereira JB, Rito T, Samuels DC, Pereira L: Fine time scaling of purifying selection on human nonsynonymous mtDNA mutations based on the worldwide population tree and mother-child pairs. *Human mutation* 2015, 36:1100-1111.
51. Watanabe H, Orii KE, Fukao T, Xiang-Qian S, Aoyama T, IJlst L, Ruiters J, Wanders RJ, Kondo N: Molecular basis of very long chain acyl-CoA dehydrogenase deficiency in three Israeli patients: identification of a complex mutant allele with P65L and K247Q mutations, the former being an exonic mutation causing exon 3 skipping. *Human mutation* 2000, 15:430.
52. Shteyer E, Saada A, Shaag A, Kidess R, Revel-Vilk S, Elpeleg O: Exocrine pancreatic insufficiency, dyserythropoietic anemia, and calvarial hyperostosis are caused by a mutation in the COX4I2 gene. *The American Journal of Human Genetics* 2009, 84:412-417.

53. Plaitakis A, Latsoudis H, Kanavouras K, Ritz B, Bronstein JM, Skoula I, Mastorodemos V, Papapetropoulos S, Borompokas N, Zaganas I: Gain-of-function variant in GLUD2 glutamate dehydrogenase modifies Parkinson's disease onset. *European Journal of Human Genetics* 2010, 18:336.
54. Alston CL, Davison JE, Meloni F, van der Westhuizen FH, He L, Hornig-Do H-T, Peet AC, Gissen P, Goffrini P, Ferrero I: Recessive germline SDHA and SDHB mutations causing leukodystrophy and isolated mitochondrial complex II deficiency. *Journal of medical genetics* 2012, 49:569-577.
55. Martins RG, Nunes JB, Máximo V, Soares P, Peixoto J, Catarino T, Rito T, Soares P, Pereira L, Sobrinho-Simões M: A founder SDHB mutation in Portuguese paraganglioma patients. *Endocrine-related cancer* 2013, 20:L23-L26.
56. Ni Y, Zbuk KM, Sadler T, Patocs A, Lobo G, Edelman E, Platzer P, Orloff MS, Waite KA, Eng C: Germline mutations and variants in the succinate dehydrogenase genes in Cowden and Cowden-like syndromes. *The American Journal of Human Genetics* 2008, 83:261-268.
57. Meng F, Cang X, Peng Y, Li R, Zhang Z, Li F, Fan Q, Guan AS, Fischel-Ghosian N, Zhao X: Biochemical evidence for a nuclear modifier allele (A10S) in TRMU (methylaminomethyl-2-thiouridylate-methyltransferase) related to mitochondrial tRNA modification in the phenotypic manifestation of deafness-associated 12S rRNA mutation. *Journal of Biological Chemistry* 2017:jbc. M116. 749374.
58. Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza A, Elsas LJ, Nikoskelainen EK: Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science* 1988, 242:1427-1430.
59. Holt IJ, Harding AE, Morgan-Hughes JAJN: Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. 1988, 331:717.

60. Witters P, Saada A, Honzik T, Tesarova M, Kleinle S, Horvath R, Goldstein A, Morava E: Revisiting mitochondrial diagnostic criteria in the new era of genomics. *Genetics in Medicine* 2018, 20:444.
61. Niyazov DM, Kahler SG, Frye RE: Primary mitochondrial disease and secondary mitochondrial dysfunction: importance of distinction for diagnosis and treatment. *Molecular syndromology* 2016, 7:122-137.
62. Plutino M, Chaussonot A, Rouzier C, Ait-El-Mkadem S, Fragaki K, Paquis-Flucklinger V, Bannwarth S: Targeted next generation sequencing with an extended gene panel does not impact variant detection in mitochondrial diseases. *BMC medical genetics* 2018, 19:57.
63. Alcázar-Fabra M, Trevisson E, Brea-Calvo G: Clinical syndromes associated with Coenzyme Q10 deficiency. *Essays in biochemistry* 2018, 62:377-398.
64. Heeringa SF, Chernin G, Chaki M, Zhou W, Sloan AJ, Ji Z, Xie LX, Salviati L, Hurd TW, Vega-Warner V: COQ6 mutations in human patients produce nephrotic syndrome with sensorineural deafness. *The Journal of clinical investigation* 2011, 121:2013-2024.
65. Blackburn PR, Selcen D, Gass JM, Jackson JL, Macklin S, Cousin MA, Boczek NJ, Klee EW, Dimberg EL, Kennelly KD: Whole exome sequencing of a patient with suspected mitochondrial myopathy reveals novel compound heterozygous variants in RYR1. *Molecular genetics & genomic medicine* 2017, 5:295-302.
66. Görlach A, Bertram K, Hudecova S, Krizanova O: Calcium and ROS: a mutual interplay. *Redox biology* 2015, 6:260-271.
67. Zhou H, Jungbluth H, Sewry CA, Feng L, Bertini E, Bushby K, Straub V, Roper H, Rose MR, Brockington M: Molecular mechanisms and phenotypic variation in RYR1-related congenital myopathies. *Brain* 2007, 130:2024-2036.
68. Abath Neto O, Martins CdA, Carvalho M, Chadi G, Seitz KW, Oliveira ASB, Reed UC, Laporte J, Zanuteli E: DNM2 mutations in a cohort of sporadic patients with centronuclear myopathy. *Genetics and molecular biology* 2015, 38:147-151.

69. Shahin H, Walsh T, Sobe T, Rayan AA, Lynch ED, Lee MK, Avraham KB, King M-C, Kanaan M: Mutations in a novel isoform of TRIOBP that encodes a filamentous-actin binding protein are responsible for DFNB28 recessive nonsyndromic hearing loss. *The American Journal of Human Genetics* 2006, 78:144-152.
70. Riazuddin S, Khan SN, Ahmed ZM, Ghosh M, Cauton K, Nazli S, Kabra M, Zafar AU, Chen K, Naz S: Mutations in TRIOBP, which encodes a putative cytoskeletal-organizing protein, are associated with nonsyndromic recessive deafness. *The American Journal of Human Genetics* 2006, 78:137-143.
71. Neveling K, Feenstra I, Gilissen C, Hoefsloot LH, Kamsteeg EJ, Mensenkamp AR, Rodenburg RJ, Yntema HG, Spruijt L, Vermeer S, Rinne T, van Gassen KL, Bodmer D, Lugtenberg D, de Reuver R, Buijsman W, Derks RC, Wieskamp N, van den Heuvel B, Ligtenberg MJ, Kremer H, Koolen DA, van de Warrenburg BP, Cremers FP, Marcelis CL, Smeitink JA, Wortmann SB, van Zelst-Stams WA, Veltman JA, Brunner HG, Scheffer H, Nelen MR: A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum Mutat* 2013, 34:1721-1726.
72. Haack TB, Haberberger B, Frisch E-M, Wieland T, Iuso A, Gorza M, Strecker V, Graf E, Mayr JA, Herberg U: Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. *Journal of medical genetics* 2012, 49:277-283.
73. Lieber DS, Calvo SE, Shanahan K, Slate NG, Liu S, Hershman SG, Gold NB, Chapman BA, Thorburn DR, Berry GT: Targeted exome sequencing of suspected mitochondrial disorders. *Neurology* 2013, 80:1762-1770.

Figure legends

Figure 1: Demographic information for 212 paediatric patients. **A:** Summary of the ethnicity distribution. Caucasian, Asian, and colored patients are collectively referred to as non-African patients in text. **B:** Summary of the age categories representing the age at symptom onset for each of the 212 patients. **C:** Summary of the clinical features and findings in patient cohort.

Figure 2: Summary for the biochemical RC enzyme deficiency distribution. **A:** Combined versus isolated enzyme deficiency for 127 patients. **B:** Summary for CI-CIV enzyme deficiency in 64 patients with isolated RC deficiency.

Appendix A.2

Table 1: Summary of mtDNA and nDNA variants of interest identified in South African patients with clinically confirmed mitochondrial disease.

Mitochondrial variants									
Gene	Variant position	Amino acid change	Mutpred score	ACMG classification*	GenBank frequency, variant allele frequency	Patient, ethnicity, sex	Clinical profile	Biochemical profile	Reference
<i>MT-ND6</i>	m.14484T>C [‡]	p.Met64Val	0.627	Likely benign	0.19, 53%	S014, A, M	DR, CNS, Eye, M, L	CI	Brown et al (1992) ³⁷
Nuclear variants									
Gene	Variant position (allele call)	Amino acid change	RefSNP ID	ACMG classification*	ExAC frequency	Patient, ethnicity, sex	Clinical profile	Biochemical profile	Reference
<i>ETFDH</i> [†]	c.1448C>T (+/-)	p.Pro483Leu	rs377656387 [‡]	Pathogenic	None	S057, NA, F	DD, CNS, M, L, S	CI, CIII, CIV, CII+CIII	Van der Westhuizen et al (2017) ³⁸
	c.1067G>A (+/-)	p.Gly356Glu	Novel	Pathogenic	None				
<i>SURF1</i> [†]	c.575G>A (+/-)	p.Arg192Gln	rs782021521	Likely pathogenic	0.00001	S085, A, F	FTT, DD, DR, CNS, M, E	CIV	Tanigawa et al (2012) ³⁹
	c.754_755delAG (+/-)	p.Ser252HisfsTer39	rs782007828 [‡]	Likely pathogenic	0.00002				
<i>COQ6</i> §	c.41G>A (+/-)	p.Trp14Ter	rs17094161	Likely pathogenic / risk factor	0.07047	S002, A, F	Mac, DD, Eye, M	CII, CII+CIII	Louw et al (2018) ¹⁶
	c.859G>T (+/-)	p.Ala49Ser	rs61743884		0.03162				

Appendix A.2

<i>RYR1</i> §	c.8342_8343delTA (+/-)	p.Ile2781ArgfsX49	rs75858007 5 [‡]	Likely pathogenic	None	S032, A, F	FTT, DD, Eye, M	CIV, CII+CIII	Wilmshurst et al (2010) ⁴⁰
	c.11926C>T (+/-)	p.His3976Tyr	rs14877285 4	Likely pathogenic	0.01287				
	c.14524G>A (+/-)	p.Val4842Met	rs19392287 9 [‡]	Likely pathogenic	0.001935	S033, A, F	FTT, DD, Eye, M, R	CI	Wilmshurst et al (2010) ⁴⁰
	c.11193+1G>A (+/-)	-	rs11198631 6	Likely pathogenic	None				
<i>STAC3</i> §	c.851G>C (+/+)	p.Trp284Ser	rs14029109 4 [‡]	Likely pathogenic	0.0011	S011, A, F	Mac, FTT, DD, Dys, Eye, M, R	CIII	Horstick et al (2013) ⁴²
<i>ALAS2</i> §	c.1757A>T (+/+)	p.Tyr586Phe	rs13959686 0	Likely pathogenic / VUS	0.009364	S117, A, M	M, Skin	CI, CIII, CIV	To-Figueras et al (2011) ⁴³ Balwani et al (2013) ⁴⁴
<i>TRIOBP</i> §	c.3232C>T (+/+)	p.Arg1078Cys	rs20035970 8	VUS	0.00030	S059, A, M	DD, CNS, Eye, D, M, GIT	CI, CIII, CIV, CII+CIII	None to date

*Variant classification using ACMG criteria, including phenotypic evaluation.

