Curriculum Vitae

Haichen Zhang

Ph.D. Candidate

Human Genetic Track

Graduate Program in Life Sciences, University of Maryland, Baltimore

CONTACT INFORMATION

Hzhang_nov@outlook.com

EDUCATION

PhD University of Maryland School of Medicine — Baltimore MD
Human Genetics — July 2021(expected)

MS The George Washington University — Washington DC
Bioinformatics — May 2015

BS China Agricultural University — Beijing, China
Biology Science — July 2013

RESEARCH EXPERIENCE

University of Maryland School of Medicine, Baltimore MD — 2015 to current
Graduate Research Assistant

- Sequenced, analyzed sequencing data and curated variants of 295 patients for Personalized Diabetes Medicine Program (PDMP).
- Constructed and clinically validated NGS pipeline of whole-exome sequencing data for the Patient ENgaged Genomic Understanding and INterpretation (PENGUIN) study.
- Served as bioinformatician and biocurator in ClinGen Monogenic Diabetes Expert Panel and assisted in the collaboration of more than 30 institutions.
- Performed NGS data annotation, filtration and classified variants of 14 monogenic diabetes genes for 3,071 patients from Progress of Diabetes Genetics in Youth (ProDiGY) study.

George Washington University, Washington DC — 2013 to 2015
Graduate Research Assistant

- Performed TCGA cancer cell line NGS data analysis using various platforms and constructed phylogenetic tree of variants.

- Helping constructed and individually completed manual curation of RNA-seq database BioXpress.

PUBLICATIONS

Published

Zhang, H., Colclough, K., Gloyn, A. L., & Pollin, T. I. (2021). Monogenic diabetes: a gateway to precision medicine in diabetes. *Journal of Clinical Investigation*, *131*(3), e142244.

Wan, X., Perry, J., **Zhang, H.**, …, Fu, M. Heterozygosity for a pathogenic variant in SLC12A3 that causes autosomal recessive Gitelman syndrome is associated with lower serum potassium. *Journal of the American Society of Nephrology*, *32*(3), 756-765.

Zhang, H., & Pollin, T. I. (2018). Epigenetics Variation and Pathogenesis in Diabetes. *Current diabetes reports*, *18*(11), 121.

Wan, Q., Dingerdissen, H., Fan, Y., Gulzar, N., Pan, Y., Wu, T. J., Yan, C., **Zhang, H.** & Mazumder, R. (2015). BioXpress: an integrated RNA-seq-derived gene expression database for pan-cancer analysis. *Database*, *2015*.

Pan, Y., Karagiannis, K., **Zhang, H.**, Dingerdissen, H., Shamsaddini, A., Wan, Q., ... & Mazumder, R. (2014). Human germline and pan-cancer variomes and their distinct functional profiles. *Nucleic acids research*, 42(18), 11570-11588.

Xu, X.L., Yi, F., Pan, H.Z., Duan, S.L., Ding, Z.C., Yuan, G.H., Qu, J., **Zhang, H.C**. and Liu, G.H., 2013. Progress and prospects in stem cell therapy. *Acta Pharmacologica Sinica*, 34(6), pp.741-746.

Accepted

Goodrich J, Singer-Berk M, Son R, Sveden A, Wood J, England E, Cole JB, Weisburd B, Watts N, Zappala Z, **Zhang H**. Determinants of penetrance and variable expressivity in monogenic metabolic conditions across 77,184 exomes. Accepted by Nature Communications

Manuscript

Zhang, H., J.W. Kleinberger, et al. The Personalized Diabetes Medicine Program: identification, diagnosis and treatment of monogenic diabetes. (In preparation)

Zhang, H., Maloney, K.A., et al. Adaption of The ACMG/AMP Variant Interpretation Guidelines For GCK, HNF1A, HNF4A-MODY: Recommendations From The ClinGen Monogenic Diabetes Expert Panel. (In preparation)

Todd, J., Kleinberger, J., **Zhang, H.**, et al. Monogenic diabetes in the Progress for Diabetes Genetics in Youth (ProDiGY) Collaboration. (Submitted to Diabetes Care)

Poster Abstract (first author)

Zhang, H., et al. "1453-P: Adaption of the ACMG/AMP Variant Interpretation Guidelines for GCK, HNF1A, HNF4A-MODY: Recommendations from the ClinGen Monogenic Diabetes Expert Panel." Poster presented at: American Diabetes Association 80th Scientific session; 2020 June 12-16; Virtual meeting
Zhang, H., et al. "Gene-specific Variant Interpretation Guidelines In GCK, HNF1A, And HNF4A: Update From The ClinGen Monogenic Diabetes Expert Panel" Poster presented at: 2020 ACMG Clinical Genetics Meeting; 2020 March 17-21; Virtual meeting
Zhang, H., Pollin, T.I., Does 23andMe Raw Data Accurately Detect GCK-MODY? Poster presented at: The American Society of Human Genetics 2019 Annual Meeting: 2019 Oct 15-20; Houston, TX
Zhang, H., J.W. Kleinberger, et al. Cascade Screening of Personalized Diabetes Medicine Program (PDMP) Probands is Efficient in Diagnosing At-Risk Individuals of Monogenic Diabetes. Poster presented at: ACMG Annual Clinical Genetics Meeting; 2019 Apr 2-6; Seattle, WA. (**Top Rated**)
Zhang, H., J.W. Kleinberger, et al. The Personalized Diabetes Medicine Program: identification, diagnosis and treatment of monogenic diabetes. Poster presented at: American Society of Human Genetics; 2018 Oct 16-20; San Diego, CA.

OTHER ACADEMIC ACTIVITIES

Reviewer: Scientific Report, Oncology Letters

Conference moderator: 2019 Annual Meeting of the American Society of Human Genetics, Platform session #40: Methods and Resources in Large-scale Population Data

SERVICES

Vice President of UMB Chinese Student and Scholar Association 2016 to 2017

SKILLS

Programming: R, Python, SQL, Linux

Language: Mandarin, English

HONORS AND AWARDS

China Agricultural University Outstanding Graduate 2013

George Washington University Excellence in Biochemistry 2013

Abstract

Title of Dissertation: Genomic Medicine in Diabetes: Improving the Diagnostic Rate of Monogenic Diabetes

Haichen Zhang, Doctor of Philosophy, 2021

Dissertation directed by:

Toni I. Pollin, M.S., Ph.D., CGC, Associate Professor, Department of Medicine; Track Leader, Human Genetics and Genomic Medicine, Program in Epidemiology and Human Genetics, Graduate Program in Life Sciences

Monogenic diabetes is an uncommon type of diabetes caused by genetic defects in one of several genes, and it accounts for 1-2% of all diabetes. The primary subtypes are Maturity Onset Diabetes of the Young (MODY), neonatal diabetes, and syndromic diabetes. The correct treatment of each subtype of monogenic diabetes depends on the corresponding disease etiology that can only be confirmed by genetic testing. However, the diagnostic rate of monogenic diabetes is inadequate, mainly due to the overlapping phenotype of monogenic diabetes with type 1 diabetes and type 2 diabetes and lack of awareness among patients and physicians. To improve the diagnostic rate of monogenic diabetes, this project focuses on three aspects: 1) systematically screening of patients for genetic testing; 2) comprehensively re-analyzing next-generation sequencing (NGS) data from multiple diabetes cohorts; 3) assessing the ability of Direct-to-Consumer Genetic Testing (DTC-GT) raw data in detecting *GCK*-MODY variants. The Personalized Diabetes Medicine Program (PDMP) screened 2,522 patients with diabetes with a simple questionnaire, assigned patients to different algorithm criteria groups based on clinical

features, and performed genetic testing on suspected patients. Overall, 38 of 313 patients suspected of monogenic diabetes were tested positive for causative variants. The group of patients diagnosed before age 30 who were not treated with insulin had the highest pick-up rate. The comprehensive re-analysis of NGS panel data in PDMP, including re-classification and updating variant calling algorithm, improved the diagnostic rate from 11.82% to 13.10%. Also, the comparison between exome sequencing (ES) and NGS panel or Sanger sequencing of the Progress for Diabetes Genetics in Youth samples showed ES failed to identify all MODY-causing variants, but re-analysis of ES unfiltered data discovered the missing variants. By analyzing the *GCK* variants in the 23andMe DTC-GT raw data from 3,044 anonymous volunteers and calculating the ancestry-specific allele frequency of *GCK*-MODY variants, some of the variants showed higher-than-expected minor allele frequency compared with the large population database. Such inconsistency suggests customers should not use DTC-GT as a supplementary method of clinical genetic testing for *GCK*-MODY. In a summary, these studies provide practical approaches to improve the diagnostic rate of monogenic diabetes.

Genomic Medicine in Diabetes: Improving the Diagnostic Rate of Monogenic Diabetes

by
Haichen Zhang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Acknowledgements

The six and half year of pursuing the PhD degree was a long and lonely journey. I was extremely fortunate to work with, be friend with and learn from so many excellent people and I would like to take this opportunity to express my appreciation and respect.

Firstly, I want to thank my mentor, Dr. Toni Pollin, for her continuous patient guidance and support over the years. I am lucky to follow her path into the research of monogenic diabetes and become a researcher in the translational medicine field. This fulfills my expectation of a dream job that I could use my expertise to help people. As a mentor, she offered all the training experience I asked for and supported my ideas and decision all the time. As a scientist, she showed me the principles of research and establishing collaborations with other investigators. As a wonderful person, she taught me how to live a meaningful life by devoting her time to caring and helping others. I am grateful for everything I have learned from her.

Also, I would like to thank Kristin Maloney and Dr. Linda Jeng for their mentoring role during my Ph.D. time. I learned their meticulous attitude on work and kindness in life. They are always there when I need help and support. Dr. Jeffery Kleinberger, who taught me how to perform sequencing with patience, is another person I want to acknowledge. I will never forget the Friday afternoons when we sat together and curated variants.

This dissertation's completion attributes teamwork from everyone who has given time and effort to move these studies forward. I want to thank all the study leadership, coordinators, and staff of both the ProDiGY study, the PDMP, and the ClinGen Monogenic Diabetes Expert Panel. I also need to thank the administration and personnel

of the University of Maryland Translational Genomics Lab, where I performed the sequencing. I appreciate Devon Nwaba, Kathleen Palmer, Ruth Cosentino, Danielle Sewell for their ongoing support of the projects and my never-ending requests.

Additionally, I would like to thank my other Dissertation committee members for their support and guidance. I want to thank Dr. Elizabeth Streeten for her guidance in clinical genetics and for allowing me to shadow her in the clinic for several weeks. I want to thank Dr. Timothy O'Connor for mentoring me during my rotation and during my thesis. I want to thank Dr. Christy Chang for her encouragement and advice during my Ph.D. time. And I want to thank Dr. Yuji Zhang for teaching me how to be successful in graduate school.

It is also important to thank everyone in the Epidemiology and Human Genetics Graduate Program and the Department of Endocrinology, Diabetes, and Nutrition. I had the great opportunity to work closely with almost everyone in the dry labs, and I want to thank Brady Gaynor, Kathy Ryan, and James Perry for their selfless help.

Finally, I want to thank my parents for supporting me over the past eight years, financially and emotionally, though I could not be at home when needed. And I want to thank my friends who cared about me all these years. Special thanks to my two cats, Peanut and Biscuit, for their accompany and belief in me. Finally, I want to thank my boyfriend, Yafei Ou, who motivated me to be a better person with his endurance and unconditional love from hundreds of miles away. He supported and respected my decisions and cheered me up when I felt upset and desperate. I am deeply grateful for meeting him and having him in my life.

Table of Contents

Acknowledgements iii

Chapter 1 : Introduction 1

1.1 Monogenic Diabetes, a Gateway to Precision Medicine in Diabetes* 1

1.1.1 Abstract 1

1.1.2 Introduction 2

1.1.3 Conclusion 30

1.2 Summary and objective 31

Chapter 2 : Methods 33

2.1 Next-generation sequencing 34

2.2 NGS data analysis 35

2.3 Genotyping using DNA microarray 36

2.4 Methods for specific aims 37

Chapter 3 : Identifying a Model for Successful Integration of Monogenic Diabetes Diagnosis into Routine Care: The Personalized Diabetes Medicine Program* 44

3.1 Abstract 44

3.2 Introduction 45

3.3 Methods 48

3.4 Results 54

3.5 Discussion 71

3.6 Conclusion and Recommendations 73

Chapter 4 : Comprehensive Re-analysis of DNA Sequencing Data Improves Diagnostic Yield of Monogenic Diabetes 75

4.1 NGS Panel for the Diagnosis of Monogenic Diabetes: Comprehensive Re-analysis Improves Diagnostic Rate* 75

4.1.1 Abstract 75

4.1.2 Introduction 76

4.1.3 Materials and methods 81

4.1.4 Results 85

4.1.5 Discussion 91

4.2 Exome Sequencing in Identifying Monogenic Diabetes in Youth with Type 2 Diabetes: Lessons from the Progress for Diabetes Genetics in Youth (ProDiGY) Collaboration 96

4.2.1 Abstract 96

4.2.2 Introduction 97

4.2.3 Materials and methods 100

4.2.4 Results 104

4.2.5 Discussion 111

4.3 Summary 112

Chapter 5 : Direct-to-Consumer Genetic Testing Raw Data Has Limited Clinical Utility in Detecting *GCK*-MODY Variants* 114

5.1 Abstract 114

5.2 Introduction 115

5.3 Materials and methods 119

5.4 Results 122

5.5 Discussion 126

Chapter 6 : Discussion 129

6.1 Summary 129

6.2 Implications 132

6.3 Limitations 136

6.4 Recommendations and future directions 138

References 141

List of Tables

Table 1.1 Genetic causes of maturity-onset diabetes of the young[A] 5

Table 1.2 Genetic causes of neonatal diabetes with current ISPAD testing guidelines* . 13

Table 1.3 Biomarkers for monogenic diabetes detection 24

Table 2.1 Comparing different NGS technologies for clinical use 34

Table 2.2 Tools for sequencing data analysis 35

Table 3.1 PDMP inclusion criteria 48

Table 3.2 PDMP intake questionnaire 50

Table 3.3 Genes on the sequencing panel 53

Table 3.4 Distribution of individuals screened for the PDMP by recruitment method and site 56

Table 3.5 Demographic of PDMP participants (excluding missing data) 56

Table 3.6 Screening questionnaire responses of all PDMP participants 57

Table 3.7 Comparison of enrollment categories between African American and European American 58

Table 3.8 Demographics of individuals sequenced by 40-gene panel 59

Table 3.9 Total variants reviewed by gene and variants effects 59

Table 3.10 PDMP novel variants 60

Table 3.11 Screening questionnaire responses of patients with and without monogenic diabetes 64

Table 3.12 PDMP patients with positive genetic testing results 65

Table 3.13 Hit rate by recruitment methods and algorithm criteria 70

Table 4.1 Summary of published re-analysis studies on NGS data 80

Table 4.2 PDMP pathogenic or likely pathogenic variants identified through re-analysis 87
Table 4.3 PDMP benign or likely benign variants identified through re-analysis..... 87
Table 4.4 Comparing performance of splicing effect predictors on non-canonical splicing variants 89
Table 4.5 Results of GATK filtration based on decision tree cutoff 90
Table 4.6 Results of GATK filtration based on decision tree cutoff 90
Table 4.7 ProDiGY collaboration sample distribution 101
Table 4.8 Pathogenic or likely pathogenic variants filtered out by initial ES pipeline in TODAY and SEARCH study 107
Table 4.9 Classification and validation of unfiltered-data only variants 108
Table 5.1 Types and classification of *GCK* variants on 23andMe microarray chips 122
Table 5.2 Pathogenic or likely pathogenic *GCK* variants in PDMP participants included on 23andMe microarray chips 123
Table 5.3 *GCK* variants with more than one allele in openSNP samples..... 124
Table 5.4 Number of *GCK* variants in surveyed PDMP participants 125

List of Figures

Figure 1.1 Schematic representation of glucose-induced insulin secretion and MODY-associated genes. 10

Figure 1.2 Proposed diagnostic algorithm for monogenic diabetes. 22

Figure 2.1 Sequencing data analysis workflow 35

Figure 2.2 Ion Torrent platforms and sequencing workflow 39

Figure 2.3 Calling SNP/CNV alleles from raw data. 42

Figure 3.1 PDMP criteria algorithm 49

Figure 3.2 Flowchart of screened PDMP participants 55

Figure 4.1 Workflow of re-analyzing PDMP samples using the GATK pipeline 82

Figure 4.2 Decision trees of determining quality score cutoffs of GATK filtering. 82

Figure 4.3 Sanger sequencing confirmation of *GCK*:c.898G>A; p.(Glu300Lys) 91

Figure 4.4 Screenshot of BAM file of *GCK* p.(Glu300Lys) 95

Figure 4.5 Distribution of variants quality score of SNVs in ProDiGY 109

Figure 4.6 Distribution of variants quality score of indels in ProDiGY 110

Figure 5.1 Distribution of *GCK* variants on 23andMe microarray chips 122

Figure 5.2 Venn diagrams of *GCK* variants interpreted by third-party tools 126

Figure 6.1 Proposed approaches to improve the diagnostic rate of monogenic diabetes 138

List of Abbreviations

ABMGG: American-Board of Medical Genetics and Genomics

ACMG: American College of Medical Genetics

AD: Autosomal dominant

ADA: American Diabetes Association

AMP: Association for Molecular Pathologists

AR: Autosomal recessive

BDD: Better Diabetes Diagnosis

BMI: Body mass index

BVAMC: Baltimore Veterans Administration Medical Center

CAP: College of American Pathologists

CGL: Congenital generalized lipodystrophy

CHI: Congenital hyperinsulinemic hypoglycemia

CLIA: Clinical Laboratory Improvement Amendments of 1988

ClinGen: Clinical Genome

CNS: Central nervous system

DAA: Diabetes autoantibodies

DEND: Developmental delay, and early-onset epilepsy

DM: Diabetes Mellitus

DS: Donohue syndrome

DTC-GT: Direct-to-consumer genetic testing

ExAC: Exome Aggregation Consortium

eGFR: estimated Glomerular filtration rate

EHR: Electronic health record

ER: Endoplasmic reticulum

ES: Exome sequencing

ESP: Exome Sequencing Project

FPLD: Familial partial lipodystrophy

GAD: Glutamic acid decarboxylase

GATK: Genome Analysis Tool Kit

GDM: Gestational diabetes mellitus

GHS: Geisinger Health System

GIAB: Genome in a Bottle

GS: Genome sequencing

GWAS: Genome-wide association study

HbA1c: Hemoglobin A1c

HDL: High-density lipoprotein

HGDP: Human Genome Diversity Project

HIPAA: Health Insurance Portability and Accountability Act of 1996

IA-2: Islet antigen 2

IGNITE: Implementing Genomics in Practice Network

Indels: Insertions and deletions

IPEX: Immunodysregulation, polyendocrinopathy, and enteropathy, X-linked

ISPAD: International Society for Pediatric and Adolescent Diabetes

IUGR: Intrauterine growth retardation

LADA: Latent autoimmune diabetes of adulthood

LDL: Low-density lipoprotein

LOF: Loss-of-function

MAF: Minor Allele Frequency

MAVEs: Multiplexed assays of variant effect

MDEP: Monogenic Diabetes Expert Panel

MIDD: Maternally-inherited diabetes and deafness

MODY: Maturity-onset diabetes of the young

NDM: Neonatal diabetes mellitus

NGS: Next-generation sequencing

OHA: Oral antihyperglycemic agents

OR: Odds ratio

PDMP: Personalized Diabetes Medicine Program

PGM: Personalized Genome Machine

PNDM: Permanent neonatal diabetes mellitus

PPV: Positive predictive value

ProDiGY: Progress for Diabetes Genetics in Youth

QC: quality control

RCAD: Renal cysts and diabetes Syndrome

RMS: Rabson-Mendenhall syndrome

SBS: Sequencing-by-synthesis

SBL: Sequencing-by-ligation

SD: Standard deviation

SNP: Single nucleotide polymorphism

SNV: Single nucleotide variant

T1D: Type 1 diabetes

T2D: Type 2 diabetes

TGL: Translational Genomics Lab

TMAP: Torrent Mapping Alignment Program

TNDM: Transient neonatal diabetes mellitus

TODAY: Treatment Options for type 2 Diabetes in Adolescents and Youth

TVC: Torrent Variant Caller

Type A-IR: type A insulin resistance

UM-CDE: University of Maryland Center for Diabetes and Endocrinology

VQSR: Variant Quality Score Recalibration

VUS: Variant of uncertain significance

WS: Wolfram syndrome

ZFP: Zinc finger protein

Chapter 1 : Introduction

1.1 Monogenic Diabetes, a Gateway to Precision Medicine in Diabetes[1]*

1.1.1 Abstract

Monogenic diabetes refers to diabetes mellitus (DM) caused by a mutation in a single gene and accounts for approximately 1-5% of diabetes. Correct diagnosis is clinically critical for certain types of monogenic diabetes, since the appropriate treatment is determined by the etiology of the disease (e.g., oral sulfonylurea treatment of *HNF1A/HNF4A* diabetes vs. insulin injections in type 1 diabetes [T1D]). However, achieving a correct diagnosis requires genetic testing, and the overlapping of the clinical features of monogenic diabetes with those of T1D and type 2 diabetes (T2D) has frequently led to misdiagnosis. Improvements in sequencing technology are increasing opportunities to diagnose monogenic diabetes, but challenges remain. In this review, we describe the types of monogenic diabetes, including common and uncommon types of maturity onset diabetes of the young (MODY), multiple causes of neonatal DM, and syndromic diabetes such as Wolfram syndrome and lipodystrophy. We also review methods of prioritizing patients undergoing genetic testing and highlight existing challenges facing sequence data interpretation that can be addressed by forming collaborations of expertise and by pooling cases.

[1] Zhang H, Colclough K, Gloyn AL, Pollin TI. Monogenic diabetes: a gateway to precision medicine in diabetes. *J Clin Invest.* 2021;131(3):e142244. doi:10.1172/JCI142244. Format has been modified to meet the dissertation requirement. License number of reprint permission: 1128340-1
*HZ drafted the manuscript.

1.1.2 Introduction

Monogenic diabetes is caused by a single defect in one of over 40 genes[1,2]. Since MODY (maturity onset diabetes of the young) was named by Fajans for the T2D-like presentation in young people with an autosomal dominant pattern of inheritance[3,4], our understanding of phenotypic and genetic heterogeneity in monogenic diabetes has increased. The major monogenic diabetes categories are MODY, neonatal diabetes mellitus (NDM), and syndromic diabetes[5]. Misdiagnosis is frequent because of overlapping of phenotypes with T1D, such as young onset and leanness, and with T2D, such as preserved β-cell function and family history. Tailored treatment of some monogenic diabetes depends on the disease's underlying etiology, e.g., oral sulfonylurea treatment of *HNF1A*/*HNF4A* MODY, and requires genetic testing to diagnose. Here we will describe monogenic diabetes types, etiologies, diagnosis, management, and strategies to improve diagnosis.

Monogenic vs. polygenic diabetes

Monogenic and polygenic diabetes are traditionally considered distinct, with monogenic diabetes resulting from one highly penetrant variant in one gene in a given individual, and polygenic diabetes resulting from the contribution of several variants with smaller effects in the context of environmental/lifestyle factors. In T1D, autoimmune dysfunction is the prominent mechanism, with variation in the major histocompatibility locus and other genomic factors combining with apparent environmental triggers to result in beta cell loss and diabetes. In monogenic diabetes, highly penetrant variants, mostly causing extremely impaired beta cell development and insulin secretion, cause diabetes regardless of other risk factors. T2D, sometimes considered a diagnosis of exclusion, is a

heterogeneous group of disorders involving smaller genetic effects on multiple mechanisms including insulin secretion and insulin sensitivity, combining with environmental and lifestyle factors, mostly impacting insulin sensitivity. While this distinction is important both scientifically and clinically, emerging studies of the genetic architecture of diabetes reveal more of a spectrum with respect to the penetrance of genetic variants and their relative role in diabetes. For example, the *HNF4A* variant p.R114W, found in 0.02% of non-Finnish Europeans, has been shown to be over-represented in patients with MODY (OR = 30.4 vs. public variant databases) but to have a distinct clinical phenotype (including lack of sulfonylurea response) and much lower penetrance than other *HNF4A* MODY mutations (54% vs. 71% by age 30[6]). In Mexican Americans, *HNF1A* variant p.E508K (NM_000545.8, rs483353044) was associated with T2D with a much greater effect size (OR = 4.2) than most polygenic T2D variants, with diabetic carriers and non-carriers having similar onset age and BMI[7]. T2D polygenic risk scores have also shown evidence of modifying age at diagnosis of monogenic diabetes[8]. Finally, while lack of features of either autoimmunity or obesity/metabolic syndrome raises the likelihood of monogenic diabetes, these features can co-exist with monogenic diabetes, particularly obesity given its high prevalence especially in youth. In the Treatment Options for Diabetes in Adolescents and Youth (TODAY) clinical trial in which overweight or obesity was required for the newly diagnosed youth with T2D enrolled, at least 4.5% were identified as having MODY. Those with *HNF4A*-MODY had poor response to metformin, representing a previously missed opportunity for optimal treatment[9]. In summary, monogenic and polygenic forms of diabetes exist along more of a continuum than previously appreciated. Therefore, knowledge about

monogenic diabetes not only provides opportunities for etiology-based treatment of the minority of individuals with highly penetrant variants, but also informs broader understanding of diabetes etiology.

Types of monogenic diabetes

Maturity onset diabetes of the young (MODY)

MODY comprises most monogenic diabetes cases, with classical characteristics of young diagnosis age, family history of diabetes in an autosomal dominant pattern of transmission, and insulin independence, with some types having additional features (Table 1.1). While 14 genes have now been designated as MODY genes in OMIM and/or the literature, three of these (*BLK*, *PAX4*, and *KLF11*) have been proposed for elimination based on a recent study[10] (https://search.clinicalgenome.org/kb/affiliate/10016?page=1&size=25&search= , see Table 1.1 for remaining 11 along with *RFX6*, recently proposed as an additional MODY gene[11]). Variants in *GCK*, *HNF1A,* and *HNF4A* are responsible for most MODY cases, followed by *HNF1B*[12]. Given the known genetic etiology of most MODY cases the increased frequency of pediatric T2D due to increased childhood overweight and obesity prevalence, it has been suggested that this term be abandoned in favor of terms describing the etiology of the type of diabetes, such as transcription factor diabetes for MODY caused by mutations in the transcription factors genes *HNF1A*, *HNF4A*, *HNF1B,* and others[13]. Moreover, it can be argued that any diabetes designation is unsuitable for the usually benign condition of heterozygous GCK deficiency, which is characterized by only mildly elevated glucose levels often not reaching the diabetic range and, more to the point, generally does not lead to diabetic micro- and macrovascular complications[14].

Table 1.1 Genetic causes of maturity-onset diabetes of the young[A]

Gene/ MODY Number (OMIM Number)	Age of onset	Treatment	Distinguishing features	Pathophysiology	Reference
HNF4A/ MODY1 (125850)	Pre-adolescence to young adulthood	Low-dose sulfonylureas	Neonatal macrosomia and hyperinsulinemic hypoglycemia	Transcription factor defect disrupting beta-cell development and function	[15]
GCK/ MODY2 (125851)	Birth	None (except during pregnancy when insulin may be required if fetus does not inherit variant)	Lifelong mild fasting hyperglycemia, low post-prandial glucose increment; usually neither responds to nor requires treatment. No increased risk of micro and macrovascular complications.	Impaired beta-cell glucose sensing	[16]
HNF1A/ MODY3 (600496)	Pre-adolescence to young adulthood	Low-dose sulfonylureas	Low renal threshold for glucose	Transcription factor defect disrupting beta-cell development and function	[15]
PDX1/ MODY4 (606392)	Young adulthood	OHA or insulin	pancreatic agenesis in homozygotes/ compound heterozygotes (rare)	Transcription factor defect disrupting beta-cell development and function	[15,17-20]
HNF1B/ MODY5 (137920)	Pre-adolescence to young adulthood	Insulin	Renal structural abnormalities, genital-tract malformations, pancreatic hypoplasia, hypomagnesemia, abnormal liver function, intellectual disabilities	Transcription factor defect disrupting beta-cell and renal cell development and function	[21]
NEUROD1 / MODY6 (606394)	Young adulthood	Insulin	Reduced penetrance for diabetes.	Transcription factor defect disrupting beta-cell development and function	[22]

Table 1.1 Continued

CEL/ MODY8 (609812)	Young adulthood	Insulin	pancreatic exocrine dysfunction in childhood and diabetes in later adulthood with multiple pancreatic cysts	Unclear	[23-25]
INS/ MODY10 (176730)	Childhood to young adulthood	Diet or insulin	Specific mutations cause insulin dependence due to ER stress mediated beta cell apoptosis	Proinsulin misfolding and ER stress or impaired insulin activity	[26-32]
ABCC8/ MODY12 (600509)	Pre-adolescence to young adulthood	High-dose sulfonylureas	Neurological abnormalities in some cases	beta-cell ATP-sensitive potassium channel defect	[30-32]
KCNJ11/ MODY13 (616329)	Childhood	High-dose sulfonylureas	Neurological abnormalities in some cases	beta-cell ATP-sensitive potassium channel defect	[23-25,30-32]
RFX6	Varies	Diet, insulin or OHA		Transcription factor defect disrupting beta-cell function	[11]

A: *BLK*, *PAX4* and *KLF11,* although classified as MODY genes (#11, #9, and #7 respectively) in OMIM, are not listed as MODY-causing because of recently disputed or refuted gene-disease relationships (see section "Rare types of MODY"). *APPL1* was proposed as MODY14 based on two families reported in 2015 [33], but evidence is otherwise limited. *RFX6* does not have a MODY number in OMIM, but is included here as multiple loss-of-function variants were recently implicated in a phenotype very similar to that of other MODY genes but with lower penetrance [11]. OHA: Oral Hypoglycemia Agents

Common types of MODY-classified monogenic diabetes

HNF1A-MODY and *HNF4A*-MODY are caused by variants in genes encoding HNF1 homeobox A and hepatic nuclear factor 4α, respectively. These transcription factors play essential roles in transcription of genes related to β-cell development and insulin secretion. *HNF1A* variants decrease expression of HNF1A target genes[34]. Among patients diagnosed with diabetes, *HNF1A*-MODY is the most common MODY. To date over 400 *HNF1A* variants and 100 *HNF4A* variants have been discovered from MODY families[15].

HNF1A/HNF4A-MODY is usually diagnosed in adolescence or early adulthood. Compared with T2D, *HNF1A*-MODY and *HNF4A*-MODY occur at younger ages with lower BMI, lower HbA1c, and triglycerides, and similar risk for microvascular complications. Approximately 50% of patients with *HNF4A*-MODY are macrosomic, which is attributed paradoxically to transient neonatal hyperinsulinemic hypoglycemia at birth[35]. Hyperinsulinemic hypoglycemia was also recently observed in some patients with *HNF1A*-MODY[36].

Individuals with *HNF1A*- and *HNF4A*-diabetes have increased sensitivity to sulfonylureas, an insulin-stimulating class of drug[37,38] such that low doses are effective, and more typical T2D doses cause hypoglycemia. Sulfonylreas bind to the subunit of the K_{ATP} channel to depolarize the β-cell and release insulin. In a randomized clinical trial, low doses of sulfonylureas (e.g., 20-40 mg gliclazide daily) produced better glucose control than metformin in *HNF1A*- and *HNF4A*-MODY[39]. In an observational study, most patients with presumed T1D who were subsequently found to have *HNF1A*-diabetes gained glycemic control when treatment was changed from insulin to sulfonylureas[37].

Glucagon-like peptide-1 receptor agonists monotherapy[40] or sulfonylurea in combination with dipeptidyl peptidase-4 inhibitor [41] were recently demonstrated to achieve good glycemic control in *HNF1A*-MODY with reduced or no hypoglycemia, suggesting possible utility as *HNF1A*/*HNF4A*-diabetes first line treatment.

GCK encodes glucokinase, an enzyme catalyzing glucose phosphorylation at glycolysis initiation. GCK is a pancreatic β-cell glucose sensor; genetic defects change the glucose-stimulated insulin secretion threshold[16,42]. In the United Kingdom, the prevalence of *GCK*-hyperglycemia was estimated at 0.1% among white Europeans[43] - higher than that of *HNF1A*-diabetes because the lack of symptoms keeps many cases from coming to medical attention. Further studies are needed in other populations. *GCK*-hyperglycemia has limited phenotypic heterogeneity; most patients have lifelong mild, persistent, and asymptomatic fasting hyperglycemia within the prediabetes range[44]. Their HbA1c values usually do not exceeding 7.5% (60 mmol/mol)[45], though some have glucose levels that meet the DM criteria, and a few have T2D and related complications, likely due to additional genetic and environmental risk factors[46,47]. Glucose levels are resistant to lowering by insulin or oral agents. Moreover, since *GCK*-hyperglycemia does not appear to be associated with significant microvascular and macrovascular diabetes complications[14,48], patients with *GCK*-hyperglycemia usually do not require glucose-lowering medication, except possibly during pregnancy. Maternal *GCK* mutations increase risk for macrosomia and associated perinatal complications similarly to gestational or pre-gestational diabetes of any type owing to the excess insulin secretion in response to a hyperglycemic intrauterine environment. Fetal *GCK* mutations decrease birthweight as a result of poor insulin response. A paternally inherited fetal mutation

places the fetus at risk for low birthweight in a normoglycemic intrauterine environment. A maternal mutation creates a hyperglycemic intrauterine environment for fetal insulin secretion needed for normal growth of a GCK-deficient fetus, and thus attempts at normalizing maternal glucose may result in harm. In pregnant women with *GCK* mutations, it is recommended that at minimum fetal growth be monitored by serial ultrasound to guide treatment, but it is ideal to know the fetal mutation status early in pregnancy[49]. A non-invasive technique for determining fetal *GCK* mutation status from cell-free DNA in maternal circulation is being developed that will enable women with a mutation-positive fetus to be discharged from high-risk antenatal care[50].

HNF1B variants are estimated to account for less than 1% of MODY[51]. Patients with *HNF1B* defects may exhibit early-onset DM only; diabetes with renal, pancreas, or liver phenotypes (renal cysts and diabetes [RCAD] syndrome); or other features with or without diabetes, such as neurodevelopmental disorders[52,53] and hypomagnesemia. *HNF1B* genotype-phenotype correlation is currently unclear, with clinical heterogeneity even among family members with the same variant. However, renal outcome as measured by estimated glomerular filtration rate [eGFR]) has been reported to be better in deletion versus nondeletion variants[53,54], hypothesized to result from a dominant-negative effect[55,56]. Some *HNF1B*-MODY initially responds to sulfonylurea or repaglinide[53] but may ultimately require insulin.

<u>Rare MODY-classified monogenic diabetes types</u>

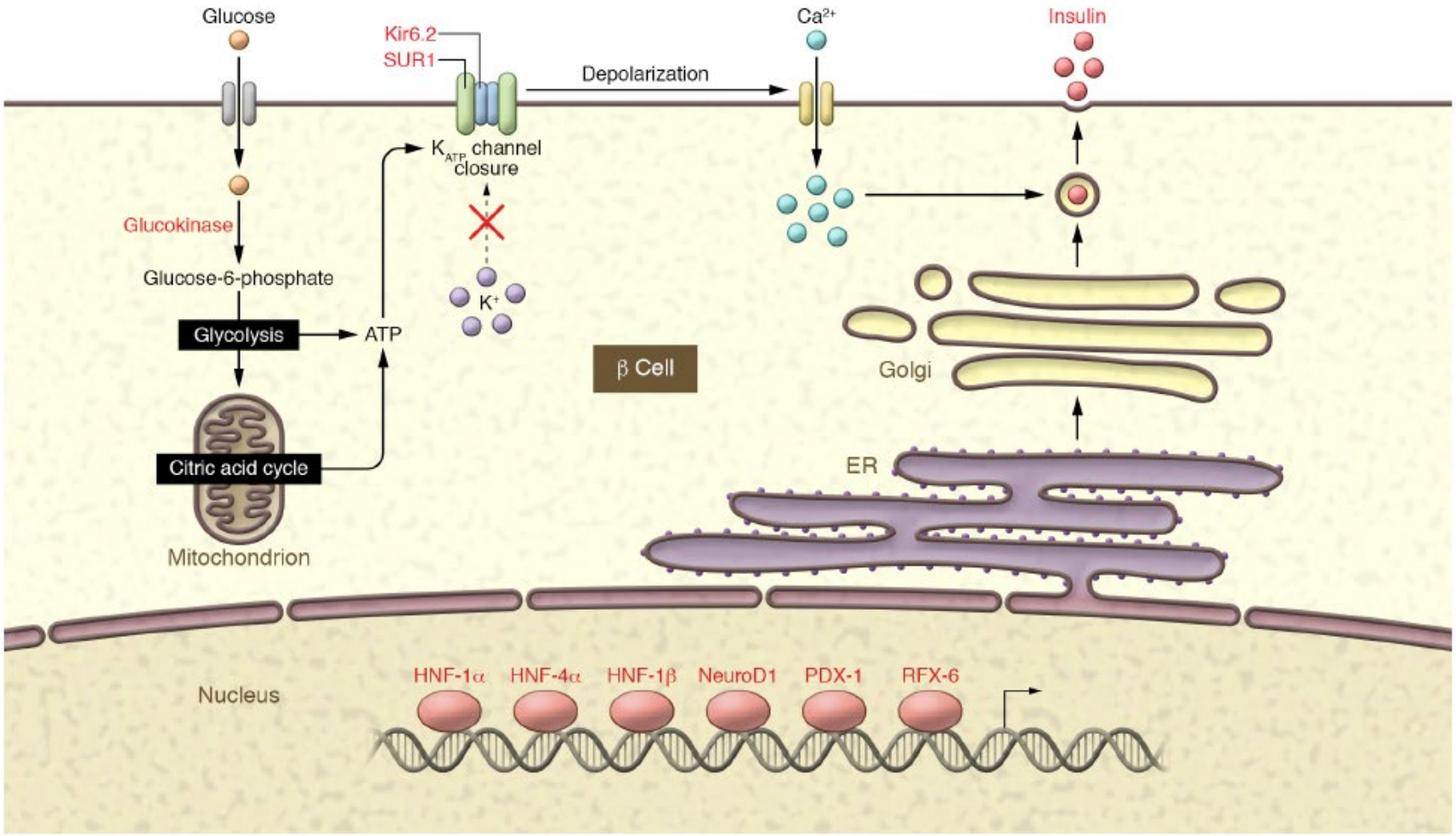
ATP-sensitive potassium channel (K_{ATP} channel) Diabetes

Pathogenic variants in *ABCC8* and *KCNJ11*, the genes encoding the sulfonylurea receptor 1 (SUR1) and the inward rectifying potassium channel 11 (Kir6.2), subunits of

the ATP-sensitive potassium channel (K_{ATP}) found in β-cells (Figure 1.1), are common causes of NDM (either permanent or transient, see below) but also can occasionally cause diabetes with later childhood or young adult onset (sometimes referred to as MODY12[57] and MODY13[58], respectively). K_{ATP} diabetes is discussed further in the NDM section below.

