Western Norway University of Applied Sciences

BACHELOR'S THESIS

Whole genome sequencing in individuals with suspected MODY from a population of mixed ancestry from Cape Town, South Africa.

Helgenomsekvensering i individer med mistanke om MODY fra en populasjon med blandede aner fra Cape Town, Sør-Afrika.

Aurora A. Løvaas, Tomas B. Hersdal, Victoria A. M. Markaskard
BIO160
Western Norway University of Applied Science
Cape Peninsula University of Technology, South Africa

Host institution mentor: Shanel Raghubeer and Tandi Matsha-Erasmus
Home institution mentor: Lise Bjørkhaug Gundersen and Ingvild Aukrust

29th of May 2020

Cape Peninsula University of Technology

ABSTRACT

Pathogenic variants in the genes *HNF1A* and *GCK* (encoding hepatocyte nuclear factor 1 alpha and glucokinase, respectively) are associated with a dominant inherited subgroup of monogenic diabetes; Maturity-Onset Diabetes of the Young (MODY). Some patients with MODY are misdiagnosed as type 1 or type 2 diabetes, because of overlapping clinical features and lack of knowledge about MODY disease.

In this study, *HNF1A* and *GCK* genes were analyzed in 15 individuals with suspected MODY from a mixed ancestry population from Cape Town, based on whole genome sequencing data. Protein coding variants that had a MAF <0.01% (more likely to be pathogenic and be the cause of rare diagnosis like MODY) and the most common missense variants was further examined in population frequency databases such as gnomAD and clinical variant databases such as ClinVar. Accordingly, relevant literature from PubMed concerning the rare variants was investigated, to support the suspicion of variants being pathogenic and cause/risk for MODY/other diabetes, in affected individuals. Two suspected variants (MAF<0.1%) in *HNF1A* was found: (c.1124delG; p.(Gly375AlafaTer9) and c.1243G>A; p.(Gly415Arg)). p.(Gly375AlafaTer9) was found in two individuals with impaired glucose tolerance and normoglycemic, respectively, and p.(Gly415Arg) was found in one individual with impaired glucose tolerance. Based on this and other clinical features, these variants seems not a cause of MODY, however might be risk factor for type 2 diabetes (p.(Gly415Arg)). As only rare variants in MODY genes are known to be cause of early-onset dominantly inherited diabetes, the probability of finding rare variants in few individuals, in only two MODY genes, was very low. To determine the prevalence of MODY in such a population cohort, more individuals with certain early-onset diabetes should be screened.

Key words: *HNF1A* gene, *GCK* gene, MODY, type 2 diabetes, MAF

SAMMENDRAG

Patogene varianter i genene *HNF1A* og *GCK* (koder for hepatocytt nukleær faktor 1 alfa og glucokinase, henholdsvis) er assosiert med en dominant arvelig undergruppe av monogen diabetes; Maturity-Onset Diabetes of the Young (MODY). Flere pasienter med diagnosen MODY blir feildiagnostisert med type 1 eller type 2 diabetes, på grunn av overlappende kliniske indikasjoner og manglende kunnskap rundt sykdommen MODY.

I dette studiet ble *HNF1A* og *GCK* gener analysert i 15 individer med mistanke om MODY, fra en populasjon med blandede aner fra Cape Town, basert på helgenomsekvenseringsdata. Proteinkodende varianter som hadde en MAF <0,01% (større sannsynlighet for å være patogene og årsak til sjeldne diagnoser som MODY) og de mest forekommende missense varianter ble videre undersøkt i populasjonsfrekvensdatabaser som gnomAD og kliniske variantdatabaser som ClinVar. Tilsvarende ble relevant litteratur i databasen PubMed om de sjeldne variantene undersøkt, som støtte til mistanken om at disse variantene var patogene og kunne føre til/øke risiko for MODY/annen diabetessykdom i rammede individer. To suspekte varianter (MAF>0,1%) i *HNF1A* ble funnet: (c.1124delG; p.(Gly375AlafaTer9) og c.1243G>A; p.(Gly415Arg)). p.(Gly375AlafaTer9) ble funnet i to individer med nedsatt glukose toleranse og normoglykemisk, og p.(Gly415Arg) ble funnet i ett individ med nedsatt glukose toleranse. Basert på dette og andre kliniske opplysninger ser det ikke ut til at disse variantene forårsaker MODY, men kan være risikofaktorer for type 2 diabetes (p.(Gly415Arg)). Ettersom kun sjeldne varianter i MODY gener er kjent for å forårsake tidlig dominant arvelig diabetes, var sannsynligheten for å finne sjeldne varianter hos få individer, i kun to MODY gener, meget lav. For å fastslå forekomsten av MODY i en slik populasjonsgruppe, bør flere personer med tidlig debut av diabetes screenes.

Nøkkelord: *HNF1A*-gen, *GCK*-gen, MODY, type 2 diabetes, MAF

PREFACE

Our bachelor thesis was intended carried out in the laboratory of Prof. Tandi Matsha, at the Department of Biomedical Sciences, Cape Peninsula University of Technology. However, due to the Covid-19 pandemic we completed the study in Norway. The duration of the project was from the 11th of March to the 29th of May.

Throughout this period, we have learned to work independently, as well as how to collaborate on a research project with an international research group from Cape Peninsula University of Technology, South Africa.

Our sincere thanks go to our external supervisors Prof. Shanel Raghubeer and Tandi Matsha-Erasmus, for giving us the opportunity to be involved in and to complete the study, although we had to do it in Norway. We are sad that this study could not be performed on campus in Cape Town, and that we never got the chance to meet, but we are glad you trusted us in analyzing the data you retrieved by performing the whole genome sequencing.

Additionally, we sincerely thank our main supervisor in Norway Prof. Lise Bjørkhaug Gundersen and co-supervisor Dr. Ingvild Aukrust for answering all our questions in the best way, motivating, and guiding us during this project. We would not have made it without your help.

Norway, 29th of May 2020

_____	_____	_____
Aurora A. Løvaas	Tomas B. Hersdal	Victoria A. M. Markaskard

TABLE OF CONTENT

ABBREVATIONS

GCK	Glucokinase (gene)
GnomAD	Genome Aggregation Database
HNF1A	Hepatocyte Nuclear Factor-1 Alpha (gene)
IFG	Impaired fasting glucose
IGT	Impaired glucose tolerance
MAF	Minor Allele Frequency
MODY	Maturity-onset Diabetes of the Young
NGS	Next Generation Sequencing
OGTT	Oral Glucose Tolerance Test
SNV	Single Nucleotide Variant
T1D	Type 1 diabetes
T2D	Type 2 diabetes
VUS	Variant of Unknown Significance
WGS	Whole Genome Sequencing
WHO	World Health Organization
WT	Wild-type

1. INTRODUCTION

1.1 Diabetes Mellitus

Diabetes mellitus (DM) is a metabolic disorder characterized by chronic hyperglycemia due to defects in insulin secretion and/or insulin action. This occurs when the pancreas does not produce enough insulin, or the body does not effectively use the insulin it produces (World Health Organization, 2018). Insulin is the primary hormone responsible for regulating levels of glucose in plasma by stimulating entry of glucose into the cell and by increasing glycogenesis, lipogenesis and glycolysis (Bishop , Fody, & Schoeff, 2018). DM is a multisystem disease that carries significant morbidity and mortality from its chronic macrovascular and microvascular complications (Rees, Levy, & Lansdown, 2017).

A diabetes diagnosis is given if the plasma glucose is ≥7.0 mmol/L after fasting, or ≥11.1 mmol/L two hours after an oral glucose tolerance test (OGTT) (see Table 1). In addition, HbA1c can be measured to confirm the diagnosis of diabetes, using a technique that is certified by National Glycohemoglobin Standardization Program and standardized to the Diabetes Control and Complications Trial Assay (WHO and International Diabetes Federation, 2006). Diabetes is indicated with a HbA1c level ≥47 mmol/mol (WHO and International Diabetes Federation, 2006). However, measurement of HbA1c is not common practice in low-income countries.

According to the World Health Organization (WHO) 422 million adults worldwide have diabetes. The steadily increase for the past three decades are correlated to the increased prevalence of overweight and obesity among people. According to WHO 1.6 million people die every year directly attributed to diabetes.

According to the WHO, prediabetes is a state of intermediate hyperglycemia using two different specific parameters: impaired fasting glucose (IFG) and impaired glucose tolerance (IGT) (WHO and International Diabetes Federation, 2006). IFG is defined as abnormal fasting plasma glucose and occurs when the liver releases too much glucose into the bloodstream overnight (Lin, et al., 2007). IGT is defined as normal fasting plasma glucose, but an abnormal two-hour plasma glucose after an OGTT, and occurs when the insulin does not work like it should, if there is not released adequate insulin too meet the demand, or a combination of both (Lin, et al., 2007; WHO and International Diabetes Federation, 2006). Until recently, this type of diabetes was seen only in adults, but it is now also occurring increasingly frequently in children (World Health Organization, 2018).

Table 1: Diagnostic criteria and classification of diabetes and intermediate hyperglycemia. The reference values are given by WHOs Technical Report relating to diabetes from 1999 (World Health Organization, 1999).

		REFERENCE VALUES	
		Fasting glucose plasma (mmol/L)	**2hr. OGTT (mmol/L)**
STATE	**Normoglycemia**	< 6.1	< 7.8 implied
	Prediabetes (IGT)	< 7.0	and ≥ 7.8 - < 11.1
	Prediabetes (IFG)	≥ 6.1 - < 7.0	and < 7.8
	Diabetes Mellitus	≥ 7.0	≥11.1

1.2 Classification of diabetes mellitus

1.2.1 Type 1 diabetes

Type 1 diabetes (T1D), previously known as insulin-dependent or juvenile diabetes, is characterized by inadequate insulin production and requires daily administration of insulin (World Health Organization, 2018). T1D is caused by complete lack of insulin secretion due to autoimmune destruction of pancreatic beta-cells (Atkinson & Eisenbarth, 2001). Multiple predisposing genetic and environmental factors are thought to be the cause of T1D, however, the complete basis for T1D disease development is still not completely understood (Rees, Levy, & Lansdown, 2017). Genetic studies have made it possible to explain 80% of the genetic architecture of T1D (Stankov, Benc, & Draskovic, 2013). Familial clustering has a prevalence of 6% in siblings, however, more than 85% of the disease cases do not have a family history (Risch, 1987; Szablewski, 2011). There is also a difference in which parent who has the disease. If the mother has diabetes, the risk increases to 1-2% and 3-7% if the father has diabetes (Haller, Atkinson & Schatz, 2005; Warram, Krolewski & Kahn, 1988).