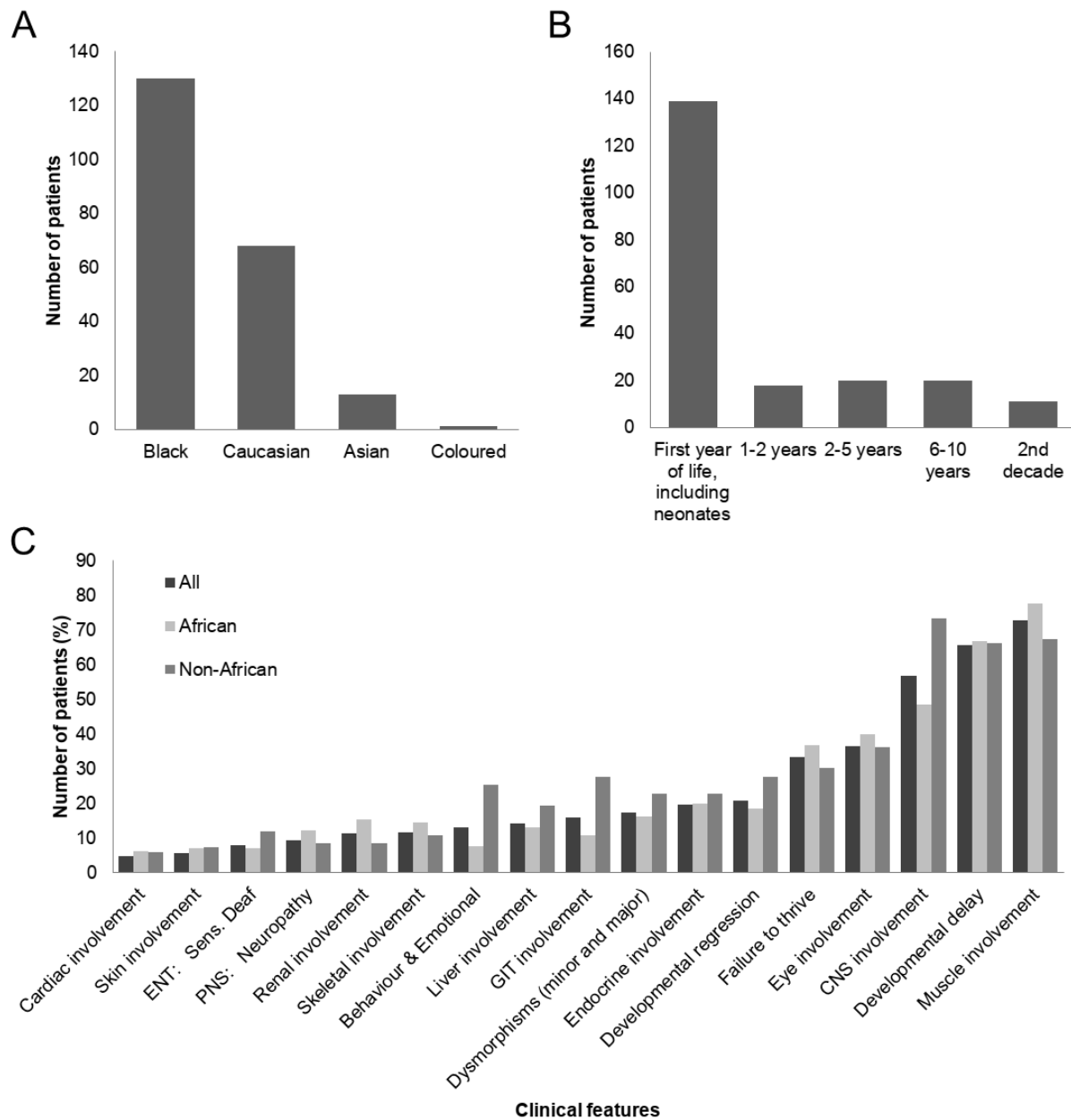
† Variants identified using panel NGS.

‡ Previously reported pathogenic variants.

§Variants identified using WES.

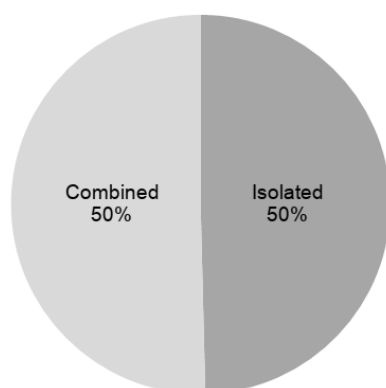
Appendix A.2

+/+ : Homozygous; +/- : Heterozygous; F: Female; M: Male; A: African; NA: Non-African; VUS: variant of uncertain significance; Mac: Macrocephaly; FTT: Failure to thrive; DD: Developmental delay; DR: Developmental regression; CNS: Central nervous system involvement; Eye: Eye involvement; D: Sens. Deafness; M: Muscle involvement; GIT: Gastro-intestinal involvement; R: Renal involvement; C: Cardiac involvement; E: Endocrine abnormalities; L: Liver involvement; S: Skeletal involvement; Skin: Skin involvement; VUS: variant of uncertain significance.

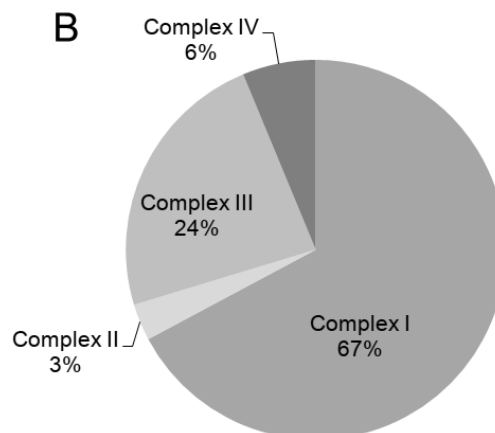


Appendix A.2

A



B





Contents lists available at ScienceDirect

Journal of the Neurological Sciences

journal homepage: www.elsevier.com/locate/jns

A novel mutation in *ETFDH* manifesting as severe neonatal-onset multiple acyl-CoA dehydrogenase deficiency

Francois H. van der Westhuizen^{a,*}, Izelle Smuts^b, Engela Honey^c, Roan Louw^a, Maryke Schoonen^a, Lindi-Maryn Jonck^a, Marli Dercksen^a

^a Human Metabolomics, North-West University, Potchefstroom, South Africa

^b Department of Paediatrics, Steve Biko Academic Hospital, University of Pretoria, South Africa

^c Department of Genetics, University of Pretoria, South Africa

ARTICLE INFO

Keywords:

Multiple acyl-CoA dehydrogenase deficiency

MADD

Glutaric aciduria type II

ETFDH

ABSTRACT

Neonatal-onset multiple acyl-CoA dehydrogenase deficiency (MADD type I) is an autosomal recessive disorder of the electron transfer flavoprotein function characterized by a severe clinical and biochemical phenotype, including congenital abnormalities with unresponsiveness to riboflavin treatment as distinguishing features. From a retrospective study, relying mainly on metabolic data, we have identified a novel mutation, c.1067G > A (p.Gly356Glu) in exon 8 of *ETFDH*, in three South African Caucasian MADD patients with the index patient presenting the hallmark features of type I MADD and two patients with compound heterozygous (c.1067G > A + c.1448C > T) mutations presenting with MADD type III. SDS-PAGE western blot confirmed the significant effect of this mutation on *ETFDH* structural instability. The identification of this novel mutation in three families originating from the South African Afrikaner population is significant to direct screening and strategies for this disease, which amongst the organic acidemias routinely screened for, is relatively frequently observed in this population group.

1. Introduction

Multiple acyl-CoA dehydrogenase deficiency (MADD, OMIM: 231680), also known as glutaric aciduria type II (GAII), is an autosomal recessive metabolic disorder. Mutations in the *ETFA*, *ETFB* and *ETFDH* genes result in the deficient function of alpha and beta subunits of the electron transfer protein (ETF) or ETF-coenzyme Q oxidoreductase, respectively. Dysfunction in either of these two flavoproteins leads to compromised fatty acid and amino acid oxidation as well as choline metabolism.

The MADD phenotype can be divided into three clinical classes: two neonatal-onset forms with (type I) or without (type II) congenital anomalies or a mild/delayed late-onset form (type III), which shows a variable degree of response to riboflavin treatment [1–3]. The neonatal-onset form has a poor prognosis, presenting within the first week(s) of life with severe hypoglycemia, mild to severe hyperammonaemia, metabolic acidosis, with or without dysmorphic features, hepatic-, cardiac- and renal involvement. The phenotypic presentation of type III is highly variable, characterized by episodic lethargy, vomiting, hypoglycemia, metabolic acidosis with or without hyperammonaemia and

hepatomegaly. Painful myopathy can be progressive and abnormal lipid storage and free carnitine depletion is commonly observed [1,3].

The biochemical characterization of MADD includes organic acid and acylcarnitine profiling, which indicates, increased levels of aliphatic mono- and dicarboxylic acids, acylglycine conjugates as well as an increase of C4-C18 acylcarnitines in blood [1]. Although mutations are most frequently encountered in the *ETFDH* gene in the late-onset form of MADD, for the most severe (type I) neonatal form of the disease mutations in *ETFDH*, *ETFA*, and *ETFB* are found more or less in equal frequency [2,3,4–8].

As mutation screening is not currently offered routinely in our (South African) population, and since biochemical data through routine metabolic screening suggests that MADD may be in the same order of prevalence as other organic acidemias in South African populations [9,10] we've begun to retrospectively investigate the molecular genetics of MADD with metabolic data as the denominator. From the initial explorative investigations in three Caucasian patients, we've identified a novel pathogenic variant present in all three cases, with the homozygous form in the index case presenting as a (type I) severe neonatal onset form of the disease.

* Corresponding author.

E-mail address: Francois.vanderWesthuizen@nwu.ac.za (F.H. van der Westhuizen).

¹ Postal Address: Human Metabolomics, Private Bag × 6001, North-West University, Potchefstroom 2520, South Africa.

Table 1
Summary of clinical, biochemical and genetic features of patients.

	Patient 1 (Index)	Patient 2	Patient 3
Gender	Male	Female	Female
Age of onset/death	Week 1/Day 9	Year 2/undetermined (at least past 8 years)	Year 4/23
Main clinical features	Intra-uterine growth retardation Multiple congenital anomalies: • Cardiac lesions • Hydronephrosis Riboflavin unresponsive	Progressive muscle weakness Hepatomegaly Migraine Exercise intolerance Hypermobility Riboflavin responsiveness undetermined	Progressive muscle weakness Hepatomegaly Migraine Partially responsive to riboflavin
Routine Biochemistry:			
• Metabolic acidosis	Present	Not present	Present
• Glucose	Normal-↓	Normal	Normal-↓
• Ammonia	↑-↑↑	Not performed	↑-↑↑
• Urine ketones	Absent	Absent	Absent
• Lactate/pyruvate	Normal-↑	↑↑	Normal-↑
• Transaminases	↑↑↑GGT, AST	↑AST, ALT	↑AST, ALT
• Creatine kinase	Not performed	Normal	Normal-↑↑
• CBC findings	Pancytopenia	Normal	Normal
Histology	Not performed	Muscle: cytoplasmic lipid accumulation in myofibres	Liver: macrovascular steatosis, abnormal lipid storage
Metabolic findings in urine:			
• Organic acids	↑Ethylmalonic acid ↑↑↑Dicarboxylic acids ↑↑↑Glycine conjugates ↑↑2-Hydroxyglutaric acid ↑↑↑Lactic acid ↑Krebs cycle intermediates	↑Ethylmalonic acid ↑Dicarboxylic acids ↑↑↑Glycine conjugates ↑2-Hydroxyglutaric acid	↑Ethylmalonic acid ↑Dicarboxylic acids ↑↑Glycine conjugates ↑2-Hydroxyglutaric acid Normal-↑Lactic acid
• Acylcarnitines	↑C4-, C5, C5-DC Low free carnitine	↑C4-, C5, C5-DC Low free carnitine	↑C4-, C5, C5-DC, C8 Normal free carnitine
• Amino acids	General amino aciduria with ↑↑ Sarcosine	↑↑Sarcosine, ↑↑Glycine	↑Sarcosine, ↑↑Glycine
Specialized investigations	None	Muscle respiratory chain enzymes deficient (CI, CIII, CII + III, CIV); Muscle CoQ10 levels reduced	Oxidation of C16 and C14 fatty acids reduced in fibroblasts (20% vs. controls mean; C16:C14 = 0.5) and leukocytes (5% vs. controls mean; C16:C14 = 0.96)
ETFDH variant/s (dbSNP)	c.1067G > A +/+; p.Gly356Glu (novel)	c.1067G > A +/-; p.Gly356Glu c.1448C > T +/-; p.Pro483Leu (377656387)	c.1067G > A +/-; p.Gly356Glu c.1448C > T +/-; p.Pro483Leu (377656387)

C4: butyryl-/isobutyrylcarnitine; C5: isovalerylcarnitine; C5-DC: glutarylcarnitine; C8: octanoylcarnitine; CBC: complete blood count; AST: aspartate aminotransferase; ALT: alanine aminotransferase; GGT: gamma-glutamyltransferase.

2. Patients and methods

Informed consent was obtained as part of current studies with ethical approval numbers 91/98 (University of Pretoria) and NW-00170-13-S1 (North-West University). Table 1 summarizes the clinical, biochemical and genetic features of the three patients.