The prominent and rarer types of MODY and their genetic and clinical features are summarized in Table 1.1. Emerging findings obtained through next-generation sequencing (NGS) to identify new causes of MODY have suggested potential roles of *APPL1*[33] and *PCBD1*[59] in MODY.

Figure 1.1 Schematic representation of glucose-induced insulin secretion and MODY-associated genes. The pancreatic KATP channel directly regulates insulin secretion. It is a hetero-octamer formed by four subunits of the inward rectifier potassium channel 11 (Kir6.2, encoded by KCNJ11) and four sulfonylurea receptor 1 (SUR1, encoded by ABCC8) subunits. Glucose enters the β cell and glucokinase phosphorylates glucose to glucose-6-phosphate, which further breaks down in the glycolysis and citric acid cycle to produce ATP. The increased ATP/MgATP ratio leads to the closure of the KATP channel and causes depolarization of the β cell membrane and subsequent activation of voltage-gated calcium channels. Calcium flows into the cell through activated voltage-gated calcium channel and triggers the insulin to be released from the β cell. Transcription factors (HNF1A, HNF4A, HNF1B, NEUROD1, PDX1, and RFX6) constitute a network that regulates the expression of insulin and β cell development and proliferation. The MODY-associated genes are labeled in red.


Glucose
Kir6.2
SUR1
Depolarization
Ca2+
Insulin
KATP channel closure
Glucokinase
Glucose-6-phosphate
Glycolysis
ATP
K+
β Cell
Golgi
Citric acid cycle
ER
Mitochondrion
Nucleus
HNF-1α
HNF-4α
HNF-1β
NeuroD1
PDX-1
RFX-6

NDM

NDM is defined as diabetes diagnosed within the first 6 months of age and can either be permanent (PNDM) or transient (TNDM). Clinical features of NDM also include intrauterine growth retardation (IUGR), failure to thrive, polyuria, and severe dehydration [60,61]. Depending on the genetic etiology, some patients can also have birth defects and neurological disorders[62]. It affects 1/90000 to 1/260000 livebirths[63,64], with 50% being PNDM and 50% being TNDM[60].

The diabetes phenotype in TNDM results from inadequate insulin production presenting at the first week of life and resolving by 18 months[60], but 50% of patients relapse during early adulthood[62]. Approximately 60-70% of TNDM is caused by overexpression of paternally expressed imprinted genes on chromosome 6q24 (hereafter referred to as 6q24-TNDM) resulting from paternally inherited duplications or paternal disomy for the region or chromosome (both copies inherited from the father)[65]. The remaining cases mostly result from mutations in K_{ATP} channels: *KCNJ11*[66] and *ABCC8*[67], which tend to be functionally less severe than those causing PNDM [68]. There are also rare occurrences attributed to mutations in *INS* (encoding insulin precursor molecule preproinsulin)[67,69,70], *HNF1B*[21,71], and other genes (Table 1.2). It remains undetermined why only some TNDM patients relapse later, but theories about b-cell function and the development of insulin resistance at puberty represent possible explanations[72].

Table 1.2 Genetic causes of neonatal diabetes with current ISPAD testing guidelines*

Gene	Phenotype	Inheritance	Other features	Pathophysiology	Reference
ABCC8	PNDM, TNDM	AD	Developmental delay, epilepsy (DEND)	β-cell dysfunction	30-32
EIF2AK3	PNDM	AR	Wolcott-Rallison syndrome	β-cell destruction	73
FOXP3	IPEX	XLR	IPEX syndrome	β-cell destruction	74
GATA4	PNDM, TNDM	AD	Pancreatic agenesis, congenital cardiac defects, developmental delay, neurocognitive defects	Abnormal pancreatic development	75
GATA6	PNDM	AD	Pancreatic agenesis, congenital cardiac defects, congenital biliary tract anomalies	Abnormal pancreatic development	76,77
GCK	PNDM	AR		β-cell dysfunction	16,78-80
GLIS3	PNDM	AR	Congenital hypothyroidism, IUGR, polycystic kidney disease	Abnormal pancreatic development	81-83
HNF1B	PNDM, TNDM	AD	Pancreatic hypoplasia and renal cyst	Abnormal pancreatic development	21,71
IER3IP1	PNDM	AR	Microcephaly, simplified gyral pattern, severe epilepsy	β-cell destruction	84
INS	PNDM, TNDM	AD, AR		β-cell destruction	27-29,69,85-89
KCNJ11	PNDM, TNDM	AD	Developmental delay, epilepsy (DEND)	β-cell dysfunction	30-32
MNX1	PNDM	AR	Developmental delay, sacral agenesis, imperforate anus, IUGR	Abnormal pancreatic development	90
NEUROD1	PNDM	AR	Developmental delay, cerebellar hypoplasia, sensori-neural deafness, and visual impairment	β-cell dysfunction	22
NEUROG3	PNDM	AR	Malabsorptive diarrhea	Abnormal pancreatic development	91-93

Table 1.2 Continued

NKX2-2	PNDM	AR	Developmental delay, hypotonia, short stature, deafness, constipation	Abnormal pancreatic development	[90]
PAX6	PNDM	AR	Brain anomalies, microphthalmia	Abnormal pancreatic development	[94-96]
PDX1	PNDM	AR	Pancreatic agenesis (common)	β-cell dysfunction	[17,18,97-102]
PLAGL1/HYMAI	TNDM	Imprinting	Macroglossia, umbilical hernia	Abnormal pancreatic development	[103]
PTF1A	PNDM	AR	Pancreatic agenesis, cerebellar agenesis	Abnormal pancreatic development	[104-106]
RFX6	PNDM	AR	Pancreatic hypoplasia, intestinal atresia, and gallbladder aplasia or hypoplasia (Mitchell-Riley syndrome)	Abnormal pancreatic development	[107-110]
SLC2A2	PNDM	AR	Fanconi-Bickel syndrome	β-cell dysfunction	[111]
SLC19A2	PNDM	AR	Roger syndrome	β-cell dysfunction	[112]
WFS1	PNDM	AD, AR	Wolfram syndrome	β-cell destruction	[113,114]
ZFP57	TNDM	AR	IUGR, microglossia, facial dysmorphism, cardiac anomalies, umbilical hernia, and developmental delay	Abnormal pancreatic development	[115]

TNDM; transient neonatal diabetes mellitus; PNDM: permanent neonatal diabetes mellitus; AD: autosomal dominant; AR: autosomal recessive; XLR: X-linked recessive; IUGR: intrauterine growth restriction. *The Exeter diabetesgenes.org website maintains an up to date, annotated list of genes sequenced for monogenic diabetes in their UK national laboratory which currently includes 71 genes, including some in putative/research status of which 35 are evaluated for neonatal diabetes, including those listed here. In addition to MODY and NDM genes, the list also contains genes for syndromic subtypes diagnosed outside of the neonatal period. The list is maintained at https://www.diabetesgenes.org/tests-for-diabetes-subtype

Some of the same genes implicated in TNDM, including *ABCC8*, *KCNJ11*, and *INS,* also have variants more commonly causing PNDM. Homozygous or compound heterozygous inactivating *GCK* mutations cause PNDM[16] (Table 1.2). PNDM can also be part of IPEX (immune dysregulation, polyendocrinopathy, enteropathy, X-linked) syndrome caused by variants in *FOXP3* gene, Wolcott-Rallison syndrome caused by variants in *EIF2AK3,* and others [116]. Unlike more common types of PNDM, the PNDM of IPEX syndrome (and a few other rare types of monogenic diabetes [117,118]) is autoimmune, as FOXP3 is crucial in maintaining regulatory T cells' normal function of inhibiting proliferation and cytokine production of other T cells[74]. PNDM in Wolcott-Rallison syndrome is possibly due to increased pancreatic β-cell apoptosis that is regulated by *EIF2AK3*[73]. Genes implicated in NDM and associated phenotypes are listed in Table 1.2. Since a genetic cause has only been identified in 82% of patients with NDM[119], the search continues through exome sequencing, most recently implicating *YIPF5* in autosomal recessive neonatal diabetes and microcephaly[120].

Patients diagnosed with diabetes in the first 6 months of life are recommended to have immediate genetic testing to identify the subtype, since T1D is extremely rare in this subgroup. Approximately 80-85% of NDM cases have an identifiable genetic cause[119], half of which have K_{ATP}-diabetes caused by *KCNJ11* or *ABCC8* mutations, treatable with high dose sulfonlyureas rather than insulin[66,121,122]. The benefit of identifying patients with K_{ATP}-diabetes is thus considerable, and many studies have attempted to establish genotype-phenotype correlation[123] to facilitate the prediction of patients' clinical courses based on genetic data.

NDM caused by pathogenic variants in K_{ATP} channels

Activating K_{ATP} channel gene variants cause NDM by decreasing ATP's ability to achieve channel closure in multiple ways[124]. Whether variants will cause PNDM or TNDM (or rarely, MODY), is determined in part by the functional severity of the mutation as well as which gene is involved, with *KCNJ11* variants mainly associated with PNDM and most *ABCC8* variants linked to TNDM[67,125]. Diabetes severity could be partially explained by the extent to which the variant impacts ATP sensitivity[126]; however, the same variant in one family could cause both NDM and MODY in different patients (e.g., the *KCNJ11* C42R variant[127]), suggesting that other mechanisms influence the development of clinical presentation. Loss-of-function (LOF) mutations in both genes cause an increase in insulin secretion and present as congenital hyperinsulinemic hypoglycemia (CHI) when found in the homozygous or compound heterozygous state[128,129] and when dominant LOF mutations are found in the heterozygous state. In addition, paternally-inherited recessive LOF mutations in combination with somatic loss of maternal 11p15.5 chromosomal region cause focal hyperinsulinism[130].

K_{ATP}-NDM is autosomal dominantly inherited but often (60%-84%) arises de novo[31,131]. Some individuals with NDM have neurological features in addition to DM[132], as K_{ATP} channels are expressed in other tissues, including muscle and brain. Common K_{ATP}-NDM features include muscle weakness, developmental delay, and early-onset epilepsy (DEND syndrome), while those with intermediate DEND (iDEND) syndrome do not have epilepsy[133]. Treatment of K_{ATP}-NDM caused either *KCNJ11* or *ABCC8* variants with high-dose sulfonylureas has proven safe and effective for both short-term and long-term glycemic control and may resolve CNS features[134-137]. 90% of K_{ATP}-NDM

patients could switch from insulin therapy to sulfonylurea successfully[134,138], with mutation severity [132,139,140] and diabetes duration before the transition[141] predicting the likelihood of success. For patients who cannot completely transfer to sulfonylurea, combining insulin and sulfonylurea has shown favorable results[142].

6q24-TNDM

Although patients with 6q24-TNDM always present with growth retardation and hyperglycemia during the neonatal period, different etiologies, including paternal uniparental disomy, partial duplication of paternal origin, or a methylation defect of maternal origin on 6q24, all lead to the overexpression of *PLAG1* and *HYMAI*, encoding a zinc finger protein (ZFP) and long noncoding RNA, respectively. In other cases, *ZFP57* variants cause hypomethylation of multiple imprinted loci, including at the 6q24 locus. The treatment for the first onset of diabetes is insulin, and many are treated with insulin during remission, while some are successfully treated with sulfonylureas or a combination of sulfonylureas and insulin[143,144]. Compared with K_{ATP}-TNDM patients, patients with 6q24-TNDM were observed to have lower birthweight and earlier presentation[145]. Some patients with 6q24 TNDM may also experience hyperinsulinemic hypoglycemia following diabetes remission[146]. Certain congenital abnormalities, such as macroglossia, are characteristic of 6q24-TNDM and thus could help to distinguish this type of TNDM from other types when considering testing strategies.

Syndromic diabetes

In addition to RCAD syndrome due to *HNF1B* variants as described above, other forms of monogenic syndromes include DM as one of the clinical features. We describe the most well-characterized of these syndromes below.

Wolfram syndrome

Two types of Wolfram syndrome (WS) corresponding to two causative genes have been identified to date. Wolfram syndrome 1 (WS1), characterized by diabetes insipidus, DM, optic atrophy, and deafness, is a rare autosomal recessive disease caused by variants in wolframin ER transmembrane glycoprotein (*WFS1*). Severe cases with dominant heterozygous variants are also reported. Often, patients' first manifestation is DM at an average age of 6 years. Though most WS1 patients require daily insulin as therapy, the high morbidity and mortality rate as well as low average age of death make an accurate and timely diagnosis essential. Recently, a presentation similar to WS1 in many *WFS1* mutation-negative patients was linked to variants in CDGSH iron sulfur domain 2 (*CISD2*) and thus named Wolfram syndrome 2 (WS2). Clinical features of patients with WS2 resemble WS1 but without diabetes insipidus and with the addition of peptic ulcer bleeding and defective platelet aggregation[147]. In addition, there are some *WFS1* mutations that cause isolated diabetes with significantly reduced or non-penetrance for other WS related features.[113,114]

Insulin resistance due to insulin receptor defects

Genetic defects in insulin receptor (*INSR*) result in several insulin resistance syndromes, which are distinguished from typical insulin resistance not only by their severity but by normal lipid profiles because the etiology is directly due to defects in

insulin receptor signaling rather than obesity and its sequelae [148]. The most common type is type A insulin resistance syndrome (type A-IR), which has autosomal dominant (AD) and autosomal recessive (AR) forms. Type A-IR affects predominantly non-obese females and presents with extreme insulin resistance, acanthosis nigricans, hirsutism, and polycystic ovarian disease[149,150]. Rabson-Mendenhall syndrome (RMS) is an intermediate form of insulin resistance with AR inheritance. Patients with RMS have clinical features of extreme insulin resistance, acanthosis nigricans, hirsutism, dental precocity, thick nails, pineal hyperplasia, genital enlargement in both males and females, abdominal distension, and other distinctive dysmorphic features[151,152]. The most severe form is Donohue syndrome (DS), an AR disorder in which patients present with failure to thrive, severe hyperinsulinemia, and fasting hypoglycemia. Patients with DS seldom survive infancy[153]. LOF variants in the fibronectin type III (FnIII) domain are proposed to be associated with more severe DS, and there are genotype-phenotype and structure-phenotype correlations of *INSR* variants[154].

Lipodystrophy

Monogenic lipodystrophy is a group of diseases featuring a complete or partial lack of adipose tissue and adipose tissue-derived hormones, which results in insulin resistance and other metabolic complications. Unlike insulin receptor defects, the lack of adipose tissue in lipodystrophy leads to dyslipidemia and insulin resistance due to spillover of fat into ectopic areas, paradoxically similar to the consequences of obesity[148] . Based on the loss of adipose tissue, this disease can be divided into congenital generalized lipodystrophy (CGL) and familial partial lipodystrophy (FPLD). CGL is an AR disease; pathogenic variants in genes encoding 1-acylglycerol-3-phosphate

O-acyltransferase 2 (*AGPAT2*) and Berardinelli-Seip congenital lipodystrophy 2 (*BSCL2*) account for most CGL cases, with rare cases being caused by pathogenic variants in *CAV1* and *PTRF*. CGL patients show common features, such as generalized lipodystrophy, muscular appearance, DM, and dyslipidemia; however, patients with pathogenic *BSCL2* variants display lower serum leptin levels than patients with pathogenic *AGPAT2* variants[155] but a higher rate of developing intellectual disability[156]. The majority of FPLD cases are caused by pathogenic variants in lamin A/C (*LMNA*) or PPARg (*PPARG*), and there are also other rarer forms caused by pathogenic variants in *PLIN1, AKT2*, *LIPE*, *CIDEC,* and *PCYT1A*. Body fat deficiency in FPLD is found on limbs, buttocks, and hips. Patients with pathogenic variants of either *LMNA* or *PPARG* appear to benefit similarly from leptin replacement therapy with metreleptin[157] in terms of improved glycemia and cardiometabolic outcomes.

Mitochondrial diabetes

Mitochondrial diabetes, also known as maternally inherited diabetes and deafness (MIDD), is caused by pathogenic variants in mitochondrial DNA, mostly tRNA variant m.3243A>G. Patients often present with diabetes in adulthood, but a greater proportion of mutated mitochondrial genomes in the affected tissues is associated with a younger age of diagnosis of diabetes in some studies[158]. Other clinical features such as macular pattern dystrophy, nephropathy, and neurological symptoms are more common in rarer forms of mitochondrial diabetes than the classical form[159]. The penetrance of mitochondrial diabetes is estimated to be nearly 100% by the age of 70 years. The disease etiology determined that patients have impaired insulin secretion, and insulin treatment is eventually required for most patients. The effects of other treatments, such as coenzyme

Q10 and PPARg agonists, were only evaluated in single cases, thereby requiring caution for application. To better screen patients suspected to have mitochondrial diabetes, clinical features including diabetes and hearing loss on the maternal side are key. Tian et al. established a mitochondrial diabetes score system with good performance (100% sensitivity, 69.9% specificity) to select patients diagnosed with T2D for genetic testing in a Chinese cohort[160], although this system needs validation in other populations.

Challenges in identifying and diagnosing monogenic diabetes

The broad application of personalized medicine to patients with monogenic diabetes faces challenges in two aspects: detecting patients suspected of having monogenic diabetes to pursue etiology-based therapies, and accurately interpreting sequence variants of monogenic diabetes genes.

Monogenic diabetes detection methods

At present, there have been limited practical guidelines for systematic screening for monogenic diabetes. The International Society for Pediatric and Adolescent Diabetes (ISPAD) has recommended testing for NDM in all patients diagnosed with diabetes before the age of 6 months as well as in patients diagnosed with diabetes before the age of 12 months with negative islet antibodies. This recommendation not only has the potential to dramatically improve care at the individual level when K_{ATP}-diabetes is diagnosed but has been shown to be cost-effective in this population[161]. However, adult and pediatric populations with T1D and T2D, which also contain misdiagnosed patients with monogenic diabetes[162], are more challenging to screen routinely for MODY[161] and can be challenging especially for clinicians with limited experience diagnosing MODY. More complex screening criteria based on age of onset, family history, endogenous

insulin secretion, non-obesity, and absence of pancreatic autoantibodies and/or needed to achieve cost-effectiveness and an ideal balance of sensitivity and specificity[163-165]. The American Diabetes Association recommends some scenarios for considering testing individuals not fitting into the T1D or T2D classifications[166]. A proposed algorithm to increase sensitivity is shown in Figure 1.2; clinicians are referred to the primary source[167], as well as current ADA[166] and ISPAD guidelines[1] for further guidance; additional development is needed and is ongoing in this area.

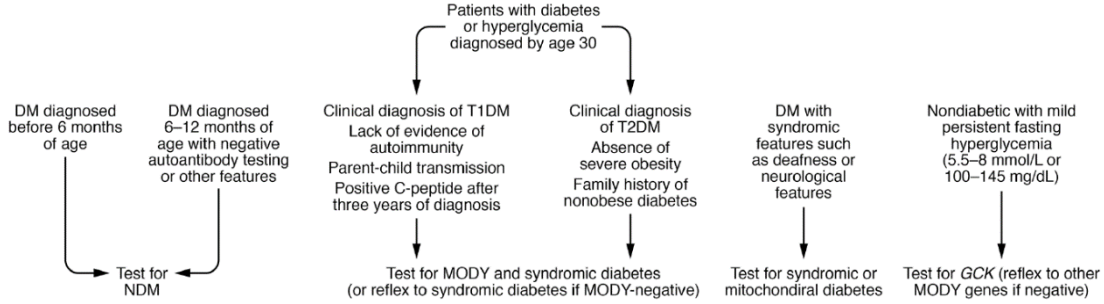


Figure 1.2 Proposed diagnostic algorithm for monogenic diabetes.
Though the majority of patients diagnosed between 6 and 12 months have T1D, NDM can exist in these patients; genetic testing should be considered if they test negative for autoantibody, have extra-pancreatic features, or have unusual family history [1,2]. High prevalence of MODY was observed in C-peptide–positive T2D diagnosed before 30 years regardless of metabolic syndrome status [167].

Biomarkers or derived scores avoid reliance on clinical judgments and arbitrary cutoffs and establish a quantitative evaluation that could be validated and replicated across cohorts. The Swedish Better Diabetes Diagnosis (BDD) study showed that absence of glutamic acid decarboxylase (GAD), islet antigen-2, zinc transporter 8 antibodies, and insulin autoantibodies could be a good discriminator, since in this study, MODY patients were only identified from the antibody-negative group, and 15% of antibody-negative patients had MODY[165]. However, other studies have shown that 1-2% of patients

diagnosed with MODY are GAD-antibody positive[168], reducing the antibody's sensitivity as a screen. Meanwhile, the types of autoantibodies tested on each patient may vary depending on the clinic; thus, using negative antibodies as a screening method may not be practical without standardization. Table 1.3 summarizes published biomarkers other than pancreatic antibodies that have been utilized to distinguish monogenic diabetes subtypes from T1D or T2D. Limited by the low prevalence of monogenic diabetes, these biomarkers were developed in selected populations to differentiate the most common types of MODY.

In addition to biomarkers, Shields et al. established a MODY calculator predicting the possibility of testing positive for MODY given a set of common clinical criteria [169]. In the initial cohort of white European patients who were diagnosed before the age of 35, the cutoff of probability at 40% yields sensitivity of 96% and specificity of 91% in differentiating MODY from T2D and yields 87% sensitivity and 88% specificity for MODY versus T1D. Validations in other cohorts with different ancestral backgrounds show variable outcomes, suggesting room for improvement, including the need for a more ancestrally diverse reference database.

Table 1.3 Biomarkers for monogenic diabetes detection

Standard biomarkers	**Cut-off value**	**Differentiation**	**Sensitivity**	**Specificity**	**Reference**
Fasting C-peptide	≥0.62 ng/ml	MODY vs T1D in children and adolescents	93%	90%	[170]
Random or glucagon-stimulated C-peptide	≥0.2 ng/ml	*HNF1A*- and *HNF4A*-MODY vs T1D in adults	100%	96%	[167]
Autoantibodies	GADA< 99th percentile	*HNF1A*, *HNF4A* and *GCK*-MODY vs T1D	99%	62%	[168]
	IA-2 < 99th percentile	*HNF1A*, *HNF4A* and *GCK*-MODY vs T1D	100%	57%	
	GADA and IA-2 < 99th percentile	*HNF1A*, *HNF4A* and *GCK*-MODY vs T1D	99%	82%	
	GADA and/or IA-2 and/or ZnT8A < 99th percentile	*HNF1A*, *HNF4A* and *GCK*-MODY vs not-known MODY in children	100%	88%	[165]
	GADA and/or IA-2 and/or ZnT8A and/or IAA < 99th percentile	*HNF1A*, *HNF4A* and *GCK*-MODY vs not-known MODY in children	100%	89%	
Proposed biomarkers	**Cut-off value**	**Differentiation**	**Sensitivity**	**Specificity**	**Reference**
Serum 1,5 anhydroglucitol (1,5AG)	>11µg/ml	*GCK*-MODY vs T2D	75%	75%	[171]
	>7.5µg/ml	*GCK*-MODY vs *HNF1A*-MODY	86%	84%	
Highly sensitive C-reactive protein (hsCRP) ***Standard in UK**	≤0.5 mg/l	*HNF1A*-MODY vs T2D	>74%	>68%	[172]
	≤0.75 mg/l	*HNF1A*-MODY vs T2D	79%	71%	[173]
	≤0.75 mg/l	*HNF1A*-MODY vs T1D	79%	67%	
	≤0.4 mg/l	*HNF1A*-MODY vs T2D	71%	77%	[174]
	≤0.28 mg/l	*GCK*-MODY vs T1D	67%	68%	[175]
	≤0.2 mg/l	*HNF1A*-MODY vs T1D	80%	80%	
	≤0.185 mg/l	*HNF1A*-MODY vs *GCK*-MODY	80%	75%	
	≤0.81 mg/l	*HNF1A*-MODY vs non *HNF1A*-MODY young adult–onset nonautoimmune diabetes	88%	69%	[176]

Table 1.3 Continued

Urinary C-peptide creatinine ratio (UCPCR)	≥0.2 nmol/mmol	*HNF1A/4A* MODY vs T1D	97%	96%	[177]
	≤3.1 nmol/mmol	*HNF1A/4A* MODY vs T2D	81%	44%	
	≥0.22 nmol/mmol	MODY vs T1D in children and adolescents	96%	86%	[170]
HDL cholesterol	≤1.56 mM	*GCK*-MODY vs T1D and HNF1A-MODY	86%	55%	[178]
Plasma Glycan GP30	<0.7%	*HNF1A*-MODY vs non *HNF1A*-MODY young adult–onset nonautoimmune diabetes	88%	80%	[176]

Selecting method and genes for testing

Previously, molecular diagnosis of monogenic diabetes was usually performed through Sanger sequencing of one or several genes based on clinical suspicion. With the development of NGS, tens of genes or even a patient's whole exome or even whole genome can be analyzed simultaneously. NGS panels for monogenic diabetes typically include all the MODY genes, or at least the most common ones, as well as the NDM and syndromic forms of diabetes genes. There are both advantages and disadvantages to using NGS panels. The low price of massively parallel sequencing enables the analysis of additional genes that were reported to be associated with syndromic forms of diabetes. This is useful because patients with syndromic forms of diabetes may lack or appear to lack the clinical features that would lead to testing of a single syndromic gene[179]. However, it is important that diagnostic NGS panels not include genes with weak or disputed associations with monogenic diabetes, or if included for surveillance purposes, results not be reported [180]. The yields of these panels will not only facilitate molecular diagnosis but also add rare or novel variants to the knowledge base for future studies. Sanger confirmation is sometimes needed after variant discovery in NGS panels, though

increasingly less so except in difficult regions of the genome. Regardless of methodology, it is becoming increasingly clear that evaluating only exonic regions will overlook some causal variants, as variants in the noncoding regulatory, deep intronic regions, and 5' and 3' UTRs have also been implicated in monogenic diabetes[15,181].

Searching for monogenic diabetes using exome or genome sequencing for research purposes enables novel gene discovery and also requires caution. The coverage of exome sequencing may not be complete, leading to the risk of false negatives[182,183]. In addition, as exome or genome sequencing could discover variants that are potentially important to health or reproduction, but are unrelated to the clinical indication, the reporting of such secondary findings must be addressed, with consideration of the recommendations of organizations such as the American College of Medical Genetics and Genomics (ACMG)[184]. With these caveats, this approach can serve as a powerful tool for searching for candidate genes in patients with monogenic diabetes for whom variants in known genes have not been found[185].

Variant classification

Key to diagnosing monogenic diabetes and other genetic conditions is not only identifying the variant, but distinguishing disease-causing variants from normal variation. Previous approaches to determine if a variant identified in a patient was disease-causing involved sequencing a group of matched controls (usually 100-200 people) to assess the variant's presence in the general population. This approach was limited because the sample size was too small to rule out population prevalence being too high for the disease, e.g., *HNF1A*-diabetes has an estimated population prevalence of 1/10,000. Moreover, the extent to which rare but benign genetic variation existed in the population

studied was not known and was thus probably underestimated. As NGS has begun to boom, the problem of large quantities of genetic data for interpretation has arisen for genetic diseases in general. The genetic and phenotypic heterogeneity of monogenic diabetes, as well as its overlapping features with T1D and T2D, together increase the difficulty of interpreting the pathogenicity of variants found in patients suspected to have monogenic diabetes. On the other hand, NGS emergence has led to the availability of exome and genome sequences of over 100,000 individuals of diverse ancestries in the gnomAD database, dramatically improving the ability to assess variant frequency in the general population. Additional resources have emerged including computational predictive tools[186-188], and other sources of data including phenotype specificity, familial segregation, and functional studies are also used. However, there is subjectivity in assigning pathogenicity to variants, and in the early 2010s, a lack of consistency of variant interpretation across laboratories became apparent.

In 2015, the ACMG and the Association of Molecular Pathologists (AMP) jointly published a consensus recommendation on standards and guidelines for clinical genomic variant interpretation[189]. The guidelines were developed through data sharing by a large number of American Board of Medical Genetics and Genomics (ABMGG)-certified clinical molecular geneticists and pathologists from Clinical Laboratory Improvement Amendment/ College of American Pathologists (CLIA/CAP)-accredited laboratories. The recommendations suggested that variants could be assigned to a five-tier system of classification: (i) pathogenic, (ii) likely pathogenic, (iii) uncertain significance, (iv) likely benign, or (v) benign. The proposed sets of criteria include population data, computational and predictive data, clinical data, functional data (*in vitro* studies), and

pedigree segregation. Each criterion is weighted by different levels of strength based on observed evidence and combined with other collected criteria to reach a conclusion. Since the publication of the initial ACMG/AMP guidelines, additional refinements have been published to improve rigor, including recommendations for evaluating the strength of evidence for LOF [190], standards for assessing functional studies[191], and applying a Bayesian quantitative point system[192].

Value of establishing gene-specific rules

The aim of ACMG/AMP guidelines is to provide a universal set of criteria for interpreting variants for Mendelian disease. Additionally, each gene-disease pair requires further specification to reflect the specific disease frequency, clinical features, and genotype-phenotype relationships. In 2013, the Clinical Genome (ClinGen) Resource was founded by National Human Genome Research Institute to serve as a knowledge base which defines gene-disease relationships and curate variants of genetic disease in a standardized approach and distribute information about the variants to researchers and clinicians. Since then, dozens of expert panels and working groups have been formed to examine specific gene or disease groups for determining clinical significance and constructing gene-specific standardizations. The Monogenic Diabetes Expert Panel (MDEP), established in 2017, brings together experts and data to adapt the ACMG/AMP variant interpretation guidelines for monogenic diabetes genes and classify variants using these gene-specific rules, thereby improving the accuracy of variant classification in these genes and in turn improving the ability to accurately diagnose monogenic diabetes[193].

Value of data-sharing

Establishing guidelines is fundamental to standardized and concordant interpretation of monogenic diabetes gene variants. This process calls for expertise in endocrinology, molecular genetic testing, genetic counseling, and biochemistry. To reach the full potential of precision medicine in monogenic diabetes, centralizing case-level data is important. For instance, when the evaluating variant is not observed in the general population but in affected individuals, a higher number of occurrences brings to a higher level of evidence supporting pathogenicity. However, the uncommonness of monogenic diabetes often makes it difficult for individual laboratories to acquire enough cases. By pooling case data, expert panels can achieve levels of case-based evidence for pathogenicity not possible for any single laboratory or clinic.

Value of functional evidence

Well-established functional studies on variants boost the understanding of disease mechanisms and provide evidence supporting or disputing the pathogenicity of the variants. Studies have shown that functional analyses clarify variant interpretation in *HNF1A*-MODY variants, especially when family segregation data or phenotype data are not available[194]. Caution is needed when utilizing these data because not all functional assays reflect the disease mechanism and not all variants impact the function in the same way. Fully inspecting the consequences of a variant may require multiple assays to reach a conclusion[195]. Enacting systematic validation and statistical quantification on the level of strength of pathogenicity or benignity on functional assays is recommended[191]. This approach encourages high-throughput mutation screenings, such as saturation mutagenesis[196] and systematic functional profiling of variants identified in the

population[197,198], which consist of pathogenic and benign variants. The MDEP is currently developing standards for evaluating evidence from luciferase assays for transactivation, which assess transcriptional activity of *HNF1A* and *HNF4A* variants, along with assays of DNA binding activity and protein expression[198,199]. For *GCK* variants, similar work is focused on the relative activity index of glucokinase as a measure of enzyme kinetic characteristics[200]. In the longer term, multiplexed assays of variant effect (MAVEs) could provide comprehensive catalogues of allelic effects which can be interrogated to aid variant interpretation. This approach is particularly well suited for transcription factors such as HNF1A. It is important to note that functional evidence does not single-handedly implicate a variant in disease; the functional data must be evaluated in concert with the population and clinical data to make a pathogenicity determination.

1.1.3 Conclusion

Accurate genetic diagnosis of monogenic diabetes is crucial for patients, since it helps optimize treatment, especially for some patients switching from insulin or metformin to low dose sulfonylureas (*HNF1A*-MODY and *HNF4A*-MODY), receiving no treatment (*GCK*-MODY), or switching from insulin to high dose sulfonylureas (K_{ATP}-diabetes). Additionally, accurate monogenic diabetes diagnosis leads to better familial risk management and clinical course prediction. Advancement in genetic testing technology has increased the capacity of genetic diagnosis while decreasing sequencing cost. However, before we can offer genetic testing to every patient with diabetes, prioritizing patients with high suspicion of monogenic diabetes through assessment of their biomarker profiles or probability score is more practical. Monogenic diabetes

provides an example of translating research findings into clinical practice that improves diagnosis and quality of life. Multidisciplinary expert collaboration and case sharing combined with incorporating basic science into sequence variant interpretation will lead to improved diagnosis. Establishing clear guidelines for evaluating the causality of individual variants by this process is essential for widespread diagnosis of monogenic diabetes; more broadly, routine incorporation of emerging genomic data into the care of diabetes and disease in general is needed to realize the full potential of personalized and precision medicine. And as we celebrate the 100th anniversary of insulin's discovery, it seems fitting to now celebrate and disseminate our more recently discovered ability to identify individuals who can make their own insulin once they have received the appropriate genomic diagnosis and treatment.

1.2 Summary and objective

In summary, monogenic diabetes has served as an example of implementing genetic testing in diagnosis and treatment. However, the diagnostic rate of monogenic diabetes is inadequate. Patients with monogenic diabetes are frequently misdiagnosed as T1D or T2D, and they are put on unnecessary treatment or treatment that is not the best choice based there their disease etiology. The barricades of improving the diagnostic rate come from many aspects, including but not limited to suboptimal patients screening strategies, absence of re-analyzing sequencing data of patients with negative genetic testing results, and lack of cheap and accurate genetic testing (Figure 1.3). The overlapping clinical features between monogenic diabetes and T1D and T2D determined that it is difficult for the physicians to easily pick up the patients who will test positive. Therefore, screening strategies with good sensitivity and specificity to select patients with

a high possibility of positive genetic testing results are urgently needed. Also, the patients who are suspicious of monogenic diabetes sometimes cannot access genetic testing. Some of these patients would seek options other than clinical genetic testing such as direct-to-consumer genetic testing. Finally, only a low percentage of genetically tested patients get positive results. The reasons for negative genetic testing results vary, but some of the negative results can be rescued by sequencing data re-analysis.

This dissertation proposed three aims to target aspects of the mentioned obstacles. The first aim evaluated the method of screening patients for monogenic diabetes with a simple questionnaire and genetically diagnosing the suspected patients with NGS panel in the Personalized Diabetes Medicine Program (PDMP) study. The second aim concentrated on re-analysis of sequencing data to discover disease-causing variants missed in the initial analysis of PDMP NGS panel data and Progress for Diabetes Genetics in Youth (ProDiGY) Collaboration ES data. Finally, the third aim assessed the ability of direct-to-consumer genetic testing, such as 23andMe genetic testing, in detecting monogenic diabetes-causing variants, especially *GCK*-MODY. Overall, this dissertation provides practical information to improve the diagnostic rate of monogenic diabetes.

Chapter 2 : Methods

It is fundamental for genetics and genomics to detect the exact DNA sequence or RNA sequence. The technologies that detect genetic variants, such as sequencing and genotyping technologies, are essential to further understand the relationship between human genomes and disease. Over the past several decades, Sanger sequencing (also known as first-generation sequencing), NGS, and third-generation sequencing techniques emerged and were popularized by different platforms and gradually applied to clinical diagnostics. These techniques and platforms utilized various sample preparation, processing, and signal detection methods to achieve the same goal of accurately detecting the DNA or RNA sequence, but with different benefits and limitations. The accompanying analytic tools, which process the large quantities of sequencing data and produce a readable report for users, designed by diversified algorithms, also have their own advantages and disadvantages. Briefly, there is no single sequencing or genotyping method, or bioinformatics tool could reach a perfect sensitivity and specificity. It is the user's responsibility to carefully select, optimize and validate the platforms and tools and to fully understand the limitation of the results.