Disease susceptibility is highly associated to the Human Leukocyte Antigen (HLA) alleles DR3 and DR4, as well as the associated alleles DQ2 and DQ8 (Szablewski, 2011). More than 9% of patients with T1D express either DR3DQ2 or DR4DQ8, and patients with haplotypes of the HLA can be highly susceptible to T1D.

1.2.2 Type 2 diabetes

Type 2 diabetes (T2D), formerly called non-insulin-dependent diabetes, accounts for 85-90% of diabetes cases world-wide and is mainly a consequence of obesity and inactivity (poor lifestyle) (World Health Organization, 2018). Unlike the early onset of T1D, T2D usually onset after age 40 years, and is characterized by hyperglycemia caused by disruptions of the body's use of insulin and of disproportionately high glucagon production; both during fasting and after meals (Rodger, 1991). The hormone glucagon further increases the blood sugar by stimulating the liver to release glucose into the blood stream. T2D is also characterized by insulin resistance. This form of diabetes can be present and undiagnosed for a long period without clinical symptoms because the hyperglycemia develops gradually and change over time (American Diabetes Association, 2011).

T2D is a heritable and complex disease, and a genetic component is an important contributing factor and estimated to explain 30-70% of T2D risk (Florez & Walford, 2010). Multiple genes and different combination of genes play a role in the disease development in different individuals, because of its polygenic and heterogeneous character.

1.2.3 Gestational diabetes mellitus

Gestational diabetes mellitus (GDM) is known as glucose intolerance with onset or first occurrence during pregnancy (Expert Committee on the Diagnosis and Classification of Diabetes Mellitus, 2003; Metzger & Coustan, 1998). According to McIntyre, et al, GDM is currently the most common medical complication of pregnancy, and prevalence of undiagnosed hyperglycemia and even overt diabetes in young women is increasing (McIntyre, et al., 2019). Major GDM risk factors are maternal overweight and obesity, later age at childbearing, previous history of GDM, family history of T2D and ethnicity. Diagnosis is usually performed using an OGTT. The primary treatments for GDM are diet and physical activity, but some individuals also require pharmacotherapy, in form of insulin.

1.2.4 Monogenic forms of diabetes and maturity-onset diabetes of the young

Monogenic diabetes is caused by defects in one single gene (National Institute of Diabetes and Digestive and Kidney Diseases, 2017). One of the most common forms of monogenic diabetes is Maturity-Onset Diabetes of the Young (MODY). MODY was first recognized by Tattersall in 1974 and is a dominantly inherited form of non-insulin-dependent diabetes that is typically diagnosed before the age of 25-30 (Tattersall, 1974). Today, MODY is referred to as a heterogenous group of monogenic disorders that result in a decrease in insulin secretion due to pancreatic beta-cell dysfunction (Gardner & Tai, 2012). It is also characterized by an autosomal dominant pattern of inheritance, a family history of diabetes, and absence of beta-cell autoimmunity and insulin resistance. Although rare (represents 1-3% of all diabetes worldwide), MODY is a serious disease which lasts through a person's lifetime, with an increased risk of heart disease, kidney failure and blindness. To date, according to the Online Mendelian Inheritance in Man (OMIM) (OMIM, 2016), 14 different subtypes of MODY exists (MODY1-14) (Table 2), each caused by a pathogenic variant in one of 14 individual genes (*HNF4A, GCK, HNF1A, PDX1, HNF1B, NEUROD1, KLF11, CEL, PAX4, INS, BLK, ABCC8, KCNJ11, APPL1*). Some genes are more frequently affected (*HNF4A, GCK, HNF1A, HNF1B)*, than others, and resulting in MODY1, MODY2, MODY3 and MODY5, respectively (Amed & Oram, 2016).

A correct diagnosis of MODY is important for determining treatment, as patients with specific subtypes of MODY (MODY1 and MODY3 caused by pathogenic *HNF4A* and *HNF1A* variants) can benefit from treatment with sulphonylureas (avoiding insulin injections) (Sheperd, et al., 2009). In addition, patients with MODY2 (milder form of MODY) can often benefit from a modified diet and in most cases does not need any further treatment other than this. Thus, molecular genetic testing is important for correct diabetes diagnosis and targeted treatment (Shields, et al., 2010).

MODY is challenging to diagnose for several reasons, namely, the low prevalence in a population, shared symptoms with DM, and limited awareness of MODY (Thanabalasingham & Owen, 2011). Further, genetic testing is not common practice in many countries, particularly low-middle income countries. For instance, MODY2 (caused by pathogenic *GCK* gene variants) is more commonly diagnosed in countries where glucose testing is more frequent, such as France, Spain, and Italy (Thanabalasingham & Owen, 2011). As most genetic studies have been conducted in Caucasian populations, it is important to determine the specific prevalence and diagnose MODY in populations of varied ethnicity (Shields, et al., 2012).

Table 2: MODY associated diabetes and its clinical features. Information collected from (OMIM, 2016).

MODY	**Gene**	**OMIM link no.**	**Clinical features**
MODY1	*HNF4A*	125850	Rare (<3-5%). Typically nonobese; Macrosomia; Hyperinsulinemia; Progressive decrease in insulin secretion.
MODY2	*GCK*	125851	Common (30-50%). Mild, long-lasting, stable fasting hyperglycemia; GDM; Usually asymptomatic.
MODY3	*HNF1A*	600496	Common (30-50%). Typically nonobese; Progressive decrease in insulin secretion; Renal glycosuria; Raised cardiovascular risk.
MODY4	*PDXI*	609392	Rare (<1%). Hyperinsulinemia; Pancreatic agenesis.
MODY5	*HNF1B*	137920	Rare (<5%). Progressive decrease in insulin secretion; Renal abnormalities.
MODY6	*NEUROD1*	606394	Very rare (<1%). Adult onset; Overweight.
MODY7	*KLF11*	610508	Very rare (<1%). Phenotype similar to T2D.
MODY8	*CEL*	609812	Very rare (<1%). Pancreatic exocrine dysfunction.
MODY9	*PAX4*	612225	Very rare (<1%). Associated with ketoacidosis in some cases.
MODY10	*INS*	613370	Very rare (<1%). < 20 years at onset.
MODY11	*BLK*	613375	Very rare (<1%). Increased BMI (Body Mass Index) and overweight in some cases.
MODY12	*ABCC8*	600509	Very rare (<1%). Hyperinsulinemic hypoglycemia; Noninsulin-dependent.
MODY13	*KCNJ11*	616329	Very rare (<1%). Phenotype similar to MODY1 and MODY3.
MODY14	*APPL1*	616511	Very rare (<1%). Increased BMI and weight in some cases.

1.2.5 MODY2

Heterozygous inactivating mutations in the *GCK* gene, which results in MODY2, is the second most common cause of MODY (Gardner & Tai, 2012). MODY2 is characterized by mild, stable fasting hyperglycemia and treatment is generally not required. Clinical features include raised fasting glucose levels, with a small rise in two-hour glucose (<3 mmol/L) following 75g OGTT (Gardner & Tai, 2012). MODY2 is often identified during routine screening or during screening in pregnancy (Gloyn, 2003; Stride & Hattersley, 2009). Recognizing MODY2 from other causes of diabetes is important because it will save patients from unnecessary treatment and improve quality of life (Sagen, et al., 2008).

GCK has the cytogenetic location 7p13 and the gene consists of 12 exons (Nishi, et al., 1992; Stoffel & Shih, 2001). It encodes the enzyme glucokinase, which is a key enzyme in the regulation of insulin secretion and has been termed the glucose sensor in pancreatic beta-cells (Zelent, et al., 2005). Glucokinase is expressed mainly in the pancreatic beta-cell and the liver. The expression of the pancreatic isoform (465 amino acids), encoded by exons 1a and 2-10, is driven by a beta-cell specific promoter (glucose stimulated), which is located around 10 kb upstream from the liver specific promoter (Stoffel & Shih, 2001; Tanizawa, et al. 1991). The liver isoforms are 466 and 464 amino acids long and insulin is the key regulator of liver glucokinase expression (Guo, et al., 1990).

1.2.6 MODY3

MODY3 is caused by heterozygous loss-of-function mutation in the *HNF1A* gene and is the most common cause of MODY in most populations (Colclough, et al., 2013; Yamagata, 2003). Clinical features include a large rise in two-hour glucose levels (>5 mmol/L) on 75g OGTT. MODY3 patients have a progressive beta-cell failure resulting in reduced insulin secretion and increased hyperglycemia, with treatment requirements. A pharmacogenetic effect has been shown in patients as some are sensitive to sulphonylureas (Gardner & Tai, 2012).

The *HNF1A* gene is located on chromosome 12q24.2, the long (q) arm of chromosome 12 at position 24.31 (Colclough, et al., 2013; Yamagata, 2003). It consists of 10 exons and spans about 23 kb of genomic DNA. The gene is expressed in the liver, pancreas, stomach, small intestine, and kidney (Baumhueter, et al., 1990; Frain, et.al., 1989; Miquerol, et al., 1994), where both endocrine and exocrine cells express *HNF1A* during development of the pancreas (Yamagata, 2003).

According to Fajans et al., variants in the *HNF1A* gene have been identified in all racial and ethnic backgrounds, including European, Chinese, Japanese, African, and American Indian (Fajans, et al., 2000). Although pathogenic variants have been identified in all exons of *HNF1A*, variants in exons 1-6 have been reported to give earlier onset of MODY3 than variants in exon 8-10 (Frayling, et al., 2001) (Figure 1). Before the age of 25, 63% of people with a pathogenic *HNF1A* variant have developed diabetes, and further 79% have developed diabetes before the age of 35, and 96% before the age of 55 (Amed & Oram, 2016).