2.1. Patient 1 (index patient)

The firstborn male child from non-consanguineous Afrikaner parents presented shortly after birth with severe metabolic decompensation unresponsive to standard metabolic intervention. He was born prematurely via cesarean section and needed admission to the intensive care immediately after delivery. A loud systolic murmur was audible and the chest X-ray showed a prominent right ventricle and absent pulmonary conus. The ECG displayed an extreme right QRS axis and a dysplastic pulmonary valve with a pulmonary outflow gradient measuring 57 mmHG indicated a severe pulmonary stenosis. The foramen ovale was patent with a right to left shunt and there was a large atrial septum aneurism. The renal ultrasound showed dilated calyces in the upper lobes more severely on the left indicating hydronephrosis.

After a relative normal 24 h complicated only with hyperlactaemia and suspected hypothyroidism his condition rapidly deteriorated with persistent hypoglycaemia requiring high doses of intravenous glucose, metabolic acidosis and hyperammonaemia (548 µmol/L). He needed intubation and ventilation on day 6 of life but his condition deteriorated despite administration of riboflavin

(200 mg/kg/day) and L-carnitine (100 mg/kg/day). He developed a bulging fontanel, sclerema neonatorum and convulsions and died on day 9 of life. GC-MS organic acid profiling performed on urine at day 3 of life showed dicarboxylic aciduria with the presence of various short and medium chain acylglycine conjugates suggestive of MADD. The metabolic excretion as well as early onset was indicative of a severe neonatal phenotype with a very poor prognosis.

His family subsequently had a normal male child delivered after prenatal diagnosis performed on chorionic villi showed the fetus was heterozygous for the reported c.1067G > A mutation.

2.2. Patient 2

An eight year-old female patient presented with the complaint of clumsiness, chronic fatigue, exercise intolerance, muscle weakness and severe migrainous-like headaches. She was born at term from non-consanguineous parents and had three normal siblings. Her early gross motor development was normal and she was able to walk before the age of one year. At the age of 17 months the parents observed that she had poor head control related to muscle weakness. Since the age of three years the weakness slowly progressed. Her body mass index for her age was 13.3 and fell on z-score - 1.8 at the age of eight years. She had a waddling gait and mild proximal muscle weakness, but she did not have a Gower sign. Polyminimyoelonus resembling spinal muscular atrophy was observed. The deep tendon reflexes were suppressed, but present. She had no other cerebellar signs. Hypermobility of all her joints was observed. Her general level of functioning was on the level of 6.5 years.

She had hepatomegaly with mildly elevated liver enzymes. Renal functions were normal, creatine kinase was normal and the lactate:pyruvate ratio was 16.3. The routine metabolic investigations showed general dicarboxylic aciduria including acylglycine and acylcarnitine conjugates. Spinal muscular atrophy was excluded: the patient did not have a homozygous deletion in the telomeric region of the survival motor neuron on exon seven. A muscle biopsy was performed for the determination of respiratory enzyme analyses, after which the patient and family was not responsive to any further follow up.

2.3. Patient 3

This patient was born at term from non-consanguineous parents with one healthy sibling and no previous family history of neuromuscular disease. She initially presented with hypotonia and only started to walk at the age of 20 months. The progressive myopathic phenotype of unknown cause was confirmed at the age of two year after a comprehensive clinical evaluation. Metabolic decompensation was observed at the age of three years when she was admitted with coma, severe hyperammonemia (394 $\mu\text{mol/L}$ which decreased to 90 $\mu\text{mol/L}$ after hospitalization), hepatomegaly and hypoglycemia without urinary ketones. No myoglobinuria were noted, but prerenal failure was suspected. Her liver enzymes were elevated, together with elevated creatine kinase. A liver biopsy indicated macrovascular steatosis with no necrosis. The metabolic investigations showed increased levels of urine mono- and dicarboxylic acid, acylglycine as well as short and long chain acylcarnitine conjugation. The latter in combination with general clinical presentation of myopathy and liver involvement was suggestive of MADD. She received riboflavin and L-carnitine treatment together with a low protein diet and improved partially, waking up from the coma and showing stronger motor function. Progressive deterioration of muscular function did however occur and she suffered from episodic migraines. Regular follow-up metabolic investigations were done to assess her metabolic status, which showed moderate to limited biochemical improvement after implementation of therapeutic intervention. She did not present with cognitive impairment and became a fully functional adult. She unfortunately died from a stroke-like episode at age 23 before the genetic diagnosis was confirmed.

2.4. Metabolic and biochemical investigations

Metabolic investigations were performed essentially as described before [11]. Respiratory chain enzyme activities and CoQ10 levels were analyzed in muscle homogenates (600 $\times$ g supernatants) for Patient 2 as described before [12,13] and normalized to citrate synthase (UCS). Fatty acid oxidation of [9,10- ^3H] palmitic (C16) and myristic (C14) acids were performed in fibroblasts and leukocytes for Patient 3 as described before [14].

2.5. Mutation and protein analyses

The outcomes reported here are the results from a retrospective investigation of selected cases, using next generation sequencing (NGS, Ion Torrent) of candidate genes. Primary NGS data analysis and secondary data mining were conducted using an in-house bioinformatics pipeline and web-based annotation tools such as Ensembl online VEP runner [15] and data mining tools, including GEMINI (v0.18.0) [16] OMIM, ClinVar, and available allele frequency data from the extensive Exome Aggregation Consortium (ExAC) database. The potential pathogenic variants identified and allelic distribution in all of these cases were confirmed using Sanger sequencing and, as the c.1067G > A variant is part of a homopolymeric region, also restriction length fragment polymorphism (RFLP) analyses (Fig. 2S).

SDS-PAGE/western blot analysis were used to confirm the steady state levels of muscle ETFDH and, due to the limited material available, could only be done in Patient 2. Immunodetection using the antibodies

against ETFDH (Abcam, ab131376) and β -actin (Abcam ab8227) were conducted and expression quantified against total protein load using the procedures described for Bio-Rad's V3 workflow (Bio-Rad Laboratories).

3. Results

Table 1 summarizes the clinical, biochemical and other specialized investigations for the three cases. The following key findings for metabolic, biochemical and genetic analyses are reported.

3.1. Metabolic and biochemical investigations

Routine and follow up metabolic investigations in urine throughout the history of all three cases showed some of the hallmark metabolites associated with MADD, notably dicarboxylic acids, ethylmalonic acid, glutaric acid as well as acylcarnitines- (butyryl-, isovaleryl- and glutaryl-carnitine) and acylglycine (hexanoyl- isobutyryl-, isovaleryl and suberylglycine) conjugates. Patient 1 had not undergone any biopsies at time of his death. Based on her clinical features, Patient 2 was early on selected to undergo a muscle biopsy for respiratory chain enzyme analyses, which revealed combined complex I, III, II + III and IV deficiencies. The follow up investigation of CoQ10 levels in 600 $\times$ g supernatants showed significantly reduced levels at 0.86 nmol/UCS (controls 2.680–8.470 nmol/UCS, $n = 37$). For Patient 3, with an early suspicion of MADD, fatty acid oxidation confirmed the functional deficiency with palmitic (C16) and myristic (C14) acid oxidations almost equally reduced in fibroblasts and leukocytes.

3.2. Mutation and protein analyses

A c.1067G > A (p.Gly356Glu) variant on the *ETFDH* gene was detected in homozygous form in Patient 1 (Fig. 2S). This variant do not appear on any clinical or genetic databases, nor in literature consulted and was shown to be heterozygous in both parents as well as in an unaffected sibling. This novel variant is predicted to be deleterious by the SIFT prediction software and disease causing by MutationTaster prediction software. For both Patients 2 and 3, a previously documented c.1448C > T (p.Pro483Leu) mutation and the putative deleterious c.1067G > A variant was identified in heterozygous forms. Although the patient and family of Patient 2 were lost to follow up, the carrier status for the c.1067G > A variant was confirmed in mother and sibling for Patient 3. SDS-PAGE-western blot analysis (Fig. 1) of the steady state levels for ETFDH in the only available tissue material (muscle from Patient 2) showed a significant (83%) reduction compared to controls.

4. Discussion and conclusions

Three patients presenting with clinical and metabolic features of MADD were retrospectively investigated to identify the possible genetic cause for the disease. Targeting genes possibly involved, we identified a novel pathogenic variant (c.1067G > A) in *ETFDH* in all three patients. This variant was homozygous in the index patient (Patient 1), who presented with a MADD type I phenotype, and compound heterozygous with the well-documented pathogenic variant c.1448C > T in two other (Patients 2 and 3). Although the trajectories of clinical and further investigations were different, all three patients presented with clear clinical and metabolic features indicative of MADD type I (Patient 1), or type III (Patient 2 and 3). Patient 1, homozygous for the novel c.1067G > A variant, presented in first days of life with the most severe (type I) neonatal onset clinical and metabolic features of MADD, notably with congenital features and being unresponsive to riboflavin treatment. Patient 2, presenting with a multi-systemic neuromuscular features early in life, was initially diagnosed with a combined mitochondrial respiratory chain disorder in her ninth year. A deeper

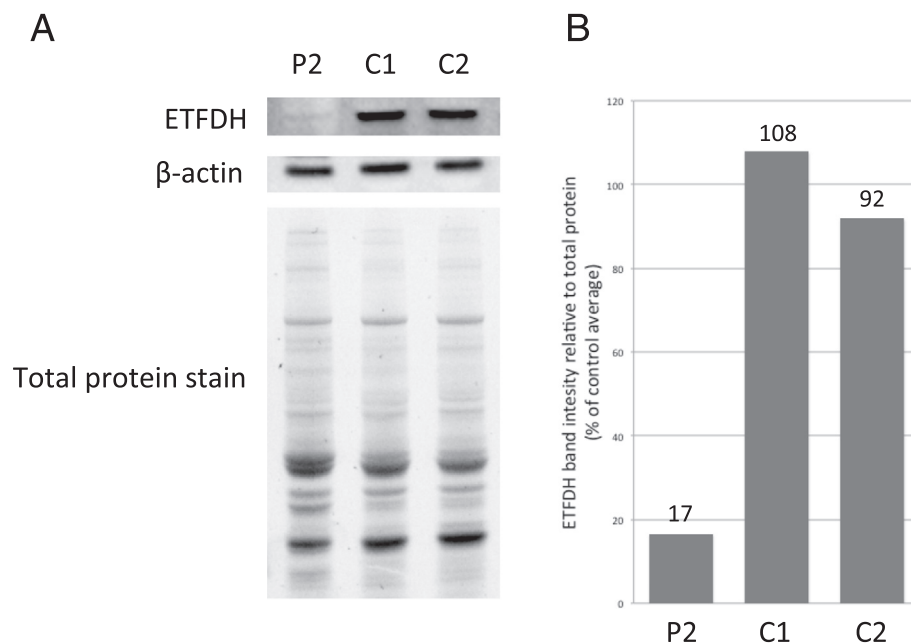


Fig. 1. Structural instability of muscle ETFDH in Patient 2. An enriched mitochondrial preparation from muscle of Patient 2 (P2, with c.1067G > A + c.1448C > T variants) and two healthy controls (C1, C2) were separated on SDS-PAGE and immune-stained for ETFDH and β -actin (top, Panel A). The band intensities for ETFDH were normalized to total protein content in each lane (bottom, Panel A), of which the quantified results are shown in Panel B.

investigation into CoQ₁₀ involvement and metabolomics resulted in the final MADD diagnosis and the genetic outcome presented here. Patient 3 where diagnosed with MADD early in life via metabolic and functional fatty acid oxidation tests, and were able to follow riboflavin treatment. In this case, having a variant associated with a riboflavin responsive (c.1448C > T) and unresponsive form (c.1067G > A), it is not surprising that this patient showed a partial responsiveness to riboflavin.

A closer look at this previously undocumented c.1067G > A variant, indicate a substitution of a highly conserved Glu for Gly at position 356, putatively affecting the FAD domain of ETFDH [17,18]. Comparing it to other type I mutations also do not highlight any evident structural similarities that can inform on any molecular basis for its pathogenicity. It is also notably the first associated with this severe phenotype identified in exon 8 of *ETFDH*. We had the limitation of having biological material for only one of the patients for further structural ETFDH analysis (Patient 2, with the compound heterozygous c.1067G > A + c.1448C > T variants). Considering that the homozygous c.1448C > T variant is known to result in significantly reduced ETFDH activity ($\pm$ 30% of WT) and steady state stability (< 10% of WT) under low riboflavin conditions [19], it can be concluded from the structural analysis (Fig. 1) that the presence of the c.1067G > A variant on the alternative allele to the c.1448C > T variant should have a similar impact on the instability of the native ETFDH protein. We propose that this result, in addition to the predictive parameters of the variant and the genetics in these families, support the suggestion that this novel variant can be added to the list of mutations causing a type I MADD phenotype.