In this dissertation, I proposed three aims focusing on screening strategies, comprehensive re-analysis, and utilizing DTC-GT to detect genetic variants. This chapter describes the sequencing and genotyping techniques and the bioinformatics analysis used in each aim and explains the limitation of these techniques and analytic tools.

2.1 Next-generation sequencing

NGS, also called massively parallel DNA sequencing, is featured by multiplexing, which means a mixed library of DNA templates can be sequenced simultaneously. It often contains several stages, including library preparation, template amplification, and sequencing. Different platforms use different technologies in each step [201]. Before the sequencing reactions, the DNA library can be constructed by multiplexed PCR (amplicon-based method) or enriched by hybridization and capture (capture-based method). The multiplexed reactions are achieved by in vitro amplification, such as bridge amplification, emulsion PCR, and rolling circle amplification. Different platforms use various sequencing-by-synthesis (SBS) or sequencing-by-ligation (SBL) strategies to generate and detect signal for DNA sequences with the amplified multiplexed DNA library. NGS panel (one gene or many genes consisting of a panel), ES (entire exome of a human genome), and GS (entire human genome) are different types of NGS to meet various needs of laboratories. Their features are summarized and compared in Table 2.1.

Table 2.1 Comparing different NGS technologies for clinical use

Sequencing method	**Genome sequencing**	**Exome sequencing**	**NGS panel**
Sequencing region	Whole genome	Customized regions targeting all exonic regions	Customized regions targeting selected exonic regions
Sequencing depth	>30X	>50X-100X	>500X
Detecting variants	SNVs, INDELs, CNVs and SVs	SNVs, INDELs, CNVs, and SVs in coding regions	SNVs and INDELs in targeted regions
Cost	High	Moderate	Low
Analytical complexity	Need a proper filter for variants and to interpret deep intronic variants	Need a proper filter for variants	Less complicated for analysis
Incidental findings	Likely	Likely	Unlikely

2.2 NGS data analysis

The sequencing data processing is critical for downstream annotation and interpretation. It usually involves reads alignment, pre-processing, variant calling, and filtration (Figure 2.2). This chapter will only focus on tools for germline variants summarized in Table 2.2, though many tools specifically look at somatic mutations. Most of these are open-source tools, but some DNA sequencer manufacturers also have their compatible bioinformatics tools to provide one-stop analysis after generation of sequencing data, such as the Ion Torrent Suite software[202].

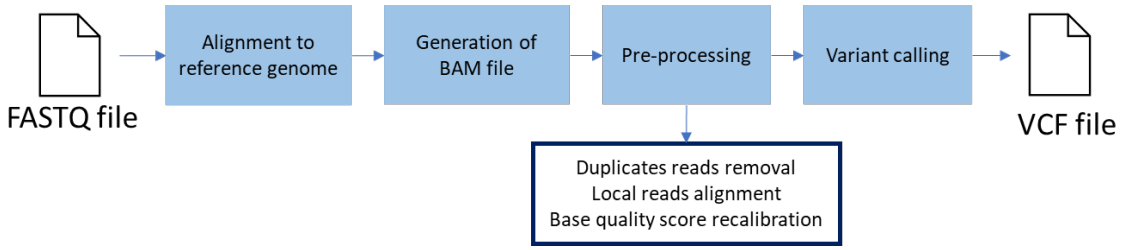


Figure 2.1 Sequencing data analysis workflow
The figure shows the steps needed from sequencing data of FASTQ format to the output VCF file. The pre-processing steps are recommended but not required. The output VCF file is then ready for annotation, filtration, and further interpretation.

Table 2.2 Tools for sequencing data analysis

Steps	**Tools**
Alignment and pre-processing	
Reads alignment	BWA, Bowtie, mrFAST and mrsFAST, Novoalign, minimap2
Marking duplicates	Picard, SAMTools, Sambamba, SAMBLASTER
BAM file creation	SAMTools, GATK
Sequencing metrics	BEDTools, Picard, QualiMap2
Sample quality control	KING, VerityBamID
Variant calling	
Germline SNVs/Indels	FreeBayes, GATK HaplotypeCaller, Platypus, SAMTools/BCFTools
Copy number variants	cn.MOPS CONTRA, CoNVEX, ExomeCNV, ExomeDepth, XHMM
Variant review/storage	
Visualization and review	Artemis, Integrative Genomics Viewer (IGV)
VCF/BCF file manipulation	BCFTools

Aligning the sequencing reads is the first step of data processing. The raw sequencing data, usually in FASTQ format, is aligned with the reference sequence

(GRCh37 or GRCh38 for the human genome) through various tools. The output file, often in BAM format, further goes through pre-processing steps before variant calling. PCR duplicates are the redundant reads produced from the same DNA molecule, which algorithms can remove and calculate the correct variant allele fraction. It is not always necessary to mark duplicates and remove them [203] because reads duplication can also arise from sampling bias, and the removal of these duplicates might cause overcorrection [204].

Variant calling is the foundation of any further analysis on the output variants. Different software may have different results, but usually, they demonstrate good sensitivity. Choosing the proper tool based on the purpose of the laboratory and sequencing platforms is accepted but may need orthogonal validation from another tool. To evaluate software accuracy for variant calling, resources such as Genome in a Bottle (GIAB)[205] provide sequencing data of several individuals from multiple platforms and reference genomes. The true variants of these individuals have already been determined and confirmed.

2.3 Genotyping using DNA microarray

DNA microarray technology has a long history in genetics and genomics research, starting from the early 1980s. Its application expands from SNPs analysis to gene expression analysis, pharmacogenomic research, infectious disease diagnostics, and forensic genetics identification purposes. Mainly, disease-related polymorphisms, ancestry-specific markers, and genome-wide association studies (GWAS) greatly rely on the SNP array.

The DNA microarrays are made of a substrate such as glass, silicone, or plastic with immobilized single-stranded DNA probes. The areas on the microarray with probes are called test sites, and a single array could have different sizes (10 to 500 microns) and numbers (up to one hundred million) of test sites. By hybridization, the target DNA with a fluorophore binds to the probe complementarily. The molecules bound to the probes determine the strength of the signal, and the hybridization pattern could be scanned or imaged to collect the data for analysis. In the SNP array, the hybridization signals are translated into genotypes by sophisticated algorithms.

With the large quantities of genotypes provided by the SNP array, it is convenient to infer an individual's ancestry with software like *structure*[206] and ADMIXTURE[207]. ADMIXTURE, similar to *structure*, is a model-based approach that assumes the samples contain K subpopulations and each population has different population frequencies in the genotyped loci.

2.4 Methods for specific aims

In this dissertation, DNA sequencing was mostly performed by NGS, including NGS panel and ES. However, the first-generation sequencing technology, Sanger sequencing, also played important roles in the PDMP and ProDiGY studies. Sanger and Coulson developed the Sanger sequencing in 1977[208]; many scientists helped automate the procedure to the current platforms[209]. Automated Sanger sequencing has led to many accomplishments, such as the first human genome sequence[210]. The process utilizes the target DNA sequence as the template and starts the DNA elongation by DNA polymerase at the 3' end of the labeled DNA primer. Once one of the four dideoxynucleotides (ddATP, ddGTP, ddCTP, or ddTTP) adds to the chain, the elongation process of DNA

ends. In this way, each of the four extension processes produces fragmented DNAs with the specific ddNTP at the end. Dideoxynucleotides tagged with different labels will be excited by a laser, and signals will be detected by a sensor to be further processed to translate to the DNA sequence. Unlike NGS, Sanger sequencing can typically test DNA sequences of 500-600bp with high accuracy, and it is not bothered by the computational difficulty of aligning reads in repeated regions. It is regarded as the gold standard of DNA sequencing and is often recommended to fill in areas with low coverage in ES. However, the low throughput of Sanger sequencing has determined it is laborious and expensive to perform large-scale testing.

In the first aim, the PDMP study was designed to screen for patients suspected of monogenic diabetes with a simple questionnaire and diagnose them with genetic testing. The major genes that have been associated with monogenic diabetes are well known. Disease-targeted sequencing is more suitable than exome or genome sequencing, considering the cost and turnaround time. However, the NGS panel approach limits the ability to re-analyze additional genes when the results returned negative. And the NGS panel approach has limited application in detecting CNVs and SVs.

The PDMP study used two platforms from Ion Torrent, the Personal Genome Machine (PGM) Dx system and the GeneStudio S5 system. Consistently, both platforms used amplicon-based methods for DNA library construction, emulsion PCR for in vitro amplification, and detect signals of pH change caused by releasing of H^+ when dNTP is added to the DNA chain. (Figure 2.3) Since there signal change is not proportional to the number of dNTPs, it is error-prone in the polymer regions.

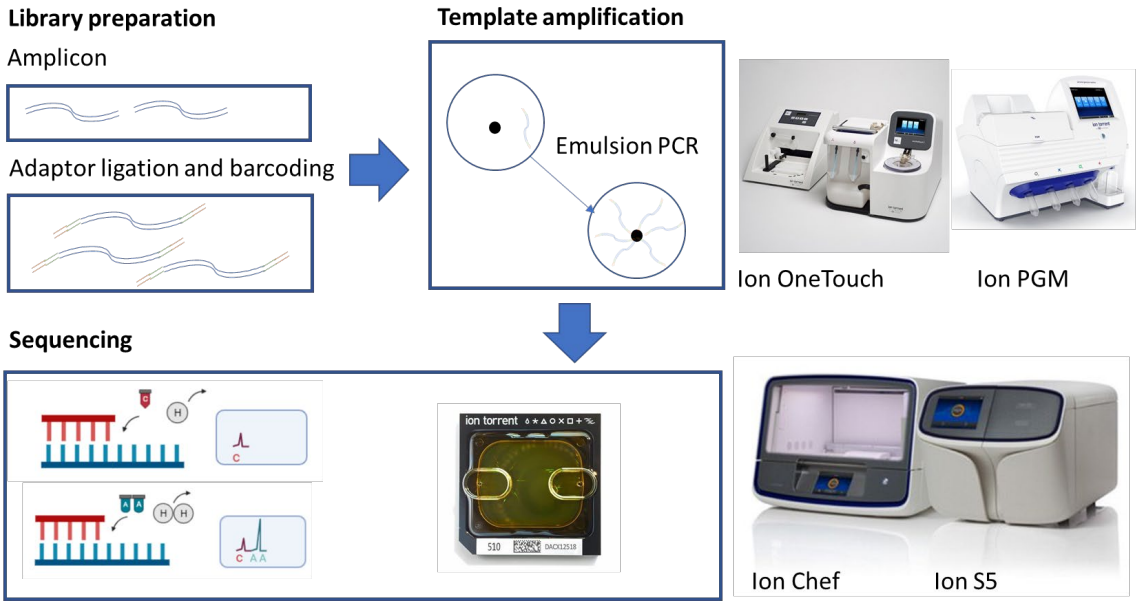


Figure 2.2 Ion Torrent platforms and sequencing workflow
The figure shows the sequencing steps of Ion Torrent PGM and S5 platforms used in the PDMP. The library was constructed by Ion AmpliSeq PCR assay that targets the 40 genes associated with monogenic diabetes with designed primers. Then the adaptors with barcodes were ligated on the amplicon to label the reads from the same individual. Constructed library with barcode went through purification before enrichment and emulsion PCR by Ion OneTouch for in vitro amplification. An amplified multiplexing library was then added to a chip for sequencing. In the Ion S5 system, the enrichment and emulsion PCR were combined and automated by the Ion Chef.

Other problems also exist for NGS, such as uneven coverage in GC-rich regions and overall higher error rates than Sanger sequencing. It is often routine for many laboratories to perform Sanger sequencing validation on variants from NGS. The comparison between NGS and Sanger sequencing listing their advantages and disadvantages has been reviewed extensively by Shendure et al. [211]. Sanger sequencing, though accurate, has low throughput and can be expensive and time-consuming when sequencing hundreds of genes on large quantities of samples. Previous studies have suggested that orthogonal validation of NGS results using Sanger sequencing is unnecessary when certain conditions are applied (capture based NGS, coverage above 100X, meets a certain quality score threshold) [212]. However, no study has provided evidence indicating an appropriate quality score cutoff for the specific sequencing

platform and genes. In addition, the clinically actionable results of the PDMP and TODAY study need to be returned to the patients and their physicians and such results must be performed in a Clinical Laboratory Improvement Amendments (CLIA) certified laboratory. Thus, any variants that were determined as pathogenic or likely pathogenic in the PDMP and TODAY study were confirmed by Sanger sequencing.

In the second aim of comprehensively re-analyzing the sequencing data for the PDMP study, the primary consideration is that orthogonal confirmation from another set of sequencing data analysis tools will help capture missing variants. Previous research has shown that five Illumina pipelines (SOAP, BWA-GATK, BWA-SNVer, GNUMAP, and BWA-SAMtools) had only 57.4% concordance of SNVs across all 15 exomes tested. Indels had even worse concordance that was as low as 26.8% between three pipelines. Furthermore, De Summa et al. have demonstrated that the TMAP aligner and GATK suite combination had better sensitivity in detecting SNVs and indels in real datasets from targeted gene panels and simulated datasets[213]. Therefore, the PDMP data re-analysis adopted the same strategy of switching to the GATK suite in the variant calling step and expected to discover the variants missed by the TVC.

Another part of the second aim, re-analyzing the sequencing data for the ProDiGY study, focused on discrepancies between ES and NGS panel from the same samples. The different results from different sequencing methods in the ProDiGY study could result from different sample preparation, library construction, sequencing platforms, and analytical tools. The beginning step is to assess the analytical tools before digging further into other factors that require extra cost and effort. Previous studies mainly used NGS panel as the first-tier test and ES only when the NGS panel had

negative results[214,215]. Some studies compared ES and simulated disease-specific panels from the ES data[216]. Other studies compare the diagnostic yield between NGS panel with different numbers of genes and ES-only[217]. Rarely studies compare NGS panel and ES on the same samples and same sequencing platforms[218]. Additionally, the majority of these studies compared the sensitivity and specificity without researching the reasons for the discrepancies. Therefore, this dissertation evaluated the possible causes of the differences, especially the variant calling step, which has many tunable parameters that could lead to different yields.

In the third aim of determining DNA microarray's ability to detect disease-causing variants, the concerns over the quality of microarray data are based on DNA microarray mechanisms. The 23andMe company uses the Illumina platform, on which several possible genotypes (AA, AB, BB, --, A- or B-) appear as clusters (Figure 2.4). The quality of a genotype is determined by the quality of the automated clustering. Consequently, the genotypes with rare variants are difficult to form a good cluster, especially variants observed only once in a population (singletons). The uncommonness of monogenic diabetes determined that the disease-causing variants are often only rarely seen or even absent in the population, thus it is necessary to evaluate the quality of genotyping before making inference from the DNA microarray data.

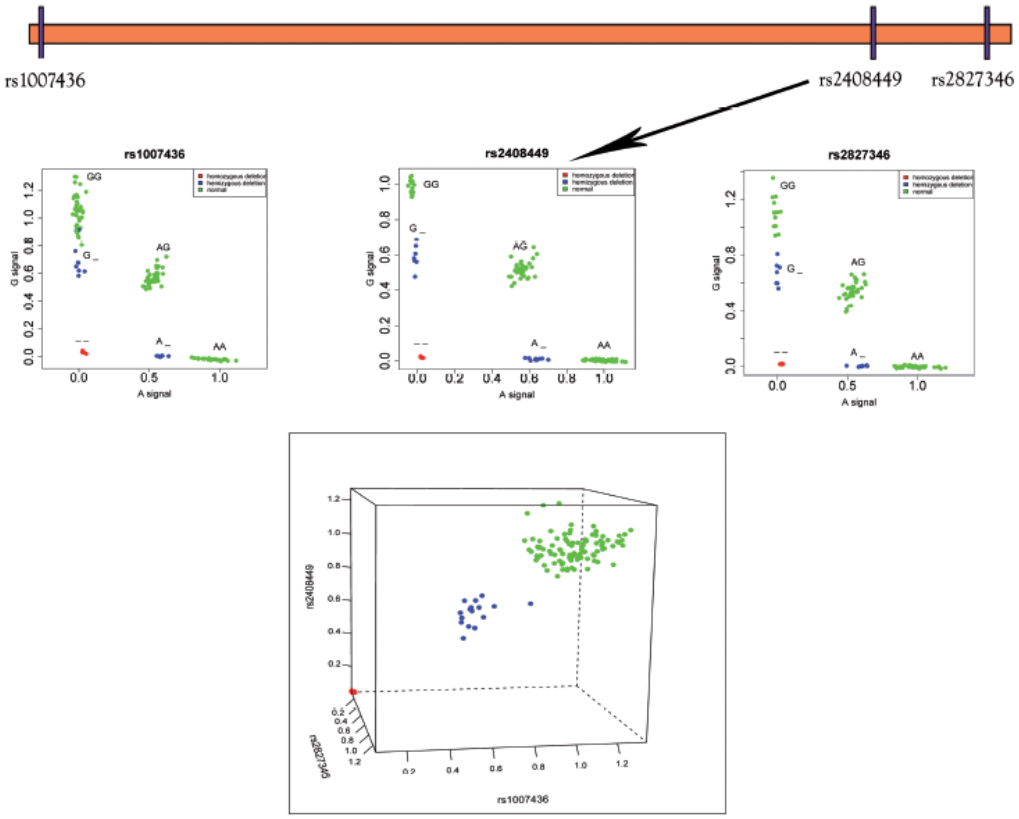


Figure 2.3 Calling SNP/CNV alleles from raw data.
All three SNPs shown here on chromosome 21 have alleles A and G. All plots show A allele and G allele intensity values from Illumina HumanHap550 data for 112 HapMap samples. The top three panels show each of the three SNPs individually along with their generalized genotypes. The bottom panel shows the total raw copy number sums (A signal + G signal) plotted, with each axis representing one of the SNPs. Note that the samples clearly separate into homozygous deletions (red), hemizygous deletions (blue), and normal (green). From LaFramboise 2009 Nucleic Acids Research [219].[2]

The third aim used the 23andMe raw data from openSNP, an open-source dataset containing data from anonymous volunteers. The raw data is not the same as the microarray output and they were generated by different versions of DNA microarray at different times. Consequently, the quality control (QC) step in this dissertation cannot directly replicate the conventional batch-based methods that require many parameters that were not included in the 23andMe raw data from the openSNP[220,221]. Alternatively, this

[2] Permission to reuse granted by Oxford University Press, license number 5095290016634.

dissertation compared the minor allele frequency between the openSNP samples and subjects from gnomAD, the current largest population frequency dataset. The allele frequency assessment, like other studies have used as supplementary QC steps, also demonstrates the quality of the genotyping [220,221]. Since the allele frequency is sensitive to ancestry, this dissertation stratified the allele frequency based on SNPs-inferred ancestry [222].

Chapter 3 : Identifying a Model for Successful Integration of Monogenic Diabetes Diagnosis into Routine Care: The Personalized Diabetes Medicine Program[3]*

3.1 Abstract

Objective: To implement, disseminate and evaluate a sustainable method for identifying, diagnosing, and promoting individualized therapy for monogenic diabetes.

Research Design and Methods: Patients were recruited into the implementation study through a screening questionnaire completed in the waiting room or through the patient portal, clinician or self-referral, or manual chart review by a study clinician. Patients suspected of monogenic diabetes underwent next-generation sequencing for 40 genes implicated in monogenic diabetes and related conditions.

Results: Three hundred thirteen probands with suspected monogenic diabetes enrolled. Sequencing identified 38 individuals with monogenic diabetes, with most variants found in *GCK* or *HNF1A*. Most (21/38, 55%) patients had a prior diagnosis of type 2 diabetes. Hit rates for ascertainment methods were 3.1% in clinic screening, 5.3% in EHR portal screening, 16.5% for physician recognition, and 32.4% for self-referral. The pre-defined criteria of non-type 1 diabetes before age 30 had an overall hit rate of 15.0%, and several individuals with young-onset pre-diabetes recruited by clinicians were also diagnosed with monogenic diabetes.

---

[3] Haichen Zhang, MS, Jeffery W. Kleinberger, PhD, Trevor J. Mathias, BS, Yue Guan, PhD, Kristin A. Maloney, MS, MGC, Elizabeth A. Streeten, MD, Kristina Blessing, MSW, Mallory N. Synder, RT, Lee Bromberger, BS, Jessica Goehringer, MS, Amy Kimball, MS, Coleen M. Damcott, PhD, Casey L. Overby Taylor, PhD, Michaela Nicholson, MHS, Devon Nwaba, MPH, Kathleen Palmer, BSN, Nicholas Ambulos, PhD, Linda J.B. Jeng, MD, PhD, Alan R. Shuldiner, MD, Philip Levin, MD, David J. Carey, PhD, Toni I. Pollin, PhD. Identifying a Model for Successful Integration of Monogenic Diabetes Diagnosis into Routine Care: The Personalized Diabetes Medicine Program. In preparation for submission.
*HZ performed genetic testing on 247 participants, analyzed sequencing data of 247 participants, involved in variant classification process, and designed primers for Sanger sequencing and performed PCR reaction. HZ drafted the manuscript and generated tables and figures.

Conclusions: The high hit rate of monogenic diabetes in self-referred individuals suggests that educating the lay public in addition to clinicians may be the most effective way to increase the diagnosis rate of monogenic diabetes. Follow-up of these patients to assess the effects of genomic diagnosis of monogenic diabetes on diagnosis, treatment, clinical outcomes, healthcare utilization/cost-effectiveness of screening, and cascade testing is in progress. Results will be used to inform a clinical service model to assure that all individuals and families with monogenic diabetes have access to diagnosis and customized treatment and more broadly access to personalized medicine across disease areas.

3.2 Introduction

Diabetes mellitus is currently classified into four major categories: (1) type 1 diabetes, absolute insulin deficiency usually caused by autoimmune destruction of the insulin-producing pancreatic beta cells; (2) type 2 diabetes, a heterogeneous group of diseases caused by a combination of relative insulin deficiency and insulin resistance; (3) specific types of diabetes due to other causes, including defects in beta cell function and other components of the glucose homeostasis system caused by highly penetrant genetic mutations as well as forms of diabetes secondary to specific exposures and/or other diseases; and (4) gestational diabetes mellitus (GDM)[2].

At least 1-2% of diabetes is monogenic [1] or caused by a highly penetrant mutation in a single gene and falling into the third category, which amounts to at least 300,000 individuals in the United States alone. The most well-known and apparently most common monogenic forms are traditionally known as “maturity onset diabetes of the young,” or MODY, a term coined in 1974 to describe non-obese individuals with non-

insulin requiring diabetes diagnosed before the age of 25 years, usually with an autosomal dominant pattern of inheritance[3]. Most cases of MODY have now been attributed to mutations in *HNF1A* (MODY3[223]), *HNF4A* (MODY1[224]) or *GCK* (MODY2[225]), but several other genes are implicated in a small number of cases each[17,24,28,29,57,58,226-230]. Three genes implicated in the rarer forms of MODY, *ABCC8*, *KCNJ11*, and *INS*, account for the majority of cases of neonatal diabetes, which also has several other rare forms[231]. Several other forms of diabetes are caused by a single gene mutation, often manifesting in a syndrome that includes extra pancreatic features, including Wolcott-Rallison syndrome (*EIF2AK3*[232])), extreme insulin resistance and lipodystrophy syndromes (*INSR*, *AGPAT2*, *LMNA*, *PPARG*, and others reviewed in a review paper[1] (see Chapter 1)), Wolfram syndrome[233], and mitochondrial diabetes and deafness[234].

Distinguishing monogenic diabetes from type 1 diabetes or type 2 diabetes often directly leads to more effective and less invasive treatment, as well as more accurate prediction of prognosis and familial risk. While metformin is generally the first line treatment for individuals diagnosed with type 2 diabetes and insulin is required for treatment of type 1 diabetes, sulfonylureas are the first line therapy for individuals with *HNF1A*-MODY and *HNF4A*-MODY and offer a less expensive and less invasive therapy and improved glucose control compared with insulin[37]. *GCK*-MODY, a nonprogressive form of mild fasting hyperglycemia rarely associated with complications, usually does not require treatment, thereby removing the risks of adverse events associated with pharmacotherapy[235]. Neonatal diabetes, found in approximately 1/100,000 infants, is now known in most cases to be monogenic and in nearly half of cases to result from

mutations in *KCNJ11* and *ABCC8*. Most individuals with mutations in *KCNJ11* or *ABCC8* can be treated exclusively with high doses of sulfonylureas, which in addition to being less invasive than insulin, control glucose levels more effectively and confer less risk of dangerous hypoglycemic episodes in these patients[236], including over the long term[135]. While there are emerging recommendations for the use of genetic testing in diabetes diagnosis[1,2], most cases are missed and consequently receive inappropriate treatment[237]. Barriers to systematic diagnosis include lack of awareness among providers, and among those who are aware, difficulty obtaining testing due to expense and complexity and difficulty obtaining insurance coverage, all of which results from a lack of a standard protocol for identifying individuals for testing and incorporating results into medical care. For example, current guidelines such as ADA standard of care[166] and ISPAD guidelines[1] about patients identification only provides recommendations rather than practical strategies. With the growing availability of next generation sequencing (NGS), several investigators in countries outside the United States have recently reported on systematic clinical approaches to identifying and diagnosing individuals with monogenic diabetes, including France[238], Turkey[239], China[240], Switzerland[241], and the United Kingdom[242]. As part of the NIH-Funded IGNITE (Implementing Genomics in Practice) Network, we set out to develop a comprehensive approach to address these barriers by designing an implementation project to systematically identify individuals likely to have monogenic diabetes from the general diabetes population, perform cost-effective molecular testing using a targeted NGS panel, and disclose and incorporate results into the electronic health record.

3.3 Methods

Inclusion criteria and algorithm criteria

In stage one, only patients suspected of having monogenic diabetes were enrolled. Eligible patients had to meet at least one of initially six and later seven (updated to include nondiabetic individuals with hyperglycemia during pregnancy) criteria (Table 3. 1a), which were based on previously published studies and guidelines[1,2,167]. The eligible patients were then evaluated for fitness to algorithm criteria based on their medical and family history, and lab results (antibodies and C-peptide); Figure 3.1). In stage two, we modified the inclusion criteria (Table 3.1b) to include any patient with diabetes while still using the algorithm to track those who met algorithm criteria and who did not.

Table 3.1 PDMP inclusion criteria

1a. Inclusion criteria at Stage 1
Has at least one of the following characteristics suggesting high likelihood of monogenic diabetes:
Diagnosed with diabetes < 1 year of age, or
Diagnosed with type 1 diabetes and has a parent or child with type 1 diabetes, or
Diagnosed with type 2 diabetes at < 30 years old, or
Diagnosed with type 2 diabetes at ≤ 45 years old and not obese at diagnosis and two or more relatives with diabetes diagnosed ≤ 50 years old, or
Presence in patient of diabetes plus extra-pancreatic features fulfilling diagnostic criteria for a genetic diabetes syndrome (e.g. urological abnormalities, optic atrophy, deafness), or
Fasting glucose ≥ 100 mg/dl during a current or past pregnancy without pre-existing diabetes together with a pre-pregnancy BMI<25, or*
High suspicion of highly penetrant genetic form of diabetes
* Criterion added 5/27/2016

1b. Inclusion criteria at Stage 2
Meets at least one of the following criteria:
Diagnosed with diabetes mellitus (any type), or
History of persistent mild hyperglycemia, or
Fasting glucose ≥ 100 mg/dl during a current or past pregnancy without pre-existing diabetes together with a pre-pregnancy BMI<25, or
Clinical suspicion of highly penetrant genetic form of diabetes based on clinician referral or review of available lab results.

Figure 3.1 PDMP criteria algorithm

Criteria	Description	Intake Questionnaire Indicators	Further investigation (Visit 1)	Further investigation	Meets Criteria
1	Diagnosed with diabetes before the age of 1 year old	# 1 = Yes →		IA-2 ab testing. Test C-peptide if on insulin	Yes
2	Diagnosed with T1DM + a parent or child with T1DM	# 5 = Yes AND # 6 = Yes →	Test antibodies If IA-2 ab neg →	Test C-peptide if on insulin If C-peptide positive →	Yes
3a	Diagnosed with non-T1DM at ≤ 30 years old ***Not on insulin***	# 5 = No AND # 2 = Yes →	Test antibodies If IA-2 ab neg →		Yes
3b	Diagnosed with non-T1DM at ≤ 30 years old ***On insulin***	# 5 = No AND # 2 = Yes →	Test antibodies If IA-2 ab neg →	Test C- peptide If C-peptide positive →	Yes
4	Diagnosed with T2DM at ≤ 45 years old + not extremely overweight at diagnosis + two or more relatives with DM diagnosed ≤ 50 years old	# 5 = No AND #2 = Dx ≤ 45 AND # 3 = No AND # 7 = Yes If Fam Hx shows 2 or more relatives with DM diagnosed ≤ 50 years old →	Pedigree & Medical Hx Test antibodies →	If Fam Hx shows 2 or more 1st or 2nd degree relatives (same lineage) with T2DM diagnosed ≤ 50 years old AND If IA-2 ab neg, Test C-peptide if on insulin If C-peptide positive and/or at the discretion of the team →	Yes
5	Hearing, visual, or cognitive impairment, birth defects or kidney disease in childhood	# 4 = Yes Team Assessment = Yes to Visit 1 →	Pedigree & Medical Hx Test antibodies →	If IA-2 ab neg Test C-peptide if on insulin If C-peptide positive and/or at the discretion of the team →	Yes
6	Non-diabetic with a fasting glucose ≥100 mg/dl during a current or past pregnancy and pre-pregnancy BMI < 25	Gestational DM = Yes AND *Confirm documentation of BMI pre-pregnancy (<25) AND glucose during pregnancy ≥100 mg/dl before scheduling Visit 1* →	Pedigree & Medical Hx Test antibodies →	If IA-2 ab neg Test C-peptide if on insulin If C-peptide positive *Or* at discretion of team →	Yes
7	Clinical suspicion of highly penetrant genetic form of diabetes based on clinician referral or review of available lab results	Complete Intake Questionnaire Team Assessment = Yes to Visit →	Pedigree & Medical Hx Test antibodies →	If IA-2 ab neg Test C-peptide if on insulin If C-peptide positive and/or at the discretion of the team →	Yes

Antibody Negative = Islet Antigen 2 antibody negative

Positive C-peptide = ≥0.6 ng/mL (which converts to ≥0.2 nmol/L, depending on units)

Group Assignment: _____ Meets Criteria Group

_____ Unselected Group

Inits/Date:_____

Sites

Patients were recruited from endocrinology clinics at four distinct settings, the University of Maryland Center for Diabetes and Endocrinology (UM-CDE, a university medical center), the Baltimore Veterans Administration Medical Center (BVAMC), the Geisinger Health System (GHS, an integrated health system), and Bay West Endocrinology Associates (a suburban private practice).

Recruitment

Patients were screened in the clinic waiting room and through the GHS patient portal using a simple questionnaire (Table 3.2) with yes/no questions designed to identify red flags raising suspicion of monogenic diabetes. Self-referrals and physician recognition, which includes physician-referral at all sites and manual chart review at UM-CDE, were used to identify cases. The EHRs for patients recruited from the main sites was reviewed under a HIPAA waiver for prior antibody and C-peptide measures.

Table 3.2 PDMP intake questionnaire

Have you been diagnosed with the following? (check all that apply)
- ☐ Type 1 diabetes
- ☐ Type 2 diabetes
- ☐ Gestational diabetes
- ☐ Pre-diabetes or high blood sugar

1. Were you diagnosed with diabetes or high blood sugar before 1 year of age?
   ☐ Yes ☐ No
2. Were you diagnosed with diabetes or high blood sugar at age 30 or younger?
   ☐ Yes ☐ No

How old were you when you were diagnosed?_____

3. Were you extremely overweight when you were diagnosed?
   ☐ Yes ☐ No
4. As a child, did/do you have hearing or vision problems, intellectual disability (for example, learning disabilities, mental retardation, autism), birth defect(s) or kidney disease?
   ☐ Yes ☐ No If yes, describe: _____
5. Do you have type 1 diabetes (if unsure, were you on insulin at diagnosis and have been ever since)?
   ☐ Yes ☐ No
6. Do you have a parent or a child with type 1 diabetes?
   ☐ Yes ☐ No
7. Do you have 2 or more people related to you by blood with diabetes?*
   ☐ Yes ☐ No * If yes, please complete Family History Questionnaire.

Study visits

During the first visit, informed consent and a detailed medical history were obtained by a study coordinator, and a three-generation pedigree focusing on details of diabetes was obtained by a genetic counselor using a customized pedigree form. Blood was collected for measurement of GAD65 and IA-2 and C-peptide, next-generation sequencing, banking of serum and plasma and extraction and holding in the Translational Genomics Laboratory (TGL) for future clinical variant confirmation.

Next-generation sequencing

Research grade next-generation sequencing was performed on samples from patients chosen for testing in Stage 1 and all patients in Stage 2 using two platforms, the Ion Torrent Personal Genome Machine (PGM) for the first 271 samples and the Ion Torrent Chef system/S5 sequencer for the remainder. The consistency of the two platforms was validated during the transition. The next-generation testing was performed using a custom gene panel including 795 amplicons covering the coding and flanking intronic regions of 40 genes implicated in monogenic diabetes or related conditions. The panel consists of genes that are known to cause forms of monogenic diabetes, including MODY, NDM, syndromic forms of monogenic diabetes (e.g., Wolfram Syndrome, Roger's Syndrome, etc.), lipodystrophy, hyperinsulinemic hypoglycemia (on the possibility that variants with reverse effects could cause diabetes as is already known for a subset), and monogenic forms of obesity (Table 3.3). The sequencing methods and variant classification based on the 2015 American College of Medical Genetics/ Association of Molecular Pathologists Standards and Guidelines[189] has been described previously[9]. Variant classification into the five ACMG/AMP categories (pathogenic,

likely pathogenic, variant of uncertain significance [VUS], benign, or likely benign) was performed by a team consisting of the PI, TGL director, TGL genetic counselor, and 1-2 PhD students using the ACMG/AMP standards and guidelines as implemented through a custom interface they developed[243]. Personal communication with other colleagues involved in monogenic diabetes genetic testing was sought to increase available evidence for VUS with suspicion of pathogenicity.

Clinical confirmation and disclosure of pathogenic and likely pathogenic variants

Pathogenic and likely pathogenic results were considered to be clinically relevant and were confirmed in the CLIA-certified and CAP-accredited TGL at the University of Maryland School of Medicine in order to make them reportable to patients and providers. A clinical report, developed by the variant classification team in collaboration with the TGL, was issued to the study site physician, who then in collaboration with the genetic counselor, disclosed the result to the patient. After the disclosure session, a physician clinic note containing recommendations for treatment based on both the genetic testing result and the patients history, a genetic counseling letter, and a copy of the laboratory report were mailed to the patient. These documents were also uploaded to the EHR and routed to the physician treating the patient's diabetes, or mailed to the patient's physician if referred from outside the main sites.

If no variants were identified or if only benign variants, likely benign variants, or VUS were identified, then the genetic counselor called the patient and reported that no clinically significant variants were found and followed with a letter summarizing this and indicating that the result should not prevent clinical testing for monogenic diabetes if indicated, since the sequencing was not performed under a clinical protocol. If one or

more VUS were identified, the patient was also told that it might be useful to improve our understanding of the causes of diabetes to recruit the patient's family members.

Table 3.3 Genes on the sequencing panel

Gene name	MODY type	Neonatal diabetes type	Diabetes syndrome
ABCC8	MODY 12	-	-
AGPAT2	-	-	CGL type 1
ALMS1	-	-	Alstrom syndrome
BLK	MODY11	-	-
BSCL2	-	-	CGL type 2
CAV1	-	-	CGL type 3
CEL	MODY8	-	MODY with exocrine dysfunction
CISD2	-	-	Wolfram syndrome 2
EIF2AK3	-	-	Wolcott-Rallison syndrome
FOXP3	-	-	Immunodysregulation, polyendocrinopathy, and enteropathy, X-linked (IPEX)
GATA6	-	-	Pancreatic agenesis and congenital heart defects
GCK	MODY2	PNDM	-
GLIS3	-	-	Neonatal diabetes mellitus with congenical hypothyroidism
HNF1A	MODY3	-	-
HNF1B	MODY5	-	Renal Cysts And Diabetes (RCAD) syndrome
HNF4A	MODY1	-	-
INS	MODY10	PNDM	-
INSR	-	-	Type A insulin resistance, Donohue syndrome, Rabson-Mendenhall syndrome
PDX1	MODY4	-	Pancreatic agenesis
KCNJ11	MODY13	PNDM, TNDM	PNDM with neurological features
KLF11	MODY7	-	-
LMNA	-	-	FPL type 2
NEUROD1	MODY6	-	-
PAX4	MODY9	-	Ketosis-prone diabetes
PLIN1	-	-	FPL type 4
PPARG	-	-	FPL type 3, digenic severe insulin resistance
PPP1R3A	-	-	Digenic severe insulin resistance
PTF1A	-	-	Pancreatic (and cerebellar) agenesis
PTRF	-	-	CGL type 4
RFX6	-	-	Mitchell-Riley syndrome, Martinez-Frias syndrome
SLC19A2	-	-	Thiamine-responsive megaloblastic anemia syndrome (Rogers syndrome)
SLC2A2	-	-	Fanconi-Bickel syndrome (symptom of prolonged hyperglycemia in response to glucose loads)
WFS1	-	-	Wolfram syndrome
ZFP57	-	TNDM	-
GLUD1	-	-	Hyperinsulinism-hyperammonemia syndrome
HADH	-	-	Familial hyperinsulinemic hypoglycemia type 4
MC4R	-	-	Autosomal dominant obesity
LEP	-	-	Obesity due to leptin deficiency
LEPR	-	-	Obesity due to leptin receptor deficiency
SIM1	-	-	Severe Obesity

Statistical analysis

A χ^2test was used to test whether there was a significant difference in the distribution of enrollment subset between European and African American. Either Chi-square test or Fisher's exact test (when the sample size is smaller than 5) was used to compare the performance of questionnaire between patients with or without monogenic diabetes. Two-tailed t-test was used to test whether there is a significant difference of the age at diagnosis between patients with or without monogenic diabetes. The test was considered statistically significant when the p-value was less than 0.05.