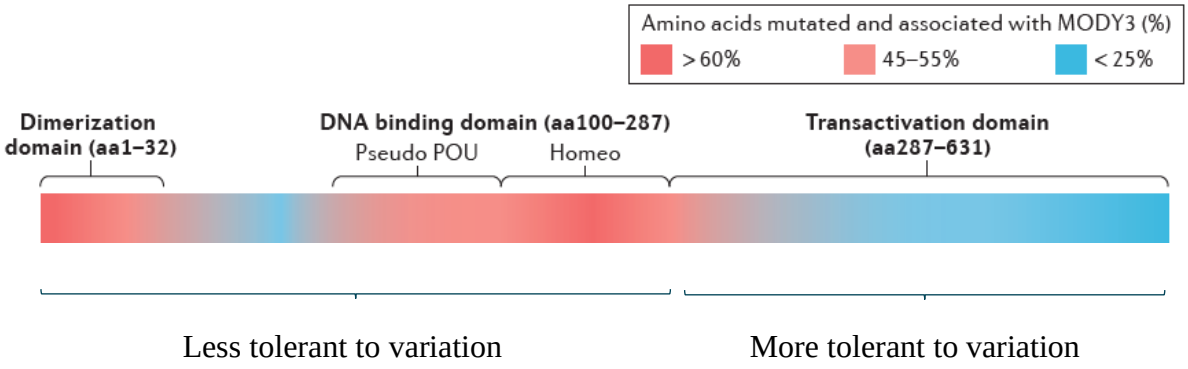

Figure 1: An outline of the HNF1A pancreatic protein isoform and functional domains, including the dimerization-, DNA binding- and transactivation domain. The heatmap illustrates the frequency of reported variants associated with MODY3 where the domain areas in red shows the accumulation of most variants associated with MODY3, and blue accumulation of fewer MODY3 associated variants (Harries, Ellard, Stride, Morgan, & Hattersley, 2006). Figure collected with permission from Prof. Lise Bjørkhaug Gundersen.

The protein encoded by pancreatic *HNF1A* is a 631 amino acid long transcription factor. As Figure 1 shows, the HNF-1A protein is composed of three functional domains: a dimerization domain (amino acids 1-32), a DNA binding domain (homeo domain amino acids 100-184; POU domain amino acids 198-287) and a transactivation domain (amino acids 287-631) (Ryffel, 2001; Vaxillaire, et al., 1999). A total of three different protein isoforms of HNF-1A has been reported produced (HNF-1A A-C), due to differential splicing and polyadenylation (Bach & Yaniv, 1993), of which one is termed the pancreatic isoform (A) and two the liver isoforms (B and C). As illustrated in Figure 1, pathogenic variants leading to MODY3 are more frequently associated with variants located in the dimerization domain and the DNA binding domain, compared to the transactivation domain.

1.2.7 MODY *versus* T2D and disease association

An increased overlap between traits and phenotypes of monogenic and complex genetic diseases provides a platform to understand the disease biology better (Tallapragada, Bhaskar, & Chandak, 2015). Most of the T2D susceptibility genes identified in the pre-GWAS era are key players in the development of some forms of monogenic diabetes. Later, studies have identified additional monogenic diabetes loci including *HNF1A, HNF1B, GCK*, and others, with variants associated with T2D risk. Of importance, several variants in these loci influence intermediate glycemic and anthropometric traits associated with T2D like fasting plasma glucose (FPG), insulin and lipid related traits, HbA1c levels, C-reactive protein, homocysteine, and waist circumference.

Patients with T2D have a more variable blood glucose profile than patients with MODY2. This variability may affect the complication prevalence and severity rates in T2D. MODY2 is infrequently related with vascular complications, and affected patients do not require treatment (Steele, et al., 2014). Patients with MODY2 are younger than patients diagnosed with T2D later in life with the same duration of hyperglycemia. Age as well as duration are significant factors for complication development (Steele, et al., 2014). However, it is stated in Amed and Orams study (Amed & Oram, 2016), that "these patients have the same risk as the general population for developing polygenic T2D".

MODY3 patients demonstrate mild increase in fasting blood glucose levels during childhood, and then later develop progressive hyperglycaemia and finally diabetes in adulthood (Fajans, Bell, & Polonsky, 2001). MODY3 patients are at similar risk as T2D patients for developing cardiovascular disease (CVD) (Fajans, Bell, & Polonsky, 2001) and carries a two to six times risk of death from cardiovascular aetiologies (Wong, Fan, & Glovaci, 2019). Microvascular complications, especially those involving the retina or kidneys, are as common in these patients as in patients with T1D and T2D and are probably determined by the degree of glycemic control (Fajans, Bell, & Polonsky, 2001). In South Africa, CVD is the second leading cause of death after HIV/AIDS (Matsha, et al., 2012).

1.3 Next-Generation Sequencing

Next-generation sequencing (NGS) is a high-throughput methodology that enables rapid sequencing of the base pairs in DNA or RNA samples (Thermo Fisher Scientific, n.d.). NGS is driving discovery and enabling the future of personalized medicine, as well as gene expression profiling, chromosome counting, molecular analysis, and detection of epigenetic changes. General NGS methods include whole genome sequencing and targeted sequencing, which is further subdivided into exome sequencing and gene or region-specific panels.

Whole genome sequencing (WGS) is the most comprehensive method and was the method used for screening individuals in this study. It enables an in-depth analysis of entire genomes, including exons, non-coding regions, and structural variants (Thermo Fisher Scientific, n.d.). WGS does not require prior knowledge of the genome sequence being analyzed. Even though WGS has advanced discovery and human health, certain complex regions of the genome are difficult to analyze by this approach. This may result in a population sequencing bias, and existing WGS-based databases are noted to be neither complete nor accurate (Thermo Fisher Scientific, n.d.).

1.4 Aim of the project

A large number of patients with MODY are often misdiagnosed as T1D or T2D, due to similar clinical features and lack in knowledge of MODY as a diabetes type (Gardner & Tai, 2012). To confirm a clinical diagnosis of MODY, genetic screening for rare pathogenic variants in MODY genes is necessary, however not a common practice in for instance low-middle income countries. A correct diagnosis can adjust the treatment protocols and the patient's life quality, as MODY patients can benefit from oral treatment by sulfonylureas instead of insulin. The prevalence of MODY by rare pathogenic *HNF1A* and *GCK* gene variants in Africa is currently unknown.

The overall aim of this project was to investigate the presence of MODY by *HNF1A* and *GCK* rare variants in individuals with diabetes, pre-diabetes, or normotolerant (controls), from a mixed-ancestry population in Cape Town, South Africa, by screening and evaluating WGS data from 15 analyzed individuals.

2. MATERIALS AND METHODS

The following chapter describe and present the sample material (2.1), instruments and software's (2.2) and NGS methodology (2.3) used in the study.

2.1 Sample material

Bellville South is located within the northern suburbs of Cape Town, South Africa. The township was formed in the late 1950s and largely consists of a mixed-ancestry population. According to South African census data of 2011, the population is comprised of 76.0% mixed-ancestry, 18.5% black, 1.0% Asian, 0.5% Caucasian, and 4.0% individuals from other ethnicities (Statitics South Africa, 2011). A study conducted by Erasmus, et al., reported on the increase in diabetes incidence in this population, thus providing the ideal setting to investigate MODY prevalence (Erasmus, et al., 2012).

The term "township" has no formal definition but is commonly understood to refer to the underdeveloped, usually (but not only) urban, residential areas that during Apartheid were reserved for non-whites (Africans, Coloreds and Indians) (Pernegger & Godehart, 2007).

DNA samples from 15 randomly selected individuals underwent WGS and sequence data on *HNF1A* and *GCK* was provided by Cape Peninsula University of Technology (CPUT) (prof. Tandi Matsha). Among these, five were characterized with diabetes, six with prediabetes and four as normotolerant. In this assignment, *HNF1A* and *GCK* genes from these 15 individuals was evaluated and variations as cause of MODY was considered. Anthropometric measurements like diagnosis, gender, age (years), BMI, glucose fasting blood (mmol/L), glucose 2 hours post prandial (mmol/L), glycosylated Hb (HbA1c) (%), insulin fasting (mIU/L) in these 15 individuals, were also provided by CPUT.

2.2 Instruments and software's

Table 3 and 4 gives an overview of the instruments and software used, supplier, and their application.

Table 3: Analytical instruments.

Name	Supplier	Application
Ion GeneStudio S5 System	Applied biosystems by Thermo Fisher Scientific	NGS
Ion Chef™ System	Applied biosystems by Thermo Fisher Scientific	NGS

Table 4: Analytical software.

Name	Supplier	Application
Torrent Suite™ Software	Applied biosystems by Thermo Fisher Scientific	NGS
Ion Reporter™ Software	Applied biosystem by Thermo Fisher Scientific	NGS

2.3 Methodology

Theory of methods and approach used in the study is presented in the following chapter.

2.3.1 Ion Torrent Technology

According to Thermo Fisher Scientific, the Ion Torrent NGS technology uses the fact that addition of a dNTP to a DNA polymer releases a hydrogen ion. It measures the pH change resulting from the hydrogen ions using semiconductors. These changes are measured simultaneously millions of times in different wells on a single chip (e.g. Ion 5 Series chip) to determine the sequence of each fragment (Thermo Fisher Scientific A., n.d.). The technology is able to translate chemically encoded information such as Adenine (A), Cytosine (C), Guanine (G) and Thymine (T) into digital information on a semiconductor chip (Thermo Fisher Scientific B., n.d.). This contributes to information about inherited disease, oncology, infectious disease, human identification (HID), human leukocyte antigen (HLA) typing and agrigenomics (Thermo Fisher Scientific A., n.d.).

2.3.2 Ion GeneStudio S5 System

The Ion GeneStudio S5 system uses the Ion Torrent technology to enable simple targeted sequencing workflow (Thermo Fisher Scientific C., n.d.). Cartridge-based reagents make the system simple to use and makes it possible to run all targeted sequencing or microbial applications on a single next-generation sequencer. The system uses Ion 5 Series chips to more quickly and cheaply read large amounts of sequences (Churko, et al., 2013). The chips are capable of sequencing from 2-80 million reads per run which makes it possible to analyze both small and large projects on a single system (Thermo Fisher Scientific C., n.d.). As little as 1 ng input DNA or RNA from for instance formalin-fixed paraffin-embedded (FFPE) tissue, retrospective samples from fine needle aspirates, and cell-free DNA extracted from blood (cfDNA) can be used.