Finally in conclusion, from the limited family histories and genetics that could be gathered from these three cases, no relation between these families could as yet be established. Nevertheless, we believe it is reasonable to suspect that a common founder for the c.1067G > A mutation may have originated within this population group. The reasons for this is, firstly, that all three of these patients originate from Caucasian Afrikaner families, which is a southern African population group having descended from Europe since the 17th century; and, secondly, that this mutation has not yet been reported in the better-screened European population groups from which the Afrikaner population originated. Considering the possibility that this mutation, in addition to the milder, riboflavin-responsive c.1448C > T mutation, may have a relatively high frequency in MADD patients of this

population group, we propose that this mutation should be included in screening strategies for families with a history of MADD in this and other related population groups.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jns.2017.11.012>.

Conflict of interest

There are no conflicts of interest.

Acknowledgements

Our thanks go out to the families of patients for their assistance in this follow up study, the referring physicians, Drs M Lippert and M Engelbrecht.

Funding

This work was in part supported by the Medical Research Council of South Africa under project title: *Investigating the aetiology of South African pediatric patients diagnosed with mitochondrial disorders*.

References

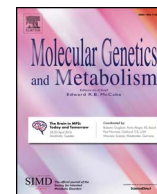
- [1] F.E. Frerman, S.I. Goodman, Defects of electron transfer flavoprotein and electron transfer flavoprotein-ubiquinone oxidoreductase: glutaric acidemia type II, in: C.R. Scriver, W.S. Sly, B. Childs, A.L. Beaudet, D. Valle, K.W. Kinzler, B. Vogelstein (Eds.), *The Metabolic and Molecular Bases of Inherited Disease*, 2001.
- [2] W.C. Liang, A. Ohkuma, Y.K. Hayashi, L.C. López, M. Hirano, I. Nonaka, S. Noguchi, L.H. Chen, Y.J. Jong, I. Nishino, ETFDH mutations, CoQ10 levels, and respiratory chain activities in patients with riboflavin-responsive multiple acyl-CoA dehydrogenase deficiency, *Neuromuscul. Disord.* 19 (2009) 212–216.
- [3] S.C. Grünert, Clinical and genetical heterogeneity of late-onset multiple acyl-coenzyme A dehydrogenase deficiency, *Orphanet J. Rare Dis.* 22 (2014) 117.
- [4] R.K. Olsen, S.E. Olpin, B.S. Andresen, Z.H. Miedzybrodzka, M. Pourfarzam, B. Merinero, F.E. Frerman, M.W. Beresford, J.C. Dean, N. Cornelius, O. Andersen, A. Oldfors, E. Holme, N. Gregersen, D.M. Turnbull, A.A. Morris, ETFDH mutations as a major cause of riboflavin-responsive multiple acyl-CoA dehydrogenation deficiency, *Brain* 130 (2007) 2045–2054.
- [5] Y. Yotsumoto, Y. Hasegawa, S. Fukuda, H. Kobayashi, M. Endo, T. Fukao, S. Yamaguchi, Clinical and molecular investigations of Japanese cases of glutaric acidemia type 2, *Mol. Genet. Metab.* 94 (2011) 61–67.
- [6] J. Xi, B. Wen, J. Lin, W. Zhu, S. Luo, C. Zhao, D. Li, P. Lin, J. Lu, C. Yan, Clinical features and ETFDH mutation spectrum in a cohort of 90 Chinese patients with late-onset multiple acyl-CoA dehydrogenase deficiency, *J. Inherit. Metab. Dis.* 37 (2014) 399–404.

- [7] K. Yamada, H. Kobayashi, R. Bo, T. Takahashi, J. Purevsuren, Y. Hasegawa, T. Taketani, S. Fukuda, T. Ohkubo, T. Yokota, M. Watanabe, T. Tsunemi, H. Mizusawa, H. Takuma, A. Shioya, A. Ishii, A. Tamaoka, Y. Shigematsu, H. Sugie, S. Yamaguchi, Clinical, biochemical and molecular investigation of adult-onset glutaric acidemia type II: characteristics in comparison with pediatric cases, *Brain and Development* 38 (2015) 293–301.
- [8] Fu HX, X.Y. Liu, Z.Q. Wang, M. Jin, D.N. Wang, J.J. He, M.T. Lin, N. Wang, Significant clinical heterogeneity with similar ETFDH genotype in three Chinese patients with late-onset multiple acyl-CoA dehydrogenase deficiency, *Neurol. Sci.* 37 (2016) 1099–1105.
- [9] M. Dercksen, M. Duran, L. Ijlst, L.J. Mienie, C.J. Reinecke, J.P. Ruiter, H.R. Waterham, R.J. Wanders, Clinical variability of isovaleric acidemia in a genetically homogeneous population, *J. Inherit. Metab. Dis.* 35 (2012) 1021–1029.
- [10] G. van der Watt, E.P. Owen, P. Berman, S. Meldau, N. Watermeyer, S.E. Olpin, N.J. Manning, I. Baumgarten, F. Leisegang, H. Henderson, Glutaric aciduria type 1 in South Africa-high incidence of glutaryl-CoA dehydrogenase deficiency in black South Africans, *Mol. Genet. Metab.* 101 (2010) 178–182.
- [11] C.J. Reinecke, G. Koekemoer, F.H. van der Westhuizen, R. Louw, J.Z. Lindeque, L.J. Mienie, I. Smuts, Metabolomics of urinary organic acids in respiratory chain deficiencies, *Metabolomics* 8 (2012) 264–283.
- [12] I. Smuts, R. Louw, H. du Toit, B. Kloppe, L.J. Mienie, F.H. van der Westhuizen, An overview of a South African cohort of patients with mitochondrial disorders, *J. Inherit. Metab. Dis.* 33 (Suppl. 3) (2010) S95–104.
- [13] O. Ikonen, A. Suomalainen, U. Turpeinen, Mitochondrial coenzyme Q10 determination by isotope-dilution liquid chromatography-tandem mass spectrometry, *Clin. Chem.* 59 (2013) 1260–1267.
- [14] M. Brivet, A. Slama, J.M. Saudubray, A. Legrand, A. Lemonnier, Rapid diagnosis of long chain and medium chain fatty acid oxidation disorders using lymphocytes, *Ann. Clin. Biochem.* 32 (1995) 154–159.
- [15] W. McLaren, L. Gil, S.E. Hunt, H.S. Riat, G.R. Ritchie, A. Thormann, P. Flicek, F. Cunningham, The Ensembl variant effect predictor, *Genome Biol.* 17 (2016) 122.
- [16] U. Paila, B.A. Chapman, R. Kirchner, A.R. Quinlan, GEMINI: integrative exploration of genetic variation and genome annotations, *PLoS Comput. Biol.* 9 (2013) e1003153.
- [17] J. Zhang, F.E. Frerman, J.J. Kim, Structure of electron transfer flavoprotein-ubiquinone oxidoreductase and electron transfer to the mitochondrial ubiquinone pool, *Proc. Natl. Acad. Sci. USA* 103 (2006) 16212–16217.
- [18] N.J. Watmough, F.E. Frerman, The electron transfer flavoprotein: ubiquinone oxidoreductases, *Biochim. Biophys. Acta* 1797 (2010) 1910–1916.
- [19] N. Cornelius, F.E. Frerman, T.J. Corydon, J. Palmfeldt, P. Bross, N. Gregersen, R.K. Olsen, Molecular mechanisms of riboflavin responsiveness in patients with ETF-QO variations and multiple acyl-CoA dehydrogenation deficiency, *Hum. Mol. Genet.* 21 (2012) 3435–3448.



Contents lists available at ScienceDirect

Molecular Genetics and Metabolism

journal homepage: www.elsevier.com/locate/ymgmeThe dilemma of diagnosing coenzyme Q₁₀ deficiency in muscleRoan Louw^{a,*}, Izelle Smuts^b, Kimmey-Li Wilsenach^a, Lindi-Maryn Jonck^a, Maryke Schoonen^a, Francois H. van der Westhuizen^a^a Human Metabolomics, North-West University (Potchefstroom Campus), Potchefstroom, South Africa^b Department of Paediatrics and Child Health, Steve Biko Academic Hospital, University of Pretoria, Pretoria, South Africa

ARTICLE INFO

Keywords:

Coenzyme Q10 deficiency

Complex II + III

Electron transport chain

OXPHOS

Reference range

ABSTRACT

Background: Coenzyme Q₁₀ (CoQ₁₀) is an important component of the mitochondrial respiratory chain (RC) and is critical for energy production. Although the prevalence of CoQ₁₀ deficiency is still unknown, the general consensus is that the condition is under-diagnosed. The aim of this study was to retrospectively investigate CoQ₁₀ deficiency in frozen muscle specimens in a cohort of ethnically diverse patients who received muscle biopsies for the investigation of a possible RC deficiency (RCD).

Methods: Muscle samples were homogenized whereby 600 × g supernatants were used to analyze RC enzyme activities, followed by quantification of CoQ₁₀ by stable isotope dilution liquid chromatography tandem mass spectrometry. The experimental group consisted of 156 patients of which 76 had enzymatically confirmed RCDs. To further assist in the diagnosis of CoQ₁₀ deficiency in this cohort, we included sequencing of 18 selected nuclear genes involved with CoQ₁₀ biogenesis in 26 patients with low CoQ₁₀ concentration in muscle samples.

Results: Central 95% reference intervals (RI) were established for CoQ₁₀ normalized to citrate synthase (CS) or protein. Nine patients were considered CoQ₁₀ deficient when expressed against CS, while 12 were considered deficient when expressed against protein. In two of these patients the molecular genetic cause could be confirmed, of which one would not have been identified as CoQ₁₀ deficient if expressed only against protein content.

Conclusion: In this retrospective study, we report a central 95% reference interval for 600 × g muscle supernatants prepared from frozen samples. The study reiterates the importance of including CoQ₁₀ quantification as part of a diagnostic approach to study mitochondrial disease as it may complement respiratory chain enzyme assays with the possible identification of patients that may benefit from CoQ₁₀ supplementation. However, the anomaly that only a few patients were identified as CoQ₁₀ deficient against both markers (CS and protein), while the majority of patients where only CoQ₁₀ deficient against one of the markers (and not the other), remains problematic. We therefore conclude from our data that, to prevent possibly not diagnosing a potential CoQ₁₀ deficiency, the expression of CoQ₁₀ levels in muscle on both CS as well as protein content should be considered.

1. Introduction

The mitochondrion is at the center of energy production and referred to as the “energy hub” of the eukaryotic cell. The oxidative phosphorylation (OXPHOS) system is situated in the inner mitochondrial membrane and incorporates the four complexes (CI–IV) of the respiratory chain (RC) as well as complex V (ATP-synthase) in the transduction of energy molecules into the energy releasing adenosine triphosphate (ATP) [18]. CoQ₁₀ plays a vital role in this process by

acting as an electron transport molecule from CI and CII (and other dehydrogenases) to CIII of the RC [9]. Therefore a CoQ₁₀ deficiency may result in RC dysfunction with diverse clinical manifestations [16]. Within the group of respiratory chain enzyme deficiencies, which has a minimum live birth prevalence of 1 in every 5000–10,000 ([1,2,19], Schaefer et al. 2008, [5,6]), the presence of the combined CI + III and/or CII + III RC deficiencies may be indicative of a CoQ₁₀ deficiency [3]. CoQ₁₀ deficiencies are the most readily treatable subgroup of mitochondrial disorders, and CoQ₁₀ supplementation has also shown

Abbreviations: CI to CIV, Respiratory chain enzyme complexes I to IV respectively; CS, Citrate synthase; RCD, Respiratory chain deficiency; CRC, Clinically referred controls; PDSS1, Prenyl (decaprenyl) diphosphate synthase, subunit 1; PDSS2, Prenyl (decaprenyl) diphosphate synthase, subunit 2; COQ2, Coenzyme Q2, polyprenyltransferase; COQ3, Coenzyme Q3, methyltransferase; COQ4, Coenzyme Q4; COQ5, Coenzyme Q5, methyltransferase; COQ6, Coenzyme Q6, monooxygenase; COQ7, Coenzyme Q7, hydroxylase; COQ8A, Coenzyme Q8A; COQ8B, Coenzyme Q8B; COQ10A, Coenzyme Q10A; COQ10B, Coenzyme Q10B; APTX, Aprataxin; ETFDH, Electron Transfer Flavoprotein Dehydrogenase; ETFB, Electron Transfer Flavoprotein Alpha subunit; ETFB, Electron Transfer Flavoprotein Beta subunit; BRAF, B-Raf proto-oncogene, serine/threonine kinase; APOE, Apolipoprotein E

* Corresponding author at: Human Metabolomics, North-West University (Potchefstroom Campus), Potchefstroom 2531, South Africa.