3.4 Results

Target population

Over 3.5 years, 2,190 individuals were screened across four sites, including 598 at UM-CDE, 252 at BVAMC, 1061 at GHS, and 278 at Bay West (Table 3.4). In addition, there were 222 patients recognized by physicians either as individual physician referrals (both within and outside the sites) or by manual chart review by one of us (EAS) of patients coming in for clinical visits at UM-CDE. Finally, there were 109 self-referrals. Of the 2,522 individuals considered, the majority were European American (66.2%) and African American (29.7%) (Table 3.5). More than half (65.5%) of the target population had more than two blood relatives with diabetes, about a third (32.6%) of the screened individuals were diagnosed with type 1 diabetes, and more than a third (39.0%) of the screened individuals were diagnosed with high blood sugar or diabetes before age of 30 (Table 3.6).

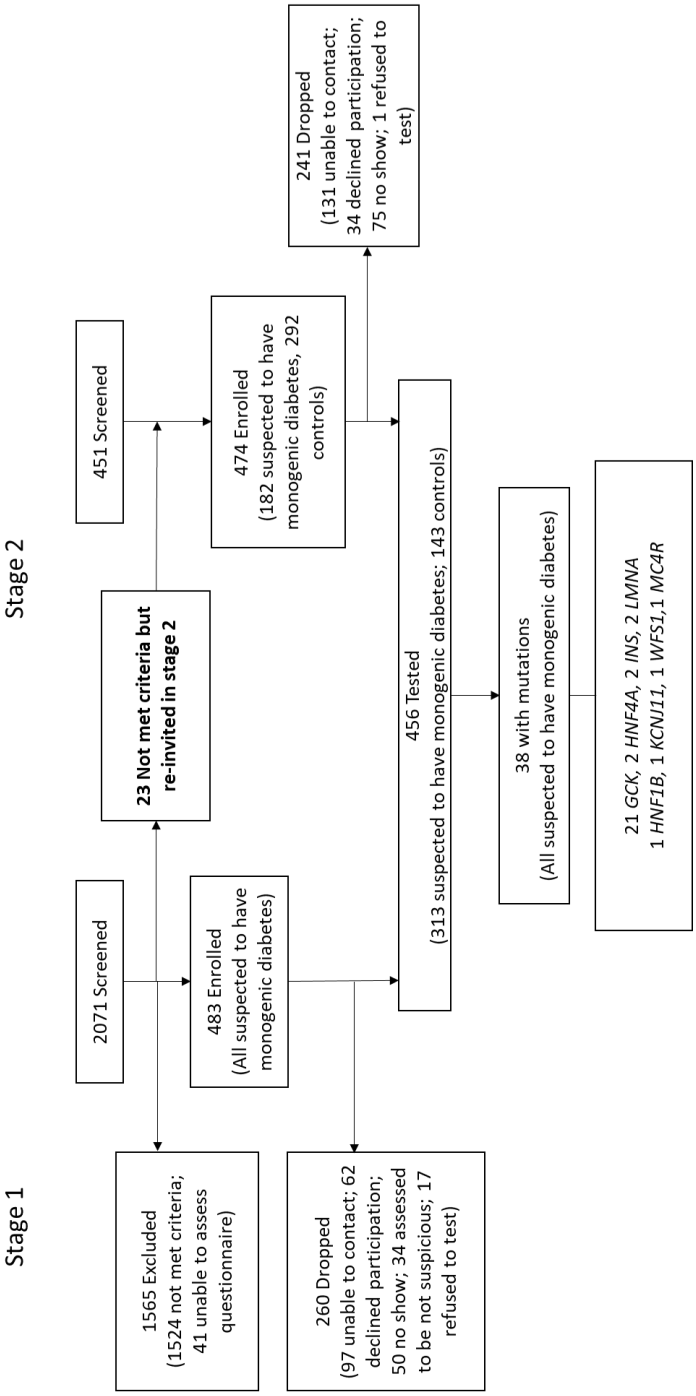


Figure 3.2 Flowchart of screened PDMP participants

During stage 1, 2071 participants were screened and 483 enrolled into the study based on inclusion criteria and all of the enrolled participants were suspected to have monogenic diabetes. Twenty-three of the excluded participants in stage 1 were re-invited to join the study during stage 2. With extra 451 screened participants, the enrolled 474 participants included 182 suspected participants and 292 controls. A total of 501 enrolled participants dropped for various reasons, 260 during stage 1 and 241 during stage 2. Overall, 456 participants went through genetic testing, including 313 suspected participants and 143 controls. Finally, 38 out of the 313 suspected participants were tested positive for pathogenic and likely pathogenic variants in eight genes.

Table 3.4 Distribution of individuals screened for the PDMP by recruitment method and site

Participant sources	Bay West	BVAMC	Geisinger	UM-CDE	Total	Percentage
Clinical Screen	278	252	777	598	1905	75.5%
Physician Recognition	25	6	14	177	222	8.8%
Self-Referral	0	0	1	108	109	4.3%
EHR Portal Screen	0	0	286	0	286	11.3%
Total	303	258	1078	883	2522	-

Table 3.5 Demographic of PDMP participants (excluding missing data)

Characteristic	**UM-CDE (n=874)**	**Geisinger (n=1076)**	**Bay West (n=302)**	**BVAMC (n=256)**	**Total (n=2508)**
Age±SD (y)	43.8±18.3	50.6±19.3	54.4±15.9	60.6±10.5	49.7±18.6
	UM-CDE (n=881)	**Geisinger (n=1079)**	**Bay West (n=306)**	**BVAMC (n=258)**	**Total (n=2517)**
Female(%)	58.2% (513/881)	56.0% (602/1075)	63.5% (193/304)	8.6% (22/257)	52.8% (1330/2517)
	UM-CDE (n=681)	**Geisinger (n=976)**	**Bay West (n=270)**	**BVAMC (n=162)**	**Total (n=2089)**
Ethnicity Hispanic (%)	2.8%(19/681)	2.0%(20/976)	1.1%(3/270)	4.9%(8/162)	2.4%(50/2089)
	UM-CDE (n=864)	**Geisinger (n=1061)**	**Bay West (n=302)**	**BVAMC (n=250)**	**Total (n=2477)**
Ancestry (%)					
African American	56.5% (488/864)	1.9% (20/1061)	21.5% (65/302)	65.2% (163/250)	29.7% (736/2477)
European American	38.5% (333/864)	95.4% (1012/1061)	70.9% (214/302)	32.4% (81/250)	66.2% (1640/2477)
Asian	3.0% (26/864)	0.6% (6/1061)	3.6% (11/302)	0% (0/250)	1.7% (43/2477)
Pacific Islander	0.1% (1/864)	0.1% (1/1061)	0.3% (1/302)	0% (0/250)	0.1% (3/2477)
Native American	0.5% (4/864)	0.4% (4/1061)	0.3% (1/302)	0% (0/250)	0.4% (9/2477)
Other	1.4% (12/864)	1.7% (18/1061)	3.3% (10/302)	2.4% (6/250)	1.9% (46/2477)

Table 3.6 Screening questionnaire responses of all PDMP participants

PDMP Screening Questions	Percent Answered Yes
1. Were you diagnosed with diabetes or high blood sugar before 1 year of age?	2.0% (49/2512)
2. Were you diagnosed with diabetes or high blood sugar at age 30 or younger? How old were you when you were diagnosed?	39.0%(977/2507)
3. Were you extremely overweight when you were diagnosed?	30.3%(682/2253)
4. As a child, did.do you have hearing or vision problems, intellectual disability (for example, learning disabilities, mental retardation, autism), birth defect(s) or kidney disease?	19.4%(482/2484)
5. Do you have type 1 diabetes (if unsure, were you on insulin at diagnosis and have been ever since)?	32.6%(811/2488)
6. Do you have a parent or a child with type 1 diabetes?	12.6%(283/2239)
7. Do you have 2 or more people related to you by blood with diabetes? If yes, please list relationship, age of diagnosis, and type of diabetes.	65.5%(1633/2493)

Enrollment

During stage one, 479 individuals suspicious for monogenic diabetes based on the initial screen were invited to enroll in the study and 273 enrolled. During stage two, 472 individuals were invited without regard to suspicion for monogenic diabetes and 234 were enrolled. Of the 507 (273 + 234) enrolled, 313 were found to meet algorithmic criteria and were sequenced and 51 were determined not to be tested after team assessment of laboratory results or medical history at stage one. An additional 143 individuals enrolled at Stage 2 did not meet algorithmic criteria and were sequenced as controls. More than half (53.0%) of the invited individuals were enrolled. The major reason for not enrolling was the inability to contact (24.0%). African American individuals were more likely than European American individuals to decline to enroll when contacted (OR=2.37, p=0.0003), and more likely to cancel the enrollment visit (OR=3.87, $p=1.6\times10^{-7}$) (Table 3.7).

Table 3.7 Comparison of enrollment categories between African American and European American

Enrollment categories	African-American (n=299)	European American(n=576)	χ^2-square test P-value	Odds Ratio (95% CI)
Unable to contact	66 (22.1%)	145 (25.2%)	0.5	0.84 (0.60,1.17)
Declined participation	58 (19.4%)	53 (9.2%)	0.0003	2.37 (1.59, 3.55)
Cancelled visit	54 (18.1%)	31 (5.4%)	1.6×10^{-7}	3.87 (2.43, 6.18)
Enrolled	121 (40.5%)	347 (60.2%)	0.002	0.45 (0.34, 0.60)

Genetic testing

In total, 456 (223 meeting algorithmic criteria in stage one, 90 meeting algorithmic criteria in stage two, and 143 controls from stage 2) of 507 enrolled patients underwent genetic testing. The numbers of screened, enrolled participants at stage 1 and stage 2 and total genetically tested participants are illustrated in Figure 3.2. Most sequenced participants were European American (68.4%) or African American (25.4%), with 6.2% from other racial groups, and 1.3% having Hispanic ethnicity. Table 3.8 shows these individuals stratified by suspected monogenic diabetes vs. unselected controls. After filtering for coding and consensus splice site variants with <5% frequency in 1000 Genomes, ExAC, or NHLBI-ESP, a total of 449 missense variants, three nonsense variants, nine splice-site variants, and 17 small (less than 20bp) insertions/deletions were analyzed according to the ACMG/AMP guidelines for variant interpretation during the variant classification meetings (Table 3.9). A total of 107 novel variants were discovered (not found in 1000 Genomes, NHLBI-ESP, ExAC, dbSNP, or T2D-Genes Exome database) in 98 tested participants (Table 3.10).

Table 3.8 Demographics of individuals sequenced by 40-gene panel

a. Suspected monogenic diabetes

Characteristic	UM-CDE (n=195)	Geisinger (n=78)	Bay West (n=34)	BVAMC (n=6)	Total (n=313)
Age ± SD (y)	38.6±17.9	47.1±19.5	46.8±10.7	47.3±14.3	42.5±16.8
Female %(n)	62.6% (122)	62.8% (49)	85.3% (29)	0	63.9% (200)
Hispanic %(n)	2.6% (5)	1.3% (1)	0	0	1.9% (6)
African American	40.0% (78)	0	17.6% (6)	66.7% (4)	28.1% (88)
White	54.9% (107)	98.7% (77)	79.4% (27)	33.3% (2)	38.1% (213)
Asian	4.0% (8)	0	2.9% (1)	0	2.9% (9)
Pacific Islander	0	0	0	0	0
Native American	1.0% (2)	0	0	0	0.6% (2)
Other	0	1.3% (1)	0	0	0.3% (1)

Table 3.8 Continued

b. Unselected diabetic controls (Stage 2 only)

Characteristic	UM-CDE (n=64)	Geisinger (n=52)	Bay West (n=23)	BVAMC (n=4)	Total (n=143)
Age ± SD (y)	34.7±20.5	47.1±18.2	44.4±11.3	57.5±14.1	39.9±21.5
Female %(n)	57.8% (37)	69.2% (36)	60.9% (14)	25.0% (1)	61.5% (88)
Hispanic %(n)	0	0	0	0	0
African American	32.8% (21)	0	13.0% (3)	75.0% (3)	21.1% (27)
White	50.0% (32)	100% (52)	69.6% (16)	0	69.9% (100)
Asian	12.5% (8)	0	17.4% (4)	0	8.4% (12)
Pacific Islander	0	0	0	0	0
Native American	3.1% (2)	0	0	25.0% (1)	2.1% (3)
Other	1.5% (1)	0	0	0	0.7% (1)

Table 3.9 Total variants reviewed by gene and variants effects

Gene	Total variants	Missense	Nonsense	Splice	Insertion/Deletion
LEPR	11	11	0	0	0
SLC19A2	4	4	0	0	0
KLF11	11	10	0	0	1
ALMS1	63	63	0	0	0
EIF2AK3	8	7	0	1	0
NEUROD1	2	2	0	0	0
PPARG	1	1	0	0	0
SLC2A2	3	3	0	0	0
WFS1	47	45	0	0	2
HADH	6	6	0	0	0
SIM1	3	3	0	0	0
RFX6	12	12	0	0	0
GCK	19	14	1	2	2
PPP1R3A	23	23	0	0	0
CAV1	1	1	0	0	0
PAX4	9	9	0	0	0
LEP	1	1	0	0	0
BLK	11	10	0	0	1
GLIS3	23	23	0	0	0
CEL	6	6	0	0	0
AGPAT2	14	12	0	1	1
PTF1A	6	6	0	0	0
GLUD1	3	3	0	0	0

Table 3.9 Continued					
INS	2	2	0	0	0
KCNJ11	6	6	0	0	0
ABCC8	10	9	1	0	0
BSCL2	10	7	0	1	2
HNF1A	13	10	0	1	2
PDX1	8	7	0	0	1
PLIN1	12	11	0	1	0
HNF1B	6	6	0	0	0
GATA6	8	8	0	0	0
MC4R	7	6	1	0	0
INSR	10	10	0	0	0
HNF4A	7	7	0	0	0
FOXP3	2	2	0	0	0
ZFP57	7	7	0	0	0
LMNA	6	6	0	0	0

Table 3.10 PDMP novel variants

Gene	**rsID**	**Transcript**	**Nucleotide Change**	**Amino Acid Change**	**Classification**
LEPR	rs1057524881	NM_001198687	c.G1055T	p.C352F	VUS
LEPR	rs937591370	NM_001198687	c.A1292G	p.N431S	VUS
LMNA	rs267607555	NM_170707	c.C1045T	p.R349W	Likely pathogenic
LMNA	.	NM_170707	c.G1504A	p.A502T	VUS
LMNA	rs267607613	NM_170707	c.C1633T	p.R545C	VUS
KLF11	rs1057524882	NM_003597	c.608_609AG	.	VUS
ALMS1	.	NM_015120	c.G112A	p.D38N	Likely Benign
ALMS1	.	NM_015120	c.C895T	p.P299S	VUS
ALMS1	rs1057524883	NM_015120	c.G2209A	p.E737K	Likely Benign
ALMS1	.	NM_015120	c.G2970T	p.K990N	Likely Benign
ALMS1	rs1057524884	NM_015120	c.G6034T	p.A2012S	Likely Benign
ALMS1	rs370622410	NM_015120	c.C9109G	p.P3037A	Likely Benign
ALMS1	rs745518051	NM_015120	c.A9776C	p.D3259A	Likely Benign
EIF2AK3	rs756231274	NM_004836	c.T833A	p.F278Y	VUS
EIF2AK3	.	NM_004836	c.309-1G>C	.	Likely Pathogenic
PPARG	.	NM_138712	c.C838T	p.R280C	VUS
SLC2A2	.	NM_000340	c.1170+1G>C	.	Likely Pathogenic
SLC2A2	.	NM_000340	c.A971G	p.Y324C	VUS
WFS1	rs1057524887	NM_006005	c.G172A	p.A58T	VUS
WFS1	rs1057524889	NM_006005	c.C667A	p.L223M	VUS
WFS1	rs876657735	NM_006005	c.1059_1061del	p.353_354del	VUS
WFS1	.	NM_006005	c.G1135T	p.D379Y	VUS
WFS1	rs373862003	NM_006005	c.C1820T	p.P607L	VUS
WFS1	rs758746685	NM_006005	c.T1868A	p.M623K	VUS
WFS1	rs1057524890	NM_006005	c.G1896T	p.M632I	VUS
WFS1	.	NM_006005	c.C2570A	p.T857N	VUS
ZFP57	.	NM_001109809	c.G288C	p.K96N	VUS
SIM1	.	NM_005068	c.C1894T	p.H632Y	VUS
RFX6	.	NM_173560	c.A878C	p.H293P	VUS
RFX6	rs1018054636	NM_173560	c.A1319G	p.Y440C	VUS
RFX6	.	NM_173560	c.C1930T	p.P644S	Likely Benign
RFX6	rs867520268	NM_173560	c.G1955A	p.R652Q	Likely Benign

Table 3.10 Continued

GCK	rs1057524900	NM_000162	c.C1361A	p.A454E	Likely Pathogenic
GCK	rs1057524901	NM_000162	c.1344delC	p.G448fs	Pathogenic
GCK	.	NM_000162	c.1254-2A>G	.	Likely Pathogenic
GCK	rs1057524902	NM_000162	c.1253+2T>A	.	Pathogenic
GCK	.	NM_000162	c.C1113G	p.C371W	Likely Pathogenic
GCK	rs1057524903	NM_000162	c.A1016G	p.E339G	Pathogenic
GCK	rs193922331	NM_000162	c.T787C	p.S263P	Likely pathogenic
GCK	rs1057524904	NM_000162	c.C748T	p.R250C	Likely Pathogenic
GCK	.	NM_000162	c.A718G	p.N240D	Likely pathogenic
GCK	rs1057524905	NM_000162	c.680-1G>A	.	Pathogenic
GCK	rs148311934	NM_000162	c.G676A	p.V226M	Pathogenic
GCK	rs1085307455	NM_000162	c.C571T	p.R191W	Pathogenic
GCK	rs1064793998	NM_000162	c.G184A	p.V62M	Pathogenic
GCK	rs193922286	NM_000162	c.C146A	p.T49N	Pathogenic
GCK	.	NM_000162	c.C127T	p.R43C	Pathogenic
GCK	rs1057524906	NM_000162	c.T122C	p.M41T	Likely Pathogenic
GCK	rs1064794268	NM_000162	c.A113C	p.Q38P	Likely pathogenic
GCK	rs762263694	NM_000162	c.C106T	p.R36W	Pathogenic
GCK	rs193922325	NM_000162	c.T74A	p.L25Q	VUS
PPP1R3A	.	NM_002711	c.T2990C	p.F997S	VUS
PPP1R3A	rs1057524893	NM_002711	c.C416G	p.S139C	Likely Benign
PAX4	rs1057524894	NM_006193	c.T104C	p.I35T	VUS
BLK	rs915462095	NM_001715	c.T2G	p.M1R	VUS
BLK	.	NM_001715	c.53_54insC	p.K18fs	VUS
GLIS3	rs1057524895	NM_001042413	c.A2765T	p.Q922L	VUS
GLIS3	rs977500969	NM_001042413	c.C1973G	p.A658G	Likely benign
GLIS3	rs377094207	NM_001042413	c.C1464G	p.D488E	VUS
AGPAT2	.	NM_006412	c.646_647TT	.	VUS
AGPAT2	.	NM_006412	c.T455C	p.V152A	VUS
AGPAT2	.	NM_006412	c.T274C	p.C92R	VUS
PTF1A	.	NM_178161	c.C265A	p.L89I	VUS
PTF1A	.	NM_178161	c.C343T	p.P115S	VUS
PTF1A	.	NM_178161	c.C440T	p.A147V	VUS
PTF1A	.	NM_178161	c.C520T	p.R174W	VUS
INS	.	NM_000207	c.C290G	p.T97S	Likely pathogenic
INS	rs1057524907	NM_000207	c.A278G	p.E93G	Likely Pathogenic
KCNJ11	.	NM_000525	c.G902C	p.R301P	VUS
KCNJ11	.	NM_000525	c.G794T	p.S265I	VUS
KCNJ11	.	NM_000525	c.C697T	p.L233F	Likely pathogenic

Table 3.10 Continued

ABCC8	.	NM_000352	c.A4043G	p.N1348S	VUS
ABCC8	.	NM_000352	c.A3817T	p.R1273W	VUS
ABCC8	rs761862121	NM_000352	c.A2666C	p.K889T	VUS
ABCC8	rs777461294	NM_000352	c.C1127T	p.S376F	VUS
BSCL2	.	NM_001122955	c.G969T	p.W323C	VUS
BSCL2	rs1057524896	NM_001122955	c.631-1G>C	.	VUS
BSCL2	rs1057524897	NM_001122955	c.A199C	p.N67H	VUS
HNF1A	.	NM_000545	c.G85C	p.A29P	VUS
HNF1A	.	NM_000545	c.143delA	p.E48fs	Pathogenic
HNF1A	.	NM_000545	c.G152T	p.G51V	VUS
HNF1A	rs1057524898	NM_000545	c.T616A	p.W206R	VUS
HNF1A	rs1057524908	NM_000545	c.694dupC	p.T231fs	Pathogenic
HNF1A	rs1057520504	NM_000545	c.G788A	p.R263H	Pathogenic
HNF1A	rs137853238	NM_000545	c.G815A	p.R272H	Pathogenic
HNF1A	.	NM_000545	c.956-1G>C	.	Pathogenic
PDX1	.	NM_000209	c.G208A	p.V70M	VUS
PDX1	.	NM_000209	c.C455T	p.A152V	VUS
PDX1	.	NM_000209	c.C719G	p.P240R	VUS
PDX1	.	NM_000209	c.C728G	p.P243R	VUS
HNF1B	.	NM_000458	c.A1630G	p.T544A	VUS
HNF1B	.	NM_000458	c.A1180G	p.N394D	VUS
HNF1B	.	NM_000458	c.1002_1006del	p.S334fs	Pathogenic
HNF1B	.	NM_000458	c.G499C	p.A167P	VUS
HNF1B	.	NM_000458	c.A262G	p.T88A	VUS
GATA6	rs924843423	NM_005257	c.A172G	p.S58G	VUS
MC4R	rs369841551	NM_005912	c.C466T	p.Q156X	Pathogenic
MC4R	rs372794914	NM_005912	c.A31G	p.T11A	VUS
INSR	.	NM_000208	c.C1741T	p.R581W	VUS
INSR	.	NM_000208	c.A1459C	p.K487Q	VUS
HNF4A	.	NM_001287183	c.G769T	p.D257Y	Likely Pathogenic
HNF4A	rs920346355	NM_001287183	c.A892G	p.I298V	VUS
FOXP3	rs1057524899	NM_014009	c.G409A	p.A137T	VUS
LEPR	rs1805094	NM_001198687	c.G1968T	p.K656N	Benign
GCK	.	NM_000162	c.C1113A	p.C371X	Pathogenic
HNF1B	.	NM_000458	c.1001_1002insCCC	p.S334delinsSP	VUS

Monogenic diabetes diagnosis

Thirty-eight patients were discovered to have 35 pathogenic or likely pathogenic (P/LP) variants in target genes. Based on the screening/intake questionnaire, the patients with mutations were less likely to be extremely overweight when diagnosed with diabetes (2.9% vs. 23.4%, p=0.01), and were less likely to have a prior diagnosis of type 1 diabetes (7.9% vs. 32.1%, p=0.01) (Table 3.11). The patients with monogenic diabetes

(22.7±13.3 years) were diagnosed at a significantly younger age compared to those without monogenic diabetes (27.4±14.7 years, p=0.04).

Of the patients identified to have P/LP variants, most had *GCK* variants (21 patients with 20 different variants) or *HNF1A* variants (seven patients with six different variants) (Table 3.12). For *GCK*, the two individuals with the same variant (c.680-1G>A) were siblings who enrolled in the study independently. For *HNF1A*, there were two individuals with the same variant (p.Arg272His) who were unrelated; notably this variant was previously found in at least 20 families, making it one of the most common known *HNF1A*-MODY variants [15]. An additional individual had two variants, missense variant (p.Ala29Pro, previously reported [15] but currently a VUS) and a frameshift variant (p.Glu48fs, classified as pathogenic), in *HNF1A*. The remaining variants were found in *HNF4A* (two possibly related but independently enrolling individuals with p.Asp260Tyr), *HNF1B* (in an individual with diabetes and congenital absence of one kidney), *INS* (two variants in two individuals), *KCNJ11* (one variant in one individual), *LMNA* (two variants in two individuals), *MC4R* (one variant in one individual), and *WFS1* (one individual homozygous for the p.Arg558Cys variant previously associated with non-syndromic young onset diabetes in the homozygous state[114].

Table 3.11 Screening questionnaire responses of patients with and without monogenic diabetes

PDMP Screening Questions	Monogenic diabetes	No monogenic diabetes	p-value
1. Were you diagnosed with diabetes or high blood sugar before 1 year of age?	0.0% (0/38)	23.6%(11/467)	1.00
2. Were you diagnosed with diabetes or high blood sugar at age 30 or younger? How old were you when you were diagnosed?	81.6% (31/38)	59.8%(278/465)	0.27
3. Were you extremely overweight when you were diagnosed?	2.9%(1/35)	23.4%(102/436)	**0.01**
4. As a child, did.do you have hearing or vision problems, intellectual disability (for example, learning disabilities, mental retardation, autism), birth defect(s) or kidney disease?	13.2%(5/38)	27.5%(129/469)	0.18
5. Do you have type 1 diabetes (if unsure, were you on insulin at diagnosis and have been ever since)?	7.9%(3/38)	32.1%(150/467)	**0.01**
6. Do you have a parent or a child with type 1 diabetes?	14.7%(5/34)	14.2%(61/431)	1.00
7. Do you have 2 or more people related to you by blood with diabetes? If yes, please list relationship, age of diagnosis, and type of diabetes.	89.2%(33/37)	83.5%(391/468)	0.89

Table 3.12 PDMP patients with positive genetic testing results

#	Gene	Ancestry	Sex	Age (y)	Prior Dx (Age Dx)	Variant	Current Treatment	ClinVar# (UMB Classification)	rs# (Refs or Novel)
1	*GCK*	AA	M	18	PreDM (18)	c.106C>T/ p.Arg36Trp	None	431973 (P)	rs762263694[16,244,245]
2	*GCK*	W	M	32	PreDM (5)	c.113A>C/ p.Gln38Pro	None	420070 (LP)	rs1064794268 [246,247]
3	*GCK*	W	M	41	T2D (26)	c.122T>C/ p.Met41Thr	OHA	393454 (LP)	rs1057524906[16,248,249]
4	*GCK*	W	F	13	PreDM (11)	c.127C>T/ p.Arg43Cys	None	585911 (LP)	rs1486280029[16,250,251]
5	*GCK*	W	F	59	DM 1.5 (54)	c.128G>A/ p.Arg43His	OHA	393453 (LP)	rs764232985 [252,253]
6	*GCK*	W	F	23	PreDM (13)	c.146C>A/ p.Thr49Asn	OHA	36204 (P)	rs193922286 [253] *
7	*GCK*	W	F	14	PreDM (14)	c.184G>A/ p.Val62Met	None	419624 (P)	rs1064793998[16,248,254,255]
8	*GCK*	W	F	15	DM Unknown Type (15)	c.571C>T/ p.Arg191Trp	None	426122 (LP)	rs1085307455 [16,245,246,256-260]
9	*GCK*	W	F	30	T2D (30)	c.676G>A/ p.Val226Met	Insulin	36243 (P)	rs148311934 [46,195,260-267]
10	*GCK*	W	F	22	T2D (22)	c.680-1G>A	OHA	393452 (P)	rs1057524905[260]
11	*GCK*	W	F	26	PreDM (25)	c.680-1G>A	None	393452 (P)	rs1057524905[260]
12	*GCK*	W	F	39	T2D (20)	c.718A>G/ p.Asn240Asp	OHA	617647 (LP)	rs1562715574[268]
13	*GCK*	W	F	66	PreDM (20) T2D (65)	c.748C>T/ p.Arg250Cys	None	393451 (LP)	rs1057524904[16,269,270]
14	*GCK*	W	M	12	PreDM (11)	c.787T>C/ p.Ser263Pro	None	36258 (LP)	rs193922331[271]
15	*GCK*	W	F	34	PreDM (16) T2D (21)	c.1016A>G/ p.Glu339Gly	OHA	393450 (P)	rs1057524903[258,272]
16	*GCK*	A	F	27	T2D (27)	c.1113C>A/ p.Cys371*	OHA	617645 (P)	rs556581174[266]
17	*GCK*	W	F	57	T2D (30)	c.1113C>G/ p.Cys371Trp	OHA	447379 (LP)	rs556581174[16]
18	*GCK*	AA	F	51	PreDM (30s) DM (45)	c.1253+2T>A	OHA	393449 (P)	rs1057524902 (Novel)
19	*GCK*	A	F	34	PreDM (19) T2D (27)	c.1254-2A>G	Insulin, OHA	LP	NA[16]

Table 3.12 Continued

20	*GCK*	W	F	28	PreDM (19)	c.1344delC/ p.Ala449Argfs*165	None	393448 (P)	rs1057524901 (Novel)
21	*GCK*	W	F	26	GDM (20) T2D (26)	c.1361C>A/ p.Ala454Glu	Insulin, OHA	393447 (LP)	rs1057524900[263,270,273]
22	*HNF1A*	W	F	31	T2D (22)	c.85G>C/ p.Ala29Pro c.143delA/ p.Glu48Glyfs*107	OHA	VUS/P	NA[15]/ NA (Novel)
23	*HNF1A*	W	F	44	T2D (19)	c.694dupC/ p.Leu232Profs*7	Insulin	380275 (P)	rs1057524908 (Novel)
24	*HNF1A*	W	F	42	T2D (14)	p.788G>A/ p.Arg263His	OHA	379138 (P)	rs1057520504 [274-277]
25	*HNF1A*	AA	F	29	T2D (29)	c.815G>A/ p.Arg272His	None	14931 (P)	rs137853238[274-277]
26	*HNF1A*	AA	F	20	T1D (20)	c.815G>A/ p.Arg272His	None	14931 (P)	rs137853238[274-277]
27	*HNF1A*	W	F	8	DM (8)	c.956-1G>C	None	617646 (P)	rs1565886545 (Novel)
28	*HNF1A*	W	F	9	PreDM (8)	c.1330_1331delCA/ p.Gln444Glufs*104	None	P	rs776793516[278-280]
29	*HNF4A*	W	F	29	T2D (28)	c.778G>T/ p.Asp260Tyr	None	549555 (LP)	rs1236613475 (Novel)
30	*HNF4A*	W	F	60	GDM (29) T2D (43) T1D (early 20s)	c.778G>T/ p.Asp260Tyr	OHA	549555 (LP)	rs1236613475 (Novel)
31	*HNF1B*	A	M	46	T2D (40s) Congenital absence of one kidney	c.1002_1006del/ p.His336Profs*22	Insulin	P	NA (Novel)
32	*INS*	W/H	F	5	T1D (4)	c.278A>G/ p.Glu93Gly	Insulin	393455 (LP)	rs1057524907 (Novel)
33	*INS*	W/A	M	31	T2D (27)	c.290C>G/ p.Thr97Ser	OHA	617648 (LP)	rs1564911425 (Novel)
34	*KCNJ11*	W	F	9	T1D (15 mos.)	c.697C>T/ p.Leu233Phe	Insulin	617651(LP)	rs1564865302[281]
35	*LMNA*	W	M	50	Post-transplant DM/lipodystrophy (50)	c.1045C>T/ p.Arg239Trp	Insulin	66762 (LP)	rs267607555[282,283]
36	*LMNA*	W	F	39	T2D (18)	c.1445G>A/ p.Arg482Gln	OHA	14486 (P)	rs11575937 [284]
37	*MC4R*	AA/W	F	38	T2D (25)	c.466C>T/ p.Gln156*	Insulin, OHA	562356 (P)	rs369841551[285]
38	*WFS1*	W	M	52	T1D (30)	c.1672C>T/ p.Arg558Cys (homozygous)	Insulin, OHA	198835 (LP)	rs199946797[114,286-288]

* Novel at the time of analysis.

We were able to identify pathogenic or likely pathogenic (P/LP) monogenic diabetes variants in 12.1% (38 out of 313) of individuals suspected of having monogenic diabetes (Table 3.13). The likelihood of identifying a mutation was strongly dependent on the method of recruitment.

Clinic Screening

Screening of 1,904 individuals in the clinic waiting room, evaluation of existing laboratory records, after accounting for refusals, and detailed review of family and medical history yielded 130 (6.8%) individuals meeting algorithmic criteria for monogenic diabetes. Of these 130, four patients, or 5.3%, had mutations. All four patients met algorithm criteria 3, non-type 1 diabetes diagnosed by age 30, and comprised 5.1% (4/78) of screened clinic patients meeting these criteria.

EHR Portal Screening

Through sending electronic questionnaires to 286 individuals in MyGeisinger Portal identified as having diabetes, and later evaluation of laboratory results and family and medical history, 19 (6.6%) individuals meeting algorithmic criteria were tested. One patient (7.7%) meeting algorithm criteria 3 had a mutation.

Physician Recognition

Participating physicians recognized patients for PDMP in two ways: direct referrals by physicians at four sites and manual chart review by EAS at UM-CDE. Overall, of 108 patients referred by their own physicians and passing algorithm criteria, 19 patients had mutations, for a hit rate of 17.6% in physician referred cases. Of these 19, just over half (10/19, or 52.6%) qualified on the basis of non-type 1 diabetes diagnosed by age 30 and the remainder (9/19, or 47.4%) met no criteria other than

clinical suspicion. Of these nine patients, five had prediabetes by age 30: one had a normal BMI and was diagnosed with type 2 diabetes at age 45 and then had two healthy adult children with incidental hyperglycemia; one was diagnosed with type 1 diabetes, wa suspected of monogenic diabetes because her diabetes was picked up incidentally on a routine screen, and was antibody negative; one was diagnosed with GDM at 29 and T2D at age 43, with a near-normal BMI; and one was diagnosed with T2D at age 50, but with low BMI (17.2 kg/m^2).

Manual chart review at the UM-CDE yielded 19 additional individuals concluded by the study team to meet criteria, and two of those individuals were found to have disease causing variants, for a hit rate of 10.5%. One was diagnosed with type 1 diabetes at age 15 months, had a parent with diabetes, and was found to have a variant in *KCNJ11*. One had been diagnosed with type 1 and 2 diabetes at different stages, was born with only one kidney, and was found to have a variant in *HNF1B*, which causes MODY5, also known as renal cysts and diabetes (RCAD) and frequently associated with urogenital anomalies. Moreover, four patients were recognized by both their physician and manual chart review at UM-CDE as meeting criteria, but none were found to have disease-causing mutations. An additional 25 patients were recognized in stage 2 and did not meet algorithm criteria but were sequenced as controls and no mutations were found.

Self-referrals

Of 37 self-referring patients passing algorithm criteria, 12 patients had mutations, for a hit rate of 32.4% in self-referrals. Of these 12 patients, seven had non-type 1 diabetes diagnosed by age 30. Of the remaining five cases, which did not meet pre-specified criteria, two were diagnosed at older ages (one at age 54 and one at age 65 but

with recollections of mild hyperglycemia in her 20s), two were diagnosed with prediabetes before age 30 (one at age 5 and one at age 25), and one had type 1 diabetes with periods of non-insulin dependence. An additional 31 patients were self-referred in stage 2, did not meet criteria, and no mutations were found.