2.3.3 Ion Chef™ system

The Ion Chef™ system, using Ion AmpliSeq technology, automates and incorporates all steps of library preparation such as targeted PCR amplification with respective primer panel, partial digestion of primer sequences, adapter/barcode ligation, purification, library equalization and pooling (Thermo Fisher Scientific D., n.d.). It also automates all steps in template preparation and chip loading, making one or two chips that are ready for sequencing on the Ion GeneStudio S5 systems. In doing so, it integrates several manual and instrumental steps into one single process and by this greatly cuts down on time used on manual work. The system uses pre-packaged reagent kits that combined with the fact that it automates many of the processes greatly helps reduce the chance for user-introduced variability and errors.

Figure 2: The figure illustrates the Ion Chef™ system and Ion GeneStudio S5 system (One Lambda). The system is equipped with a visual system that provides barcode reading and identification of samples, chips and reagents, as well as preforming the setup check before runs, a load and unload check, calibration and diagnostics (Thermo Fisher Scientific D., n.d.).

2.3.4 Torrent Suite™ software

Prior to starting a run on the Ion Chef™ and Ion S5 systems, Torrent™ Suite software grants the possibility of reviewing the accuracy and quality of the run, as well as planning, tracking, monitoring and viewing sequencing run results (Thermo Fisher Scientific E., n.d.). The software is also used in detecting single nucleotide polymorphisms, as well as performing variant calling (Thermo Fischer Scientific G., 2019). The software analyzes Ion Torrent raw data, does alignments and variant calling (Thermo Fisher Scientific E., n.d.). The Torrent Suite™ Software works well with Ion Reporter™ Software to make a smooth workflow from start to finish (Thermo Fisher Scientific F., n.d.).

2.3.5 Ion Reporter™ software

Ion Reporter™ Software uses over 20 public and proprietary databases to simplify the analysis and identify relevant variants from the analysis data provided by Ion S5 systems and the Torrent Suite™ software. The use of parameters copy number baselines and filters makes is possible to determine how the data is analyzed and to remove variants that are not relevant to the research being done (Thermo Fisher Scientific H., 2020).

2.3.6 Minor Allele Frequency

Minor allele frequency (MAF) refers to the frequency at which the second most common allele occurs in a population. It is widely used in population genetics studies because it provides information on how to differentiate between common and rare variants in the population. If a single nucleotide polymorphism (SNP) for instance has an allele G frequency of 0.40, it implies that 40% of the population has the G allele versus the major allele which is found in 60% of the population. Common variants have a MAF $\geq$ 5% and indicates an effect size that is very small and insufficient to independently cause disease (Mani, 2017). The common variants reside mainly in the introns of the genome and are unpowered to show disease causality. Rare variants with a MAF $\leq$ 0.1% have a large effect size and are sufficient to independently cause disease.

In this study, rare variants, either in the exon or within ±8 base pairs from the intron/exon boundary of the gene, with a MAF <0.01% and the most common missense variants were further examined for occurrence and prevalence in population databases (e.g. gnomAD), reported disease association based on clinical variant databases (e.g. ClinVar), and literature (e.g. PubMed).

2.3.7 Genome Aggregation Database

The Genome Aggregation Database (http://www.gnomad.broadinstitute.org), or gnomAD, is a database where researchers and investigators from all over the world can submit and harmonize their sequencing projects and genomic data. The database consists of a collection of exome and genome sequencing data, which is available for the wider scientific community (gnomAD, 2020).

GnomAD (v.2.1.1) spans exomes from 125 748 individuals and genomes from 15 708 individuals, in total 141 456 individuals, from a broad specter of disease-specific and genetic studies from multi-ethnic populations (Francioli, et al., 2018). GnomAD is aligned against the GRCh37 reference (gnomAD, 2020).

Further, gnomAD houses information on genes, regions, and variants, by allele counts (counts of each alternate allele for each site across all samples), allele number (number of times the exome/genome has been sequenced), age distribution, and MAF in individual populations. MAF data is collected from approximately 18 million variants identified through exome or genome sequencing from the 141 456 unrelated individuals tested. The database is therefore useful obtaining estimates of MAF for variants and relevant in evaluating disease relevance for individual variants (Karczewski, et al., 2019).

2.3.8 Clinically Relevant Sequence Variations

Clinically Relevant Sequence Variations (ClinVar) (http://www.ncbi.nlm.nih.gov/clinvar/) is an open database with a public archive of reports about genomic variations evidence and its relationship to human health. The database thereby promotes access to and communication about the relationships claimed between human variation and noted health status, and the history of that interpretation (NCBI, 2020). It also provides information about the variant (e.g.

SNV) and which amino acids that are affected in the encoded protein. Each variant is provided with an accession number with the format SCV000000000.0, where all submitted data on the specific variant is gathered, and when possible harmonized to controlled vocabularies or other data standards (Landrum, et al., 2014). The database presents assertions made regarding the clinical significance of the variants and indicates an interpretation of the variants pathogenicity based on the submissions. Variant interpretation is further explained in chapter 2.3.9.

ClinVar was used to investigate whether the whole genome identified and selected variants had a known clinical significance or not, if the variant previously had been reported and whether it was reported in connection with MODY2, MODY3 or T2D.

2.3.9 Variant interpretation

To confirm a MODY diagnosis, DNA sequencing and identification of pathogenic variants in MODY genes is required (Valkovicova, et al., 2019). If a variant has previously been reported in ClinVar as MODY, based on identification in other families/individuals with a MODY phenotype, MODY suspicion can be confirmed. If a novel sequence variant is identified (previously unreported in MODY genes), the next step is the classification of the variant according to degree of predicted pathogenic effect.

In 2015 the American College of Medical Genetics and Genomics (ACMG) reported a clinical guidance for reporting and interpreting sequence variations (Richards, et al., 2015). The report is used to describe variants identified in Mendelian disorders, such as MODY, and provide a process for classification of variants into five categories. Based on the instructions of ACMG, the variants can be placed in one of the following modifiers: benign (class 1), likely benign (class 2), uncertain significance (class 3), likely pathogenic (class 4), or pathogenic (class 5).

As stated in the report from ACMG the process is based on different criteria's using a collection of information from literature, functional criteria, MAF data from population databases (e.g. gnomAD), disease databases (e.g. ClinVar, OMIM), sequence databases (e.g. NCBI), and tools for *in silico* analyses (e.g. SIFT, Poly-Phen-2) (Valkovicova, et al., 2019).

Due to insufficient information from these databases and tools, numerous variants are classified as variants of unknown significance (VUS; class 3). As reported in the ACMG guidelines, VUSs cannot be used in clinical decision making to avoid any mis-diagnosis and mis-treatment (Richards, et al., 2015).

3. RESULTS

The WGS resulted in 388 identified variants in total (both *HNF1A* and *GCK* gene) for all the 15 analyzed individuals from the cohort in Bellville South, Cape Town. Of the 388 variants, 14 were in UTRs, 355 in introns, 7 in upstream regions and 12 in exons. Further, the screening for protein coding variants in *HNF1A* and *GCK* in the 15 analyzed individuals resulted in two rare sequence variants (MAF<0.01%) and four common sequence variants (MAF>5%) in the *HNF1A* gene, and none in the *GCK* gene. The two rare *HNF1A* variants included a frameshift variant c.1124delG p.(Gly375AlafsTer9) in exon 6 and the missense variant c.1243G>A p.(Gly415Arg) in exon 6 (NM_000545.6) (Table 5). The four common missense *HNF1A* variants were c.293C>T p.(Ala98Val), c.1460G>A p.(Ser487Asn), c.79A>C p.(Ile27Leu) and c.1720A>G p.(Ser574Gly) (Table 5), and were further considered due to the consequences they may cause HNF-1A protein function and association with diabetes according to previous literature.

If an intron or UTR variant was to be evaluated, the variant had to be within ± 8 base pairs from the exon, and additionally have a MAF<0.01%. None of the introns or UTR variants in nor *HNF1A* nor *GCK* met these requirements, and therefore none of them were further considered. However, it is not possible to completely exclude the intronic variants as benign because theoretically they can influence RNA splicing and therefore function of the protein. In addition, it is challenging to evaluate the intronic variants because of missing reports in ClinVar and in literature (PubMed). Also, none of the individuals harbored protein coding (exon) variants in the *GCK* gene, which is important for such a variant to have a pathogenic effect (due to MODY2 being a mild diabetes disease). In addition, several non-significant variants were identified in *HNF1A* (p.(Thr10=), p.(Leu17=), p.(Ser70=), p.(Gly288=), p.(Leu459=), p.(Thr515=)). All these are common variants that had MAF>13%, except from p.(Ser70=); a rare variant with MAF=0.02%.

Table 5: Variants identified in the HNF1A gene and the sequence identity (based on NCBI accession number NM_000545.6). All variants are single nucleotide variants (SNV). Numbering is according to the nomenclature for description of sequence variations (den Dunnen, et al., 2016), reported frequency according to gnomAD (overall and in African specifically), and classification of pathogenicity according to ClinVar and reference reports. NA: Not Available.

Frequency of variant	**Region in *HNF1A***	**Nucleotide change**	**Amino acid change in *HNF1A***	**Total MAF (%) (gnomAD)**	**MAF in Africans (%) (gnomAD)**	**ClinVar**	**Reference**
Rare	Exon 6	c.1124delG	p.(Gly375Alafs Ter9)	NA	NA	NA	(Lehto, et al., 1999)
	Exon 6	c.1243G>A	p.(Gly415Arg)	0.001194	0.00	NA	(Flannick, et al., 2013; Yoshiuchi, et al.,1999)
Common	Exon 1	c.293C>T	p.(Ala98Val)	2.820	0.5131	Class 1/2	(Awa, et al., 2011)
	Exon 7	c.1460G>A	p.(Ser487Asn)	32.96	12.03	Class 1	(Awa, et al., 2011)
	Exon 1	c.79A>C	p.(Ile27Leu)	34.79	11.91	Class 1	(Awa, et al., 2011)
	Exon 9	c.1720A>G	p.(Ser574Gly)	99.54	95.27	Class 1	

Both rare and common variants in the *HNF1A* gene were further evaluated based on data from gnomAD and ClinVar, and by revisiting clinical data from carrier individuals. Revisiting clinical data for signs of early-onset diabetes (non-obesity etc.) is necessary in the further evaluation of whether a variant is likely to be a cause of MODY, or whether may represent a risk factor for other types of diabetes (T2D).