E-mail address: Roan.Louw@nwu.ac.za (R. Louw).

<https://doi.org/10.1016/j.ymgme.2018.02.015>

Received 19 December 2017; Received in revised form 21 February 2018; Accepted 21 February 2018

Available online 23 February 2018

1096-7192/ © 2018 Elsevier Inc. All rights reserved.

therapeutic benefit in patients with mitochondrial RCDs [6]. But, although CoQ₁₀ mainly localizes in the mitochondria (Saito et al., 2009), the exact distribution of CoQ₁₀ between the different organelles, membranes and cytoplasm remains elusive. Furthermore, no universally accepted units on how to represent muscle CoQ₁₀ status exists between researchers as CoQ₁₀ is expressed in units against mass of fresh tissue, per protein content or per unit of citrate synthase (CS).

The aim of this study was therefore to quantify CoQ₁₀ and to determine the diagnostic differences when CoQ₁₀ is expressed against CS and protein content in a cohort of patients who was previously included for investigations into underlying mitochondrial disorders. Furthermore, we aimed to determine if there is an association between CoQ₁₀ deficiency and other RC enzyme deficiencies. A panel of selected nuclear genes involved with CoQ₁₀ biosynthesis and genes involved in possible secondary CoQ₁₀ deficiencies were sequenced and the most important findings are briefly discussed. A new reference interval for CoQ₁₀ levels in 600 × g muscle supernatants, the working material often used when investigating RCDs in patients with suspected mitochondrial disease, is reported as well as the diagnostic difference for CoQ₁₀ deficiency when expressing muscle CoQ₁₀ against CS and protein.

2. Materials and methods

2.1. Patient samples

All muscle samples from the cohort of 156 pediatric patients referred to the Pediatric Neurology Unit at the Steve Biko Academic Hospital, Pretoria, South Africa, were obtained from the *Vastus lateralis* muscle and stored immediately at −80 °C until use. Patients with a Mitochondrial Disease Criterion score [23] of six or higher, or that presented with a severe clinical phenotype suggestive of one of the syndromic mitochondrial disorders, were included for OXPHOS analysis (n = 156) and genetic analysis (n = 26) in this study. The cohort consisted of 94 African, 54 Caucasian and eight Indian patients.

2.2. Muscle sample preparation

Frozen muscle samples (30–100 mg) were thawed and homogenized in 10 volumes of isotonic buffer [mannitol 210 mM; sucrose 70 mM; 4-(2-hydroxyethyl)-1-piperazine-ethanesulphonic acid (HEPES) 5 mM; ethylene glycol tetra-acetic acid (EGTA) 0.1 mM; pH 7.2] with 15 strokes in a tight-fit Potter-Elvehjem tissue grinder. Samples were centrifuged for 10 min at 600 × g at 4 °C. Supernatants were transferred to new tubes and frozen at −80 °C prior to enzyme analyses and CoQ₁₀ quantification.

2.3. Analyses of respiratory chain enzymes

Mitochondrial RC enzymes (CI to IV; EC 1.6.5.3, EC 1.3.5.1, EC 1.10.2.2, EC 1.9.3.1, respectively, as well as CII+III, i.e. succinate:cytochrome c reductase; EC. 1.3.5.1 + 1.10.2.2) and citrate synthase (CS, EC 2.3.3.1) activities were measured in muscle as described previously [21]. Protein content was determined with the bicinchoninic acid assay [20]. A respiratory chain deficiency was diagnosed when an RC enzyme activity was lower than reference values when expressed against at least two of three enzyme markers (CS, CII, or CIV), providing that these were not deficient.

2.4. Quantification of coenzyme Q₁₀

CoQ₁₀ was quantified with a modified version of the assay described by Itokonen et al. [10]. In short, 30 µl muscle sample (600 × g supernatant) was mixed with 30 µl internal standard (0.5 µg/ml CoQ₁₀[²H₆] dissolved in acetonitrile) in a glass Kimax tube for 30 s. Then 1 ml ice-cold ethanol: isopropanol (95/5; v/v) was added to precipitate protein

before the sample was mixed for 10 s. The sample mixture was then extracted twice with 2 ml *n*-hexane. The organic phase from the second extraction was added to the organic phase from the first extraction before 30 µl of ethanolic 1,4-benzoquinone (0.4 mg/ml) was added for the complete oxidation of CoQ₁₀. The sample was dried under nitrogen gas at room temperature in a light protection glass vial for 20 min. The dried residue was re-dissolved in 60 µl acetonitrile prior to LC-MS/MS analysis.

Samples were analyzed using an Agilent 6460 Triple Quadrupole mass spectrometer (MS) quipped with an Agilent 1290 Infinity Binary Pump for sample handling and mobile phase delivery. The injection volume was 1 µl and the injection needle was washed for 10 s in a flush port containing a methanol:isopropanol:water (70:10:20, v/v/v) solution to prevent carryover. The mobile phase used for the isocratic separation was 100% methanol containing 5 mM ammonium formate, with a constant flow rate of 0.3 ml/min. An Agilent ZORBAX StableBond SB-C18 narrow-bore column-cartridge (2.1 × 30 mm, 3.5 µm) was used for reversed phase chromatography and the column temperature was kept at 45 °C. Samples were delivered to the MS via electrospray ionization (ESI) in positive mode. Multiple reaction monitoring (MRM) was used for the monitoring the ammonium adducts of the target analytes (m/z 880.7 → 197.2 for CoQ₁₀, m/z 882.7 → 197.2 for CoQ₁₀H₂, and m/z 886.7 → 203.1 for CoQ₁₀[²H₆]). Sheath gas and nebulizing/drying gas temperature and flow rate were set at 300 °C and 6 l/min, respectively. Nebulizer pressure was set at 30 psi, the capillary voltage at 3500 V, the nozzle voltage at 500 V and the dwell time was set at 50 ms. Although CoQ₁₀ was oxidized using ethanolic 1,4-benzoquinone, the presence/absence of CoQ₁₀H₂ was monitored to confirm that all the CoQ₁₀ in the sample was oxidized. CoQ₁₀ content was normalized using the internal standard (CoQ₁₀[²H₆]) and expressed against citrate synthase activity as nmol CoQ₁₀/UCS (where U = µmol/min·mg^{−1} CS) or against total protein content as nmol/mg. The linear dynamic range of the CoQ₁₀ assay was established as 0.01–2 µg/ml for CoQ₁₀ ($R^2 > 0.999$). The intra- and inter-day coefficient of variance for the CoQ₁₀ assay was 1.96% (determined after sequential analysis of five QC samples) and 2.37% (determined after analysis of 18 QC samples in 18 different batches over a period of 4 months), respectively.

2.5. Genetic analysis on coenzyme Q₁₀ deficiency

The coding regions of selected genes involved in CoQ₁₀ biosynthesis and secondary CoQ₁₀ deficiencies were sequenced in selected patients identified with reduced muscle CoQ₁₀ levels, expressed on both protein and CS. The genes included for sequencing were *PDSS1* (*COQ1*), *PDSS2*, *COQ2-9*, *COQ10A*, *COQ10B*, *APTX*, *ETFDH*, *ETFA*, *ETFB*, *BRAF*, and *APOE*. The Ion PGM™ System for Next-Generation Sequencing (NGS) was used with a custom designed Ion AmpliSeq™ Targeted Sequencing Technology panel from ThermoFischer Scientific. The panel of 18 genes had a total of 304 amplicons with a combined size of 61 kb with coverage of 98% of exonic target regions. The amplicon library preparation, enrichment and sequencing were done in accordance to manufacturer's protocols.

Primary data analysis (quality assessment, read alignment against GRCh37/hg19 and variant identification) was done using the Torrent Suite™ Software for Sequencing Data Analysis (v5.0.2) and Ion Torrent™ Suite Software Plugins (v5.0) for variant calling. Secondary data analyses included variant annotation using Ensembl online VEP runner (Ensembl GRCh37 release 85, [10]) and data mining using GEMINI, an open-source light database framework for genome mining (GEMINI v0.19.1, [15]). Variants were firstly filtered based on allele frequency for specifically African and Caucasian population groups as reported in the Exome Aggregation Consortium (ExAC; [8]) followed by variant classification as either novel (coding variants with or without Loss of Function; LoF) or previously reported (coding variants with or without LoF). The variant prediction software, SIFT (sorting intolerant from tolerant) and PolyPhen-2 (Polymorphism Phenotyping, v2) were

used to sort the intolerant from tolerant variants and predicted the polymorphism impact respectively [25]. Potential disease-causing variants were further investigated using OMIM and ClinVar (<http://www.ncbi.nlm.nih.gov/clinvar/>), to determine if a known clinical phenotype (clinical significance) was associated with the identified variants. Variants of importance were confirmed and validated by Sanger Sequencing.

3. Results and discussion

3.1. The effect of ethnicity, gender and age on 600 × g muscle CoQ₁₀ levels

Samples (n = 156) were randomized and CoQ₁₀ analyzed in 18 batches with a quality control (QC) sample included in every batch. The QC sample was prepared by mixing a small volume of numerous samples used in the study. Only data from the clinical referred control (CRC; patients who initially presented with symptoms usually associated with mitochondrial disease that were referred to the clinic, but where no respiratory chain deficiency on enzymatic level was detected; n = 80) were used to investigate the potential influence of ethnicity, gender and age on muscle CoQ₁₀. As illustrated in Fig. S1, CRC samples were divided into ethnic groups [Black African (n = 38), Caucasian (n = 35) and Indian (n = 7)], gender groups (39 female and 41 male) and the following age groups: < 1 years (n = 16), 1–10 years (n = 57), 11–25 years (n = 5) and 26–60 years (n = 2). Muscle CoQ₁₀/UCS were not significantly influenced by ethnicity (P = .07; Kruskal Wallis test), gender (P = .63; Mann Whitney U test) or age (P = .60; ANOVA). When expressing the CoQ₁₀ against protein content (nmol/mg protein), no significant effect were found either for ethnicity (P = .52; Kruskal Wallis test), gender (P = .80; Mann Whitney U test) or age (P = .34; ANOVA). Thus the results indicate that muscle CoQ₁₀ normalized to CS or protein was not influenced by ethnicity, gender or age in our cohort. Itokonen reported that gender and age affected muscle CoQ₁₀ when normalized to mass of wet tissue, but when they normalized the CoQ₁₀ to CS, no significant effect of either age or gender was found (Itokonen et al. [10]). Thus although Itokonen used isolated mitochondria (15,000 × g pellet), while we used 600 × g supernatant, the same outcome was obtained: when normalizing CoQ₁₀ to CS, age group and gender had no significant effect on muscle CoQ₁₀.

3.2. CoQ₁₀ in respiratory chain deficient (RCD) patients

CoQ₁₀ levels from the CII + III defect group (which included patients with combined CII + III deficiencies; n = 29), other RCD group [which included patients diagnosed with RCD (excluding CII + III defects); n = 47] and the CRC group (n = 80) were compared (Fig. 1). Table 1 summarizes the mean, median and interquartile (25th percentile - 75th percentile) ranges for CoQ₁₀ in these groups expressed against CS activity (UCS) and protein content.

Both the median CoQ₁₀/UCS ratios of the CRC group and the Other RCD group differed significantly from the CII + III defect group (P < .001 and P = .036, respectively), although no difference was found between the other RCD group and the CRC group. When expressing the CoQ₁₀ against protein content (nmol/mg protein), the median CoQ₁₀ of the CII + III deficient group differed significantly from the CRC group (P = .018) but not from the other RCD group (P = .216). The CRC and other RCD groups also did not differ significantly from one another. This significant lower CoQ₁₀ detected in the CII + III group relative to the controls (CRC) was expected, since low CoQ₁₀ availability in the mitochondria would impair electron transfer essential for ATP production ([17], Yubero et al. 2016). However, it is likely that CoQ₁₀ must decrease to a certain threshold before oxidative phosphorylation would be impaired.