Table 3.13 Hit rate by recruitment methods and algorithm criteria

All	1	2	3a	3b	3a and 3b	4	5	6	7	Total	Hit rate	Screened	Hit rate in screened
Physician Recognition	0/7	1/1	8/19	2/21	10/40	0/11	1/5	0/2	9/61	21/127	16.50%	21/222	9.50%
Self-Referral	-	-	4/9	3/7	7/16	0/4	0/1	-	5/16	12/37	32.40%	12/109	11.00%
Clinical Screen	0/1	0/9	3/20	1/58	4/78	0/29	0/2	0/1	0/10	4/130	3.10%	4/1904	0.20%
EHR Portal Screen	-	-	0/4	1/9	1/13	0/6	-	-	-	1/19	5.30%	1/286	0.30%
Total	0/8	1/10	15/52	7/95	22/147	0/50	1/8	0/3	14/87	38/313	12.10%	38/2522	1.50%
Hit Rate	0.00%	10.00%	28.80%	7.40%	15.00%	0.00%	12.50%	0.00%	16.10%				

*Criteria (see Figure 3.1 for full algorithm):
1: Diagnosed with diabetes ≤ 1 year of age
2: Diagnosed with type 1 diabetes + a parent or child with type 1 diabetes, IA-2-, C-peptide+
3: Diagnosed with non-type 1 diabetes ≤ 30 years of age
    a: Not on insulin and IA-2-
    b: On insulin and IA-2- and C-peptide+
4: Diagnosed with type 2 diabetes < 45 years of age and 2+ relatives in same lineage diagnosed ≤ 50 years of age and IA-2 negative and C-peptide positive
5: Hearing, visual, or cognitive impairment, birth defects or kidney disease in childhood
6: Non-diabetic with fasting glucose > 100 mg/dl during a current or past pregnancy and pre-pregnancy BMI < 25 kg/m^2, IA-2- and C-peptide+
7: Clinical suspicion

3.5 Discussion

The PDMP was designed to implement, disseminate, and evaluate the impact of a systematic approach to screening, genomically diagnosing, and promoting individualized therapy for monogenic diabetes. Here, we have reported the results of the first stage of our study: initial identification of individuals with monogenic diabetes among those with diabetes or prediabetes.

Through a combination of physician recognition, clinical screening, self-referral, and EHR portal screening, we identified 313 individuals suspected of having monogenic diabetes and identified P/LP mutations in 38 of these individuals, mostly in *GCK* (21 individuals) and *HNF1A* (seven individuals). Mutations in both genes are actionable, generally indicating that either no treatment (*GCK*) or low-dose sulfonylureas (*HNF1A*) are the most appropriate therapy. The remaining ten individuals had mutations in other genes: *HNF4A* (n=2), *KCNJ11* (n=1), *INS* (n=2), *LMNA* (n=2), *HNF1B* (n=2), *WFS1*(n=1), and *MC4R* (n=1). Of these, *HNF4A* and *KCNJ11* have clear associated management strategies (low dose sulfonylureas in *HNF4A* and high dose sulfonylureas in *KCNJ11*), and all provide important information about etiology and familial risk. We are following all cases for changes in management and glycemic outcomes, as well as performing cascade testing of relatives, and will report the impact in the future. The *KCNJ11* mutation was previously reported in one individual who transitioned from insulin to high dose sulfonylureas, and our patient has already successfully done the same after nine years of insulin therapy.

Our intention in designing this study was to evaluate in multiple settings a simple method to identify individuals for testing for monogenic diabetes. The screening

questionnaire was designed to identify red flags for monogenic diabetes to select people for sequencing because it is not currently practical or cost-effective to sequence all individuals with diabetes to identify the approximately 2% with monogenic diabetes, and such a screening strategy would potentially be more scalable than training all practitioners to recognize suspicious cases by clinical judgment. Unfortunately, we were unable to achieve a high hit rate with our questionnaire as a whole, though sequencing all individuals diagnosed with non-type 1 diabetes by the age of 30 had a hit rate of 5.1% (4/78) when picked up through clinical screening and an overall hit rate of 15.0% (22/147) when physician and self-referrals were included. The 5.1% hit rate for young onset non-type 1 diabetes in the clinical screening is likely an underestimate, since the awareness raised around this study among local referring clinicians likely took some candidates out the screening pool. Physician referrals had an overall hit rate of 18.3%, but the most successful strategy for identifying cases came from sequencing self-referrals, with 32.4% of sequenced self-referrals found to have a P/LP mutation. It is also worth noticing that manual chart review by an endocrinology geneticist, as a supplement to physician referral on site, was able to pick up two cases of less common types of monogenic diabetes, emphasizing the different perspectives from different physicians when evaluating atypical diabetes. A pragmatic way, as recommended by Riddle et al [289] is to establish a regional center with specialized expertise to support additional case identification through systematic chart review.

Our data, showing that 38 individuals, mostly previously diagnosed with type 2 diabetes, have monogenic diabetes concur with previous reports that monogenic diabetes is underdiagnosed. We found that the most successful strategy in the PDMP study for

identifying cases is through self-referral, followed by physician referral, underscoring the importance of developing strategies to raise awareness of monogenic diabetes in both the provider and lay communities. In addition, our results combined with previous reports[167] also suggest that all individuals with non-type 1 diabetes diagnosed by age 30 should be evaluated for monogenic diabetes, and a large number of mutations identified in young pre-diabetic individuals suggest that those found to have hyperglycemia that does not reach the diabetic range should be included in this group as well. While the self-referral group had the highest hit rate, relying solely on this method is likely to exacerbate health disparities, given the high educational level required. Therefore, it is important to develop an infrastructure in which monogenic diabetes is considered strongly on the differential diagnosis.

3.6 Conclusion and Recommendations

Through a combination of diabetes clinic screening, provider recognition/referral, and self-referral, combined with standard biomarkers, we identified 313 individuals suspected to have monogenic diabetes and were able to identify a clinically significant mutation in 38, or 12% of these individuals. Most (31/38) had variants in genes (*GCK*, *HNF1A*, *HNF4A*, or *KCNJ11*) with clear potential implications for treatment, indicating that systematic effort to identify these patients is worthwhile. For easily implemented criteria, our study was only powered enough to contribute to the growing body of evidence that non-type 1 or autoantibody-negative, C-peptide positive diabetes diagnosed before the age of 30 should prompt evaluation for monogenic diabetes, with 15% of individuals in this category having P/LP variants. In addition, the high hit rate of physician-referrals was consistent with the agreement from the monogenic diabetes

expert forum that training healthcare providers and having centers of expertise will benefit the diagnosis of monogenic diabetes[289]. Finally, the finding that approximately one third of sequenced self-referrals tested positive emphasizes the importance and value of educating the public on diabetes heterogeneity and motivates future research to understand the characteristics of patients that enable them to self-identify their presentation are not typical types of diabetes.

Chapter 4 : Comprehensive Re-analysis of DNA Sequencing Data Improves Diagnostic Yield of Monogenic Diabetes

The DNA sequencing data generation and analysis determine what variants to be analyzed further to explain the patient phenotype. The variant interpretation and classification step decides whether suspected variants likely cause the patient phenotype. Though sequencing methods control the sequencing data generation step, the sequencing data can be re-analyzed by different tools, and the variant classification can be updated based on interpretation of new evidence. Both have the potential to improve the diagnostic yield of monogenic diabetes. This chapter demonstrated the increment of monogenic diabetes variant detection in two projects by focusing on different components of sequencing data analysis and variant classification.

4.1 NGS Panel for the Diagnosis of Monogenic Diabetes: Comprehensive Re-analysis Improves Diagnostic Rate[4]*

4.1.1 Abstract

Purpose: Genetic testing is required for the accurate diagnosis and etiology-based treatment of monogenic diabetes. However, many factors, including lack of sequence data re-analysis on patients with negative genetic testing results, result in an inadequate diagnostic rate. Comprehensive re-analysis of NGS panel data was applied in the Personalized Diabetes Medicine Program to improve the diagnostic rate.

Methods: NGS panel data of 10 Maturity Onset Diabetes of the Young (MODY) genes from 313 Personalized Medicine Diabetes Program (PDMP) participants suspected of

[4] Haichen Zhang, Kristin A. Maloney, Linda J.B. Jeng, Toni I. Pollin. NGS panel for the Diagnosis of Monogenic Diabetes: Comprehensive Re-analysis Improves Diagnostic Rate. In preparation for submission
*HZ performed the data analysis and drafted the manuscript.

monogenic diabetes were re-analyzed by the Genome Analysis Tool Kit (GATK) pipeline and re-interpreted according to gene-specific American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) sequence variant interpretation guidelines. Also, these variants were analyzed for splicing effects by multiple splicing prediction tools (SpliceAI, Human Splicing Finder, MaxEntScan, etc.)

Results: Four additional pathogenic or likely pathogenic (P/LP) variants were identified during the re-analysis in addition to the 33 P/LP variants and increased the total diagnostic yield of monogenic diabetes in the PDMP study from 11.8% (37[5]/313) to 13.1% (41/313).

Conclusion: This article highlights the importance of variant re-interpretation including sharing of the case and functional data and the use of orthogonal validation of different variant calling tools in regions with low sequencing depth.

4.1.2 Introduction

Monogenic diabetes is an uncommon type of diabetes caused by variants in one of more than 40 genes, and it accounts for less than 5% of all diabetes [1,2]. Since the first description of MODY (maturity-onset diabetes of the young, a name given based on the observation of a T2D-like presentation in young people with an autosomal dominant pattern of inheritance) by Fajans [3,4], subsequent research has added to our understanding of the phenotypic and genetic heterogeneity of this disorder. MODY is the most common form of monogenic diabetes, and 14 genes have been associated with MODY, though the

[5] The re-analysis started while the PDMP initial analysis was ongoing. The *GCK*:c.146C>A, p.(Thr49Asn) was identified in the re-analysis but been included in the Chapter 1 analysis.

number is decreasing because of following studies showing lack of support of pathogenicity of *BLK*, *KLF11* and *PAX4*[10]. The overlapping phenotype of monogenic diabetes with type 1 diabetes (T1D) and type 2 diabetes (T2D) often leads to misdiagnosis. The correct treatment of each subtype of monogenic diabetes depends on the pathogenesis of diabetes, which can only be determined by identifying the causative gene through genetic testing. However, the diagnostic rate of monogenic diabetes is inadequate, and the impacting factors include but are not limited to limitations and complexities of sequencing data analysis and challenges in variant interpretation.

In terms of the sequencing data analysis, the routine pipeline usually contains one set of analytic tools, including sequence read alignment, pre-processing, and variant calling. The output variants are then filtered to include only rare coding or canonical splice variants. Firstly, though often validated before applying to clinical diagnosis, different variant caller algorithms, built on different models aiming for particular sequencing errors, have been reported to show variable sensitivity and specificity on exome sequencing data[290], especially in complicated regions. For example, the Genome Analysis Tool Kit (GATK), developed by the Broad Institute, is regarded as the gold standard in variant calling. The GATK used Variant Quality Score Recalibration (VQSR) during the quality control step when analyzing large numbers of exome sequencing data. VQSR is a machine learning algorithm, which has poor performance when the training set cannot include enough rare variants such as monogenic diabetes-related variants, but the GATK hard filtering parameters can be manually tailored to meet the requirement of NGS panel data. Hence, orthogonal validation using different tools will help to discover the missing variants that could lead to the diagnosis. Secondly, variants that tend to be

filtered out, such as promoters and non-canonical splicing variants, have also been found to be disease-causing. There have been studies reporting non-coding variants to cause monogenic diabetes [15,16]. However, bioinformatics pipelines usually filter out synonymous variants and deep intronic variants that might influence splicing as the first step due to overall inadequate evidence supporting the pathogenicity of these variants. Recent publications have shown that RNA analysis on these types of variants could increase the diagnostic rate, though splicing prediction tools are more error-prone than RNA analysis. S prediction tool with the best performance could reach 90% accuracy (sensitivity 89.9%, specificity 91.6%) [291]. Therefore, broadening the analysis to non-canonical splicing variants by multiple splicing prediction tools has the potential to increase the diagnosis rate.

Variant interpretation has been historically ununified. The guidelines for variant interpretation among labs lack uniformity and result in discrepancies; the databases storing interpreted variants and patient information need consolidation. In 2015, the American College of Medical Genetics and Genomics (ACMG), together with the Association for Molecular Pathology (AMP) published recommendations for sequence variant interpretation, aiming for reducing the discrepancies and improving the accuracy of variant interpretation. However, the guideline was designed for Mendelian disease in general and specification of gene and/or disease is needed of application in the interpretation process. In an endeavor to improve accuracy, standardization, and concordance of variant interpretation for monogenic diabetes, the Clinical Genome Resource (ClinGen) Monogenic Diabetes Expert Panel (MDEP) was formed to develop gene/disease-specific guidelines for monogenic diabetes. The ClinGen MDEP is one of

many ClinGen expert panels dedicated to specifying the gene-disease and variant-disease relationships aiming for precision medicine in clinics and research.

Recently, the findings that re-analysis of NGS data could improve diagnostic rate has been reported. Most of these studies were performed on WGS or ES data and only a few on targeted panels. The data used to improve diagnostic rate were mostly newly discovered gene-disease relationships, newly published data, and including CNV in the analysis. (Table 4.1)

The goal of this study was to utilize multiple methods to comprehensively re-analyze NGS panel data from the Personalized Diabetes Medicine Program (PDMP) and thus improve the diagnosis rate of monogenic diabetes. The PDMP, as a part of the NIH IGNITE (Implementing GeNomics In PracTicE) Network, was designed to implement, disseminate, and evaluate a sustainable method for identifying, genomically diagnosing, and promoting individualized therapy for monogenic diabetes.

Table 4.1 Summary of published re-analysis studies on NGS data

Phenotype	NGS data	Sample size	Initial diagnostic rate	Re-analysis diagnostic rate	Major reasons of improving diagnostic rate	Reference
Intellectual disability	Panel/ES	33	42.00%	57.00%	Newly discovered gene-disease relationship; Include CNV calling; Updated literature; updated phenotype data	292
Neurodevelopmental disorder	Panel	116	0.00%	5.20%	Updated variant interpretation guidelines; Updated variant-disease relationship; Updated gene-disease relationship; updated phenotype data	293
Encephalopathy	Panel	384	0.00%	1.80%		293
Suspected genetic disorder	GS	48	0.00%	4.20%	Correction of manual error; updated literature; Functional studies	294
Suspected genetic disorder	ES	40	0.00%	10.00%	Updated gene-disease and variant disease association	295
Intellectual disability	ES	50	0.00%	12.00%	Improvement in bioinformatics tools; updated databases and literature; updated phenotype data	296
Suspected genetic disorder	GS	100	34.00%	41.00%	Updated gene-disease association and genotype-phenotype association	297
Suspected genetic disorder	ES	54	30.00%	41.00%	Improvement in bioinformatics tools; updated gene-disease association; updated phenotype data	298
Congenital anomalies and intellectual disability	ES	156	0.00%	15.40%	Updated literature; include CNV calling; correction for misinterpretation	299
Developmental disorders	ES	1133	27.00%	43.00%	Improved bioinformatics pipeline, updated variant annotation; updated filtering strategies; updated gene-disease association	300
Suspected genetic disorder	ES	250	24.80%	46.80%	Updated gene-disease association; variant reclassification; include CNV calling; familial studies; clinical update	301
Suspected genetic disorder	ES	2000	25.20%	36.70%	Updated gene-disease association; variant reclassification; include CNV calling; familial studies; clinical update	301
Suspected genetic disorder	ES	240	0.00%	15.80%	Updated gene-disease association; updated genotype-phenotype association; updated variant classification	302
Hearing loss	Panel	210	39.00%	43.00%	Updated literature; updated gene-disease association; include CNV calling	303

Panel; NGS panel; ES: exome sequencing; GS: genome sequencing; CNV: copy number variants

4.1.3 Materials and methods

Patient cohort and sequencing data collection

The detailed patient recruitment and sequencing data collection of PDMP were described in Chapter 3. To summarize, over 2015-2018, the PDMP screened 2522 patients with diabetes from four participating clinics and through self-referring. The PDMP enrolled 507 patients meeting selection criteria and sequenced 456 patients who were suspected of having monogenic diabetes (313 patients) or as the control group (143 patients).

Variant detection

In the initial analysis, the sequence reads of PDMP were aligned to hg19 by Torrent Mapping Alignment Program (TMAP) and then called for variants by Torrent Variant Caller (TVC) which were parts of the Thermo Fisher's Torrent Suites Software.

The re-analysis analyzed the alignment data of the PDMP in BAM format (binary version of sequence alignment map file) using the GATK pipeline. Hard filtering based on data quality was performed separately on single nucleotide variants and insertions/deletions. The parameter was selected based on the decision tree generated from the Sanger confirmed variants from the initial PDMP analysis. (Figure 4.1, 4.2)

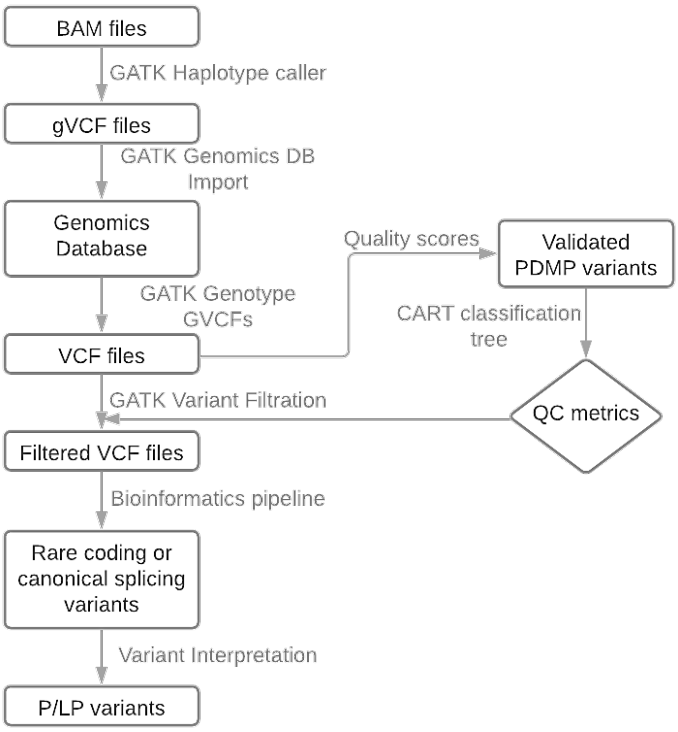
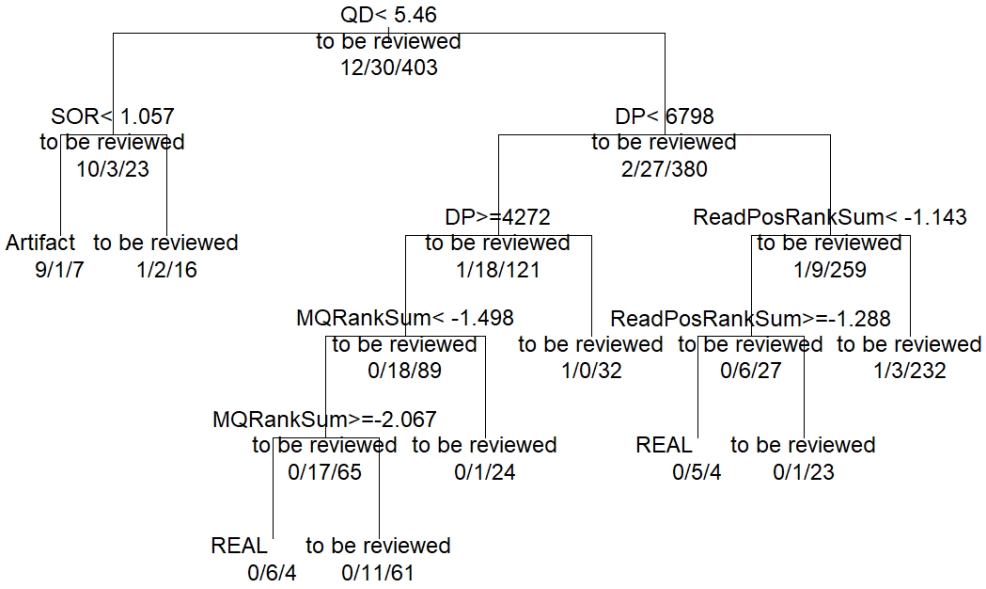


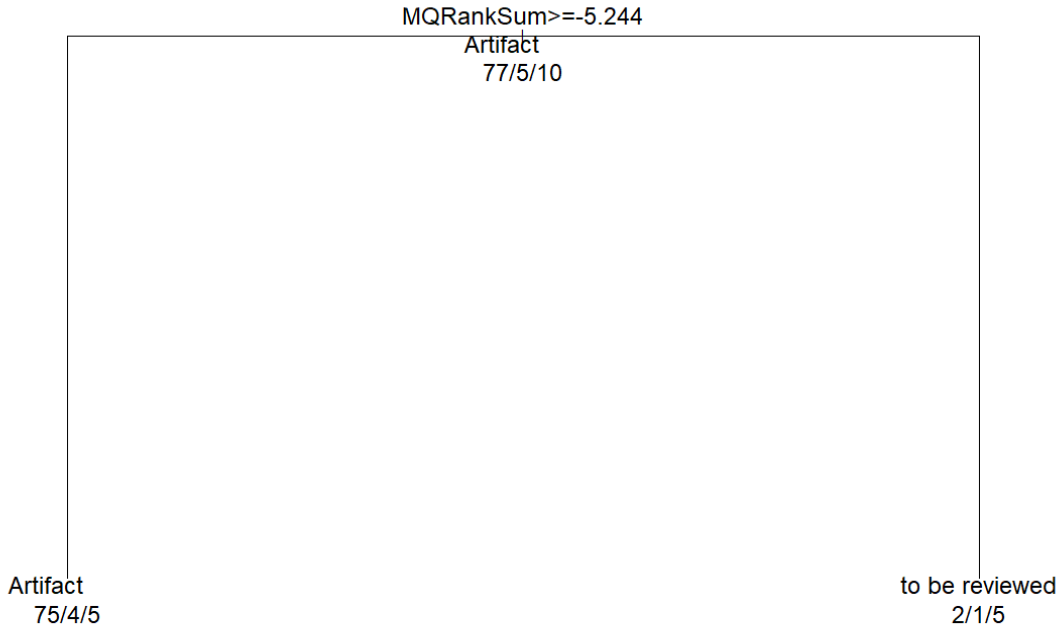
Figure 4.1 Workflow of re-analyzing PDMP samples using the GATK pipeline

Figure 4.2 Decision trees of determining quality score cutoffs of GATK filtering.
a. SNV classification tree. b. Indel classification tree. The classification step started from the top node. The cutoffs of each node were shown at the top, the category with the maximum number of variants in the middle and the numbers of variants in each category at the bottom, in the format of "Artifact/Real/to be reviewed" QD, Quality by Depth; SOR, Strand Odds Ratio; DP, Depth of coverage; ReadPosRankSum, Read Pos Rank Sum Test; MQRankSum, Mapping Quality Rank Sum Test.

a.SNV classification tree



b.Indel classification tree



Variant classification

This project developed a pipeline that includes variants annotation and filtration. The pipeline annotated the sequencing data using Annovar[304], a package to functionally annotate genetic variants integrating multiple up-to-date resources, with population frequency (gnomAD, ExAC, 1000 Genome, NHLBI GO Exome Sequencing Project (ESP)), in silico predictors (REVEL, SIFT, Polyphen, LRT, MutationTaster, MutationAssessor, FATHMM, PROVEAN, VEST, CADD, DANN, FATHMM-MKL, MetaSVM, MetaLR, GERP++, phyloP, and SiPhy), dbSNP, and ClinVar. The database of numbers of heterozygotes and homozygotes for specific variants was obtained from the Type 2 Diabetes knowledge portal [305] as part of the annotation to evaluate whether the evaluated variants were enriched in T2D patients versus controls. The resulting variants were filtered, and only low frequency (<5% minor allele frequency [MAF] in ExAC, 1000 Genome, and ESP all population) coding or canonical splice-site variants remained for the initial interpretation.

In the initial analysis, the PDMP variant classification team interpreted the annotated and filtered variants for pathogenicity according to American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) guidelines[189]. These guidelines incorporated computational predictions, clinical phenotypes, functional assessments, and population frequencies into consideration and categorized the variants into five categories (Pathogenic (P), Likely Pathogenic (LP), Variant of Unknown Significance (VUS), Likely Benign (LB), or Benign (B)). Later in the re-analysis, the variants of PDMP participants were re-classified by HZ based on gene-specific ACMG/AMP guidelines being established by the ClinGen MDEP [306].

Splicing effect prediction

In the re-analysis, the variants of PDMP participants were retrieved and screened by SpliceAI, a deep learning splicing prediction tool. Variants that met the following criteria were kept for further analysis using Human Splicing Finder (HSF), MaxEntScan (MES), and NNSplice.: 1) Δscore above 0.2 [307]; 2) Popmax filtering allele frequency in gnomAD below 0.01%; 3) Intronic variant and missense or synonymous variant which alters the last three bases of an exon preceding a donor splice site or the first three bases of an exon following a splice acceptor site; 4) Seen in fewer than three participants (to eliminate platform-related artifacts).

Statistical analysis

Categorical values were compared by χ^2 test or, where expected counts were below 5, by Fisher's exact test. Decision tree was generated using CART package. All statistical analysis was performed using R 3.6.3 [308].

4.1.4 Results

Re-classification of PDMP variants identified new monogenic diabetes-causing variants

The re-classification starting from Jan 2019 included 97 variants in 10 MODY genes that were reviewed by the PDMP variant classification team. *PAX4, KLF11,* and *BLK* were removed because of lack of evidence supporting the gene-disease relationship because of higher population frequency than MODY prevalence and lack of rare variant enrichment in the Exeter MODY cohort [10]. The re-analysis resulted in re-classification of three more variants from VUS to pathogenic or likely pathogenic, which could lead to the

new diagnosis. (Table 4.2) In addition, the re-analysis helped to re-classify a total of eight VUSs to benign or likely benign. The re-classification helped to resolve the clinical significance of VUSs. (Table 4.3)

Table 4.2 PDMP pathogenic or likely pathogenic variants identified through re-analysis

Gene	Transcripts	BaseChange	AminoAcidChange	ACMG criteria (Updated criteria in bold)	Classification	Reason for classification change
GCK	NM_000162.5	c.146C>A	p.(Thr49Asn)	PP2, PM2_Supporting, **PP1_strong, PS3**	Pathogenic	New case report including variant-disease co-segregation data and functional data
HNF1A	NM_000545.8	c.616T>A	p.(Trp206Arg)	PM1, PM2_supporting, PP3, **PP4_Moderate**	Likely Pathogenic	Gene-specific rules on disease-specific phenotype
HNF1B	NM_000458.4	c.499G>C	p.(Ala167Pro)	PM2_supporting, PP3, PM1, PP4, **PS3_Moderate**	Likely Pathogenic	Functional data from collaborator

Table 4.3 PDMP benign or likely benign variants identified through re-analysis

Gene	dbSNP ID	Transcripts	BaseChange	AminoAcidChange	ACMG criteria (Updated criteria in bold)	Classification	Reason for classification change
ABCC8	rs200563930	NM_000352.6	c.3976G>A	p.(Glu1326Lys)	**BA1**	Benign	Gene-specific rules on population frequency evidence: cutoffs for population frequencies that are too common to be disease-causing.
HNF1A	rs772222326	NM_000545.8	c.142G>A	p.(Glu48Lys)	**BS1**, BS3_Supporting	Likely benign	
HNF1A	rs202039659	NM_000545.8	c.1541A>G	p.(His514Arg)	**BA1**, PP3	Benign	
HNF1B	rs140562402	NM_000458.4	c.244G>A	p.(Asp82Asn)	**BA1**, PP3	Benign	
NEUROD1	rs201293992	NM_002500.5	c.750C>A	p.(Ser250Arg)	**BA1**	Benign	
PDX1	rs936474667	NM_000209.4	c.28G>A	p.(Ala10Thr)	**BA1**, BP4	Benign	
HNF1A	rs140491072	NM_000545.8	c.965A>G	p.(Tyr322Cys)	**BA1**	Benign	

The improvement of diagnostic rate in PDMP participants was made possible by the newly published literature and updated guidelines. The c.146C>A; p.(Thr49Asn) variant in *GCK* (NM_000162.5) was classified as VUS in 2017 due to lack of evidence. In 2018, Liu et al. published a case report of an affected family with six informative meioses and functional studies showing reduced enzyme activity and thermostability of GCK p.(Thr49Asn), thus enabled the usage of PP1_Strong and PS3_Moderate, which promoted the classification from VUS to pathogenic. In addition to the up-to-date evidence, the re-classification process was greatly attributed to the guidance and resources from the ClinGen MDEP. The ClinGen MDEP established rule specifications for monogenic diabetes genes, which aims to improve the accuracy and concordance of variant classification in monogenic diabetes genes and improve the ability to diagnose monogenic diabetes accurately. The c.616T>A; p.(Trp206Arg) in *HNF1A* (NM_000545.8) was upgraded from VUS to likely pathogenic because of the clear definition of PP4_Moderate by MDEP, which evaluated the phenotype data that was specific for the disease etiology. Another important aspect was that MDEP specified gene-specific MAF cutoffs, which were based on the most common MODY-causing variants in each gene and the estimated disease prevalence. The observation of a variant having a MAF in an extensive population database (e.g., gnomAD) being higher than the gene-specific cutoff disputes variant pathogenicity; the variant being absent in the database supports variant pathogenicity. Such specification was the major reason for the re-classification of eight VUSs to benign or likely benign variants in PDMP. The ClinGen MDEP also facilitated sharing of data and expertise, which contributes to both rule specification and variant classification. The c.499G>C; p.(Ala167Pro) in *HNF1B*

(NM_000458.4) was upgraded to Likely Pathogenic because of the functional studies performed by Norway collaborator who proved that Ala167Pro reduced transactivation activity and reduced protein expression [309].

Splicing prediction tools perform differently on PDMP variants

A total of 2,530 variants were screened for effect on splicing by SpliceAI. Nineteen variants were included in the further analysis by HSF, SPiCE-SSF, SPiCE-MES [310], and NNSplice among the 48 variants with spliceAI Δscore above 0.2. Seventeen variants were single nucleotide change, and two spanned multiple nucleotides. Variants were located across 12 genes in total, with particularly high numbers of variants in *ABCC8* (three) and *GCK* (three). Nine of the 19 variants were previously discovered canonical splicing variants, and all had Δscore above 0.95; four of them (three from *GCK* and one from *HNF1A*) were previously classified as pathogenic or likely pathogenic.

The comparison between splicing effects predictors (Table 4.4) showed among the 10 predicted non-canonical splicing sites by spliceAI, all other tools missed the acceptor gain sites, three of them missed the donor gain sites, and all missed one or two of the acceptor loss sites.

Table 4.4 Comparing performance of splicing effect predictors on non-canonical splicing variants

Scoring metric	Acceptor gain	Acceptor loss	Donor gain
HSF (2%)	0	2	1
SpliceAI (0.2)	4	4	2
SPiCE-SSF (5%)	0	2	0
SPiCE-MES (10%)	0	3	0
NNSplice (5%)	0	2	0

Variant calling using GATK identified new pathogenic variant in PDMP variants

The GATK haplotype caller and genotype caller identified all P/LP variants which were detected by TVC in the initial analysis. In addition, the GATK variant calling step identified 280 new SNVs and 6,043 new Indels. (Table 4.5, Table 4.6)

Table 4.5 Results of GATK filtration based on decision tree cutoff

	TVC	GATK	Overlaps	GATK newly discovered after filtration
All	705	6862	539	89
Exonic	691	6721	526	87
Splicing	14	141	13	2
SNVs	520	727	447	49
INDELs	185	6135	92	40

Table 4.6 Results of GATK filtration based on decision tree cutoff

SNVs	Number of variants	Mean QD + SD
Newly identified	268	3.14+3.9
Newly identified QD>=5.46	50	8.91+3.48
Overlaps	447	12.77+4.62
Confirmed P/LP	30	13.69+4.01
INDELs	**Number of variants**	**Mean QD + SD**
Newly identified	6044	18.11+780.87
Newly identified SOR<1.285	412	212.07+2987.39
Newly identified QD>=9.48	642	148.66+2393.59
Newly identified SOR<1.285 and QD>=9.48	40	2161.45+9472.80
Overlaps	92	10.37+7.80
Confirmed P/LP	5	15.33+7.17

Overall, there are 89 variants that were newly discovered by GATK haplotype caller and genotype caller. After evaluation of BAM files for quality and variant interpretation, one *GCK* missense c.898G>A;p.(Glu300Lys) was determined to be LP and confirmed by research Sanger. (Figure 4.3) The QD of this variant is high (13.39) compared to the mean QD of confirmed PDMP pathogenic or likely pathogenic variants.

Figure 4.3 Sanger sequencing confirmation of *GCK*:c.898G>A; p.(Glu300Lys)
The figure shows chromatograms of NM_000162.5(*GCK*):c.898G>A; p.(Glu300Lys). The upper part shows the reference sequence from Varsome.com, and the lower part shows the automated sequence predictions from the chromatogram (GeneStudio). The arrow shows the target variant.

4.1.5 Discussion

Re-analysis of NGS panel data increased the diagnostic rate of monogenic diabetes in PDMP

This study implemented comprehensive methods in re-analysis of NGS panel data and improved the diagnostic rate of PDMP participants who were suspected of having monogenic diabetes from 11.82% to 13.10% (37 to 41 out of 313). The small increment of diagnostic rate is consistent with 1-2% of improvement from previous reports of targeted gene panel re-analysis [293,303].

The Ion Torrent TVC, as part of a commercial software suite, has the strength of being user-friendly and well-designed to streamline and simplify the workflow of bioinformatics analysis, though it is more expensive compared to open-source bioinformatics software. Open-source bioinformatics software that is free and flexible to mix and match requires experience setting up the computational environment and

patience exploring appropriate parameters, which is not easy to satisfy by all clinical and research labs. Previous studies have compared the performance of Ion Torrent TVC and GATK against the gold standard variants from GIAB and concluded that GATK outperformed TVC with a better area under the precision-recall curve (SNP: 0.82 v.s. 0.78, Indel: 0.10 v.s. 0.05) [311]. The reason that TVC did not detect *GCK*:c.898G>A; p.(Glu300Lys) is unknown, but the quality metrics of this variant compared to artifacts and confirmed variants (Figure 4.2, Figure 4.4) hints that differences in mapping quality (MQRankSum) and base quality (BaseQRankSum) in comparison to the confirmed variants could be the reason. This was possibly caused by the low depth at this locus that only reads from one direction were retained and mapped to the reference genome.

Expanding the analysis to non-canonical splicing variants identified possible disease-causing variant in PDMP participants

The prediction model of each splicing prediction tool has determined that spliceAI was able to pick up more variants influencing splicing. SpliceAI took short- and long-range effects on splicing into account when developing. Tiulpakov et al. recently published research evaluating 20 novel synonymous and intronic *GCK* variants using minigene splicing assessment [312]. Two of the synonymous variants (c.339C>T;p.(Asp113=) and c.600G>A;p.(Val200=)) were assessed not to cause splicing effects. These two variants were also observed in PDMP participants and predicted not to cause splicing change by spliceAI, which is consistent with the *in vitro* evaluation. The accuracy of the prediction relies on evidence from RNA studies; thus, this project cannot provide the estimation of sensitivity or specificity of each tool.

One variant worth discussion is NM_000352.5 (*ABCC8*):c.3989-9G>A, also known as NM_001287174.2(*ABCC8*):c.3992-9G>A (rs151344623); both nomenclatures were used in publications. This variant was reported to be pathogenic in homozygotes or compound heterozygotes for the recessive disorder hyperinsulinism (HI) in ClinVar by Chicago, LabCorp, Athena, Invitae, Emory, CHOP, and Counsyl. It is more common in the Ashkenazi Jewish population compared with other populations (gnomAD AF= 0.0060 in Ashkenazi Jewish vs. 0.00012 in gnomAD non-Finnish European). The carrier frequency is 1 in 60 in Ashkenazi Jewish[313]. Many studies have shown this variant to cause an impact on splicing, and it resulted in a 7 bp addition, 20 bp deletion, or a 30 bp deletion in vitro[314]. According to Glaser 2011, this variant is one of the two most common mutations found in the Ashkenazi Jewish population with diffuse or focal congenital hyperinsulinism (CHI) [313]. A PDMP patient is a heterozygote of this variant, and he is of Ashkenazi Jewish origin for the paternal and maternal side. His paternal side has a family history of diabetes; however, there is no mention of CHI in the medical history of this family. This patient has already been found to have two *ABCC8* VUSs, p.Y356C and p.H125Q, which are located upstream of the splicing variant. Family studies and functional research has shown p.Y356C could be potentially contributing to or cause modestly early-onset diabetes, but its effect combining with the splicing variant has not been reported yet[315,316]. Furthermore, because these three variants are far away from each other (16910 bp and 56173 bp), we are not able to identify whether they are in cis or in trans from the BAM files. Further study needs to be done to evaluate whether this non-canonical splicing variant contributes to the diabetes phenotype of the proband and his other relatives, because the splicing variant may cause abnormal potassium

channel in addition to the p.Y356C effect when in trans and worsen the phenotype. Previous association studies have shown that two variants *ABCC8* p.S1396A and *KCNJ11* p.E23K, which are in strong linkage disequilibrium, predispose homozygotes to T2D[317]. Functional studies have not clearly explained the mechanism but suggested slight gain-of-function, which leads to decreased ATP sensitivity[318,319]. This could suggest a combination of p.Y356C as a potential gain-of-function variant and the Ashkenazi Jewish founder splicing variant together modifies the risk of developing diabetes, such as lowering the age of onset of diabetes due to deficiency of normal channels when they are in trans.