Table 6 gives an indication of how rare or common the different variants are based on how many of the 15 analyzed individuals they were identified in. The common variants c.1720A>G (p.(Ser574Gly)) was identified in all individuals, c.79A>C (p.(Ile27Leu)) in 10

individuals, c.293C>T (p.(Ala98Val)) in 2 and c. 1460G>A (p.(Ser487Asn)) in 9. The rare variants c.1124delG (p.(Gly375AlafsTer9)) was identified in 2 individuals and c.1243G>A (p.(Gly415Arg)) in 1 individual.

Table 6: The table gives an overview of which variants in HNF1A that are identified in which of the 15 individuals.

		Nucleotide change					
Sample ID	**9**	c.1720A>G	c.79A>C		c. 1460G>A		
	16						
	17						
	18						c.1243G>A
	44						
	75						
	86		c.79A>C				
	104				c. 1460G>A		
	109						
	333						
	1467		c.79A>C	c.293C>T	c. 1460G>A		
	1484						
	1488		c.79A>C		c. 1460G>A	c.1124delG	
	1500						
	1508			c.293C>T		c.1124delG	
Protein change		p.(Ser574Gly)	p.(Ile27Leu)	p.(Ala98 Val)	p.(Ser487Asn)	p.(Gly375Alafs Ter9)	p.(Gly415Arg)

The clinical information and biometric data on the 15 individuals are presented in Table 7 and sorted according to whether individuals are carriers of rare and common variants. The 15 individuals are placed in one of three categories of DM: Normotolerant, IGT or DM, which is based on gender, age, BMI and values of; glucose fasting blood, 2hr-OGTT, HbA1c and fasting insulin.

WHO defines overweight as BMI ≥25, and obesity as a BMI ≥30. The three individuals that carry a rare variant in *HNF1A* has BMI of 21.7, 30.5 and 32.5. This indicates that one of the individuals is normal (BMI<25) and two have obesity (BMI≥30). The individuals that are carriers of common variants has BMI between 17.9 and 34.8, which indicates that three individuals are normal (<25), three have overweight (25≤30) and five have obesity (≥30). Only one individual (1484) of the 15 individuals show indications of insulin resistance, based on the insulin fasting value.

Table 7: Clinical data on the 15 individuals selected for WGS. The three categories from WHO is either normotolerant, IGT, or DM. IGT: Impaired fasting glucose. Reference values in Table 1. BMI: Body Mass Index. NA: Not Available. Reference value: 18.5-25 indicates normal weight.

Frequency of Variants	Sample ID	Diagnosis WHO: DM (3 categories)	Gender	Age	BMI	Glucose Fasting Blood (Ref.value: <5.6mmol/L)	2hr OGTT (Ref.value: 8.5mmol/L)	Glycated HB (HbA1c) (Ref.value: 4-6%)	Insulin Fasting (Ref.value: <25mIU/L)
Rare	1488	IGT	Male	59	21.7	3.9	8.0	6	1.3
	1508	Normotolerant	Female	51	30.5	4.9	4.0	6.1	2.9
	18	IGT	Female	52	32.5	4.7	7.08	5	5.7
Common	9	DM	Female	49	28.6	NA	NA	10.7	17.9
	16	DM	Female	41	NA	7.0	5.7	6.1	18.7
	17	DM	Female	38	32.2	7.7	5.1	5.6	23.0
	44	DM	Female	30	34.8	7.0	10.3	7.1	12.2
	75	IGT	Male	62	23.5	4.4	9.8	5.3	2.9
	86	Normotolerant	Female	38	28.8	4.9	3.6	5.5	4.5
	104	IGT	Female	36	21.9	5.2	7.8	5.7	2.4
	109	Normotolorant	Female	25	17.9	3.2	4.6	5.1	3.3
	333	DM	Female	46	29.4	4.6	12.5	6.6	4.8
	1467	Normotolorant	Female	36	31.2	4.2	7.7	6.0	12.0
	1484	IGT	Female	26	30.5	5.8	8.2	6.2	32.5
	1500	IGT	Female	65	31.7	4.7	7.9	5.7	4.3

In order to further investigate whether the HNF-1A variant p.(Gly415Arg) affects an amino acid that is evolutionary conserved between species, sequence alignment by Clustal Omega was performed. This analysis gives an indication whether the amino acid change, based on being conserved, may have an impact on the structure, and hence the function of a specific protein. Since the p.(Gly375AlafsTer9) variant is a frameshift variant causing early termination, only alignment of the variant p.(Gly415Arg) was performed. As shown in Figure 3, all the different species analyzed have the same amino acid glycine in position 415, which indicates that this amino acid is highly conserved in HNF-1A.

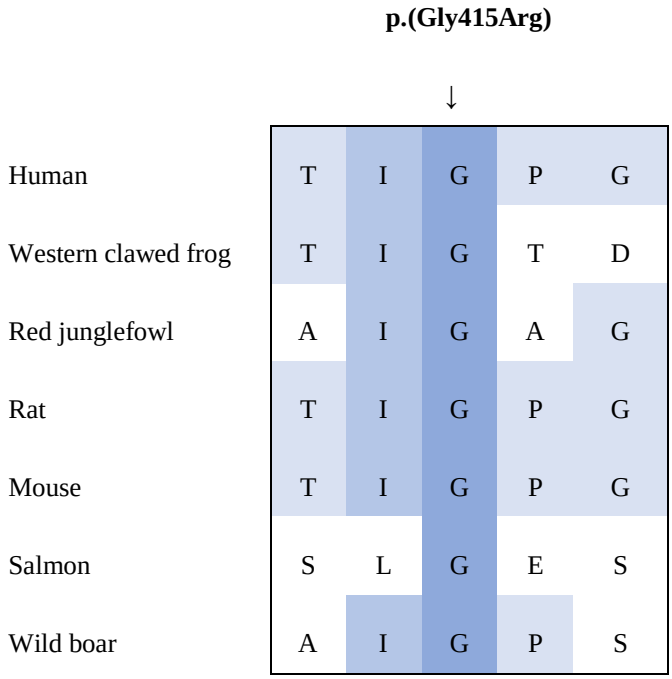

Figure 3: The figure illustrates an alignment between human HNF1A amino acid sequence and selected species including Gly415 (changed in individual ID18 by p.(Gly415Arg) due to c.1243G>A). The alignment is retrieved from the Clustal Omega software (Madeira, et al., 2019; Multiple Sequence Alignment, n.d). Highly conserved amino acids are indicated in dark blue, while moderately conserved amino acids are colored lighter blue, and low conserved amino acids are colored white.

4. DISCUSSION

The overall aim of the project was to investigate the presence of MODY by *HNF1A* and *GCK* rare variants in individuals with diabetes, pre-diabetes, or normotolerant (controls), from a mixed-ancestry population in Cape Town, South Africa, by screening and evaluating WGS data from 15 analyzed individuals.

Based on the WGS data we identified two rare variants of interest, one frame shift variant c.1124delG p.(Gly375AlafsTer9) and one missense variant c.1243G>A p.(Gly415Arg) in *HNF1A*, with MAF <0.01% in the mixed-ancestry population from Cape Town, South Africa. In addition, four common *HNF1A* missense variants were identified among the 15 analyzed individuals.

The c.1124delG p.(Gly375AlafsTer9) is as mentioned a frame shift variant, which means that one guanine (G) nucleotide gets deleted (del) and causes a frame shift causing an early termination of protein sequence. A deletion occurs when a part of a DNA molecule is not copied during DNA replication. When a nucleotide gets deleted, the following amino acid sequence will also be a target for frame shift. If a frameshift occurs in an early exon (except the last two exons), the mRNA will in most cases be degraded in a process called nonsense mediated mRNA decay (NMD) (Nagy & Maquat, 1998; Zhang, et al., 2009). Consequently, the protein will not be produced. The exception is if the frameshift occurs in the second last exon (55 nucleotides from last exon/intron transition), or if it occurs in the last exon. In these cases, NMD will in most cases not take place. The protein will be produced, but with some defects. In the case of p.(Gly375AlafsTer9), where the frameshift takes place in exon 6/10 (Table 5), NMD will most likely occur. This signalizes that the mRNA will be degraded and an end of translation of the HNF-1A protein. p.(Gly375AlafsTer9) is located in the transactivation domain. Based on reports on tolerance for variability in HNF-1A, the transactivation domain is more tolerant to variability in terms of MODY3, compared to the dimerization and DNA binding domain (Bellanne-Chantelot, et al., 2008). This variant has been previously reported in a MODY family (Lehto, et al., 1999).

In our study, the p.(Gly375AlafsTer9) was identified in two individuals (sample ID 1488 and 1508). Based on clinical data (Table 3) from carrier individuals in our study, they were characterized with IGT (1488) or was normotolerant (1508). For other indicators like BMI and fasting insulin, individual 1508 had high BMI (30.5) and slightly elevated HbA1c levels (6.1). Based on the clinical data, these two carriers of p.(Gly375AlafsTer9) indicated no sign of early-onset diabetes, which is a requirement for MODY. Age of individuals being 59/51 years for 1488 and 1508, respectively, this further supports that this variant is not associated with MODY in these individuals. Based on this variant being previously associated with MODY, it is unexpected that this frameshift variant seems neutral in our carrier individual.

The c.1243G>A (p.(Gly415Arg)) *HNF1A* missense variant, identified in individual 18, is reported in three alleles in gnomAD and none reported in Africans. The variant is located in the transactivation domain (exon 6) (Table 5), which is more tolerant to variations in terms of cause of MODY (Ellard, et al., 2008; Flannick, et al., 2016; Harries, et al., 2006). Unpublished data communicated by supervisors have shown that the p.(Gly415Arg) effect on HNF-1A transactivation activity was measured to ~73% in transfected HeLa cells, which is lower than WT HNF-1A (set to 100%). Variants that are associated with MODY3 usually display transactivation activity of 20-30% (Balamurugan, et al., 2016; Bjørkhaug, et al., 2003; Malikova, et al., 2020), which is much lower than what is communicated for p.(Gly415Arg). Further, the DNA binding activity of p.(Gly415Arg) showed a non-significant reduction (reduced to 84% compared to WT HNF-1A of 100% by unpublished data). Sequence alignment of Gly415 shows strong conservation between numerous species (Figure 3). This may indicate that Gly415 is structurally important and that the variant may lead to altered HNF-1A structure and potentially loss of some protein function.