3.3. Correlation of muscle CoQ₁₀ concentrations with complex II + III RC enzyme activity

Firstly, no significant correlation was found between CoQ₁₀/UCS levels and complex II + III enzyme activity in the CII + III defect group (Spearman's correlation = 0.02, P < .92). This was however expected since (i) the group of CII + III patients was relatively small (only 29 patients) and (ii) not all CII + III defects are caused by deficient CoQ₁₀, some patients might have a genetic defect resulting in altered structure/function of CII and/or CIII. However, when investigating the Other RCD and CRC groups, muscle CoQ₁₀/UCS strongly correlates with CII + III enzyme activity (Spearman's correlation = 0.62) as illustrated in Fig. S2. The correlation was also significant (P < .001). This finding is consistent with the report by Yubero et al. (2016), where they showed a strong correlation between CII + III (normalized to CS) and CoQ₁₀ (also normalized to CS) in a large cohort (n = 435) of patients investigated for mitochondrial disease. However, when expressing CoQ₁₀ against protein (nmol/mg), no significant correlation was found in our cohort (Spearman's correlation = 0.202).

3.4. A central 95% reference interval for CoQ₁₀ in 600 × g muscle supernatants

A central 95% reference interval (i.e. the 2.5th percentile to 97.5th percentile) for 600 × g muscle supernatants was calculated according to the International Federation of Clinical Chemistry (IFCC) guidelines (using the CRC samples; n = 80) and determined to be 1.71–8.46 nmol/UCS, with a median ratio of 3.606 nmol/UCS. This observed reference interval for muscle CoQ₁₀ was similar to the reference interval (2.68–8.40 nmol/UCS) reported for skeletal muscle 600 × g supernatants in a large cohort (n = 436) (Yubero et al. 2016). When CoQ₁₀ was expressed against protein, a reference interval of 0.213–1.387 nmol/mg protein was calculated with a median of 0.547 nmol/mg protein.

3.5. Patients identified with muscle CoQ₁₀ deficiency

The new reference intervals were subsequently used to identify patients in the cohort with decreased CoQ₁₀ levels. Decreased CoQ₁₀ levels were considered as “deficient” in those samples which were below the lowest limit (2.5th %) of the reference interval. Using this interval, CoQ₁₀ levels was found to be deficient (< 1.710 nmol/UCS) in six Black African patients and three Caucasian patients for this particular cohort when expressed against CS (Table 2). However, when expressing CoQ₁₀ against protein content, 12 patients were identified to have deficient CoQ₁₀ levels (below 0.213 nmol/mg protein). A striking observation when comparing the two tables is that there are only four patients that overlaps between the two groups (P2, P14, P24, P29), thus being detected as CoQ₁₀ deficient when expressed against both CS and protein. It is also notable, but possibly coincidental, that none of the Caucasian patients were identified as deficient when expressing CoQ₁₀ levels against total protein (Table 3).

The conundrum as to why only four patients were CoQ₁₀ deficient against both CS and protein, while the rest of the patients were only CoQ₁₀ defective against one of the normalization markers, persisted until the CS activity of the homogenates were examined. Citrate synthase, an enzyme found in the mitochondrial matrix, is routinely used as a normalization factor for mitochondrial content (Reisch et al. 2007). Protein content, on the other hand, as measured in this study, is not due to mitochondrial content, but all proteins in the sample/homogenate.

When we investigated the patients CoQ₁₀ deficient against protein, but not CS, many of these homogenates displayed very low CS activity (in nmol/min·mg⁻¹ protein). These homogenates can be considered as having a low mitochondrial content. Therefore, when the CoQ₁₀ concentration is normalized to the low CS (a small value), a higher value is obtained for CoQ₁₀/UCS, falling within the reference range and the

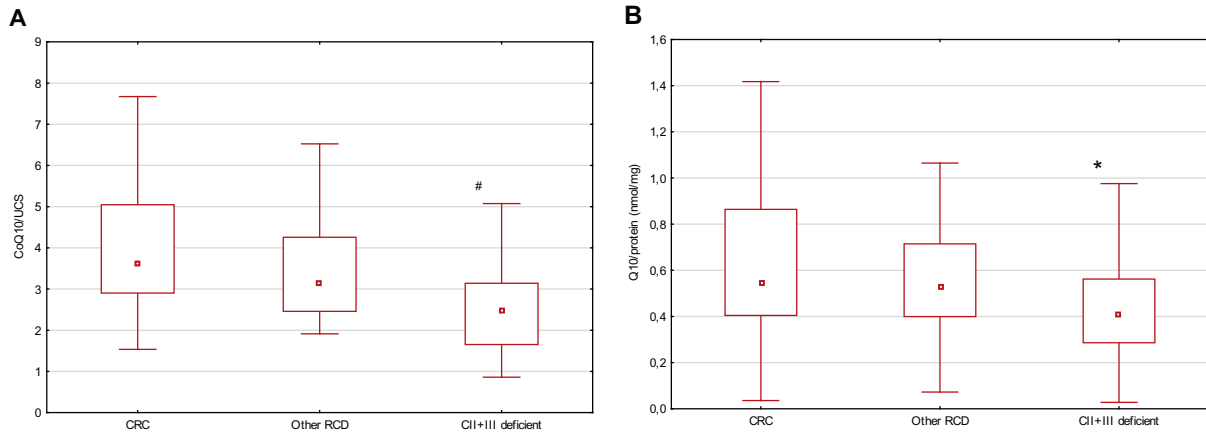


Fig. 1. Muscle CoQ₁₀ concentrations in a cohort of South African patients who received muscle biopsies for the investigation of a possible RCD. Patients were divided into clinically referred controls (CRC, *n* = 80), Other RCD (patients diagnosed with a muscle RCD on enzyme level, but excluding CII + III deficient patients, *n* = 47) and patients with CII + III deficiency (*n* = 29). (A) When expressing the CoQ₁₀ against CS activity (CoQ₁₀/UCS), the median CoQ₁₀ levels of the CII + III deficient group differed significantly (#) from the CRC group (*P* < .001) and the Other RCD group (*P* = .019). CRC and Other RCD groups did not differ significantly from one another. (B) When expressing the CoQ₁₀ against protein content (nmol/mg protein), the median CoQ₁₀ of the CII + III deficient group differed significantly (*) from the CRC group (*P* = .018) but not from the Other RCD group (*P* = .216). CRC and Other RCD groups also did not differ significantly from one another.

patient is thus not flagged as being deficient. On deeper inspection of the patients being CoQ₁₀ deficient against CS but not protein, most homogenates (> 80%) displayed very high CS values. Normalization to CS thus resulted in very low CoQ₁₀/UCS, indicating defective CoQ₁₀ levels, but still normal CoQ₁₀/protein. No significant correlation (Spearman's correlation < 0.3) was detected between CS and protein. Therefore it cannot be assumed that a CoQ₁₀ deficiency would inevitably be identified when using only one of the two normalization approaches. The question remains as to why the homogenates display such a wide range in CS activity, and why CS does not correlate with protein content. This could be due to a number of reasons, including: (1) excessive proliferation of mitochondria observed in mitochondrial disease (DiMauro et al. 2004), resulting in relative high CS activity; (2) massive rhabdomyolysis in patients, leading to muscle protein depletion [13], also resulting in elevated CS activity relative to protein content; (3) mitochondrial DNA depletion syndrome is expected to lead to reduced CS activity relative to protein content (Montero et al. [14]); (4) the quality of the muscle biopsy; and (5) the efficiency of the homogenization process of the tissue. It is thus evident that expression of CoQ₁₀ levels and choice of normalization should be mindful of pathology and sample preparation.

Another aspect to keep in mind is subcellular localization: CoQ₁₀ is not only found in mitochondria, but also in the Golgi vesicles, lysosomes, microsomes, peroxisomes, plasma membranes and cytosol [4]. Although the exact subcellular distribution of CoQ₁₀ remains elusive, some reports mention that approximately 50% of cellular CoQ₁₀ is present in the mitochondria (Yubero et al. 2014). Therefore the mitochondrial content of a homogenate might greatly influence the eventual [CoQ₁₀] of the homogenate. This, as well as the great variation in CS might, in part, explain the anomaly why most patients are only CoQ₁₀ deficient against one of the two markers used in this study.

When considering correlation of CoQ₁₀ levels with enzymology,

more often than not, combined CI + III and CII + III enzyme deficiencies are impaired by lower CoQ₁₀ levels and therefore the presence of these combined RCDs in skeletal muscle are especially predictive of a CoQ₁₀ deficiency [17]. The results in this study support this notion since eight of the nine (89%) patients identified with CoQ₁₀ deficiency when expressed against CS, (< 1.710 nmol/UCS, i.e., below the 2.5th % of the reference interval) also had CII + III deficiency (Table 2). Three of the nine patients with decreased CoQ₁₀/UCS ratios had isolated CII + III deficiency, while two patients had associated multiple RCDs, and three patients had CII + III deficiencies with a concomitant CII, CIII and CIV deficiency, respectively. This reiterates the importance of determining CoQ₁₀ even when a complex II-III deficiency is accompanied by other respiratory chain enzyme defects. Only one patient found with deficient CoQ₁₀/UCS was not diagnosed with CII + III deficiency. This particular patient however displayed very low CII + III activity that was just above the cut-off value used to diagnose a CII + III deficiency. However, when the CoQ₁₀ levels are expressed against protein levels (Table 3), only seven of the twelve patients (58%) also had a CII + III deficiency. From the remaining five patients, we were able to identify and detect another RCD in two patients but not in the remaining three patients.

3.6. Genetic analysis

We further investigated the coding regions of selected genes involved in CoQ₁₀ biosynthesis and secondary CoQ₁₀ deficiencies in selected patients (*n* = 26). The most significant variants detected in this cohort are summarized in Tables 2 and 3. All variants were detected as heterozygous, while only two patients (P78, P2) presented with compound heterozygous variants, reported here as disease-causing. The remaining variants could not contribute in resolving the etiology since autosomal recessive inheritance patterns require disease-causing variants on both alleles and not only affecting one allele. From this data is

Table 1

The mean, median and interquartile ranges calculated for patient muscle CoQ₁₀ levels expressed against CS activity and protein content.

RCD group	CoQ ₁₀ /UCS mean (nmol/UCS)	CoQ ₁₀ /UCS median (nmol/UCS)	CoQ ₁₀ /UCS Interquartile range (25th % - 75th %)	CoQ ₁₀ /protein mean (nmol/mg)	CoQ ₁₀ /protein median (nmol/mg)	CoQ ₁₀ /protein Interquartile range (25th % - 75th %)
CRC	4.133	3.606	2.902–5.048 nmol/U	0.642	0.547	0.405–0.864 nmol/mg
Other RCD	3.483	3.171	2.460–4.259 nmol/U	0.561	0.528	0.400–0.715 nmol/mg
CII + III defect	2.552	2.474	1.654–3.140 nmol/U	0.432	0.410	0.287–0.562 nmol/mg

CRC: clinically referred controls, *n* = 80; Other RCD: respiratory chain deficient patients excluding CII + III defects, *n* = 47; CII + III defect *n* = 29.

Table 2
2CoQ₁₀/UCS ratios and RCDs in patients identified with CoQ₁₀ deficiency.

Patient number	Ethnicity	Gender	CoQ ₁₀ /UCS concentration (nmol/UCS)	RCD diagnosed in patient	Sample group	CoQ ₁₀ genetic variant*		
						Gene	Variant	RefSNP
P2	African	Female	1.559	CII, CII + III	CII + III defect	COQ6**	c.41G > A (het) + c.859G > T (het)	rs17094161 rs61743884
P14	African	Female	1.638	CII + III	CII + III defect	ND		
P24	African	Female	1.204	CI, CII, CIII, CIV, CII + III	CII + III defect	COQ6	c.41G > A (het)	rs17094161
P29	African	Male	1.607	CII + III	CII + III defect	APTIX	c.597C > T (het)	rs150886026
P54	Caucasian	Male	1.535	–	CRC	COQ6	c.283G > A (het)	rs61743864
P78	Caucasian	Female	0.861	CI, CIII, CIV, CII + III	CRC	COQ9	c.304C > T (het)	rs143043228
P95	African	Female	1.654	CIV, CII + III	CRC	ND		
P98	African	Female	1.229	CII + III	CII + III defect	ETFDH**	c.1067G > A (het) + c.1448C > T (het)	novel
P99	Caucasian	Male	1.391	CIII, CII + III	CII + III defect	ND		
						COQ6	c.41G > A (het)	rs17094161
						ETFA	c.513G > A (het)	rs1801591
						COQ10A	c.523C > T (het)	novel

95% central reference interval for 600 × g muscle CoQ₁₀/UCS = 1.710–8.460 nmol/UCS. CRC = clinical referred control

* Possible disease-causing variants of genes involved in CoQ10 biosynthesis

** Protein deficiency confirmed. ND = none detected.

was clearly evident that more variants were found in the group where CoQ₁₀ was expressed against CS (6 of 9 cases) compared to CoQ₁₀ expressed against protein (4 of 11 cases). This warrants further investigation.