To summarize, this study comprehensively re-analyzed the NGS panel data from the PDMP by updating the variant calling algorithm, re-classified variants with new evidence and guidelines, and expanded the analysis to non-canonical splicing variants. The comprehensive re-analysis improved the diagnostic rate of 313 PDMP patients suspected to have monogenic diabetes from 11.82% to 13.10%. Routine re-analysis of sequencing data could help resolve cases with negative genetic testing results in monogenic diabetes and possibly apply to other monogenic disorders.

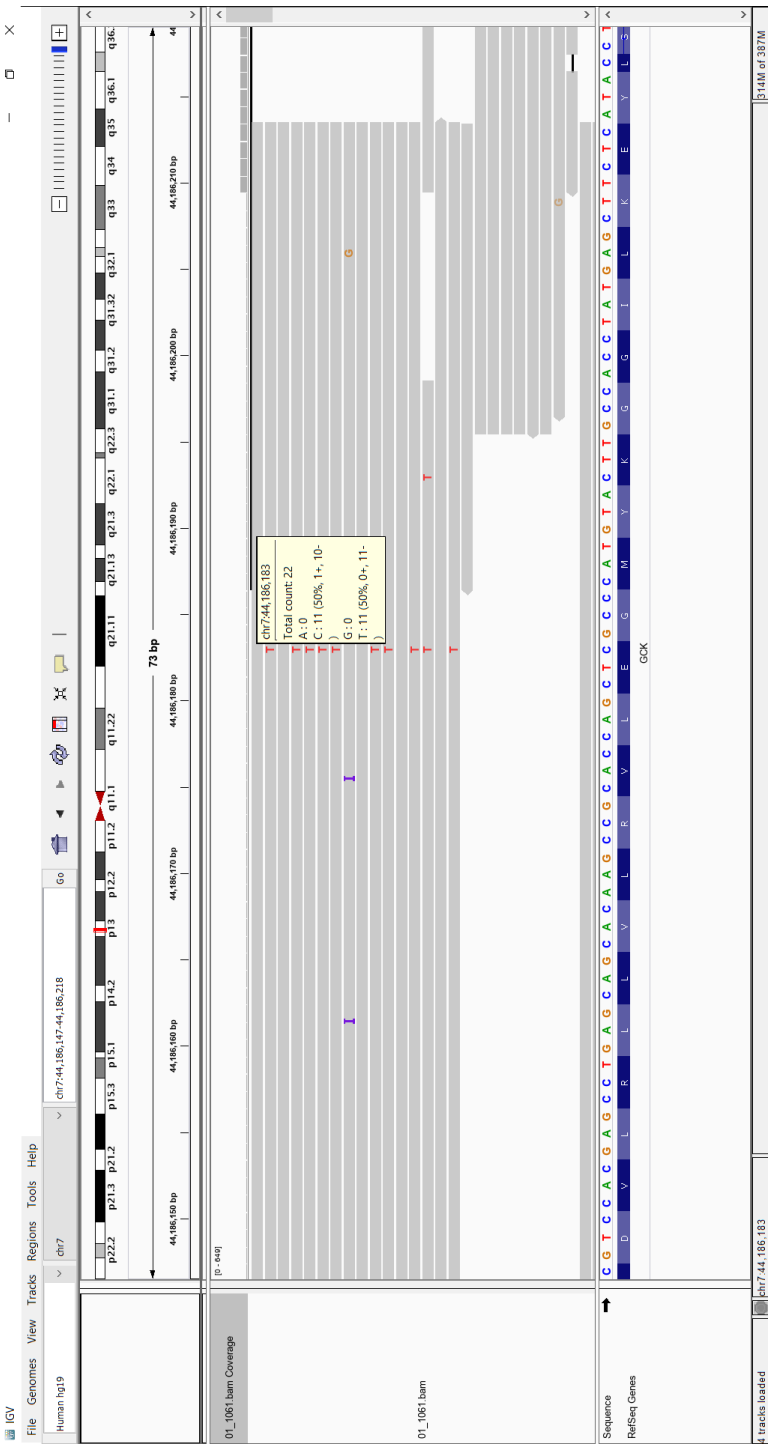


Figure 4.4 Screenshot of BAM file of *GCK* p.(Glu300Lys)
The variant is shown in the middle as the T in red. The yellow box shows the reads alignment. At this position, 11 reads (50%) contained the variant and all in reverse direction.

4.2 Exome Sequencing in Identifying Monogenic Diabetes in Youth with Type 2 Diabetes: Lessons from the Progress for Diabetes Genetics in Youth (ProDiGY) Collaboration[6]

4.2.1 Abstract

Purpose: Monogenic diabetes is often misdiagnosed as type 1 (T1D) or type 2 diabetes (T2D). The correct diagnosis requires genetic testing and could lead to changes of treatment based on disease etiology. NGS panel and Sanger sequencing are currently preferred in diagnosing monogenic diabetes, but the potential of exome sequencing (ES) in screening has rarely been evaluated. This study utilized ES data from the Progress for Diabetes Genetics in Youth (ProDiGY) Collaboration, which consists of young people (age 20 or younger) with a clinical diagnosis of T2D and aims to characterize the genetic architecture of pediatric T2D. NGS panel or Sanger sequencing have been performed in two subsets of ProDiGY collaboration, the Treatment Options for Type 2 Diabetes in Adolescents and Youth (TODAY) study and SEARCH for Diabetes in Youth (SEARCH) study to assess the prevalence of monogenic diabetes in each cohort.

Methods: This study compared the yield of disease-causing variants of 10 MODY genes from 488 TODAY samples with ES and NGS panel. This study also compared the yield of disease-causing variants of 492 SEARCH samples with ES and Sanger sequencing on the three most common monogenic diabetes genes.

Results: The initial analysis of ES filtered data missed nine variants identified by the NGS panel and five variants identified by the Sanger sequencing. Re-analysis of ES unfiltered data after removing the variant calling filtration step recovered all the missing variants and enabled increased sensitivity to detect new variants in the ES-only data.

[6] HZ performed the data analysis and drafted the manuscript.

Conclusion: The ES unfiltered data contains all monogenic diabetes variants identified in the NGS panel or Sanger sequencing in the ProDiGY samples. However, the data filtration step without proper parameters could remove some of the disease-causing variants and decrease the variants yield. It is essential to select appropriate filtering parameters based on study purposes: detecting rare variants for diagnosis requires high sensitivity.

4.2.2 Introduction

The advancement of massive parallel sequencing techniques such as next-generation sequencing (NGS) has revolutionized genetic testing. Three NGS approaches are utilized for various purposes: NGS panels, exome sequencing (ES), and genome sequencing (GS). Each approach has its advantages and limitations. Inexpensive NGS panel enables the assembly of gene panels of interest and usually has high coverage at the targeted regions. ES and GS have the advantages of covering all exons or whole genome of an individual but often have low sequence depth. The rapidly evolving research about newly discovered disease-related genes determined that ES and GS could be routinely re-analyzed when new findings came out while the NGS panel data will be outdated. It was estimated that exome sequencing might achieve a diagnostic yield similar to panel-based testing for Mendelian diseases. However, in the actual practice, many factors prevent the ES from reaching its full potential, such as an exome enrichment technique, sequencing reads alignment, and variant calling pipeline.

Monogenic diabetes has been frequently misdiagnosed as type 1 diabetes (T1D) or type 2 diabetes (T2D) because of their overlapping clinical features. Maturity-Onset Diabetes of the Young (MODY) is the most common subtype of monogenic diabetes.

The correct diagnosis of monogenic diabetes through genetic testing is essential to the etiology-based treatment and risk management of probands and family members. When the term MODY was first coined by Fajans when he observed the families with young-onset T2D diabetes and inherited in an autosomal dominant pattern [4], the list of MODY genes expanded slowly with the three most common genes as *HNF1A*, *HNF4A,* and *GCK.* When looking at one or all three genes based on patient phenotype, Sanger sequencing was the only option before NGS became available. With the research on monogenic diabetes proliferating new genes' findings, NGS panel that can simultaneously include all the MODY and neonatal diabetes mellitus (NDM) genes and syndromic forms of diabetes genes before the disease presentation are becoming popular in clinical practice and research. When NGS panel as the first-tier test returned with negative results, researchers often perform ES or WGS to search for variants in unknown genes in undiagnosed patients[33,58].

As the costs of ES decrease, health care providers and laboratories increasingly chose ES as the first-tier test for diagnosing Mendelian disorders, weighing in its comprehensiveness and the complexity of patients' phenotype [217,320]. However, some studies reported less satisfying diagnostic yields of ES than NGS panel, mainly due to low depth [321,322]. Studies that used NGS panel to search for monogenic diabetes in suspected patients have demonstrated pick-up rates varying from 17.5% to 25% [265,323-325]. ES has reached similar pick-up rates (patients with causative variants over patients who went through genetic testing) ranging from 10.7% to 55% when used as a first-tier test on patients highly suspected to have monogenic diabetes[179,182,183,185,240,326,327]. When used as a second-tier test on patients who have tested negative for major monogenic diabetes

genes, ES has found causative variants missed by NGS panel or Sanger sequencing in case reports[328,329] or found new causative genes[58]. Likewise, these studies also warned about the low depth of some regions and difficulties in interpreting large quantities of variants.

Currently, no studies to our knowledge have directly compared the pick-up rate of NGS panel and ES on unselected patients with DM, which should provide insights into evaluating the ability of ES as the first-tier genetic testing method for monogenic diabetes. This project utilized data from the Progress for Diabetes Genetics in Youth (ProDiGY) Collaboration[162], and the detected disease-causing variants from the same samples with different sequencing methods were compared. Four hundred and eighty-eight samples from the Treatment Options for Type 2 Diabetes in Adolescents and Youth (TODAY) study (see Participants) were tested with NGS panel and ES to search for monogenic diabetes, especially MODY, in young adolescents with T2D (age between 10-17)[9]. Also, 492 samples from the SEARCH for Diabetes in Youth (SEARCH) study (see Participants) were tested with Sanger sequencing on the most common MODY types (*HNF1A*-, *HNF4A*-, and *GCK*-MODY) and ES. The ProDiGY Collaboration is a collaboration to study genetic variants related to T2D risk in youth. It contains young-onset T2D cases (clinically diagnosed before age 20) from the TODAY study, SEARCH study, and additional youth with T2D from the ancillary TODAY Genetics study [330-332] to increase the power of GWAS.

4.2.3 Materials and methods

Participants

The ProDiGY collaboration includes TODAY, SEARCH for Diabetes in Youth, and TODAY Genetics (an ancillary study of TODAY that enrolled additional participants with pediatric T2D for association studies only) and tries to identify genetic variants predisposing to youth-onset type 2 diabetes[333].

Briefly, TODAY enrolled 699 participants with T2D between 2004 and 2009. The participants were aged between 10 to 17 years and overweight or obese (BMI≥85th percentile, adjusted for age, sex, and height). They were negative for pancreatic autoantibodies (GAD-65 and tyrosine phosphatase) and had a fasting C-peptide concentration >0.6 ng/mL. The TODAY clinical trial randomized the participants to one of three arms—metformin alone, metformin plus rosiglitazone, and metformin plus lifestyle intervention—and followed for an average of 3.86 years. TODAY Genetics which is distinct from the original TODAY clinical trial, also enrolled participants diagnosed at an early age (<18 years) and overweight or obese at the time of diagnosis. Twenty-five clinical sites participated in data and blood sample collection for DNA extraction, glucose, C-peptide, and autoantibodies analysis.

SEARCH is a multicenter study based in the U.S. prospectively ascertain diabetes in youth that recruit clinically diagnosed T2D participants before 20 years. The participants selected from SEARCH for Sanger sequencing were negative for both glutamic acid decarboxylase 65 (GAD65) and islet-antibody 2 (IA2) diabetes autoantibodies and had fasting c-peptide (FCP) of 0.8 ng/mL or greater.

Sequencing methods

In summary, ProDiGY has ES data for 3,333 youth with T2D who were negative for pancreatic autoantibody, as shown in Table 4.7. A subset of TODAY samples (n=488) were sequenced by ES and NGS panel on 40 monogenic diabetes-associated genes (Table 3.3) previously[9], not including the MODY14 gene, *APPL1*. A subset of SEARCH study samples (n=492) were sequenced by ES and Sanger sequencing on *HNF1A*, *HNF4A,* and *GCK*.

Table 4.7 ProDiGY collaboration sample distribution

	Total sample size	Sequenced sample size	Sequencing methods
TODAY	699	488* 511* (including above 488)	Targeted 40 gene penal ES by ProDiGY
TODAY Genetics	2571	2330**	ES by ProDiGY
SEARCH	586	492* 586***	ES by ProDiGY Sanger sequencing

*Excluding participants without available DNA for analysis.
**Excluding participants with positive pancreatic autoantibody.
***In the SEARCH study, different criteria were applied when choosing genes for sequencing. All participants were tested for *HNF1A*, 569 participants were tested for *HNF4A*, and only 357 participants were tested for *GCK*, excluding participants with HbA1c above 7.5%.

The ES of ProDiGY was performed by the Broad Institute as described [333]. Briefly, the DNA libraries were constructed using Agilent SureSelect Human All Exon 44Mb v2.0 and sequenced on Illumina GAII, or HiSeq 2000 sequencers to ~56x mean coverage. Sequence data were processed and aligned to hg19 using the Picard (http://broadinstitute.github.io/picard/) and BWA [334]. The variant calling was performed by the GATK pipelines with the Variant Quality Score Recalibration (VQSR) tool (https://gatk.broadinstitute.org/hc/en-us/articles/360035531612-Variant-Quality-Score-Recalibration-VQSR-). The VQSR is a machine learning algorithm established on multiple population sequencing databases by incorporating quality data of various dimensions and cluster the variants based on the quality metrics. The VQSR looked at the

databases for a corresponding cutoff based on users' required sensitivity and applied the cutoff on new sequencing data. The initial analysis of ES data went through GATK VQSR filtering steps, and only variants scored with a sensitivity of 99.6% for single nucleotide variants (SNVs) and 95% for insertion and deletions (Indels) were retained, and the resulting variants went through further analysis.

The investigators at the Broad Institute extracted variants from 14 genes previously identified as causal for MODY (*HNF4A*, *GCK*, *HNF1A*, *PDX1*, *HNF1B*, *NEUROD1*, *CEL*, *INS*, *PAX4*, *KLF11*, *BLK*, *KCNJ11*, *ABCC8*, and *APPL1*) for data annotation and interpretation. Since recent studies showing a high prevalence of *PAX4*, *KLF11*, *BLK* variants in the population and lack of rare variants enrichment of the three genes in the MODY patients, the three genes were disputed/refuted by MDEP as MODY-causing genes and thus excluded from further analysis[10]. Later in the unfiltered data analysis, the output of variant calling from GATK without the VQSR calibration step was similarly extracted of variants in 11 MODY genes, annotated, and interpreted.

The Department of Molecular Genetics, Royal Devon and Exeter National Health Service Foundation Trust performed the Sanger sequencing, and detailed methods were described previously [237]. Briefly, the coding exons and conserved splice sites of *HNF1A*, *HNF4A*, and *GCK* were amplified by PCR and sequenced by ABI 3730 (Applied Biosystems).

Dr. Kleinberger performed the NGS panel of the TODAY study at the University of Maryland Baltimore as described previously[9]. Briefly, the DNA libraries were constructed using Ion AmpliSeq customized panel capturing all exons of 40 genes with variants known to cause monogenic diabetes. The panel includes 13 genes known to be

MODY genes (not including *APPL1*) at that time. The constructed libraries were sequenced by the Ion Torrent Personalized Genome Machine (PGM), and sequence data were aligned to hg19 by Torrent Mapping Alignment Program (TMAP). The aligned reads were then called for variants by Torrent Variant Caller (TVC), which were parts of the Thermo Fisher's Torrent Suites Software. Only samples with ≥20x coverage depth of ≥80% of the target region (139,491bp) were used for analysis. Variants in *BLK*, *KLF11* and *PAX4* were similarly discarded during analysis.

Variant annotation and interpretation

The detailed variant annotation and interpretation process has been described before [9]. Briefly, the output vcf files after the variant calling step were annotated with gene name, transcripts, function, MAF, in silico prediction scores from multiple tools, dbSNP ID, and ClinVar (a database disease phenotype related variants) records. The variants generated from NGS panel and ES were filtered. Only low frequency (<5% minor allele frequency in ExAC, 1000 Genome, and ESP all population) coding or splice-site variants remained for the initial interpretation.

The PDMP variant classification team then interpreted the annotated and filtered variants for pathogenicity according to the 2015 American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) guidelines[189]. These guidelines incorporated computational predictions, clinical phenotypes, functional assessments, and population frequencies into consideration and categorized the variants into five categories (Pathogenic (P), Likely Pathogenic (LP), Variant of Unknown Significance (VUS), Likely Benign (LB), or Benign (B)). Later in the ES raw data analysis, the additional variants from ProDiGY were classified by HZ.

The SEARCH study's Sanger sequencing results were compared with the published reference sequences (accession numbers NM_000545.5, NM_000457.3, and NM_000162.3) using Mutation Surveyor version 3.24 (SoftGenetics). An in-house Exeter laboratory database containing *GCK*, *HNF1A*, and *HNF4A* sequence data for more than 1000 individuals was compared with resulting variants. If missense mutations or in-frame deletions were found to be novel and affecting an amino acid conserved through evolution, they were classified as highly likely to be pathogenic[237].

4.2.4 Results

ES unfiltered data contain all monogenic diabetes causative variants

In the 488 TODAY samples, 21 pathogenic or likely pathogenic variants (*GCK* (n=7), *HNF4A* (n=7), five *HNF1A* (n=5), and two *INS* (n=2)) were found in 21 participants through NGS panel (Ref [9], Table 1, excluding the *KLF11* variant[7]). Eight of these variants were not identified by ES. In the 492 SEARCH samples, 49 pathogenic or likely pathogenic variants (*HNF1A* (n=27), *HNF4A* (n=8), and *GCK* (n=14)) were found in 47 participants through Sanger sequencing (Ref[237], Supplementary Table 1). Five of these variants were not identified by ES. (Table 4.8)

After retrieving the unfiltered output VCF file after variant calling without VQSR calibration and performing the same annotation, filtration, and interpretation step, the variants were re-discovered. The quality info from the unfiltered data shows that six variants did not pass the cutoff during the VQSR step and were labeled with their VQSR

[7] The *KLF11* variant (rs121912645) was also not identified by the ES.

scores in the "Filter" column. However, eight other variants were filtered out for unknown reasons.

Removing GATK VQSR filtering identified new pathogenic variants in ES unfiltered data

The initial analysis of ES data identified 45 pathogenic or likely pathogenic variants among a total of 3,333 samples. After discovering the ES data failed to identify all pathogenic or likely pathogenic variants in TODAY and SEARCH subsets, re-analysis of ES unfiltered data removed the GATK VQSR filtering step in the VCF file and continued further annotation and interpretation.

The analysis extracted the 1,129 variants from the 11 MODY genes of all ProDiGY samples with ES data from the unfiltered VCF file. Two hundred forty-eight variants only exist in the unfiltered VCF file and continued to further analysis. Among the 248 variants, 85 variants come from the three major MODY genes: *GCK* (n=16), *HNF1A* (n=45), and *HNF4A* (n=24), and 164 variants come from the other 8 MODY genes. The 248 variants contain the eight variants that were originally discovered in NGS panel data in TODAY samples and five variants that were discovered by Sanger sequencing in previous SEARCH study samples. This project classified the additional 235 variants (248-13) according to the ACMG/AMP 2015 guidelines; 38 of them were classified as pathogenic or likely pathogenic. For each variant, the ProDiGY group evaluated the corresponding BAM file and found out 19 of the 38 variants were artifacts because of reasons including low depth, variant only existed in one direction of reads, or the variants existed in many other samples. The remaining 19 variants were considered true positives, and they were carried by 22 participants. (Table 4.9)

The distribution of the analyzed variants' quality score shows that Sanger confirmed SNVs have a similar distribution of VQSLOD, but a relatively higher BaseRankSumTest score. The artificial indels have lower VQSLOD scores, and Sanger confirmed indels have high quality by depth score (QD). (Figure 4.5 and Figure 4.6)

Table 4.8 Pathogenic or likely pathogenic variants filtered out by initial ES pipeline in TODAY and SEARCH study

Gene	dbSNP	Transcript	HGVS	Classification	Sample Source	Qual	Filter
GCK	rs104894011	NM_000162.5	c.793G>A,p.Glu265Lys	Pathogenic	TODAY	728.48	PASS
GCK	.	NM_000162.5	c.1186_1187dupAG,p.Ser396ArgfsTer7	Likely Pathogenic	TODAY	3398.48	VQSRTrancheIND EL97.00to99.00
HNF1A	rs137853244	NM_000545.6	c.391C>T, p.Arg131Trp	Pathogenic	TODAY	6153.19	PASS
HNF1A	rs779184183	NM_000545.6	c.812G>A, p.Arg271Gln	Pathogenic	TODAY	2479.48	PASS
HNF4A	rs1568724014	NM_001287183.1	c.256C>T, p.Gln86Ter	Pathogenic	TODAY	2517.48	PASS
HNF4A	rs1375557127	NM_001287183.1	c.923G>A, p.Arg308His	Likely Pathogenic	TODAY	2741.19	PASS
HNF4A	rs1555813319	NM_001287183.1	c.191G>A, p.Arg64Gln	Likely Pathogenic	TODAY	4098.67	VQSRTrancheSNP 99.60to99.80
HNF4A	rs1283736704	NM_001287183.1	c.1090delC, p.Leu364CysfsTer5	Likely Pathogenic	TODAY	419.48	VQSRTrancheIND EL99.50to99.90
GCK	rs193922254	NM_000162.5	c.1003delG, p.Val335CysfsTer18	Pathogenic	SEARCH	2992.48	VQSRTrancheIND EL97.00to99.00
HNF1A	rs137853247	NM_000545.6	c.92G>A, p.Gly31Asp	Pathogenic	SEARCH	12082.74	PASS
HNF1A	.	NM_000545.6	c.873delA, p.Pro293GlnfsTer49	Pathogenic	SEARCH	5475.67	VQSRTrancheIND EL95.00to96.00
HNF1A	rs1315462017	NM_000545.6	c.1124G>T, p.Gly375Val	Pathogenic	SEARCH	1759.48	PASS
HNF1A	rs587780357	NM_000545.6	c.608G>A, p.Arg203His	Pathogenic	SEARCH	3249.19	VQSRTrancheSNP 99.60to99.80

Table 4.9 Classification and validation of unfiltered-data only variants

Gene	Total	Artifacts among classified pathogenic/likely pathogenic variants	Validated pathogenic/likely pathogenic variants
GCK	13	0	2
HNF1A	39	8	10
HNF4A	19	2	3
HNF1B	12	0	0
INS	5	0	1
ABCC8	21	0	0
KCNJ11	16	0	0
PDX1	21	1	2
CEL	65	8	1
NEUROD1	11	0	0
APPL1	13	0	0

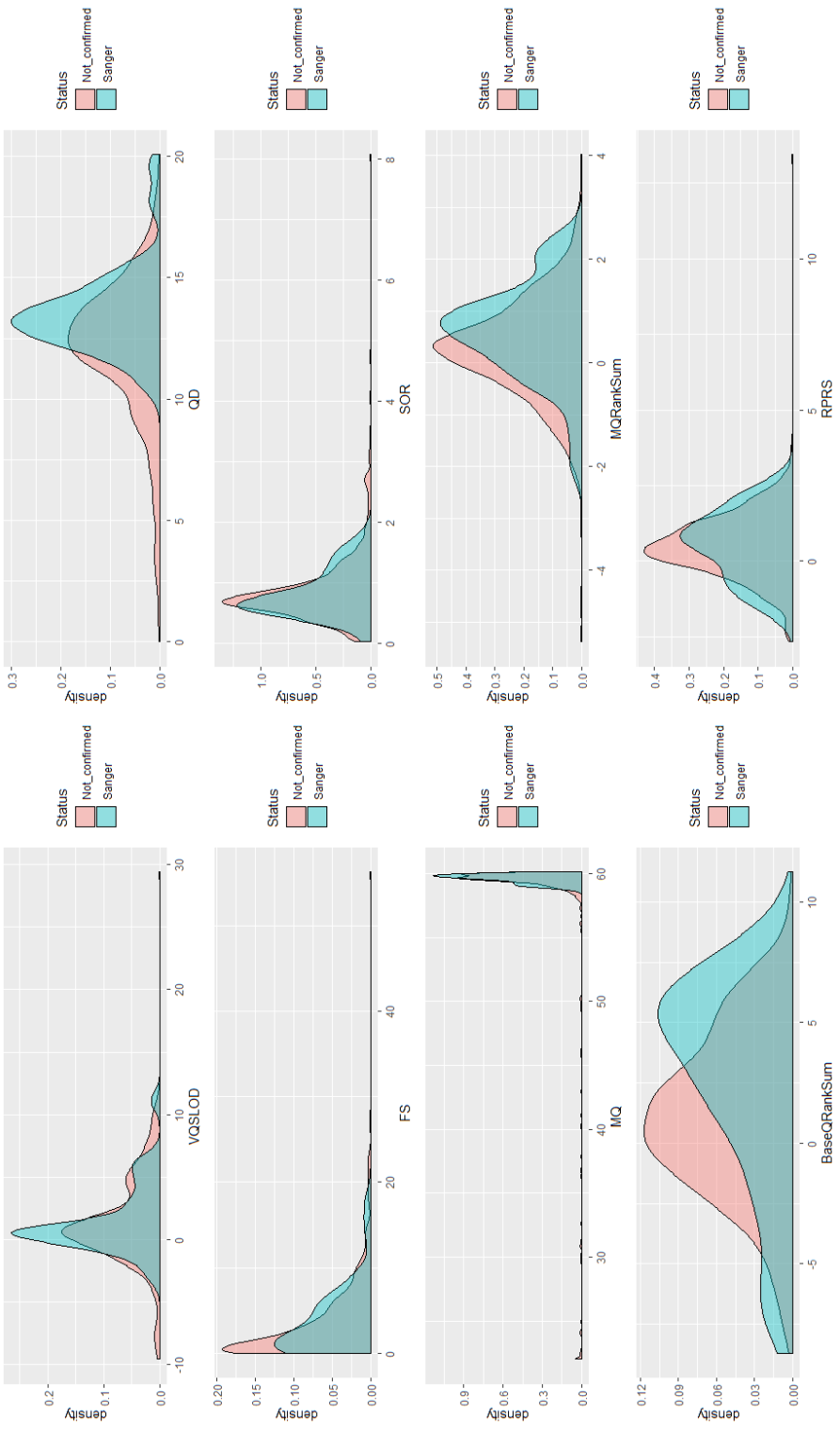


Figure 4.5 Distribution of variants quality score of SNVs in ProDiGY
Status: Sanger, confirmed by Sanger sequencing in TODAY or SEARCH study. Not confirmed, not confirmed by Sanger sequencing.

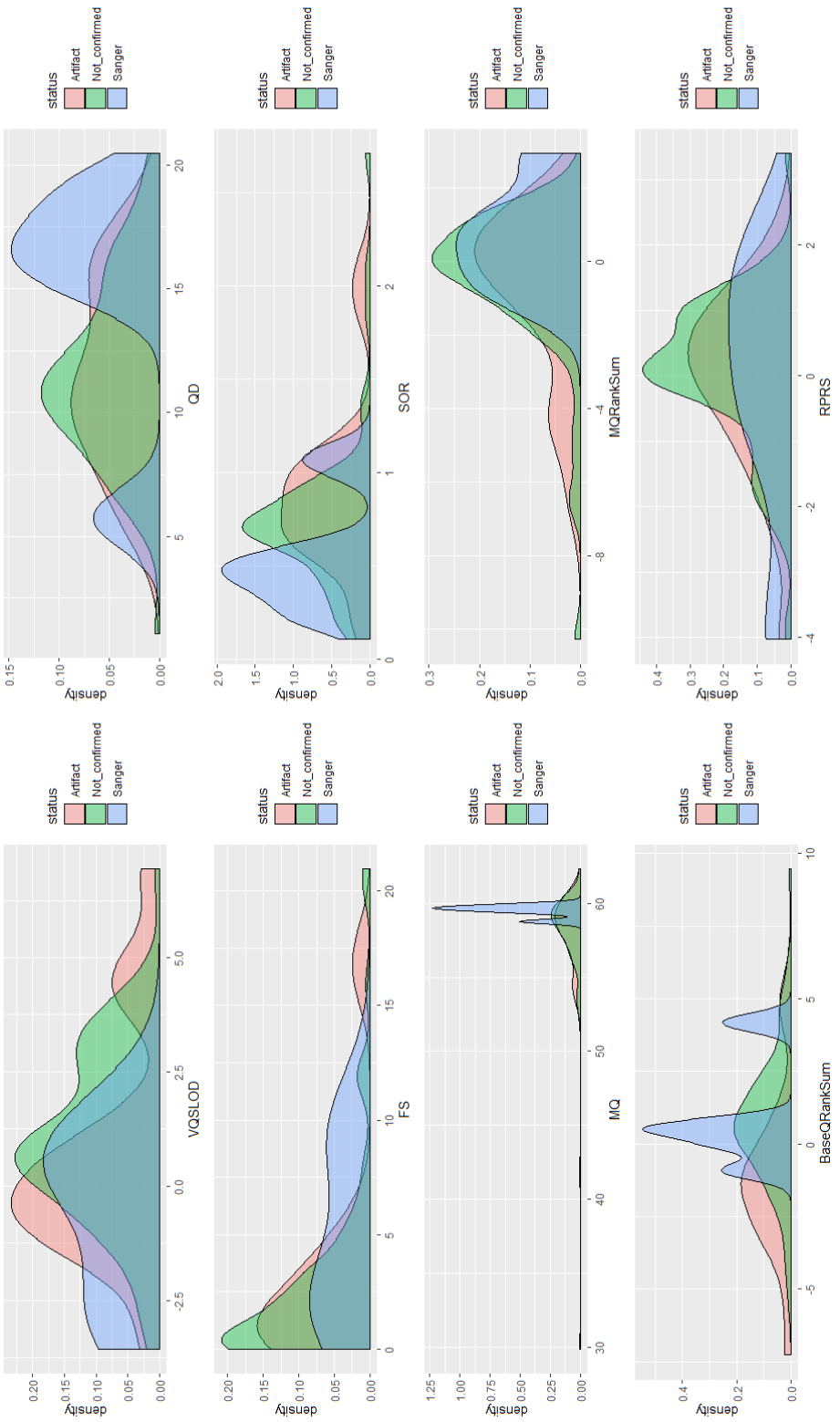


Figure 4.6 Distribution of variants quality score of indels in ProDiGY
Status: Artifacts, indel variants deemed artifacts in sequencing BAM files, Not confirmed, indel variants not confirmed by Sanger sequencing; Sanger, indel variants confirmed by Sanger sequencing

4.2.5 Discussion

This project aimed to compare the sequencing data of ES, NGS panel, and Sanger sequencing in detecting MODY-causing variants in youth with T2D. In the initial analysis of the ES filtered VCF file, ES only detected 57.1% (12/21) of MODY mutations compared to NGS panel in 488 samples and 89.8% (44/49) of *HNF1A*-, *HNF4A*- and *GCK*-MODY variants compared to Sanger sequencing in 492 samples. The ES unfiltered data captured all the pathogenic and likely pathogenic MODY variants detected through NGS panel and Sanger sequencing.

The quality control of sequencing data, especially the VQSR filtering step after the variant calling, caused some of these variants to be omitted in the final output. The possible explanation is the rareness of these variants that leads to the variant quality recalibration step to filter out the singletons. However, this does not fully explain the variants that passed the VQSR step and still got filtered out. The quality score distribution of SNVs and Indels illustrates that no single parameter of the filtered variants stood out to be the leading cause of low quality; therefore, it is difficult to use single parameters to replace the VQSR directly.

This project is limited by the small number of variants being evaluated compared to the large quantities of variants from the ES data of all 3,333 samples. Only a small percent of variants were confirmed by Sanger sequencing, and these variants showed different trends of the quality score distribution, such as QD in the SNVs and Indels. No statistical analysis or calculation of the sensitivity or specificity could be calculated.

This project demonstrated that ES unfiltered data used for the GWAS research at the Broad Institute with a mean coverage of 56x has the same ability to detect MODY

causing variants as NGS panel and Sanger sequencing. Thus, if cost and turnaround time is within a considerable range, it is optimal to use ES as a first-tier testing method for monogenic diabetes. However, this project shows that without proper filtering settings, the detection rate of ES could be impaired, especially for the purpose of diagnosis, which requires high sensitivity on rare variant detection. Though eventually all the missing variants were found in the ES unfiltered data, the time and effort in this type of re-analysis need to be considered. In this project, half (19/38) of the newly identified pathogenic or likely pathogenic variants were discarded because of the high possibility of being artificial after observing the corresponding BAM files. Considering the small number of genes in the re-analysis, it will be laborious if expanding the re-analysis into the whole exome. This project emphasized the importance of including positive controls, e.g., samples from patients with known MODY variants, in the sequencing process and validate the parameters of filtering based on their known variants.

4.3 Summary

This chapter showed the importance of comprehensive re-analysis of DNA sequencing data in improving the diagnostic rate of monogenic diabetes. The re-analysis related to several steps after sequencing data generation, including variant detection, variant filtration, and variant classification. In these studies, these steps could be improved to increase the final yield of disease-causing variants. The increased diagnostic yield from comprehensive re-analysis of NGS panel data of PDMP indicates that the re-analysis in the clinical laboratory should be planned and performed routinely. Different bioinformatics tools uncovered missing variants due to flaws in computational algorithms

and variant re-classification with new evidence and new rules of variant interpretation could resolve variants of unknown significance.

In this chapter, the re-analysis improved the diagnostic rate, but the results also apply to the initial sequencing data analysis. This chapter compared sequencing data in detecting MODY variants between ES, targeting sequencing, and Sanger sequencing data from ProDiGY, a cohort of young people diagnosed with T2D. The ES data with a mean coverage of 56x could detect the same disease-causing variants as NGS panel and Sanger sequencing through re-analysis. Notably, the ES data of ProDiGY participants was not intended to diagnose monogenic diabetes or detect rare variants but for association studies using more common SNPs. The differences in study purposes meant that different data processing strategies should be applied to the same sequencing data. The comparison demonstrated the potential of ES as the first-tier genetic testing in identifying monogenic diabetes variants. Still, it required adjustment in the data processing step to reach high sensitivity for rare variants.

Chapter 5 : Direct-to-Consumer Genetic Testing Raw Data Has Limited Clinical Utility in Detecting *GCK*-MODY Variants[8]*

5.1 Abstract

Purpose: Customers often use the information from Direct-to-Consumer genetic testing (DTC-GT) raw data and third-party tools' interpretation to inform their medical decisions despite warnings from DTC companies about safety and effectiveness and previous research results showing the high false-positive rate. In addition, evaluation of the accuracy of third-party tools' interpretation is lacking. To assess the benefits and risks to consumers from unvalidated DTC-GT raw data and uncurated third-party tools' interpretation, this project evaluated the ability of 23andMe DTC-GT raw data in detecting disease-causing variants in consumers using *GCK*-MODY as an example.

Methods: We classified the pathogenicity of *GCK* single nucleotide variants (SNVs) on the 23andMe genotyping microarray chip, collected *GCK* SNVs in 23andMe raw data from 3,040 anonymous consumers, and calculated the apparent prevalence of *GCK*-MODY in these consumers. We assessed the 23andMe raw data quality by comparing the Minor Allele Frequency (MAF) of these disease-causing variants in the consumers to the MAF from gnomAD. We also compared the interpretation accuracy of these variants among different third-party tools and attempted to validate the clinical utility in a cohort of people suspected to have monogenic diabetes.

Results: The 23andMe company designed its microarray chip to detect 316 *GCK* variants and provided the results in the raw data. Among them, 65.7% were classified as

[8] Haichen Zhang, Kristin A. Maloney, Tim O'Connor, Toni I. Pollin. Direct-to-Consumer Genetic Testing Raw Data Has Limited Clinical Utility in Detecting *GCK*-MODY Variants. In preparation for submission
*HZ performed the data analysis and drafted the manuscript.

pathogenic or likely pathogenic for *GCK*-MODY. However, the greater number of putative heterozygotes of *GCK*-MODY and inconsistency of allele frequency when compared with gnomAD suggested unreliable data quality. Furthermore, the third-party tools showed incomplete and inaccurate interpretation of *GCK* variants in 23andMe DTC-GT raw data that may cause misunderstanding in DTC-GT consumers. Conclusion: The 23andMe DTC-GT raw data is not comprehensive and is unreliable in detecting *GCK*-MODY causing variants. The third-party tools' interpretation of *GCK* variants is not accurate for consumers to understand the clinical significance of these variants.