A Japanese study detected the same *HNF1A* missense mutation with partial dysfunction, p.(Gly415Arg), in a Japanese patient with early-onset T2D (Yoshiuchi, et al., 1999). The missense mutation is also reported in a Danish cohort of T2D individuals (personal communication, unpublished data) where both patients with diabetes have overweight. The same variant is also identified in an overweight diabetic patient in the Norwegian MODY Registry, where both mother and brother are diagnosed with T2D (personal communication, unpublished data).

Based on the clinical features of individual 18 (Table 7), including prediabetes, overweight and high BMI, these observations indicate that p.(Gly415Arg) in individual 18 may present a risk factor for T2D rather than MODY, as no overt sign of diabetes is present at her age of 51. However, this individual has a normal level of HbA1c and fasting insulin, which usually is not a characteristic for T2D.

The four common missense variants identified was c.79A>C p.(Ile27Leu), c.293C>T p.(Ala98Val), c.1460G>A p.(Ser487Asn) and c.1720A>G p.(Ser574Gly). The variant p.(Ser574Gly) was identified in all 15 analyzed individuals (Table 6), which is not surprising considering the allele frequency of >95% (Table 5). The allele frequency classifies the p.(Ser574Gly) variant as benign and cannot be a sole cause of disease.

As presented in Table 6, the p.(Ile27Leu) and p.(Ser487Asn) variants were identified in the following individuals: 9, 16, 17, 18, 44, 104, 1467, 1488 and 1500. The p.(Ile27Leu) variant was also identified in individual 86. Both variants have a MAF~12% (p.(Ile27Leu)=11.91% and p.(Ser487Asn)=12.03%), which indicate that they are common variants (MAF≥5%) and cannot be the cause of MODY. The p.(Ala98Val) variant was identified in individuals 1467 and 1508 (Table 6), with MAF=0.51% in African individuals. Even though the MAF for p.(Ala98Val) is significantly lower than for p.(Ile27Leu) and p.(Ser487Asn), it is too high to be the cause of MODY. However, p.(Ile27Leu), p.(Ser487) and p.(Ala98Val) have been associated with a decrease in glucose-stimulated insulin secretion, which was found to further deteriorate over a 6-year period, reduced transcriptional activity, enhanced insulin sensitivity and increased risk of T2D in elderly overweight individuals (Holmkvist, et al., 2006).

With early genetic screening, it is possible to identify members of pedigrees who have inherited a specific variant affecting diabetes in their family, and before diabetes develops, with for instance the identification of pathogenic variants responsible for MODY (Fajans, et al., 2000). Here, periodic testing for slight abnormalities of carbohydrate metabolism is recommended, as in the case of pathogenic *HNF1A* variant and MODY3, it may have important prognostic and therapeutic significance (oral treatment by sulphonylureas). Therefore, it is important to characterize the prevalence of MODY3 in poorly genome-investigated populations like Africa, as most such studies have been performed in Europeans.

With early diagnosis of MODY3, appropriate therapy can also be instituted early in the course of their hyperglycemia, because of the risk of progression to severe hyperglycemia and insulin-requiring diabetes. Vascular and neuropathic complications can be prevented with early treatment to achieve normoglycemia (Fajans, et al., 2000). Genetic diagnosis may also be advised for patients who have been classified as having T1D and have a strong family history of diabetes.

If carbohydrate intolerance or diabetes is due to a mutation in the gene encoding *GCK*, limited therapy and follow-up are acceptable, because of the benign and nonprogressive course of diabetes in such cases (Amed & Oram, 2016; Osbak, et al., 2009).

Confirming a genetic diagnosis of MODY is important, especially since MODY3 is often misdiagnosed as T2D because of overlapping clinical features (Colclough, et al., 2013). By confirming such a diagnosis, it provides the knowledge to classify the subtype, predict the prognosis of the patient, select the precise treatment, and estimate the risk in patient relatives. MODY is however rare, and only represents 1-3% of all diabetes. Thus, expected number of MODY in the general population is about 1:1000. Thus, limitations with this study is that only 15 individuals were investigated. Among these, only 5 had diabetes, while 6 had prediabetes by IGT. Thus, the chance of identifying MODY by pathogenic variants in *HNF1A* or *GCK* is extremely small. To determine the true prevalence of MODY in this Bellville South cohort, a larger number of individuals would need to be screened.

Molecular genetic testing allows confirmation of a MODY diagnosis and defines the MODY subtype, prognosis, and relevant treatment. However, the sample material (type of clinics and number) is essential to get the best possible understanding of MODY and its prevalence in South Africa. We believe that the quality of this study would be improved if the sample material consisted of members from suspected MODY families, with clinically confirmed early-onset diabetes (not IGT or normoglycemia). By molecular genetic testing of such individuals, it would increase the chances of identifying MODY individuals in such South African population and contribute to therapeutic benefits for all possible carriers with particularly pathogenic *HNF1A* variants.

5. REFERENCES

Amed, S., & Oram, R. (2016, October). Maturity-Onset Diabetes of the Young (MODY): Making the Right Diagnosis to Optimize Treatment. *Can J Diabetes*(40(5)), pp. 449-454.

American Diabetes Association. (2011, January 1). Diagnosis and Classification of Diabetes Mellitus. *American Diabetes Association*(34), pp. 62-69.

Arslanian, S., El ghormli, L., Haymond, M. H., Chan, C. L., Chernausek, S. D., Gandica, R. G., . . . Willi, S. M. (2020, January 29). Beta cell function and insulin sensitivity in obese youth with maturity onset diabetes of youth mutations vs type 2 diabetes in TODAY: Longitudinal observations and glycemic failure. *Pediatr Diabetes*, pp. 1-11.

Atkinson, M. A., & Eisenbarth, G. S. (2001, July 21). Type 1 diabetes: new perspectives on disease pathogenesis and treatment. *The Lancet*(358), pp. 221-29.

Awa, W. L., Thon, A., Raile, K., Grulich-Henn, J., Meissner, T., Schober, E., & Holl, R. W. (2011, April). Genetic and clinical characteristics of patients with HNF1A gene variations from the German–Austrian DPV database. *European Society of Endocrinology, 164*(4), pp. 513-520. doi:10.1530/EJE-10-0842

Bach, I., & Yaniv, M. (1993, November 12). More potent transcriptional activators or a transdominant inhibitor of the HNF1 homeoprotein family are generated alternative RNA processing. *EMBO Journal*(12(11)), pp. 4229-42.

Bach, I., & Yaniv, M. (1993, November 1). More potent transcriptional activators transdominant inhibitor of the HNF1 homeoprotein family are generated by alternative RNA processing. *The Embo Journal*(12), pp. 4229-42.

Balamurugan, K., Bjørkhaug, L., Mahajan, S., Kanthimathi, S., Njølstad, P. R., Srinivasan, N., . . . Radha, V. (2016, December). Structure-function Studies of HNF1A (MODY3) Gene Mutations in South Indian Patients With Monogenic Diabetes. *Clin Genet*(90(6)), pp. 486-495.

Baumhueter, S., Mendel, D. B., Conley, P. B., Kuo, C. J., Turk, C., Graves, M. K., . . . Crabtree, G. R. (1990, March). HNF-1 shares three sequence motifs with the POU domain proteins and is identical to LF-B1 and APF. *Genes&Development*(4(3)), pp. 372-9.

Bazalova, Z., Rypackova, B., Broz, J., Brunerova, L., Polak, J., Rusavy, Z., . . . Andel, M. (2010, February 10). Three novel mutations in MODY and its phenotype in three different Czech families. *Diabetes Research and Clinical Practice*, pp. 132-138.

Bellanne-Chantelot, C., Carette, C., Riveline, J.-P., Valero, R., Gautier, F., Larger, E., . . . Laloi-Michelin, M. (2008, February). The Type and the Position of HNF1A Mutation Modulate Age at Diagnosis of Diabetes in Patients with Maturity-Onset Diabetes of the Young (MODY)-3. *American Diabetes Association*(57(2)), pp. 503-508.

Bishop , M. L., Fody, E. P., & Schoeff, L. E. (2018). *Clinical Chemistry*. Philadelphia: Wolters Kluwer.

Bjørkhaug, L., Sagen, J. V., Thorsby, P., Søvik, O., Molven, A., & Njølstad, P. R. (2003, February). Hepatocyte Nuclear factor-1 Alpha Gene Mutations and Diabetes in Norway. *J Clin Endocrinol Metab*(88(2)), pp. 920-931.

Churko, J. M., Mantalas, G. L., Wu, J. C., & Snyder, M. P. (2013, June 7). Overview of High Throughput Sequencing Technologies to Elucidate Molecular Pathways in Cardiovascular Diseases. *National Institutes of Health*(112(12)).

Colclough, K., Bellanne-Chantelot, C., Saint-Martin, C., Flanagan, S. E., & Ellard, S. (2013, May). Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha and 4 alpha in maturity-onset diabetes of the young and hyperinsulinemic hypoglycemia. *Human Mutation*(34(5)), pp. 669-685.

den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., & Roux, A.-F. (2016, March 02). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Human Mutation*(37).

Diabetes Genes. (n.d.). *Hepatic Nuclear Factor 4 Alpha (HNF4A)*. Retrieved from Diabetes Genes: https://www.diabetesgenes.org/what-is-mody/hepatic-nuclear-factor-4-alpha-hnf4a/

Diabetesforbundet. (2020, March 25). *MODY (Maturity-Onset Diabetes of the Young) og nyfødtdiabetes*. Retrieved from diabetesforbundet: https://www.diabetes.no/om-diabetes/mody/

Ekholm, E., Shaat, N., & Holst, J. J. (2012). Characterization of beta cell and incretin function in patients with MODY1 (HNF4A MODY) and MODY3 (HNF1A MODY) in a Swedish patient collection. *Acta Diabetol*(49), pp. 349-354.

Ellard, S., Bellannè-Chantelot, C., Hattersley, A. T., & MODY group, E. (2008, February 23). Best practive guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young. *Diabetologia*(51), pp. 546-553.