From the two patients that presented with disease-causing variants, the first patient (P78) presented with compound heterozygous variants (c.1067G > A and c.1448C > T) in the gene *ETFDH* as recently described by van der Westhuizen et al. (2018). This patient had the lowest CoQ₁₀ level when expressed against CS, but were not classified as deficient when CoQ₁₀ was expressed against protein content. This phenomenon was described before [13] for a patient found to be CoQ₁₀ deficient when expressed against CS, but not protein content. The investigators ascribed this discrepancy to the fact that the patient presented with massive rhabdomyolysis, leading to muscle protein depletion. In such a case the relatively low protein content in muscle could have led to a false negative result for CoQ₁₀ when normalized to protein. It is however disconcerting that P78 in our cohort would not have been diagnosed as CoQ₁₀ deficient if the CoQ₁₀ levels were only expressed against protein content like many studies do.

The second patient, P2, presented with two variants in the gene *COQ6*. Both variants, c.41G > A (p.Trp14Ter, LoF with high confidence) and c.859G > T (p.Ala49Ser) were previously reported and classified as benign and tolerated according to ClinVar and OMIM. These variants were however classified as benign from a single submitter, with no evidence of any functional tests done to support the classification. From ExAC, minor allele frequencies of 0.005 and 0.002 are reported respectively. However, to further investigate this patient, we performed functional tests on 600 × g muscle sample supernatant from this patient, including SDS page and Western Blot analyses. Results revealed lower relative total protein when compared to control samples, and lower expression of *COQ6* when compared to the controls (Fig. S3). Biochemically, this neonatal female was diagnosed with CoQ₁₀ deficiency on both CS and protein (Tables 2 and 3). We therefore classify these compound heterozygous variants as a risk factor for this patient, due to the supporting evidence mentioned. To fully understand the etiology of these compounding variants in the gene *COQ6*, we suggest segregation analysis on the non-consanguineous parents. All variants of interest were confirmed with Sanger sequencing and genetic

Table 3
2CoQ₁₀/protein concentrations and RCDs in patients identified with CoQ₁₀ deficiency.

Patient number	Ethnicity	Gender	CoQ ₁₀ /protein concentration (nmol/mg)	RCD diagnosed in patient	Sample group	CoQ ₁₀ genetic variant*		
						Gene	Variant	RefSNP
P2	African	Female	0.032	CII, CII + III	CII + III defect	COQ6**	c.41G > A (het) + c.859G > T (het)	rs17094161 rs61743884
P4	African	Female	0.199	CIII	Other RCD	ND		
P13	African	Female	0.103	–	CRC	ND		
P14	African	Female	0.132	CI, CII + III	CII + III defect	ND		
P24	African	Female	0.066	CI, CII, CIII, CIV, CII + III	CII + III defect	COQ6	c.41G > A (het)	rs17094161
						APTIX	c.597C > T (het)	rs150886026
P29	African	Male	0.176	CII + III	CII + III defect	COQ6	c.283G > A (het)	rs61743864
						COQ9	c.304C > T (het)	rs143043228
P37	African	Male	0.211	CIII, CIV, CII + III	CII + III defect	ND		
P47	African	Female	0.111	CI	Other RCD	ND		
P69	African	Male	0.157	CIII, CIV, CII + III	CII + III defect	ND		
P72	African	Male	0.028	CII + III	CII + III defect	Not performed (insufficient sample)		
P107	African	Female	0.210	–	CRC	ND		
P158	African	Male	0.212	–	CRC	APOE	c.824 T > C (het)	Novel

95% central reference interval for 600 × g muscle CoQ₁₀/protein = 0.213–1.387 nmol/mg. CRC = clinical referred control

* Possible disease-causing variants of genes involved in CoQ10 biosynthesis

** Protein deficiency confirmed. ND = none detected.

variations of importance could not be identified in the remaining 24 patients. Since only two patients presented with disease-causing variants, future studies might focus on distinguishing primary and secondary CoQ₁₀ deficiencies with the incorporation of radiolabeled substrates into CoQ₁₀, such as [³H]-mevalonate and [¹⁴C]-4-hydroxybenzoate, or that of stable isotopes in fibroblasts [24].

4. Conclusion

Although CoQ₁₀ plays a central role in energy metabolism, the prevalence of CoQ₁₀ deficiency is still unknown and the condition is at present under-diagnosed [17]. Here we report on 600 ×g muscle supernatant CoQ₁₀ levels in a cohort of ethnically diverse patients who received muscle biopsies to diagnose a possible RCD. Results indicate that muscle CoQ₁₀ was not significantly influenced by ethnicity, gender and age. Nine patients were identified with decreased muscle CoQ₁₀/UCS, eight of these patients were also CII+III deficient. A strong correlation was detected between CoQ₁₀ levels and CII+III activity, supporting the notion that CII+III deficiencies are a good indication of a CoQ₁₀ deficiency ([11,12,22], Yubero et al. 2016). However, when CoQ₁₀ was normalized to protein content, 12 patients were identified to be CoQ₁₀ deficient. Only four patients were CoQ₁₀ deficient when normalized against CS and protein, thus five patients were deficient when normalized against CS only and not when normalized against protein. Even worse – eight patients were only CoQ₁₀ deficient when normalized to protein and not CS. This anomaly remains to be further investigated, with variation in adaptive responses such as mitochondrial biogenesis and mitophagy in muscle possible having an impact on the outcome of expressing CoQ₁₀ levels.

This study is, as far as we know, the first to report a central 95% reference interval for 600 ×g muscle supernatants prepared from frozen samples in an ethnically diverse cohort. The study reiterates the importance of including CoQ₁₀ quantification as part of a diagnostic approach to study mitochondrial disease as it may complement respiratory chain enzyme assays with the possible identification of patients that may benefit from CoQ₁₀ supplementation. We conclude from our data that, to prevent possibly not diagnosing a potential CoQ₁₀ deficiency, the expression of CoQ₁₀ levels in muscle on both CS as well as protein content should be considered.

Funding

Funding was obtained from the South African Medical Research Council.

Conflict of interest

All authors declare that they have no conflict of interest.

Human and animal rights and informed consent

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from the patients or the parents of the minor patients for being included in the study.

Author contributions

Roan Louw: conception and design; drafting the article.
 Izelle Smuts: conception; revising the article.
 Kimmey-Li Wilsenach: analysis and interpretation of data; revising the article.
 Lindi-Maryn Jonck: analysis and interpretation of data; revising the article.

Maryke Schoonen: analysis and interpretation of data; revising the article.

Francois H van der Westhuizen: conception; revising the article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymgme.2018.02.015>.

References

- [1] D.A. Applegarth, J.R. Toone, R.B. Lowry, Incidence of inborn errors of metabolism in British Columbia, 1969–1996, *Pediatrics* 105 (2000) e10.
- [2] N. Darin, A. Oldfors, A.R. Moslemi, E. Holme, M. Tulinius, The incidence of mitochondrial encephalomyopathies in childhood: clinical features and morphological, biochemical, and DNA abnormalities, *Ann. Neurol.* 49 (2001) 377–383.
- [3] V. Emanuele, L.C. López, A. Berardo, A. Naini, S. Tadesse, B. Wen, E. D'Agostino, M. Solomon, S. DiMauro, C. Quinzii, M. Hirano, Heterogeneity of coenzyme Q10 deficiency: patient study and literature review, *Arch. Neurol.* 69 (2012) 978–983.
- [4] L. Ernster, G. Dallner, Biochemical, physiological and medical aspects of ubiquinone function, *Biochim. Biophys. Acta* 1271 (1995) 195–204.
- [5] G.S. Gorman, A.M. Schaefer, Y. Ng, N. Gomez, E.L. Blakely, C.L. Alston, C. Feeney, R. Horvath, P. Yu-Wai-Man, P.F. Chinnery, R.W. Taylor, D.M. Turnbull, R. McFarland, Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease, *Ann. Neurol.* 77 (2015) 753–759.
- [6] I. Hargreaves, Coenzyme Q₁₀ as a therapy for mitochondrial disease, *Int. J. Biochem. Cell Biol.* 49 (2014) 105–111.
- [7] O. Itkonen, A. Suomalainen, U. Turpeinen, Mitochondrial coenzyme Q10 determination by isotope-dilution liquid chromatography-tandem mass spectrometry, *Clin. Chem.* 59 (2013) 1260–1267.
- [8] M. Lek, K.J. Karczewski, E.V. Minikel, K.E. Samocha, E. Banks, T. Fenell, et al., Analysis of protein-coding genetic variation in 60,706 humans, *Nature* 536 (2016) 285–291.
- [9] G. Lenaz, R. Fato, G. Formigini, M.L. Genova, The role of coenzyme Q in mitochondrial electron transport, *Mitochondrion* 7 (2007) S8–S33.
- [10] W. McLaren, L. Gil, S.E. Hunt, H.S. Riat, G.R.S. Ritchie, A. Thormann, P. Flicek, F. Cunningham, The Ensembl variant effect predictor, *Genome Biol.* 17 (2016) 122.
- [11] M.V. Miles, L. Miles, P.H. Tang, P.S. Horn, P.E. Steele, A.J. DeGrauw, B.L. Wong, K.E. Bove, Systematic evaluation of muscle coenzyme Q10 content in children with mitochondrial respiratory chain enzyme deficiencies, *Mitochondrion* 8 (2008) 170–180.
- [12] R. Montero, R. Artuch, P. Briones, A. Nascimento, A. García-Cazorla, M.A. Vilaseca, J.A. Sánchez-Alcázar, P. Navas, J. Montoya, M. Pineda, Muscle coenzyme Q10 concentrations in patients with probable and definite diagnosis of respiratory chain disorders, *Biofactors* 25 (2005) 109–115.
- [13] R. Montero, J.A. Sánchez-Alcázar, P. Briones, A.R. Hernández, M.D. Cordero, E. Trevisson, L. Salviati, M. Pineda, A. García-Cazorla, P. Navas, R. Artuch, Analysis of coenzyme Q₁₀ in muscle and fibroblasts for the diagnosis of CoQ₁₀ deficiency syndromes, *Clin. Biochem.* 41 (2008) 697–700.
- [14] R. Montero, J.A. Sánchez-Alcázar, P. Briones, A. Navarro-Sastre, E. Gallardo, B. Bornstein, D. Herrero-Martín, H. Rivera, M.A. Martín, R. Martí, A. García-Cazorla, J. Montoya, P. Navas, R. Artuch, Coenzyme Q₁₀ deficiency associated with a mitochondrial DNA depletion syndrome: a case report, *Clin. Biochem.* 42 (2009) 742–745.
- [15] U. Paila, B.A. Chapman, R. Kirchner, A.R. Quinlan, GEMINI: integrative exploration of genetic variation and genome annotations, *PLoS Comput. Biol.* 9 (2013) e1003153.
- [16] C.M. Quinzii, L. López, A. Naini, S. DiMauro, M. Hirano, Human CoQ10 deficiencies, *Biofactors* 32 (2008) 113–118.
- [17] S. Rahman, C.F. Clarke, M. Hirano, 176th ENMC international workshop: diagnosis and treatment of coenzyme Q₁₀ deficiency, *Neuromuscul. Disord.* 22 (2012) 76–86.
- [18] M. Saraste, Oxidative phosphorylation at the fin de siècle, *Science* 283 (1999) 1488–1493.
- [19] D. Skladal, J. Halliday, D.R. Thorburn, Minimum birth prevalence of mitochondrial respiratory chain disorders in children, *Brain* 126 (2003) 1905–1912.
- [20] P.K. Smith, R.I. Krohn, G.T. Hermanson, A.K. Mallia, F.H. Gartner, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson, D.C. Klenk, Measurement of protein using bicinchoninic acid, *Anal. Biochem.* 150 (1985) 76–85.
- [21] I. Smuts, R. Louw, H. Du Toit, B. Klopfer, L.J. Mienie, F.H. Van der Westhuizen, An overview of a cohort of South African patients with mitochondrial disorders, *J. Inher. Metab. Dis.* 33 (2010) 95–104.
- [22] E. Trevisson, S. DiMauro, P. Navas, L. Salviati, Coenzyme Q deficiency in muscle, *Curr. Opin. Neurol.* 24 (2011) 449–456.
- [23] N.I. Wolf, J.A.M. Smeitink, Mitochondrial disorders: a proposal for consensus diagnostic criteria in infants and children, *Neurology* 59 (2002) 1402–1405.
- [24] D. Yubero, G. Allen, R. Artuch, R. Montero, The value of coenzyme Q10 determination in mitochondrial patients, *J. Clin. Med.* 6 (2017) 37.
- [25] S. Zeng, J. Yang, B.H. Chung, Y.L. Lau, W. Yang, EFIN: predicting the functional impact of nonsynonymous single nucleotide polymorphisms in human genome, *BMC Genomics* 15 (2014) 455.