5.2 Introduction

DTC-GT is defined as genetic testing offered directly to consumers apart from health care providers. Since the foundation of deCode Genetics, 23andMe, Navigenics, and other DTC companies in the early 2000s, the concept of DTC-GT has become increasingly well-known among the general public. Its market size has increased each year to an expected number of 100 million customers by 2021[335]. DTC-GT's basic products report ancestry and traits that are peripherally related to health, such as the ability to digest caffeine. In addition, some companies' products include FDA-approved health reports (e.g., Parkinson's disease and Alzheimer's disease) and predisposition reports supported by their research (e.g., 23andMe Type 2 Diabetes predisposition report). Many DTC-GT companies release raw data in the form of a text file, which can be uploaded to third-party interpretation tools such as Promethease. Importantly, these companies' websites contain disclaimers stating that raw data did not undergo the same quality control steps as reported results, and therefore, results should not be used for

medical decision-making. Indeed, recent research has challenged DTC-GT raw data quality and recommends confirmation and interpretation of variants from the raw data in a CLIA certified lab if for clinical management[336]. The DTC-GT companies perform genotyping through DNA microarray chip, which has been around since the 1980s [337]. It was initially developed for genome-wide association studies using common variants (>0.1%) and later customized to include millions of single nucleotide polymorphism (SNPs) that were reported to be related to diseases and also rare variants. Recent research also showed that genotyping of rare variants through microarray chips has low positive predictive values compared with common variants [338,339] since the mechanism of microarray chips detecting SNPs is the clustering of the identical genotypes. The signal of rare genotypes is challenging to distinguish from noise, especially when detecting singletons.

While DTC-GT enables easy access to individuals' genetic information, potentially encouraging customers' engagement in making healthcare-related decisions [340,341], reporting health information without healthcare providers' involvement has the risk of the consumer misunderstanding[342] that the DTC-GT can predict with certainty whether someone will get a disease. And in fact the FDA did not approve all health predisposition reports and many of them were inferred by the 23andMe research, .

Pathogenic and likely pathogenic variants in the glucokinase gene, *GCK*, cause mild fasting hyperglycemia starting at birth by raising the threshold for insulin release in the pancreas. Unlike pre-diabetes, defined as fasting plasma glucose level between normal (100 mg/dL) and diabetes mellitus status (126 mg/dL)[2], hyperglycemia caused by *GCK* variants is not progressive, and individuals do not require treatment; indeed, most

carriers of *GCK* variants are asymptomatic and increased blood sugars are found incidentally [44]. In fact, the term *GCK*-MODY should be replaced with *GCK*-hyperglycemia to avoid overtreatment. Many patients with hyperglycemia from *GCK* variants are wrongly diagnosed and prescribed therapies (oral hyperglycemia agents and insulin) that usually do not normalize glucose levels[44]. The observation that 23andMe's raw data contains genotypes with pathogenic and likely pathogenic *GCK* coding variants was brought to our attention by a parent with a science background who tried to see if one of her children had a *GCK*-hyperglycemia already known to be in her family. Most of the coding variants in *GCK* (and, as it turns out, other genes as noted in the snpedia.org consumer database) are masked with a 23andMe internal ID. Still, they can be unmasked using chromosome and position information either by an astute consumer or a more naïve consumer using a third-party tool such as Promethease. We speculate that the masking is meant to reduce the likelihood of consumers using the information for diagnostic purposes. The *GCK* variants in 23andMe raw data could confuse consumers without adequate knowledge of the risks of false-positive or false-negative genotypes and inaccurate variant interpretation. For example, consumers with other types of diabetes may self-diagnose with *GCK*-MODY based on a *GCK* variant in their raw data, or someone with *GCK*-MODY may wrongly attribute their diabetes to another cause if a *GCK* variant is not in their raw data.

This project annotated the pathogenicity of *GCK* variants in the 23andMe microarray chip and analyzed the *GCK* variants contained in the 23andMe raw data from 3,040 anonymous users. To evaluate the data quality of *GCK* variants of the 23andMe raw data, this project calculated the frequency of *GCK* pathogenic or likely pathogenic

variants among the 3,040 consumers and calculated the ancestry specific MAF of disease-causing SNVs, and compared them with a large population database, gnomAD. In addition, to assess the ability of 23andMe raw data in detecting *GCK* variants among patients suspected to have monogenic diabetes, we surveyed a subset of patients from the PDMP study to obtain 23andMe raw data for analysis if available and compared with next-generation sequencing (NGS) data from their PDMP results. The detailed patient recruitment and data collection of PDMP have been described in Chapter 3. To summarize, the PDMP, as a part of the NIH IGNITE (Implementing GeNomics In PracTicE) Network[343], was designed to implement, disseminate, and evaluate a sustainable method for identifying, genomically diagnosing, and promoting individualized therapy for monogenic diabetes. Over 2015-2018, the PDMP screened 2,522 patients with diabetes from four participating clinics and through self-referring. The PDMP enrolled 507 patients meeting selection criteria and sequenced 456 patients suspected of having monogenic diabetes (313 patients) or 143 patients with diabetes not meeting criteria for monogenic diabetes as part of a sensitivity evaluation. The enrolled patients went through research-level NGS of a 40-gene panel and the rare (MAF <5%) exonic or canonical splicing variants classified to be disease-causing based on the 2015 American College of Medical Genetics and Genomics (ACMG)/the Associates for Molecular Pathology (AMP) guidelines were confirmed through Sanger sequencing. Finally, this project compared the interpretations from multiple third-party tools of *GCK* variants in 23andMe raw data with the classification based on the 2015 ACMG/AMP guidelines.

5.3 Materials and methods

DTC-GT raw data collection and processing

OpenSNP (https://opensnp.org/) is a crowdsourcing website that collects DTC-GT genotypic and phenotypic data from anonymous volunteers and offers downloads to scientists for research use[344]. The users of DTC-GT and the volunteers who share their genetic data may be biased toward individuals who are more likely to be European ancestry[345,346]. Still, they were not selected based on glucose level or diabetic status and thus would not influence the estimation of *GCK*-MODY prevalence.

The DTC-GT raw data of 3,040 23andMe consumers from openSNP were downloaded in August 2019. Each line represented one genotype result in the format of " rsID (also known as dbSNP ID) or internal ID, chromosome number, chromosome position, genotype." All genotypes of *GCK* were extracted and mapped to the human reference genome GRch37 based on the chromosome number and position (chr7: 44183870-44229022). Merging all extracted *GCK* genotypes created a dictionary of all *GCK* alleles from the studied samples. A reference file containing reference genotypes of extracted *GCK* alleles was generated. The reference file was converted to the vcf format, which used the designed alternative alleles by 23andMe as the alternative alleles. This project used a python script to read each sample, compare it to the reference file, and output the *GCK* variants found in each sample.

Variant annotation and interpretation

Through Annovar (version 2017-07-17), the reference file was annotated with nucleotide change, amino acid change, dbSNP ID (avsnp147), REVEL score, ClinVar clinical significance (version 2019-03-05), and population allele frequency from

gnomAD v2.1.1. The pathogenicity of each variant was assessed according to the 2015 ACMG/AMP guidelines for variant interpretation that have been adapted to *GCK* by the ClinGen Monogenic Diabetes Expert Panel (MDEP) [189].

Admixture analysis and ancestry-specific allele frequency calculation

A reference genotyping dataset combines Yoruba (n=132), British (n=91), Han Chinese (n=88), Utah Whites (n=99), Mexican (n=64), and Peruvians (n=147) from 1000 Genomes [347], Human Genome Diversity Project (HGDP)[348] and previous studies[349]. Each 23andMe sample was converted to plink format (.bed, .bim, .fam) using plink command "plink --23file". Files that failed the conversion because of wrong format or incompleteness were regarded as ancestry undetermined. After extracting SNPs in the reference dataset for each sample, all 23andMe samples were merged with the reference dataset. The supervised admixture analysis was done by running the tool Admixture with K=4 (European, East Asian, African, Native American). The major proportion (above 50%) of ancestry composition determined the ancestry; if no major proportion was detected, the ancestry was recorded as mixed. An R script plotted the ancestry admixture of each individual.

Survey design and distribution to PDMP participants

Besides detecting *GCK* variants in random 23andMe raw data, it is also vital to examine whether confirmed *GCK* variants from individuals with *GCK-MODY* could be detected in their raw data. The PDMP study provided a useful dataset to query for such individuals. By 2020, 41 PDMP participants tested positive for monogenic diabetes, of which 22 had *GCK*-MODY.

This study surveyed PDMP participants (probands and enrolled family members of monogenic diabetes-positive participants) who enrolled at the University of Maryland, Baltimore. Participants who were older than 18, had genetic testing through PDMP, and consented to recontact for other studies were emailed an invitation letter explaining the study's purpose and a link to the survey. The survey asked whether they have undergone 23andMe genetic testing, and if they had, would they like to share the raw data with the PDMP. The participants who agreed to share were directed to a secured link to upload their raw data to a HIPAA-compliant RedCap database. The Institutional Review Board of University of Maryland, Baltimore approved this study as a modification to PDMP protocol HP-00053483.

In total, 208 PDMP participants at UMB received an invitation letter to participate in the survey through email or mail. Among these 208 participants, six participants were tested positive for *GCK*-MODY through NGS panel.

Collection of third-party tools' interpretation of *GCK* variants

A fake 23andMe raw data text file replacing all *GCK* genotypes with heterozygous genotypes of reference allele and designed alternative allele was generated from the reference file. The fake 23andMe raw data was uploaded to LiveWello, Promethease, GeneticGenie, and Codegen for variant annotation and interpretation. The interpretation of *GCK* variants pathogenicity was summarized and compared to generate Venn diagrams.

5.4 Results

23andMe microarray chips could detect genotypes containing disease-causing *GCK* variants

The 23andMe company designed its microarray chips to detect 316 *GCK* variants distributed among all *GCK* exons, some of the introns, and UTR regions (Figure 5.1). These variants include 231 missense, 34 insertion or deletion, 30 nonsense or stop-loss, one splicing and one synonymous variant in the exons, and 19 variants in the non-coding regions. (Table 5.1) Among the exonic and splicing variants, 65.7% are pathogenic or likely pathogenic (P/LP), and 33.0% are of unknown significance (VUS). Most (91.1%) of the 316 variants have been reported in a comprehensive review paper [16], clinical variants database (ClinVar), or other research papers. The 316 *GCK* variants include 10 of the 21 P/LP *GCK* variants found in PDMP participants. (Table 5.2)

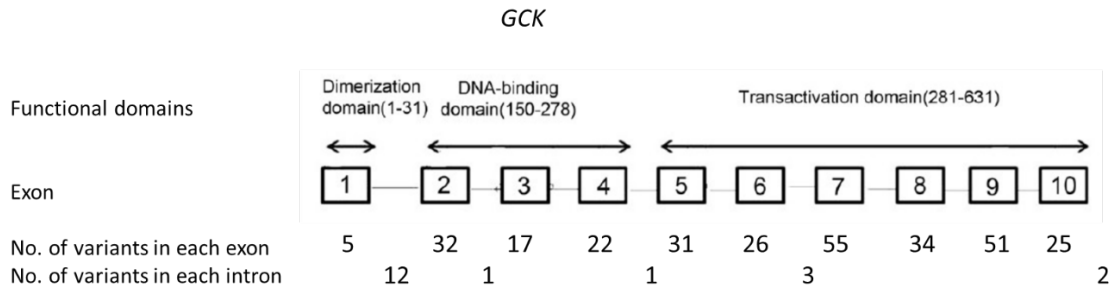


Figure 5.1 Distribution of *GCK* variants on 23andMe microarray chips

Table 5.1 Types and classification of *GCK* variants on 23andMe microarray chips

Variants	Number of Pathogenic/Likely pathogenic variants	Number of VUSs	Number of Benign/Likely benign variants
Non-coding regions: 19	-	-	-
Exonic/Splicing: 297	195	99	3
Splicing: 1	1	0	0
Missense: 231	134	95	2
Nonsense/Stop-loss: 30	29	1	0
Insertion/Deletion: 34	31	3	0
Synonymous: 1	0	0	1

Table 5.2 Pathogenic or likely pathogenic *GCK* variants in PDMP participants included on 23andMe microarray chips

Variant ID	Nucleotide change	Amino acid change	dbSNP ID	Classification
i5044928	c.1016A>G	p.(E339G)	rs1057524903	Pathogenic
i5044841	c.128G>A	p.(R43H)	rs764232985	Likely Pathogenic
i5044808	c.748C>T	p.(R250C)	rs1057524904	Likely Pathogenic
i5044782	c.113A>C	p.(Q38P)	rs1064794268	Likely pathogenic
i5044869	c.184G>A	p.(V62M)	rs1064793998	Pathogenic
i5045123	c.1113C>G	p.(C371W)	.	Likely Pathogenic
i5044872	c.676G>A	p.(V226M)	rs148311934	Pathogenic
i5044717	c.106C>T	p.(R36W)	rs762263694	Pathogenic
i5044718	c.571C>T	p.(R191W)	rs1085307455	Pathogenic
i5044748	c.122T>C	p.(M41T)	rs1057524906	Likely Pathogenic

Some disease-causing *GCK* variants in 23andMe raw data have an abnormally high minor allele frequency

Overall, 16 P/LP *GCK* variants have more than one minor allele in 3,040 openSNP samples, significantly higher than in gnomAD v.2.1.1, which had no observations. (Table 5.3) Using the allele frequency estimation tool, cardiodb (https://www.cardiodb.org/allelefrequencyapp/), the maximum number of pathogenic or likely pathogenic *GCK* variants in the 3,040 samples should be at most six. The calculation was based on a calculated maximum population allele frequency that is credible for all pathogenic or likely pathogenic variants derived from a predicted prevalence of 0.1%, allelic heterogeneity of 100%, genetic heterogeneity of 100%, and penetrance of 100%.

Table 5.3 *GCK* variants with more than one allele in openSNP samples

Variant ID	Nucleotide change; Amino acid change	Classification	Number of individuals			Minor allele frequency in openSNP samples[#$]			
			NN	NM	MM	European	African	East Asian	Native American
i5037256	c.171delG; p.(M57Ifs*30)	P	0	239	0	0.0421***	0.0714**	0.0488***	0.3333
i5044956	c.1151C>A; p.(A384E)	LP	1435	20	0	0.0034***	0.0143**	0	0.1667
i5003685	c.1132G>A; p.(A378T)	LP	2157	14	0	0.0026***	0	0	0
i5045114	c.1145G>A; p.(C382Y)	LP	1480	11	0	0.0019***	0	0	0.1667
i6011214	c.626C>T; p.(T209M)	LP	903	7	0	0.0013***	0	0	0
i5045103	c.824C>T; p.(S281F)	LP	1249	3	0	0.0006***	0	0	0
i5037262	c.485delG; p.(G162Afs*42)	LP	1664	2	0	0.0004**	0	0	0
i5000060	c.1190G>T; p.(R397L)	LP	2426	2	0	0.0004**	0	0	0
i5044946	c.1009C>T; p.(Q337X)	LP	1605	2	0	0.0004**	0	0	0
i5037204	c.549delG; p.(L184Ffs*20)	LP	1697	1	0	0.0002*	0	0	0
i5044808	c.748C>T; p.(R250C)	LP	1677	1	0	0.0002*	0	0	0
i5044897	c.917T>C; p.(L306R)	LP	1362	1	0	0	0	0	0.0833
i5044973	c.943C>T; p.(L315F)	LP	1674	1	0	0.0002*	0	0	0
i5045058	c.1235T>C; p.(V412A)	LP	1623	1	0	0.0002*	0	0	0
i6011247	c.1150G>A; p.(A384T)	LP	1303	1	0	0.0002*	0	0	0
rs193922315	c.658T>C; p.(C220R)	LP	233	1	0	0	0	0	0.1429

NN: Homozygotes of wildtype allele; NM: Heterozygotes of variant allele; MM: Homozygotes of variant allele; P: Pathogenic; LP: Likely Pathogenic; #: Allele frequencies were compared with ethnic-specific MAF in gnomAD 2.1.1 except Native American group;$ Mixed or undetermined not included. * P<0.05, **P<0.01, ***P<0.001

23andMe raw data in PDMP participants

Thirty probands and nine family members responded to the survey, eight participants had done 23andMe genetic testing, and four participants successfully uploaded their raw data. None of the four participants' raw data contained disease-causing *GCK* variants, consistent with the genetic testing results from the PDMP 40-gene panel. However, the non-coding variants discovered in the four participants were not covered on the PDMP panel, thus cannot be compared with the NGS data. (Table 5.4) Recipients of the survey included six of the ten PDMP participants who were tested positive for the *GCK*-MODY variants in Table 5.2. However, none of them had done 23andMe genetic testing.

Table 5.4 Number of *GCK* variants in surveyed PDMP participants

Participant	Number of *GCK* records	Number of detected *GCK* genotypes	Number of exonic variants	Number of non-coding variants
1	309	300	0	9
2	309	300	0	8
3	309	297	0	12
4	13	13	0	2

Different third-party tools interpreted the *GCK* variants from 23andMe raw data differently

Classifications (e.g., pathogenic, VUS, benign) of the same GCK variant differed among various third-party interpretation tools, as shown in Figure 5.2. These differences may be attributed to the sources from which the algorithms automatically pull. Tools such as LiveWello, Promethease, and Codegen only incorporated non-coding variants published in GWAS studies in their analysis. Genetic Genie listed some of the exonic variants discovered in MODY or neonatal diabetes cases and posted the links to ClinVar, PubMed, SNPedia, LitVar, and OMIM, as supplements to the results. SNPedia, unlike other third-party tools, is a database consisting of a small proportion of variants used by

major DTC-GT companies (23andMe, Ancestry, FamilyTreeDNA, etc.). It contained 134 *GCK* variants, but only 60 exonic and four non-coding variants were included in 23andMe raw data.

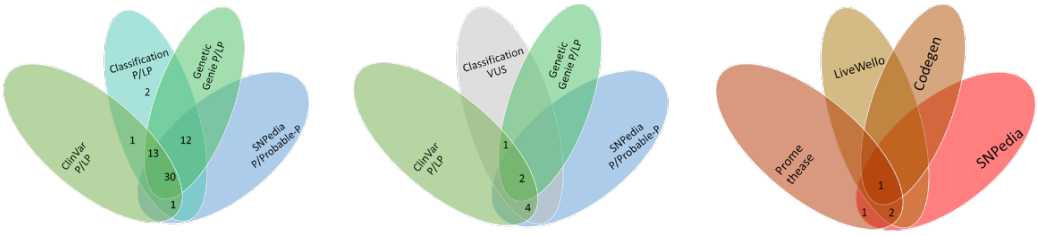


Figure 5.2 Venn diagrams of *GCK* variants interpreted by third-party tools
Numbers of overlapping variants by different tools were labeled, zeros were skipped. a. Venn diagram of P/LP variants which were also interpreted to be P/LP by ClinVar, Genetic Genie, or P/Probable-P by SNPedia. b. Venn diagram of VUS variants which were interpreted to be P/LP by ClinVar, Genetic Genie, or P/Probable-P by SNPedia. c. Venn diagram of non-coding variants analyzed by Promethease, Codegen, LiveWello, or SNPedia.

5.5 Discussion

This project attempted to evaluate the ability of DTC-GT raw data to detect disease-causing *GCK* variants, and the capability of third-party interpretation tools to correctly interpret *GCK* variants. The findings provide insights into the information accessible to DTC-GT customers when seeking interpretations of their DTC-GT raw data.

For customers who are carriers of disease-causing *GCK* variants, the chances of not having their *GCK* variants in the raw data are high since only about a third of disease-causing *GCK* variants were covered by the microarray. For example, only 10 of 21 P/LP *GCK* variants found in PDMP patients with *GCK*-MODY were included in the 23andMe raw data, meaning that the other 11 patients would not have had their disease-causing variants in their raw data if they had ordered 23andMe DTC-GT. For unaffected individuals or patients with other types of diabetes, the chance of false positivity in their

raw data exists and could lead to unnecessary testing and/or false reassurance that their phenotype was caused by *GCK* variants in their raw data.

As stated in the 23andMe raw data disclaimer, the raw data did not go through strict quality control. The MAF inconsistency between *GCK* variants from 23andMe raw data and the large population database suggests poor data quality. Previous research on genotyping very rare variants (population frequency below 0.001%) using a microarray platform in the UK biobank reached a positive predictive value (PPV) as low as 4.4%. Comparing to the 99.7% PPV of common variants (populations frequency above 1%), the DNA microarray has limited accuracy in genotyping very rare variants[338,339]. Furthermore, Tandy-Connor et al. reported that in a cohort of 49 individuals, 40% of the variants in clinically actionable genes were false positive. Also, eight variants that were stated to increase disease risk by DTC or third-party tools were classified as benign or VUS in this study[336].

In addition to issues with the DTC-GT raw data itself, third-party tools' interpretations add more complexity to understanding the consumers' raw data. Results from third-party tools are often incomplete, conflicting, and misleading[350] (Figure 5.2b). Thirty of the 55 exonic variants were consistently classified as pathogenic or likely pathogenic by SNPedia and Genetic Genie, concordant to the MDEP classification. Seven *GCK* variants classified as VUS by MDEP due to lack of evidence were called LP by SNPedia and Genetic Genie (latter derived from ClinVar). One likely benign variant NM_000162.5:c.1386G>T; p.Met462Ile (rs193922285) was classified as VUS in Genetic Genie (derived from ClinVar) and interpreted as Probable-Pathogenic in SNPedia. In fact, this variant was found in individuals with normal glucose level and were deemed not

to be disease-causing.

The project's limitation is the lack of phenotype data from the openSNP samples to calculate the sensitivity and specificity of the raw data in detecting disease-causing *GCK* variants. In addition, it is not known whether any of the variants in Table 5.2 detected in the PDMP patients with *GCK*-MODY would have been picked up by 23andMe, since the 23andMe data from these participants is not available. Further studies simultaneously evaluating DTC-GT raw data and NGS results are needed to make a more comprehensive comparison.

Overall, DTC-GT raw data's detection capability is limited for disease-causing variants in *GCK*, and genetics professionals like clinical geneticists and genetic counselors should be involved to avoid inaccurate interpretation and wrong action[351]. It seems that DTC-GT makes it easier for the public to access genetic testing; the faulted ability to detect rare variants in the presented example suggests it is not worth the risk for consumers to consider it when clinical genetic testing is inaccessible. Though other uses beyond DTC-GT's ability to predict disease do exist[352], like ancestry and relatedness, such uses should be based on reliable data and accurate interpretation.

Chapter 6 : Discussion

6.1 Summary

The overall purpose of this dissertation was to evaluate methods to improve the diagnostic rate of monogenic diabetes with various approaches, thus expanding the capability of genomic medicine in diabetes. These approaches include 1) establishing a model for the clinic to diagnose patients with monogenic diabetes routinely; 2) comprehensively re-analyzing next-generation sequencing data in the PDMP study (NGS panel) and ES data in the ProDiGY study; 3) assessing the capability of Direct-to-Consumer Genetic Testing (DTC-GT) raw data to potentially diagnose and raise awareness of monogenic diabetes among patients. The results from these studies have illustrated practical strategies to improve the diagnostic rate during patient screening, DNA sequencing, and sequencing data analysis. It also has indicated that DTC-GT cannot replace clinical genetic testing in detecting monogenic diabetes mutations. The results of each study are summarized below.

In chapter 3, the PDMP used a simple questionnaire to screen 2,522 patients with diabetes for monogenic diabetes and discovered 313 patients suspected to have monogenic diabetes. The screened patients were from various resources, including physician recognition, self-referral, clinical screening, and EHR portal screen. After sequencing the 313 suspected patients and an additional 143 patients with T1D or T2D as the control group, we identified pathogenic or likely pathogenic variants in 12.1% (38/313) of the suspected patients. Most of the variants located in the genes causing the three most common types of MODY, *GCK*(n=21), *HNF1A*(n=7), and *HNF4A*(n=2), as expected. The genetic findings would help determine the most appropriate treatment for

them: low-dose sulfonylureas for *HNF1A*-MODY (n=3) and *HNF4A*-MODY(n=1) patients and no treatment for *GCK*-MODY (n=8) patients. Another eight participants were found to have pathogenic or likely pathogenic variants in other monogenic diabetes genes: *KCNJ11*(n=1), *INS*(n=2), *LMNA*(n=2), *HNF1B*(n=1), *WFS1*(n=1) and *MC4R*(n=1). The genetic findings in other genes could also be informative in switching to more suitable treatment, such as successfully switching from insulin to high-dose sulfonylureas for the *KCNJ11*-MODY patient. I calculated the pick-up rates of different algorithm criteria groups and patient source groups and found out the high pick-up rates among the self-referred patients (32.4%), and in the physician recognized patients (18.3%), and the high pick-up rate among the patients diagnosed with non-T1D before age 30 (15.0%), especially the ones not treated with insulin (28.8%). The results could guide future studies on monogenic diabetes aiming for a high pick-up rate among selected patients.

In chapter 4, different types of next-generation sequencing data were re-analyzed by various methods to improve the diagnostic rate of monogenic diabetes. Chapter 4.1 re-analyzed the NGS panel data on 40 genes of 313 patients suspected of monogenic diabetes from the PDMP study. The updated variant calling step using a new variant caller, GATK, helped identify one *GCK*-MODY variant p.(Glu300Lys), which was not recognized by the default Torrent Variant Caller (TVC). The main reason for this missing variant is the low depth at the nearby region that lowered its mapping quality and base quality in the TVC algorithm, but not impacting the GATK algorithm. The finding of this missing variant indicates the importance of orthogonal validation of different bioinformatics tools. In addition, by re-classifying previously identified VUS variants

using evidence from updated literature, functional studies, and additional case info from collaborators, three additional pathogenic or likely pathogenic variants were identified. Identification of the three variants will potentially benefit patients with a correct diagnosis, etiology-based treatment, and risk management for family members, as mentioned before. The result suggests the significance of re-classifying VUS variants timely with new evidence from literature and collaboration. The analysis on the non-canonical splicing variants did not produce pathogenic or likely pathogenic variants with additional diagnosis. However, the analysis found one more *ABCC8* variant, that when combined with two previously identified *ABCC8* variants that could explain the patient's phenotype. The comprehensive re-analysis improved the diagnostic rate of monogenic diabetes in the PDMP study from 11.82% to 13.10% (37 to 41 out of 313). It emphasized the importance of timely re-analysis of NGS data which could lead to a new diagnosis.

Chapter 4.2 compared ES with NGS panel of 488 ProDiGY participants and Sanger sequencing of 492 ProDiGY participants to detect monogenic diabetes variants. The initial analysis of ES data failed to identify 42.9% of the pathogenic or likely pathogenic variants of 10 MODY genes found by NGS panel and 10.2% of the pathogenic or likely pathogenic variants of three MODY genes found by Sanger sequencing. The re-analysis of ES raw data removed the VQSR recalibration step during the variant calling and then retrieved the mentioned missing variants and additional 19 disease-causing variants in other ProDiGY participants. The re-analysis increased the pick-up rate from 2.2% to 2.9% (73 to 95 out of 3,333). The results showed that ES data with a mean coverage depth of 56x could reach similar sensitivity for monogenic diabetes in patients with T2D. However, for diagnosis, to make sure ES could achieve the same

detection rate as NGS panel or Sanger sequencing, it would be better not to filter out variants with low-quality metrics (depth or quality score greater than a threshold) performed by other studies[182,183,328]. Instead, researchers could check the reads' alignment at the suspected loci and determine the reason for low quality. If sequencing errors do not cause the low quality of suspected variants, researchers could continue to classify the variant. After evaluating the pathogenicity, researchers could decide whether to confirm this variant through targeted Sanger sequencing.

In chapter 5, this study found that the 23andMe raw data is not reliable in detecting *GCK*-MODY variants. The 23andMe DNA microarrays were not comprehensive in that they were designed to cover only around a third of reported *GCK*-MODY variants. And the 23andMe raw data had high false-positive of *GCK*-MODY variants because the *GCK*-MODY variants are rare and cannot be correctly detected by DNA microarray. The third-party tools' interpretations were conflicting and misleading in guiding customers to understand their raw data.

6.2 Implications

Though many studies have done systematic screening on ascertaining monogenic diabetes patients among diabetes cohorts, the shortcomings of these studies include 1) Often mainly focus on pediatric populations [165,237,353-357] 2) Only tested for the most common type of monogenic diabetes [164,165,237,353,358,359] 3) Absence of evaluation of the potential clinical significance of positive genetic testing [164,358,360-365]. The PDMP study used systematic screening on the various populations of patients with diabetes to assess the diagnosis rate of algorithm criteria that could distinguish patients with monogenic

diabetes from patients with T1D or T2D and evaluate possible treatment change after genetic diagnosis. The design of the algorithm criteria was based on previous research that looked at various populations with diabetes. It included patients diagnosed before one year of age [1,2] (group 1), patients with a family history of T1D (group 2) [1], patients with features that were consistent with monogenic diabetes syndromes (group 5), and patients who were non-obese and impaired fasting glucose (group 6)[43]. Group 1, 2, 5, and 6 had small sample sizes ($\leqslant$10); thus, limited information can be concluded about these groups. Group 4 (diagnosed with T2DM < 45 years of age, non-obese, dominant family history, and antibody-negative) did not pick up any patients with monogenic diabetes. The possible cause of diabetes in these patients is that other genes that were not covered by the sequencing panel or polygenic rather than monogenic factors contributed to the young-onset T2D diabetes. Group 3 criterion was based on previous recommendations about the findings of a high prevalence of MODY (25%) in patients with apparent T2D diagnosed before 30 years (including GAD65-antibody negativity and C-peptide positivity) [167]. PDMP adapted such results to target patients with T2D diagnosed before age 30, retained endogenous insulin production (tested positive for C-peptide or independent of insulin treatment), and negative for IA-2 autoantibody. The participants not treated with insulin in this group had the highest pick-up rate (28.8%) among all groups. Overall, the result of PDMP is consistent with previous research that testing patients diagnosed at a young age after ruling out apparent T1D could pick up patients with monogenic diabetes. In addition, the comparison between different sources of patients showed the self-referral patients were more likely to be tested positive for monogenic diabetes for the first time. This finding suggests the potential benefit of

educating patients about monogenic diabetes. Certain patients are probably more alert about unusual diabetes phenotypes and would spend money and effort to confirm their suspicion. Though these patients only represented a small percentage of all patients with diabetes, it would still be worthy of getting the knowledge about monogenic diabetes out to as many patients as possible to empower them to seek answers.

As described in chapter 4.1 introduction, the value and necessity of sequencing data re-analysis have been proven in many genetic disorders. Chapter 4.1 ascertained their findings by comprehensive re-analysis of PDMP NGS panel data. The primary reason for improving the diagnostic rate through re-analysis included updated literature and database and updated bioinformatic pipeline like other studies. The unique reason for the re-classification of VUS was adding functional evidence from collaborators. This finding suggests the importance of collaboration among various types of experts in the clinical and scientific aspects of diseases that enables the connection of different aspects of variant interpretation.

Chapter 4.2 added confidence to ES as the first-tier genetic testing for monogenic diabetes, like some of the previous research[217,320]. However, the similar sensitivity of ES in detecting disease-causing variants compared to NGS panel and Sanger sequencing was under the condition that variants from the raw data were not filtered based on quality. The ProDiGY ES was planned and executed for GWAS analysis, requiring high sensitivity and specificity of detecting common SNPs. Utilizing these data in rare variants detection and disease diagnosis requires high sensitivity in detecting rare variants. Consequently, the ES data needed adjustment in data processing and filtration to fit in different purposes. The NGS panel data analysis used in the PDMP and TODAY study illustrated

in the initial analysis that no simple threshold of quality or depth score was applied to filter out variants that could capture all possible disease-causing variants. Low-quality variants were retained for evaluation of pathogenicity and quality. However, for a cohort with a large sample size such as ProDiGY, the number of resulting variants that require evaluation will be significant, and other options to reduce the workload need investigation.

Chapter 5 supported the conclusions from previous research that DNA microarray data, including DTC-GT raw data, is not reliable in detecting very rare disease-causing variants[339]. The disclaimer in the DTC-GT raw data stating that the data was not validated for accuracy and should not be used for clinical decision cannot guarantee that the data is not being misinterpreted or misused for medical use. The consequent possible misclassification and false-positive results can cause unnecessary tests of customers and family members and emotional stress. In addition, the DNA microarray technology designed to detect variants at limited genetic loci is not a comprehensive test. The results in chapter 5 showed a high risk of false-negative in detecting *GCK* variants. As a result, customers should not use DTC-GT raw data as the solely supplementary method to clinical genetic testing. Any healthcare-related decisions should be based on validated results from a CLIA-certified lab. What's more, the variability of interpretations from different third-party interpretation tools poses a risk to customers. Their interpretations lack scientific curation built from evidence from multiple resources and could cause misunderstandings for customers. To avoid the undesirable impact of false-positive and false-negative results from DTC-GT raw data, as well as un-curated third-party

interpretations on customers, customers and healthcare providers need to be aware of the limitation of DTC-GT.

6.3 Limitations

Several limitations in this dissertation prohibited it from achieving a higher diagnostic rate for monogenic diabetes that left patients undiagnosed, and some disease-causing variants were probably undetected. This dissertation relied on the genetic data from patients or customers and the quality of that genetic data decides the maximum diagnostic yield. The design of the gene panel and amplicon-based NGS panel method used by the Ion Torrent PGM and S5 system in PDMP may result in missing variants in uncovered regions. Similarly, the exome sequencing in ProDiGY was limited to exome regions with good coverage. Such limitations cannot be rescued by improving analytical ability as applied in Chapter 4, but re-sequencing using different sequencing methods or on uncovered or poorly covered regions could help.

The 23andMe raw data in chapter 5 was not from the same batch of samples or even the same version of SNP designs of DNA microarrays. Therefore, it could not use conventional quality control methods. Typically, with raw microarray data after the genotype calls, researchers use parameters such as the call rate of a microarray chip (percentage of SNPs assigned genotypes of all SNPs on a chip) or of an SNP (percentage of samples assigned genotypes of all samples tested), or Hardy-Weinberg equilibrium of each batch to evaluate the data quality. In this study, none of these could be applied, and only the comparison of population frequency between the 23andMe data and large

population database could be used to evaluate the data quality of the 23andMe raw data from openSNP.

Additionally, some aims in this dissertation were limited by the sample sizes to reach statistical significance. The PDMP had a small sample size (≤50) in groups other than patients with non-T1D diagnosed before age 30 (group 3) or patients clinically suspected to have monogenic diabetes (group 7). Therefore, the pick-up rates of groups 1, 2, 4, 5, and 6 needed further studies to evaluate. Likewise, the 23andMe raw data were enriched with European ancestry samples, and other ancestry groups were under-represented. Furthermore, only four PDMP participants offered their raw data for analysis in chapter 5, and none of them had confirmed *GCK*-MODY variants in their NGS data. Lack of positive control (PDMP participants with *GCK*-MODY variants with 23andMe raw data uploaded) prevented the direct comparison of 23andMe raw data and NGS data in detecting *GCK*-MODY variants. On the other hand, though the ProDiGY collaboration had ES data of 3,333 participants, the analyzed variants during the primary analysis and re-analysis only consisted of a small percentage of all variants in the ES data.

6.4 Recommendations and future directions

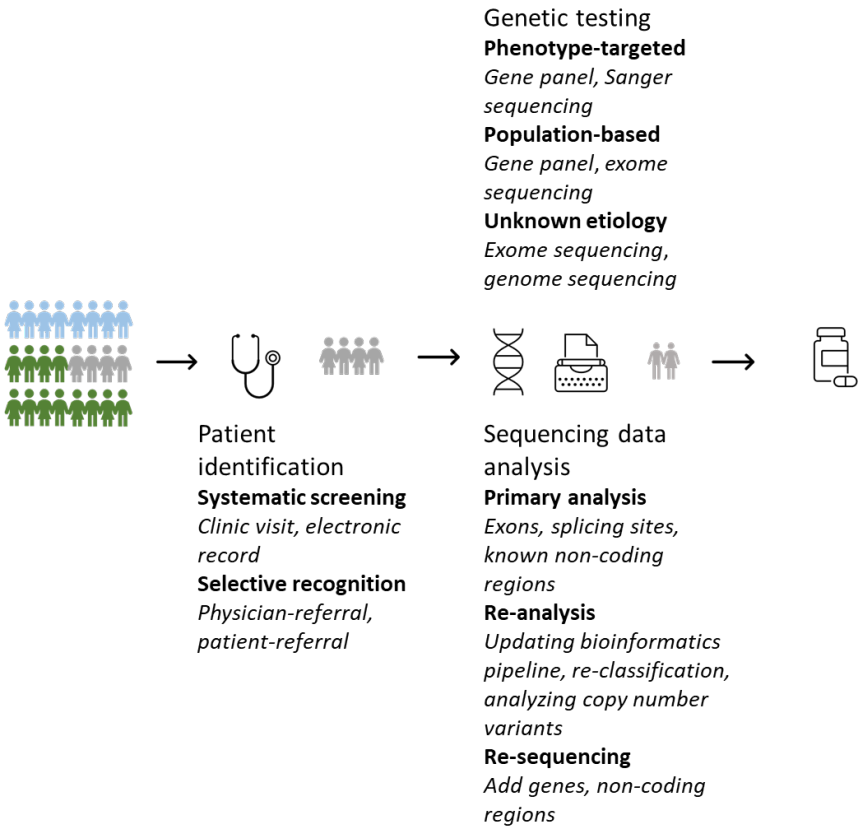


Figure 6.1 Proposed approaches to improve the diagnostic rate of monogenic diabetes

To have more patients with diabetes benefit from genomic medicine, I proposed several approaches to improve the diagnostic rate of monogenic diabetes in this dissertation, as shown in Figure 6.1. These approaches focus on patient identification, genetic testing, and sequencing data analysis. Depending on the capability and purpose, identifying patients suspected of having monogenic diabetes can be achieved through systematic screening using questionnaires during the clinic visit and electronic records searching or recognition by physicians and patients themselves. A comprehensive electronic health record system connecting patient's phenotype, lab results, and treatment information with computer algorithms will enable live and continuous screening of

patients. It is worth mentioning the importance of establishing a center of experienced physicians other health care providers such as genetic counselors, nurses, Certified Diabetes Care and Education Specialists, to recognize the unusual signals that flag the possibility of monogenic diabetes from the patient's centralized information. Further studies are demanded to evaluate the efficiency of centralized recognition compared with current localized actions. The high pick-up rate (32.4%) of monogenic diabetes among self-referred patients in PDMP provides evidence of the benefit of educating patients with diabetes about monogenic diabetes.