Erasmus, R. T., Soita, D. J., Hassan, M. S., Blanco-Blanco, E., Vergotine, Z., Kegne, A. P., & Matsha, T. E. (2012). *High prevalence of diabetes mellitus and metabolic syndrome in a South African coloured population: baseline data of a study in Bellville, Cape Town*. (102(11 Pt 1) ed.). South African medical journal.

Estrada, K., Aukrust, I., Bjørkhaug, L., Burtt, N. P., Mercader, J. M., García-Ortiz, H., . . . Fernandez-Lopez. (2014). *Association of a Low-Frequency Variant in HNF1A With Type 2 Diabetes in a Latino Population*. JAMA.

Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. (2003, January). Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care*(26), pp. 5-20.

Fajans, S. S., Bell, G. I., & Polonsky, K. S. (2001, September 27). Molecular Mechanisms and Clinical Pathophysiology of Maturity-Onset Diabetes of the Young. *The New England Journal of Medicine*. Retrieved April 20, 2020, from https://www.nejm.org/doi/full/10.1056/NEJMra002168?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub%3Dpubmed

Fajans, S. S., Bell, G. I., Herman, W. H., Polonsky, K. S., & Hatler, J. B. (2000). Natural History, Genetics and Pathogenesis of HNF-4a/MODY1. In F. Matschinsky, & M. Magnuson, *Molecular Pathogenesis of MODYs* (pp. 1-15). Basel, Switzerland: Karger.

Flannick, J., Beer, N. L., Bick, A. G., Agarwala, V., Molnes, J., Gupta, N., . . . Altshuler, D. (2013, November). Assessing the Phenotypic Effects in the General Population of Rare Variants in Genes for a Dominant Mendelian Form of Diabetes. *Nat Genet*(45(11)), pp. 1380-1385.

Flannick, J., Johansson, S., & Njølstad, P. R. (2016). Common and rare forms of diabetes mellitus: towards a continuum of diabetes subtypes. *Nature Reviews Endocrinology*.

Florez, J. C., & Walford, G. A. (2010, November). Type 2 Diabetes and Genetics: Translating Knowledge into Understanding. *Current Cardiovascular Risk Reports*(4(6)), pp. 437-445. doi:10.1007/s12170-010-0129-1

Frain, M., Swart, G., Monaci, P., Nicosia, A., Stämpfli, S., Frank, R., & Cortese, R. (1989, October 6). The liver-specific transcription factor LF-B1 contains a highly diverged homeobox DNA binding domain. *Cell*(59 (1)), pp. 145-57.

Francioli, L., Tiao, G., Karczewski, K., Solomonson, M., & Watts, N. (2018, October 17). *gnomAD v2.1*. Retrieved from MacArthur Lab: https://macarthurlab.org/2018/10/17/gnomad-v2-1/

Frayling, T. M., Evans, J. C., Bulman, M. P., Pearson, E., Allen, L., Owen, K., . . . Hattersley, A. T. (2001, February). Beta-cell genes and diabetes: molecular and clinical characterization of mutations in transcription factors. *Diabetes*(50), pp. 94-100.

Gardner, D. S., & Tai, E. S. (2012, April 27). Clinical features and treatment of maturity onset diabetes of the young (MODY). *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, pp. 101-108.

Gloyn, A. L. (2003, September 25). Glucokinase (GCK) mutations in hyper- and hypoglycemia: Maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemia of infancy. *Human Mutation*(22), pp. 353-362.

gnomAD. (2020, May 1). *About GnomAD*. Retrieved from GnomAD browser: https://gnomad.broadinstitute.org/about

Guo, L.-H., Hill, A. O., Hopper, D. J., Lawrance, G. A., & Sanghera, G. S. (1990). Direct Voltammetry of the Chromatium vinosum Enzyme, Sulfide:Cytochrome c Oxidoreductase (Flavocytochrome c552)*. *The Journal of Biological Chemistry*(265), pp. 1958-63.

Haller, M. J., Atkinson, M. A., & Schatz, D. (2005, December). Type 1 Diabetes Mellitus: Etiology, Presentation, and Management. *Pediatric Clinics of North America*(52(6)), pp. 1553-78.

Harries, L. W., Ellard, S., Stride, A., Morgan, N. G., & Hattersley, A. T. (2006, July 15). Isomers of the TCF1 gene encoding hepatocye nuclear factor-1 alpha show differential expression in the pancreas and define the relationship between mutation position and clinical phenotype in monogenic diabetes. *Human Molecular Genetics*(15(14)), pp. 2216-2224.

Holmkvist, J., Cervin, C., Lyssenko, V., Winkler, W., Anevski, D., Cilio, C., . . . Groop, L. (2006, October 11). Common variants in HNF-1 α and risk of type 2 diabetes. *Diabetalogia, 49*, pp. 2882-2891. doi:https://doi.org/10.1007/s00125-006-0450-x

Kaisaki, P. J., Menzel, S., Lindner, T., Oda, N., Rjasanowski, I., Sahm, J., . . . Bell, G. I. (1997, March). Mutations in the Hepatocyte Nuclear Factor-1a Gene in MODY and Early-Onset NIDDM: Evidence for a Mutation Hotspot in Exon 4. *American Diabetes Association*(46(3)), pp. 528-35.

Karczewski, K. J., Franciololi, L. C., Tiao, G., Cummings, B. B., Alfödi, J., Wang, Q., . . . MacArthur, D. G. (2019, January 28). The mutational constraint spectrum quantified from variation in 141,456 humans. *bioRxiv*. doi:10.1101/531210

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014, January 1). *ClinVar: public archive of relationships among sequence variation and human phenotype*. Retrieved from Oxford Academic: https://academic.oup.com/nar/article/42/D1/D980/1051029

Lehto, M., Wipemo, C., Ivarsson, S. A., Lindgren, C., Lipsanen-Nyman, M., Weng, J., . . . Groop, L. (1999, August). High Frequency of Mutations in MODY and Mitochondrial Genes in Scandinavian Patients With Familial Early-Onset Diabetes. *Diabetologia*(42), pp. 1131-37.

Lin, J. D., Wan, H.-L., Li, J.-C., Wu, C.-Z., Kuo, S.-W., Hsieh, C.-H., . . . Pei, D. (2007, July 27). Impaired Glucose Tolerance and Impaired Fasting Glucose Share Similar Underlying Pathophysiologies . *The Tohoku Journal of Experimental Medicine*, pp. 349-357.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., . . . Lopez, R. (2019, July 01). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acid Research*(47), pp. 636-641.

Malikova, J., Kaci, A., Dusatkova, P., Aukrust, I., Torsvik, J., Vesela, K., . . . Bjørkhaug, L. (2020, April 1). Functional Analyses of HNF1A-MODY Variants Refine the Interpretation of Identified Sequence Variants. *J Clin Endocrinol Metab*(105(4)).

Mani, A. (2017, December 10). The Pathogenicity of De Novo Rare Variants, Challenges and Opportunities. *Circ Cardiovasc Genet*. Retrieved from The Pathogenicity of De Novo Rare Variants, Challenges and Opportunities.

Matsha, T. E., Hassan, M. S., Kidd, M., & Erasmus, R. T. (2012, February). The 30-year cardiovascular risk profile of South Africans with diagnosed diabetes, undiagnosed diabetes, pre-diabetes or normoglycemia: the Bellville, South Africa pilot study. *Cardiovascular Journal of Africa*(23), pp. 5-11.

McDonald, T. J., & Ellard, S. (2013, September). Maturity onset diabetes of the young: identification and diagnosis. *Ann Clin Biochem*(50), pp. 403-415.

McIntyre, D., Catalano, P., Zhang, C., Desoye, G., Mathiesen, E. R., & Damm, P. (2019, July 11). Gestational diabetes mellitus. *Nature Reviews Disease Primers*(5).

McKusick, V. A., & O'Neill, M. J. (1995, April 17). *Early-Onset Diabetes of the Young, Type 3; MODY3*. Retrieved April 29, 2020, from OMIM: https://www.omim.org/entry/600496

Metzger, B. E., & Coustan, D. R. (1998, August). Summary and recommendations of the Fourth International Workshop-Conference on Gestational Diabetes Mellitus. The Organizing Committee. *Diabetes Care*(21), pp. 161-167.

Miquerol, L., Lopez, S., Cartier, N., & Tulliez, M. (1994, April). Expression of the L-pyruvate kinase gene and the hepatocyte nuclear factor 4 transcription factor in expcrine and endocrine pancreas. *Journal of Biological Chemistry*(269(12)), pp. 8944-51.

Multiple Sequence Alignment. (n.d.). Retrieved from Clustal Omega: https://www.ebi.ac.uk/Tools/msa/clustalo/

Nagy, E., & Maquat, L. E. (1998, June 1). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Elsevier Science*(23(6)), pp. 198-199.

Najmi, L. A. (2018). *Role of rare HNF1A variants function in monogenic and type 2 diabetes*. University of Bergen.

National Institute of Diabetes and Digestive and Kidney Diseases. (2017, November). *Monogenic Diabetes (Neonatal Diabetes Mellitus & MODY)*. Retrieved from NIDDK: https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/monogenic-neonatal-mellitus-mody

NCBI. (2020, May 2). *What is ClinVar?* Retrieved from ClinVar: https://www.ncbi.nlm.nih.gov/clinvar/intro/

NCBI Gene. (2020, March 13). HNF1A. NCBI.

Nishi, S., Stoffel, M., Xiang, K., Shows, T. B., Bell, G. I., & Takeda, J. (1992). Human pancreatic Beta-cell glucokinase: cDNA sequence and localization of the polymorphic gene to chromosome 7, band p 13. *Diabetologia*(35), pp. 743-747.

OMIM. (2016, August 02). *Maturity-Onset diabetes of the Young; MODY*. Retrieved from Online Mendelian Inheritance in Man: https://www.omim.org/entry/606391

One Lambda. (n.d.). *Ion Chef System*. Retrieved May 2020, from One Lambda: https://www.onelambda.com/en/product/ion-chef-system.html

Osbak, K. K., Colclough, K., Saint-Martin, C., Beer, N. L., Bellannè-Chantelot, C., Ellard, S., & Gloyn, A. L. (2009, September 29). Update on Mutations in Glucokinase (GCK), Which Cause Maturity-Onset Diabetes of the Young, Permanent Neonatal Diabetes, and Hyperinsulinemic Hypologlycemia. *Wiley InterScience*(No. 11), pp. 1512-1526.