APPENDIX B: MITOCHONDRIAL COHORT CONSENT AND ASSENT FORM

Blank versions of legal informed consent and assent forms for collection of a muscle biopsy sample from a participant.

Appendix B.1.....238

Form A

Patients: Informed consent form

Appendix B.2.....240

Form B

Patient: Assent form

Appendix B.1

PATIENTS: INFORMED CONSENT FORM

The investigation of patients with possible mitochondrial disorders in South Africa

AIM OF THE STUDY

Mitochondrial disorders (MDs) are a group of disorders that may affect the brain, muscle, liver and several other organs, because a defect exists in the energy production of the cell. It is caused by a genetic defect (primary MD) or as a consequence of medication or other reason e.g. Human immunodeficiency virus (HIV) (secondary MD). The aim of the project is to measure respiratory function in your child's muscle and/or relevant tissues in conjunction with the enzyme activity of the energy producing system as your child is identified as having a possible primary or secondary mitochondrial disorder. In confirmed cases further genetic testing for mitochondrial disorders will be performed in different stages depending of the availability of funds. The genetic tests will include different methods to identify possible mtDNA and nDNA changes that may explain the reason for the disease. The urine will also be examined for chemical compounds that may be excreted in the urine if such a disorder is present. Biochemical testing on blood specimens will be performed according to the clinical features of the patient.

No HIV testing will be performed for research purposes only. HIV testing of a child in the department of Paediatrics is only performed if it is clinically indicated and will impact on the management of the individual. The HIV patients or patients exposed to ARV in utero that will be included in the study will be recruited mainly from the HIV clinic where they have been followed up and managed routinely according to the current National HIV Management Guidelines.

The long-term goal of this project is to refine the diagnostic service of mitochondrial disorders in South Africa.

EXPECTED DURATION OF THE PROJECT

This project is a long term project and patients have been included since 2007. Blood/Tissue/Urine will be kept to avoid any additional invasive procedures for possible diagnostic testing in future when novel techniques might become available, provided that it will be discussed with the patients/parents first. It will be kept until the Principal Investigator Prof I Smuts will finally close or terminate the study. The specimens will then be discarded according to the regulations on the management of human waste.

PROCEDURES FOLLOWED DURING THIS PROJECT

The study will be explained to you and you will be allowed time to ask questions and decide whether you would prefer to give consent for your child to take part.

A family history will be taken and a complete physical examination will be performed.

In the case of a patient with suspected mitochondrial disorder appropriate special investigations, including urine, and blood will be performed to confirm the diagnosis of a mitochondrial disorder, to determine involvement of other organ systems and to monitor the course of the disease.

Not more than 15 ml of blood will be taken, depending on the age of the child. This blood will be used for appropriate biochemical and haematological investigations and a small volume will be stored away to do genetic studies when an appropriate test comes available. If it would be applicable in future, you will be contacted again for a follow up blood sample to establish cell lines to ensure a continuous supply of DNA for further analyses.

In some instances where the nervous system symptoms are unexplained it will be necessary to do a lumbar puncture. It implies that fluid surrounding the brain will be obtained through a thin needle in the back. A maximum of 5 ml cerebrospinal fluid will be needed.

Nerve conduction studies and electromyography in which the electrical functioning of the nerves and the muscle will only be done if it has not been done before and it seems to be affected.

A muscle and/or relevant tissue biopsy will be done in some cases if the blood tests were inconclusive. The biopsies will be taken under general anaesthetics by a qualified paediatric surgeon. For a muscle biopsy a 3-4 cm incision will be made on the thigh and a small piece of muscle about 1 cubic cm will be removed for the biochemical analyses and an additional smaller piece of muscle will be taken for histology. The wound will be sutured. The muscle biopsy taken for biochemical and genetic analyses will be processed and the excess will be frozen for possible future use if better biochemical or genetic tests come available. The genetic testing will be performed in South Africa.

A liver biopsy, if indicated, will be taken at the same time while under general anaesthesia by a qualified paediatric surgeon or paediatric gastroenterologist. There is always a small risk of bleeding after the procedure. Clotting ability of the blood will be tested prior to the procedure. The liver biopsy will be done with sonar guidance to ensure that large blood vessels in the liver will not be damaged. A thin needle is used to remove a 1-2 cm long piece of liver, 2-3 mm thick through a small cut in the skin. Biochemical, histological and genetic analyses will be performed on the tissue obtained. No suturing is required. After the biopsy is taken adequate pain relief will be provided and vigilant monitoring of the heart rate, blood pressure and saturation will be done to identify complications like bleeding should they occur.

A skin biopsy of 3 mm of the superficial layer of the skin will be taken with a sterile blade from an area on the arm treated with a local anaesthetic to culture cells the determination of biochemical activity and genetic analyses.

A specialised type of X-Ray, e.g. computerised tomography (CT) scan or magnetic resonance imaging (MRI) of the brain will be done to obtain a photographic picture only if it has not been done before and your doctor is of opinion it might be helpful to clarify some of the symptoms and signs.

Appendix B.1

A formal hearing and/or vision test will be performed if deafness or visual impairment is suspected.

RISKS AND POSSIBLE DISCOMFORT FOR THE PATIENT / INDIVIDUAL

Blood sampling is causing mild pain, but the risks are minimal.

A lumbar puncture will cause moderate pain and discoloration at the puncture site. The patient may suffer from a headache afterwards.

The EMG and nerve conduction study may cause pain, but the greatest care will be taken to assure the comfort of the patient as far as possible.

The muscle, liver or skin biopsies have the following risk factors: possible bleeding, discomfort including pain, possible scarring, and a remote risk of infection.

DECLARATION OF CONFIDENTIALITY

All the Information provided for the purpose of this study will be treated as highly confidential. Only individuals of the research group and officials, who are entitled to it by law, will have access to information. Data published in a medical journal will include no information that could identify a patient or his/her family.

WITHDRAWAL CLAUSE

I understand that I am entitled to withdraw from the project at any time without penalty or loss of benefits. The participation of my child, in this project is therefore on a voluntary basis until I request withdrawal in writing.

THE ABOVE MENTIONED PROJECT WAS THOROUGHLY EXPLAINED AND THE FOLLOWING ADDITIONAL INFORMATION POINTED OUT TO ME

- That the entire urine, blood and muscle biopsy and other tissue samples, including genetic material derived from it, is used for research purposes and no compensation for this will be provided.
- That I will be kept informed of significant developments in the project.
- That the withdrawal clause was explained to me and that I understand its implications.

ENQUIRIES ABOUT THIS PROJECT SHOULD BE DIRECTED TO:

- General enquiries should be directed to: Prof I. Smuts, Paediatric Neurology Unit, Department of Paediatrics and Child Health, University of Pretoria, Private Bag X323, Arcadia, 0007.
- Enquiries regarding ethical aspects of the project can be addressed to: The Chairperson of the Ethics Committee, Steve Biko Academic Hospital, Private Bag X323, Arcadia, 0007.

That I will receive a copy of this informed consent form.

DECLARATION OF CONSENT

I, _____
(Print full name and surname)

Hereby consent to the participation of my child,

_____ in the above-mentioned project.
(Print full name and surname or stick a sticker)

I confirm that I am fully aware of the content of this form and that by signing it I give the necessary consent.

Signed at: _____ on _____
(Place) (Date)

(Signature of a parent/guardian)

Witness 1: _____
(Signature)

Witness 2: _____
(Signature)

(Signature of responsible who explained the project and informed consent from to the participant.)

Appendix B.2

PATIENT: ASSENT FORM

The investigation of patients with possible mitochondrial disorders in South Africa

We are going to explain the study to you. If you understand and want to take part, we ask you to sign the form.

WHAT DO WE WANT TO DO IN THE STUDY?

Mitochondrial disorders (MDs) are diseases that may cause a problem in many parts of the body, because too little energy is made. You can be born with it or it may be caused by other things like medication or other infections like HIV. HIV will only be tested if we think it is possible that you can benefit from medication. The aim of the project is to measure if your body makes enough energy and what is the cause. We hope to be able to have a centre one day in South Africa where patients with these problems could be helped.

HOW LONG WILL THE STUDY BE?

This project runs since 2007 and will continue until Prof Smuts will finally close the program. The blood and tissue will be kept to try and make a diagnosis in future, when better tests are available, but we will discuss it with you first. Everything will be discarded if the study is finished

WHAT ARE WE GOING TO DO?

We will explain to you what we are going to do and give you time to ask your questions.

We are going to ask your parents about other people in your family and whether they are ill. We will examine your body.

If we think that you have a this possible problem, we will do some test to see if there are other parts of your body that is also ill and to see if the problem gets worse. It can be sore if we do some of the test, but then we will give you medicine. It will help you then to sleep and lie very still. It will also help to take away the pain.

Less than 15 ml blood will be taken. It may be necessary to take some of the fluid from your back. Afterwards you may have a sore head, and then you will have to lie in bed for a while. If your muscle or nerves are weak we are going to test it with a machine. It is sore for a short while. If we really do not get an answer from those tests it may be necessary to do a small operation to get a small piece of muscle, skin and/or liver. The little operation may cause bleeding, pain and leave a small mark on your e.g. Your urine will also sent to be tested in a big machine, but that is not sore. The genetic tests will be done in South Africa

Sometimes we need a picture of your brain. That is not painful at all, you just have to lie very quiet in the machine that we call a scanner. It is like taking photos of your brain.

If you cannot see or hear well, we may also test your eyes and or ears. It is not sore.

WILL ANYBODY KNOW ABOUT YOU?

Only the doctors that know you and the people in the laboratory that work with your blood, urine or muscle see the answers and your hospital number and name. If we find some answers we would like to write about it in a medical journal, but nobody that will read the article will be able to know that you took part, only the answers will be given.

WHAT HAPPENS IF YOU DO NOT WANT TO TAKE PART?

You or your parent may at any time tell us if you do not want to take part any further at any time, but your parent must please write us a letter. We will not be cross with you and we shall still carry on to treat you as good as possible.

THE ABOVE MENTIONED PROJECT WAS THOROUGHLY EXPLAINED AND THE FOLLOWING EXTRA POINTS WERE TOLD TO ME.

The blood, urine muscle or other tissue are given for research and that I will not be paid for it.

You will let us know if you find the answers.

I may decide not to take part any further in the study

Appendix B.2

WHO SHOULD WE CONTACT IF WE HAVE QUESTIONS?

- General enquiries should be directed to: Prof I. Smuts, Paediatric Neurology Unit, Department of Paediatrics and Child Health, University of Pretoria Private Bag X323, Arcadia, 0007
- Enquiries regarding ethical aspects of the project can be addressed to: The Chairperson of the Ethics Committee, Steve Biko Academic Hospital, Private Bag X323, Arcadia, 0007.

That I will receive a copy of assent form.

DECLARATION OF ASSENT

I, _____
(Print full name and surname)

Hereby assent to take part in the project.

I understand that I take part in the research project as a **patient** .
I confirm that I understand everything.

Signed at: _____ on _____.
(Place) (Date)

(Signature of a minor)

Witness 1: _____ **Witness 2:** _____
(Signature) (Signature)

(Signature of responsible who explained the project and assent to the participant.)