Pernegger, L., & Godehart, S. (2007, October). Townships in the South African Geographic Landscape - Physical and Social Legacies and Challenges. *Training for Township Renewal Initiative (TTRI)*.

Rees, A., Levy, M., & Lansdown, A. (2017). *Clinical Endocrinology and Diabetes at a Glance*. West Sussex, United Kingdom: John Wiley & Sons.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gaster-Foster, J., . . . ACMG, L. (2015, May). Standard and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommandation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*(17), pp. 405-424.

Risch, N. (1987). Assessing the Role of HLA-linked and Unlinked Determinants of Disease. *American Journal of Human Genetics*(40(1)), pp. 1-14.

Rodger, W. (1991, December 15). Non-insulin-dependent (type II) diabetes mellitus. *CMAJ Medical knowledge that matters*(145 (12)), pp. 1571-81.

Ryffel, G. U. (2001). Mutations in the human genes encoding the transcription factors of hepatocyte nuclear factor (HNF)1 and HNF4 families: functional and pathological consequences. *J. Mol. Endocrinol.*(27), pp. 11-29.

Sagen, J., Bjørkhaug, L., Molnes, J., Ræder, H., Grevle, L., Søvik, O., . . . Njølstad, P. (2008). Diagnostic screening of MODY2/GCK mutations in the Norwegian MODY Registry. *Pediatric Diabetes*(9), pp. 442-449.

Shepherd, M., Shields, B., Ellard, S., Rubio-Cabezas, O., & Hattersley, A. T. (2009, April 06). A genetic diagnosis of HNF1A diabetes alters treatment and improves control in the majority of insulin-treated patients. *Diabetic Medicine*(26).

Shields, B. M., Hicks, S., Shepherd, M. H., Colclough, K., Hattersley, A. T., & Ellard, S. (2010, May 25). Maturity-onset diabetes of the young (MODY): how many cases are we missing? *Diabetologia 53*, pp. 2504-2508.

Shields, B. M., McDonald, T. J., Ellard, S., Campbell, M. J., Hyde, C., & Hattersley, A. T. (2012, August 16). The development and validation of a clinical prediction model to determine the probability of MODY in patients with young-onset diabetes. *Diabetologia*(55), pp. 1265-72.

Stankov, K., Benc, D., & Draskovic, D. (2013, December). Genetic and Epigenetic Factors in Etiology of Diabetes Mellitus Type 1. *Pediatrics*(132(6)), pp. 1112-22.

Statitics South Africa. (2011). *Census 2011 Statistical release.*

Steele, A. M., Shields, B. M., Wensley, K. J., Colclough, K., Ellard, S., & Hattersley, A. T. (2014, January 15). Prevalence of Vascular Complications Among Patients With Gluckokinase Mutations and longed, Mild Hyperglycemia. *JAMA*(311), pp. 279-286.

Stoffel, M., & Shih, D. Q. (2001, December 4). Dissecting the transcriptional network of pancreatic islets during development and differentiation. *PNAS*(25), pp. 14189-91.

Stride, A., & Hattersley, A. T. (2009, July 08). Different genes, different diabetes: lessons from maturity-onset diabetes of the young. *Annals of Medicine*, pp. 207-216.

Szablewski, L. (2011). Glucose Homeostasis - Mechanism and Defects. In E. Rigobelo, *Diabetes - Damages and Treatments*. InTech.

Tallapragada, D. S., Bhaskar, S., & Chandak, G. R. (2015, August 7). New insights from monogenic diabetes for "common" type 2 diabetes. *Frontiers in Genetics*(6).

Tanizawa, Y., Koranyi, L. I., Welling, C. M., & Permutt, M. A. (1991, August 15). Human liver glucokinase gene: cloning and sequence determination of two alternatively spliced cDNAs. *PNAS*(88 (16)), pp. 7294-97.

Tattersall, R. B. (1974, April 2). Mild Familial Diabetes with Dominant Inheritance. *An International Journal of Medicine*, pp. 339-357.

Territory Organisations. (2012). Pre-diabetes (IGF and IGT). *A diabetes information series from Diabetes State*(29).

Thanabalasingham, G., & Owen, K. R. (2011, October 19). Diagnosis and management of maturity onset diabetes of the young (MODY). *theBMJ*(343), pp. 1-9. doi:10.1136/6044

Thermo Fischer Scientific G. (2019). *Torrent Suite^TM Software 5.12 User Guide.* Retrieved May 2020, from Thermo Fisher Scientific NGS Software Support: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0017972_031419_TorrentSuite_5_12_UG_.pdf

Thermo Fisher Scientific A. (n.d.). *Next-Generation Sequencing (NGS)*. Retrieved May 5, 2020, from Thermo Fisher Scientific: https://www.thermofisher.com/no/en/home/life-science/sequencing/next-generation-sequencing.html

Thermo Fisher Scientific B. (n.d.). *Ion Torrent Next-Generation Sequencing Technology*. Retrieved May 9, 2020, from ThermoFisher Scientific: https://www.thermofisher.com/no/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html

Thermo Fisher Scientific C. (n.d.). *Ion GeneStudio™ S5 System*. Retrieved May 2020, from Thermo Fisher Scientific: https://www.thermofisher.com/order/catalog/product/A38194#/A38194

Thermo Fisher Scientific D. (n.d.). *Ion Chef™ Instrument*. Retrieved May 5, 2020, from Thermo Fisher Scientific : https://www.thermofisher.com/order/catalog/product/4484177#/4484177

Thermo Fisher Scientific. (n.d.). *DNA Sequencing Methods*. Retrieved April 2, 2020, from ThermoFisher Scientific: https://www.thermofisher.com/no/en/home/life-science/sequencing/sequencing-education/next-generation-sequencing-basics/dna-sequencing-methods.html

Thermo Fisher Scientific E. (n.d.). *Torrent Suite Software & Ion Torrent Storage Devices*. Retrieved May 13, 2020, from Thermo Fisher Scientific: https://www.thermofisher.com/no/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-data-analysis-workflow/ion-torrent-suite-software.html

Thermo Fisher Scientific F. (n.d.). *Ion Torrent Suite Software Plugins*. Retrieved May 14, 2020, from Thermo Fisher Scientific: https://www.thermofisher.com/no/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-data-analysis-workflow/ion-torrent-suite-software-plugins.html

Thermo Fisher Scientific H. (2020, April 29). *Ion Reporter^TM Software 5.14 User Guide*. Retrieved May 2020, from Thermo Fisher Scientific NGS Software Support: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0019036_IonReporter5_14_UG.pdf

Thermo Fisher Scientific H. (n.d.). *Ion Reporter Software*. Retrieved May 13, 2020, from Thermo Fisher Scientific: https://www.thermofisher.com/no/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-data-analysis-workflow/ion-reporter-software.html

Thermo Fisher Scientific. (n.d.). *Next-Generation Sequencing (NGS)*. Retrieved April 2, 2020, from ThermoFisher Scientific: https://www.thermofisher.com/no/en/home/life-science/sequencing/next-generation-sequencing.html?gclid=EAIaIQobChMIpaKahOfJ6AIVE6WaCh30lQiZEAAYASAAEgI0EfD_BwE&cid=csd_ngs_sbu_r01_co_cp1300_pjt5043_csd00000_0se_gaw_nt_awa_ngsexact&s_kwcid=AL!3652!3!39377714

Thermo Fisher Scientifics G. (n.d.). *NGS Software Support*. Retrieved May 14, 2020, from Thermo Fisher Scientific: https://www.thermofisher.com/no/en/home/technical-resources/technical-reference-library/next-generation-sequencing-support-center/ngs-software-support.html

Valkovicova, T., Skopkova, M., Stanik, J., & Gasperikova, D. (2019). Novel insights into genetics and clinics of the HNF1A-MODY. *Endocrine Regulations*(53), pp. 110-134.

Vaxillaire, M., Abderrahmani , A., Boutin, P., Bailleul, B., Froguel, P., Yaniv, M., & Pontoglio, M. (1999). Anatomy of a homeoprotein revealed by the analysis of human MODY3 mutations. *J. Biol. Chem.*(274), pp. 35639-35646.

Warram, J. H., Krolewski, A. S., & Kahn, R. C. (1988, October). Determinants of IDDM and Perinatal Mrtality in Children of Diabetic Mothers. *American Diabetes Assosiation*(37(10)), pp. 1328-1334.

Wass, J., & Owen, K. (2014). *Oxford Handbook of Endocrinology and Diabetes* (3. ed.). Oxford, United Kingdom : Oxford University Press.

WHO and International Diabetes Federation. (2006). *Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia*. Geneva: WHO/International Diabetes Federation.

Wong, N. D., Fan, W., & Glovaci, D. (2019, March 04). Epidemiology of Diabetes Mellitus and Cardiovascular Disease. *Current Cardiology Reports*(21).

World Health Organization. (1999). *Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications: Report of a WHO Consultation. Part 1: Diagnosis and Classification of Diabetes Mellitus*. Geneva: World Health Organization.

World Health Organization. (2018, October 30). *Diabetes*. Retrieved from WHO: https://www.who.int/news-room/fact-sheets/detail/diabetes

Yamagata, K. (2003, October). Regulation of pancreatic beta-cell function by the HNF transcription network: lessons from maturity-onset diabetes of the young (MODY). *Endocrine Journal*(50(5)), pp. 491-9.

Yoshiuchi, I., Yamagata, K., Yang, Q., Iwahashi, H., Okita, K., Yamamoto, K., . . . Yamasaki T. (1999, April). Three new mutations in the hepatocyte nuclear factor 1-α gene in Japanese subjects with diabetes mellitus: clinical features and functional characterization. *Diabetologia*(42), pp. 621-626.

Zelent, D., Najafi, H., Odili, S., Buettger, C., Weik-Collins, H., Li, C., . . . Matschinsky, F. M. (2005, February). Glucokinase and Glucose Homeostatis: Proven Concepts and New Ideas. *Biochem Soc Trans*(33(Pt 1)), pp. 306-310.

Zhang, Z., Xin, D., Wang, P., Zhou, L., Hu, L., Kong, X., & Hurst, L. D. (2009, May 14). Noisy splicing, more than expression regulation, explains why some exons are subject to nonsense-mediated mRNA decay. *BMC Biology*(7).

6. APPENDIX

Appendix 1: Clinical features of participants