Based on the study purpose, the sequencing method and data analysis plan need to be wisely selected to reach desired sensitivity. It would be easy for patients with monogenic disorders to directly have Sanger sequencing on the corresponding genes when the phenotype is specific and highly distinguishable. However, monogenic diabetes is clinically heterogenous and some variants have incomplete penetrance. This determines that a gene panel that includes genes that are supported to be causing the phenotype is more beneficial in phenotype-positive patients. For research purpose but not clinical diagnosis, this panel could also include suspected genes that need further evidence supporting gene-disease relationship to search for new disease etiology. As the cost of ES decreases, it is now common to have ES data of healthy populations or other populations without diabetes. The existence of disease-causing variants in such populations requires further research and can explain the variable penetrance and expressivity of monogenic diabetes among populations. The search for new candidate genes for monogenic diabetes may require exome sequencing or genome sequencing that covers more regions than our current knowledge such as intronic regulatory regions.

Not all patients get positive results initially, and sometimes, re-sequencing or re-analyzing available sequencing data could lead to a new diagnosis. Negative genetic testing results are not the end, and as shown in chapter 4; they may be rescued by comprehensive re-analysis or potentially re-sequencing. The re-analysis should be designed to maximize the outcome of sequencing data and routinely executed after the initial analysis to make sure the guidelines and evidence are up to date. Lastly, these approaches are based on research in monogenic diabetes, but could be applied to other monogenic disorders that are similarly heterogeneous in genotype and phenotype to improve diagnostic yield.

References

1. Hattersley AT, Greeley SAW, Polak M, et al. ISPAD Clinical Practice Consensus Guidelines 2018: The diagnosis and management of monogenic diabetes in children and adolescents. *Pediatr Diabetes.* 2018;19 Suppl 27:47-63.

2. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2020. *Diabetes Care.* 2020;43(Supplement 1):S14-S31.

3. Tattersall RB. Mild familial diabetes with dominant inheritance. *Q J Med.* 1974;43(170):339-357.

4. Fajans SS, Cloutier MC, Crowther RL. Clinical and Etiologic Heterogeneity of Idiopathic Diabetes Mellitus. *Diabetes.* 1978;27(11):1112-1125.

5. Weiss RE, Refetoff S. *Genetic diagnosis of endocrine disorders.* Second edition. ed. Amsterdam Boston: Academic Press; 2016.

6. Laver TW, Colclough K, Shepherd M, et al. The Common p.R114W HNF4A Mutation Causes a Distinct Clinical Subtype of Monogenic Diabetes. *Diabetes.* 2016;65(10):3212-3217.

7. Consortium STD, Estrada K, Aukrust I, et al. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. *JAMA.* 2014;311(22):2305-2314.

8. Lango Allen H, Johansson S, Ellard S, et al. Polygenic risk variants for type 2 diabetes susceptibility modify age at diagnosis in monogenic HNF1A diabetes. *Diabetes.* 2010;59(1):266-271.

9. Kleinberger JW, Copeland KC, Gandica RG, et al. Monogenic diabetes in overweight and obese youth diagnosed with type 2 diabetes: the TODAY clinical trial. *Genet Med.* 2018;20(6):583-590.

10. Laver T, Wakeling M, Knox O, et al. Redefining the pathogenicity of Maturity Onset Diabetes of the Young (MODY) genes: BLK, PAX4 and KLF11 do not cause MODY. Paper presented at: DIABETIC MEDICINE2018.

11. Patel KA, Kettunen J, Laakso M, et al. Heterozygous RFX6 protein truncating variants are associated with MODY with reduced penetrance. *Nat Commun.* 2017;8(1):888.

12. Thanabalasingham G, Owen KR. Diagnosis and management of maturity onset diabetes of the young (MODY). *BMJ.* 2011;343:d6044.

13. Murphy R, Ellard S, Hattersley AT. Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes. *Nat Clin Pract Endocrinol Metab.* 2008;4(4):200-213.

14. Steele AM, Shields BM, Wensley KJ, Colclough K, Ellard S, Hattersley AT. Prevalence of vascular complications among patients with glucokinase mutations and prolonged, mild hyperglycemia. *JAMA.* 2014;311(3):279-286.

15. Colclough K, Bellanne-Chantelot C, Saint-Martin C, Flanagan SE, Ellard S. 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. *Hum Mutat.* 2013;34(5):669-685.

16. Osbak KK, Colclough K, Saint-Martin C, et al. Update on mutations in glucokinase (GCK), which cause maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemic hypoglycemia. *Hum Mutat.* 2009;30(11):1512-1526.

17. Stoffers DA, Ferrer J, Clarke WL, Habener JF. Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. *Nat Genet.* 1997;17(2):138-139.

18. Fajans SS, Bell GI, Paz VP, et al. Obesity and hyperinsulinemia in a family with pancreatic agenesis and MODY caused by the IPF1 mutation Pro63fsX60. *Transl Res.* 2010;156(1):7-14.

19. Caetano LA, Santana LS, Costa-Riquetto AD, et al. PDX1 -MODY and dorsal pancreatic agenesis: New phenotype of a rare disease. *Clin Genet.* 2018;93(2):382-386.

20. Yu H, Liu J, Li X, Miao F, Yang Y. Identification of a novel mutation site in maturityonset diabetes of the young in a Chinese family by wholeexome sequencing. *Mol Med Rep.* 2019;20(3):2373-2380.

21. Yorifuji T, Kurokawa K, Mamada M, et al. Neonatal diabetes mellitus and neonatal polycystic, dysplastic kidneys: Phenotypically discordant recurrence of a mutation in the hepatocyte nuclear factor-1beta gene due to germline mosaicism. *J Clin Endocrinol Metab.* 2004;89(6):2905-2908.

22. Horikawa Y, Enya M. Genetic Dissection and Clinical Features of MODY6 (NEUROD1-MODY). *Curr Diab Rep.* 2019;19(3):12.

23. Raeder H, Haldorsen IS, Ersland L, et al. Pancreatic lipomatosis is a structural marker in nondiabetic children with mutations in carboxyl-ester lipase. *Diabetes.* 2007;56(2):444-449.

24. Raeder H, Johansson S, Holm PI, et al. Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. *Nat Genet.* 2006;38(1):54-62.

25. Vesterhus M, Raeder H, Kurpad AJ, et al. Pancreatic function in carboxyl-ester lipase knockout mice. *Pancreatology.* 2010;10(4):467-476.

26. Liu M, Sun J, Cui J, et al. INS-gene mutations: from genetics and beta cell biology to clinical disease. *Mol Aspects Med.* 2015;42:3-18.

27. Stoy J, Edghill EL, Flanagan SE, et al. Insulin gene mutations as a cause of permanent neonatal diabetes. *Proc Natl Acad Sci U S A.* 2007;104(38):15040-15044.

28. Edghill EL, Flanagan SE, Patch AM, et al. Insulin mutation screening in 1,044 patients with diabetes: mutations in the INS gene are a common cause of neonatal diabetes but a rare cause of diabetes diagnosed in childhood or adulthood. *Diabetes.* 2008;57(4):1034-1042.

29. Molven A, Ringdal M, Nordbo AM, et al. Mutations in the insulin gene can cause MODY and autoantibody-negative type 1 diabetes. *Diabetes.* 2008;57(4):1131-1135.

30. Gloyn AL, Siddiqui J, Ellard S. Mutations in the genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) in diabetes mellitus and hyperinsulinism. *Hum Mutat.* 2006;27(3):220-231.

31. Flanagan SE, Clauin S, Bellanne-Chantelot C, et al. Update of mutations in the genes encoding the pancreatic beta-cell K(ATP) channel subunits Kir6.2 (KCNJ11) and sulfonylurea receptor 1 (ABCC8) in diabetes mellitus and hyperinsulinism. *Hum Mutat.* 2009;30(2):170-180.

32. De Franco E, Saint-Martin C, Brusgaard K, et al. Update of variants identified in the pancreatic beta-cell KATP channel genes KCNJ11 and ABCC8 in individuals with congenital hyperinsulinism and diabetes. *Hum Mutat.* 2020;41(5):884-905.

33. Prudente S, Jungtrakoon P, Marucci A, et al. Loss-of-Function Mutations in APPL1 in Familial Diabetes Mellitus. *Am J Hum Genet.* 2015;97(1):177-185.

34. Haliyur R, Tong X, Sanyoura M, et al. Human islets expressing HNF1A variant have defective beta cell transcriptional regulatory networks. *J Clin Invest.* 2019;129(1):246-251.

35. Pearson ER, Boj SF, Steele AM, et al. Macrosomia and hyperinsulinaemic hypoglycaemia in patients with heterozygous mutations in the HNF4A gene. *PLoS Med.* 2007;4(4):e118.

36. Yau D, Colclough K, Natarajan A, et al. Congenital hyperinsulinism due to mutations in HNF1A. *Eur J Med Genet.* 2020;63(6):103928.

37. Shepherd M, Shields B, Ellard S, Rubio-Cabezas O, Hattersley AT. A genetic diagnosis of HNF1A diabetes alters treatment and improves glycaemic control in the majority of insulin-treated patients. *Diabet Med.* 2009;26(4):437-441.

38. Pearson ER, Liddell WG, Shepherd M, Corrall RJ, Hattersley AT. Sensitivity to sulphonylureas in patients with hepatocyte nuclear factor-1alpha gene mutations: evidence for pharmacogenetics in diabetes. *Diabet Med.* 2000;17(7):543-545.

39. Pearson ER, Starkey BJ, Powell RJ, Gribble FM, Clark PM, Hattersley AT. Genetic cause of hyperglycaemia and response to treatment in diabetes. *Lancet.* 2003;362(9392):1275-1281.

40. Fantasia KL, Steenkamp DW. Optimal Glycemic Control in a Patient With HNF1A MODY With GLP-1 RA Monotherapy: Implications for Future Therapy. *J Endocr Soc.* 2019;3(12):2286-2289.

41. Christensen AS, Haedersdal S, Stoy J, et al. Efficacy and Safety of Glimepiride With or Without Linagliptin Treatment in Patients With HNF1A Diabetes (Maturity-Onset Diabetes of the Young Type 3): A Randomized, Double-Blinded, Placebo-Controlled, Crossover Trial (GLIMLINA). *Diabetes Care.* 2020;43(9):2025-2033.

42. Byrne MM, Sturis J, Clement K, et al. Insulin secretory abnormalities in subjects with hyperglycemia due to glucokinase mutations. *J Clin Invest.* 1994;93(3):1120-1130.

43. Chakera AJ, Spyer G, Vincent N, Ellard S, Hattersley AT, Dunne FP. The 0.1% of the population with glucokinase monogenic diabetes can be recognized by

clinical characteristics in pregnancy: the Atlantic Diabetes in Pregnancy cohort. *Diabetes Care.* 2014;37(5):1230-1236.

44. Carmody D, Naylor RN, Bell CD, et al. GCK-MODY in the US National Monogenic Diabetes Registry: frequently misdiagnosed and unnecessarily treated. *Acta Diabetol.* 2016;53(5):703-708.

45. Stride A, Shields B, Gill-Carey O, et al. Cross-sectional and longitudinal studies suggest pharmacological treatment used in patients with glucokinase mutations does not alter glycaemia. *Diabetologia.* 2014;57(1):54-56.

46. Wedrychowicz A, Tobor E, Wilk M, et al. Phenotype Heterogeneity in Glucokinase-Maturity-Onset Diabetes of the Young (GCK-MODY) Patients. *J Clin Res Pediatr Endocrinol.* 2017;9(3):246-252.

47. Cuesta-Munoz AL, Tuomi T, Cobo-Vuilleumier N, et al. Clinical heterogeneity in monogenic diabetes caused by mutations in the glucokinase gene (GCK-MODY). *Diabetes Care.* 2010;33(2):290-292.

48. Szopa M, Wolkow J, Matejko B, et al. Prevalence of Retinopathy in Adult Patients with GCK-MODY and HNF1A-MODY. *Exp Clin Endocrinol Diabetes.* 2015;123(9):524-528.

49. Hattersley AT, Beards F, Ballantyne E, Appleton M, Harvey R, Ellard S. Mutations in the glucokinase gene of the fetus result in reduced birth weight. *Nat Genet.* 1998;19(3):268-270.

50. Caswell RC, Snowsill T, Houghton JAL, et al. Noninvasive Fetal Genotyping by Droplet Digital PCR to Identify Maternally Inherited Monogenic Diabetes Variants. *Clin Chem.* 2020;66(7):958-965.

51. Edghill EL, Stals K, Oram RA, Shepherd MH, Hattersley AT, Ellard S. HNF1B deletions in patients with young-onset diabetes but no known renal disease. *Diabet Med.* 2013;30(1):114-117.

52. Dubois-Laforgue D, Bellanne-Chantelot C, Charles P, et al. Intellectual disability in patients with MODY due to hepatocyte nuclear factor 1B (HNF1B) molecular defects. *Diabetes Metab.* 2017;43(1):89-92.

53. Dubois-Laforgue D, Cornu E, Saint-Martin C, et al. Diabetes, Associated Clinical Spectrum, Long-term Prognosis, and Genotype/Phenotype Correlations in 201 Adult Patients With Hepatocyte Nuclear Factor 1B (HNF1B) Molecular Defects. *Diabetes Care.* 2017;40(11):1436-1443.

54. Clissold RL, Shaw-Smith C, Turnpenny P, et al. Chromosome 17q12 microdeletions but not intragenic HNF1B mutations link developmental kidney disease and psychiatric disorder. *Kidney Int.* 2016;90(1):203-211.

55. Tomura H, Nishigori H, Sho K, Yamagata K, Inoue I, Takeda J. Loss-of-function and dominant-negative mechanisms associated with hepatocyte nuclear factor-1beta mutations in familial type 2 diabetes mellitus. *J Biol Chem.* 1999;274(19):12975-12978.

56. Lindner TH, Njolstad PR, Horikawa Y, Bostad L, Bell GI, Sovik O. A novel syndrome of diabetes mellitus, renal dysfunction and genital malformation

associated with a partial deletion of the pseudo-POU domain of hepatocyte nuclear factor-1beta. *Hum Mol Genet.* 1999;8(11):2001-2008.

57. Bowman P, Flanagan SE, Edghill EL, et al. Heterozygous ABCC8 mutations are a cause of MODY. *Diabetologia.* 2012;55(1):123-127.

58. Bonnefond A, Philippe J, Durand E, et al. Whole-exome sequencing and high throughput genotyping identified KCNJ11 as the thirteenth MODY gene. *PLoS One.* 2012;7(6):e37423.

59. Simaite D, Kofent J, Gong M, et al. Recessive mutations in PCBD1 cause a new type of early-onset diabetes. *Diabetes.* 2014;63(10):3557-3564.

60. Polak M, Cave H. Neonatal diabetes mellitus: a disease linked to multiple mechanisms. *Orphanet J Rare Dis.* 2007;2:12.

61. Aguilar-Bryan L, Bryan J. Neonatal diabetes mellitus. *Endocr Rev.* 2008;29(3):265-291.

62. Busiah K, Drunat S, Vaivre-Douret L, et al. Neuropsychological dysfunction and developmental defects associated with genetic changes in infants with neonatal diabetes mellitus: a prospective cohort study [corrected]. *Lancet Diabetes Endocrinol.* 2013;1(3):199-207.

63. Iafusco D, Massa O, Pasquino B, et al. Minimal incidence of neonatal/infancy onset diabetes in Italy is 1:90,000 live births. *Acta Diabetol.* 2012;49(5):405-408.

64. Slingerland AS, Shields BM, Flanagan SE, et al. Referral rates for diagnostic testing support an incidence of permanent neonatal diabetes in three European countries of at least 1 in 260,000 live births. *Diabetologia.* 2009;52(8):1683-1685.

65. Temple IK, Shield JP. 6q24 transient neonatal diabetes. *Rev Endocr Metab Disord.* 2010;11(3):199-204.

66. Gloyn AL, Pearson ER, Antcliff JF, et al. Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. *N Engl J Med.* 2004;350(18):1838-1849.

67. Babenko AP, Polak M, Cave H, et al. Activating mutations in the ABCC8 gene in neonatal diabetes mellitus. *N Engl J Med.* 2006;355(5):456-466.

68. Gloyn AL, Reimann F, Girard C, et al. Relapsing diabetes can result from moderately activating mutations in KCNJ11. *Human Molecular Genetics.* 2005;14(7):925-934.

69. Garin I, Edghill EL, Akerman I, et al. Recessive mutations in the INS gene result in neonatal diabetes through reduced insulin biosynthesis. *Proc Natl Acad Sci U S A.* 2010;107(7):3105-3110.

70. Bonnefond A, Lomberk G, Buttar N, et al. Disruption of a novel Kruppel-like transcription factor p300-regulated pathway for insulin biosynthesis revealed by studies of the c.-331 INS mutation found in neonatal diabetes mellitus. *J Biol Chem.* 2011;286(32):28414-28424.

71. Edghill EL, Bingham C, Slingerland AS, et al. Hepatocyte nuclear factor-1 beta mutations cause neonatal diabetes and intrauterine growth retardation: support for a critical role of HNF-1beta in human pancreatic development. *Diabet Med.* 2006;23(12):1301-1306.

72. Kelsey MM, Zeitler PS. Insulin Resistance of Puberty. *Curr Diab Rep.* 2016;16(7):64.

73. Rubio-Cabezas O, Patch AM, Minton JA, et al. Wolcott-Rallison syndrome is the most common genetic cause of permanent neonatal diabetes in consanguineous families. *J Clin Endocrinol Metab.* 2009;94(11):4162-4170.

74. Bacchetta R, Barzaghi F, Roncarolo MG. From IPEX syndrome to FOXP3 mutation: a lesson on immune dysregulation. *Ann N Y Acad Sci.* 2018;1417(1):5-22.

75. Shaw-Smith C, De Franco E, Lango Allen H, et al. GATA4 mutations are a cause of neonatal and childhood-onset diabetes. *Diabetes.* 2014;63(8):2888-2894.

76. Allen HL, Flanagan SE, Shaw-Smith C, et al. GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nat Genet.* 2011;44(1):20-22.

77. De Franco E, Shaw-Smith C, Flanagan SE, et al. GATA6 mutations cause a broad phenotypic spectrum of diabetes from pancreatic agenesis to adult-onset diabetes without exocrine insufficiency. *Diabetes.* 2013;62(3):993-997.

78. Borowiec M, Mysliwiec M, Fendler W, et al. Phenotype variability and neonatal diabetes in a large family with heterozygous mutation of the glucokinase gene. *Acta Diabetol.* 2011;48(3):203-208.

79. Bennett K, James C, Mutair A, Al-Shaikh H, Sinani A, Hussain K. Four novel cases of permanent neonatal diabetes mellitus caused by homozygous mutations in the glucokinase gene. *Pediatr Diabetes.* 2011;12(3 Pt 1):192-196.

80. Esquiaveto-Aun AM, De Mello MP, Paulino MF, Minicucci WJ, Guerra-Junior G, De Lemos-Marini SH. A new compound heterozygosis for inactivating mutations in the glucokinase gene as cause of permanent neonatal diabetes mellitus (PNDM) in double-first cousins. *Diabetol Metab Syndr.* 2015;7:101.

81. Wen X, Yang Y. Emerging roles of GLIS3 in neonatal diabetes, type 1 and type 2 diabetes. *J Mol Endocrinol.* 2017;58(2):R73-R85.

82. Alghamdi KA, Alsaedi AB, Aljasser A, Altawil A, Kamal NM. Extended clinical features associated with novel Glis3 mutation: a case report. *BMC Endocr Disord.* 2017;17(1):14.

83. Fu C, Luo S, Long X, et al. Mutation screening of the GLIS3 gene in a cohort of 592 Chinese patients with congenital hypothyroidism. *Clin Chim Acta.* 2018;476:38-43.

84. Poulton CJ, Schot R, Kia SK, et al. Microcephaly with simplified gyration, epilepsy, and infantile diabetes linked to inappropriate apoptosis of neural progenitors. *Am J Hum Genet.* 2011;89(2):265-276.

85. Polak M, Dechaume A, Cave H, et al. Heterozygous missense mutations in the insulin gene are linked to permanent diabetes appearing in the neonatal period or in early infancy: a report from the French ND (Neonatal Diabetes) Study Group. *Diabetes.* 2008;57(4):1115-1119.

86. Colombo C, Porzio O, Liu M, et al. Seven mutations in the human insulin gene linked to permanent neonatal/infancy-onset diabetes mellitus. *J Clin Invest.* 2008;118(6):2148-2156.

87. Bonfanti R, Colombo C, Nocerino V, et al. Insulin gene mutations as cause of diabetes in children negative for five type 1 diabetes autoantibodies. *Diabetes Care.* 2009;32(1):123-125.

88. Rubio-Cabezas O, Edghill EL, Argente J, Hattersley AT. Testing for monogenic diabetes among children and adolescents with antibody-negative clinically defined Type 1 diabetes. *Diabet Med.* 2009;26(10):1070-1074.

89. Porksen S, Laborie LB, Nielsen L, et al. Disease progression and search for monogenic diabetes among children with new onset type 1 diabetes negative for ICA, GAD- and IA-2 Antibodies. *BMC Endocr Disord.* 2010;10:16.

90. Flanagan SE, De Franco E, Lango Allen H, et al. Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man. *Cell Metab.* 2014;19(1):146-154.

91. Rubio-Cabezas O, Jensen JN, Hodgson MI, et al. Permanent Neonatal Diabetes and Enteric Anendocrinosis Associated With Biallelic Mutations in NEUROG3. *Diabetes.* 2011;60(4):1349-1353.

92. Pinney SE, Oliver-Krasinski J, Ernst L, et al. Neonatal diabetes and congenital malabsorptive diarrhea attributable to a novel mutation in the human neurogenin-3 gene coding sequence. *J Clin Endocrinol Metab.* 2011;96(7):1960-1965.

93. Wang J, Cortina G, Wu SV, et al. Mutant neurogenin-3 in congenital malabsorptive diarrhea. *N Engl J Med.* 2006;355(3):270-280.

94. Yasuda T, Kajimoto Y, Fujitani Y, et al. PAX6 mutation as a genetic factor common to aniridia and glucose intolerance. *Diabetes.* 2002;51(1):224-230.

95. Wen JH, Chen YY, Song SJ, et al. Paired box 6 (PAX6) regulates glucose metabolism via proinsulin processing mediated by prohormone convertase 1/3 (PC1/3). *Diabetologia.* 2009;52(3):504-513.

96. Solomon BD, Pineda-Alvarez DE, Balog JZ, et al. Compound heterozygosity for mutations in PAX6 in a patient with complex brain anomaly, neonatal diabetes mellitus, and microophthalmia. *Am J Med Genet A.* 2009;149A(11):2543-2546.

97. Stoffers DA, Zinkin NT, Stanojevic V, Clarke WL, Habener JF. Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. *Nat Genet.* 1997;15(1):106-110.

98. Wright NM, Metzger DL, Borowitz SM, Clarke WL. Permanent neonatal diabetes mellitus and pancreatic exocrine insufficiency resulting from congenital pancreatic agenesis. *Am J Dis Child.* 1993;147(6):607-609.

99. Thomas IH, Saini NK, Adhikari A, et al. Neonatal diabetes mellitus with pancreatic agenesis in an infant with homozygous IPF-1 Pro63fsX60 mutation. *Pediatr Diabetes.* 2009;10(7):492-496.

100. Schwitzgebel VM, Mamin A, Brun T, et al. Agenesis of human pancreas due to decreased half-life of insulin promoter factor 1. *J Clin Endocrinol Metab.* 2003;88(9):4398-4406.

101. Nicolino M, Claiborn KC, Senee V, Boland A, Stoffers DA, Julier C. A novel hypomorphic PDX1 mutation responsible for permanent neonatal diabetes with subclinical exocrine deficiency. *Diabetes.* 2010;59(3):733-740.

102. Edghill EL, Khamis A, Weedon MN, et al. Sequencing PDX1 (insulin promoter factor 1) in 1788 UK individuals found 5% had a low frequency coding variant, but these variants are not associated with Type 2 diabetes. *Diabet Med.* 2011;28(6):681-684.

103. Temple IK, Gardner RJ, Mackay DJ, Barber JC, Robinson DO, Shield JP. Transient neonatal diabetes: widening the understanding of the etiopathogenesis of diabetes. *Diabetes.* 2000;49(8):1359-1366.

104. Sellick GS, Barker KT, Stolte-Dijkstra I, et al. Mutations in PTF1A cause pancreatic and cerebellar agenesis. *Nat Genet.* 2004;36(12):1301-1305.

105. Hoveyda N, Shield JP, Garrett C, et al. Neonatal diabetes mellitus and cerebellar hypoplasia/agenesis: report of a new recessive syndrome. *J Med Genet.* 1999;36(9):700-704.

106. Tutak E, Satar M, Yapicioglu H, et al. A Turkish newborn infant with cerebellar agenesis/neonatal diabetes mellitus and PTF1A mutation. *Genet Couns.* 2009;20(2):147-152.

107. Mitchell J, Punthakee Z, Lo B, et al. Neonatal diabetes, with hypoplastic pancreas, intestinal atresia and gall bladder hypoplasia: search for the aetiology of a new autosomal recessive syndrome. *Diabetologia.* 2004;47(12):2160-2167.

108. Chappell L, Gorman S, Campbell F, et al. A further example of a distinctive autosomal recessive syndrome comprising neonatal diabetes mellitus, intestinal atresias and gall bladder agenesis. *Am J Med Genet A.* 2008;146A(13):1713-1717.

109. Martinovici D, Ransy V, Vanden Eijnden S, et al. Neonatal hemochromatosis and Martinez-Frias syndrome of intestinal atresia and diabetes mellitus in a consanguineous newborn. *Eur J Med Genet.* 2010;53(1):25-28.

110. Smith SB, Qu HQ, Taleb N, et al. Rfx6 directs islet formation and insulin production in mice and humans. *Nature.* 2010;463(7282):775-780.

111. Sansbury FH, Flanagan SE, Houghton JA, et al. SLC2A2 mutations can cause neonatal diabetes, suggesting GLUT2 may have a role in human insulin secretion. *Diabetologia.* 2012;55(9):2381-2385.

112. Shaw-Smith C, Flanagan SE, Patch AM, et al. Recessive SLC19A2 mutations are a cause of neonatal diabetes mellitus in thiamine-responsive megaloblastic anaemia. *Pediatr Diabetes.* 2012;13(4):314-321.

113. Bonnycastle LL, Chines PS, Hara T, et al. Autosomal dominant diabetes arising from a Wolfram syndrome 1 mutation. *Diabetes.* 2013;62(11):3943-3950.

114. Bansal V, Boehm BO, Darvasi A. Identification of a missense variant in the WFS1 gene that causes a mild form of Wolfram syndrome and is associated with risk for type 2 diabetes in Ashkenazi Jewish individuals. *Diabetologia.* 2018;61(10):2180-2188.

115. Touati A, Errea-Dorronsoro J, Nouri S, et al. Transient neonatal diabetes mellitus and hypomethylation at additional imprinted loci: novel ZFP57 mutation and review on the literature. *Acta Diabetol.* 2019;56(3):301-307.

116. Rubio-Cabezas O, Klupa T, Malecki MT, Consortium C. Permanent neonatal diabetes mellitus--the importance of diabetes differential diagnosis in neonates and infants. *Eur J Clin Invest.* 2011;41(3):323-333.

117. Johnson MB, Hattersley AT, Flanagan SE. Monogenic autoimmune diseases of the endocrine system. *Lancet Diabetes Endocrinol.* 2016;4(10):862-872.

118. Johnson MB, Patel KA, De Franco E, et al. A type 1 diabetes genetic risk score can discriminate monogenic autoimmunity with diabetes from early-onset clustering of polygenic autoimmunity with diabetes. *Diabetologia.* 2018;61(4):862-869.

119. De Franco E, Flanagan SE, Houghton JA, et al. The effect of early, comprehensive genomic testing on clinical care in neonatal diabetes: an international cohort study. *Lancet.* 2015;386(9997):957-963.

120. De Franco E, Lytrivi M, Ibrahim H, et al. YIPF5 mutations cause neonatal diabetes and microcephaly through endoplasmic reticulum stress. *J Clin Invest.* 2020.

121. Vaxillaire M, Populaire C, Busiah K, et al. Kir6.2 mutations are a common cause of permanent neonatal diabetes in a large cohort of French patients. *Diabetes.* 2004;53(10):2719-2722.

122. Flanagan SE, Edghill EL, Gloyn AL, Ellard S, Hattersley AT. Mutations in KCNJ11, which encodes Kir6.2, are a common cause of diabetes diagnosed in the first 6 months of life, with the phenotype determined by genotype. *Diabetologia.* 2006;49(6):1190-1197.

123. Pipatpolkai T, Usher S, Stansfeld PJ, Ashcroft FM. New insights into KATP channel gene mutations and neonatal diabetes mellitus. *Nat Rev Endocrinol.* 2020;16(7):378-393.

124. Ashcroft FM, Puljung MC, Vedovato N. Neonatal Diabetes and the KATP Channel: From Mutation to Therapy. *Trends Endocrinol Metab.* 2017;28(5):377-387.

125. Patch AM, Flanagan SE, Boustred C, Hattersley AT, Ellard S. Mutations in the ABCC8 gene encoding the SUR1 subunit of the KATP channel cause transient neonatal diabetes, permanent neonatal diabetes or permanent diabetes diagnosed outside the neonatal period. *Diabetes Obes Metab.* 2007;9 Suppl 2:28-39.

126. Lang V, Light PE. The molecular mechanisms and pharmacotherapy of ATP-sensitive potassium channel gene mutations underlying neonatal diabetes. *Pharmgenomics Pers Med.* 2010;3:145-161.

127. Yorifuji T, Nagashima K, Kurokawa K, et al. The C42R mutation in the Kir6.2 (KCNJ11) gene as a cause of transient neonatal diabetes, childhood diabetes, or later-onset, apparently type 2 diabetes mellitus. *J Clin Endocrinol Metab.* 2005;90(6):3174-3178.

128. Aguilar-Bryan L, Nichols CG, Wechsler SW, et al. Cloning of the beta cell high-affinity sulfonylurea receptor: a regulator of insulin secretion. *Science.* 1995;268(5209):423-426.

129. Thomas P, Ye Y, Lightner E. Mutation of the pancreatic islet inward rectifier Kir6.2 also leads to familial persistent hyperinsulinemic hypoglycemia of infancy. *Hum Mol Genet.* 1996;5(11):1809-1812.

130. Glaser B, Ryan F, Donath M, et al. Hyperinsulinism caused by paternal-specific inheritance of a recessive mutation in the sulfonylurea-receptor gene. *Diabetes.* 1999;48(8):1652-1657.

131. Slingerland AS, Hattersley AT. Mutations in the Kir6.2 subunit of the KATP channel and permanent neonatal diabetes: new insights and new treatment. *Ann Med.* 2005;37(3):186-195.

132. Proks P, Antcliff JF, Lippiat J, Gloyn AL, Hattersley AT, Ashcroft FM. Molecular basis of Kir6.2 mutations associated with neonatal diabetes or neonatal diabetes plus neurological features. *Proc Natl Acad Sci U S A.* 2004;101(50):17539-17544.

133. Hattersley AT, Ashcroft FM. Activating mutations in Kir6.2 and neonatal diabetes: new clinical syndromes, new scientific insights, and new therapy. *Diabetes.* 2005;54(9):2503-2513.

134. Pearson ER, Flechtner I, Njølstad PR, et al. Switching from Insulin to Oral Sulfonylureas in Patients with Diabetes Due to Kir6.2 Mutations. *New England Journal of Medicine.* 2006;355(5):467-477.

135. Bowman P, Sulen A, Barbetti F, et al. Effectiveness and safety of long-term treatment with sulfonylureas in patients with neonatal diabetes due to KCNJ11

mutations: an international cohort study. *Lancet Diabetes Endocrinol.* 2018;6(8):637-646.

136. Mlynarski W, Tarasov AI, Gach A, et al. Sulfonylurea improves CNS function in a case of intermediate DEND syndrome caused by a mutation in KCNJ11. *Nat Clin Pract Neurol.* 2007;3(11):640-645.

137. Slingerland AS, Hurkx W, Noordam K, et al. Sulphonylurea therapy improves cognition in a patient with the V59M KCNJ11 mutation. *Diabet Med.* 2008;25(3):277-281.

138. Rafiq M, Flanagan SE, Patch AM, et al. Effective treatment with oral sulfonylureas in patients with diabetes due to sulfonylurea receptor 1 (SUR1) mutations. *Diabetes Care.* 2008;31(2):204-209.

139. Masia R, Koster JC, Tumini S, et al. An ATP-binding mutation (G334D) in KCNJ11 is associated with a sulfonylurea-insensitive form of developmental delay, epilepsy, and neonatal diabetes. *Diabetes.* 2007;56(2):328-336.

140. Sumnik Z, Kolouskova S, Wales JK, Komarek V, Cinek O. Sulphonylurea treatment does not improve psychomotor development in children with KCNJ11 mutations causing permanent neonatal diabetes mellitus accompanied by developmental delay and epilepsy (DEND syndrome). *Diabet Med.* 2007;24(10):1176-1178.

141. Babiker T, Vedovato N, Patel K, et al. Successful transfer to sulfonylureas in KCNJ11 neonatal diabetes is determined by the mutation and duration of diabetes. *Diabetologia.* 2016;59(6):1162-1166.

142. Misra S, Vedovato N, Cliff E, et al. Permanent neonatal diabetes: combining sulfonylureas with insulin may be an effective treatment. *Diabet Med.* 2018.

143. Fu JL, Wang T, Xiao XH. Relapsed 6q24-related transient neonatal diabetes mellitus successfully treated with sulfonylurea. *Chin Med J (Engl).* 2019;132(7):846-848.

144. Carmody D, Beca FA, Bell CD, et al. Role of noninsulin therapies alone or in combination in chromosome 6q24-related transient neonatal diabetes: sulfonylurea improves but does not always normalize insulin secretion. *Diabetes Care.* 2015;38(6):e86-87.

145. Docherty LE, Kabwama S, Lehmann A, et al. Clinical presentation of 6q24 transient neonatal diabetes mellitus (6q24 TNDM) and genotype-phenotype correlation in an international cohort of patients. *Diabetologia.* 2013;56(4):758-762.

146. Flanagan SE, Mackay DJ, Greeley SA, et al. Hypoglycaemia following diabetes remission in patients with 6q24 methylation defects: expanding the clinical phenotype. *Diabetologia.* 2013;56(1):218-221.

147. Rigoli L, Di Bella C. Wolfram syndrome 1 and Wolfram syndrome 2. *Curr Opin Pediatr.* 2012;24(4):512-517.

148. Semple RK, Sleigh A, Murgatroyd PR, et al. Postreceptor insulin resistance contributes to human dyslipidemia and hepatic steatosis. *J Clin Invest.* 2009;119(2):315-322.

149. O'Rahilly S, Moller DE. Mutant insulin receptors in syndromes of insulin resistance. *Clin Endocrinol (Oxf).* 1992;36(2):121-132.

150. Hegele RA. Monogenic forms of insulin resistance: apertures that expose the common metabolic syndrome. *Trends Endocrinol Metab.* 2003;14(8):371-377.

151. al-Gazali LI, Khalil M, Devadas K. A syndrome of insulin resistance resembling leprechaunism in five sibs of consanguineous parents. *J Med Genet.* 1993;30(6):470-475.

152. Kadowaki H, Takahashi Y, Ando A, et al. Four mutant alleles of the insulin receptor gene associated with genetic syndromes of extreme insulin resistance. *Biochem Biophys Res Commun.* 1997;237(3):516-520.

153. Kirkwood A, Stuart G, Harding L. Donohue syndrome: A review of literature, case series, and anesthetic considerations. *Paediatr Anaesth.* 2018;28(1):23-27.

154. Hosoe J, Kadowaki H, Miya F, et al. Structural Basis and Genotype-Phenotype Correlations of INSR Mutations Causing Severe Insulin Resistance. *Diabetes.* 2017;66(10):2713-2723.

155. Agarwal AK, Simha V, Oral EA, et al. Phenotypic and genetic heterogeneity in congenital generalized lipodystrophy. *J Clin Endocrinol Metab.* 2003;88(10):4840-4847.

156. Van Maldergem L, Magre J, Khallouf TE, et al. Genotype-phenotype relationships in Berardinelli-Seip congenital lipodystrophy. *J Med Genet.* 2002;39(10):722-733.

157. Sekizkardes H, Cochran E, Malandrino N, Garg A, Brown RJ. Efficacy of Metreleptin Treatment in Familial Partial Lipodystrophy Due to PPARG vs LMNA Pathogenic Variants. *J Clin Endocrinol Metab.* 2019;104(8):3068-3076.

158. Ohkubo K, Yamano A, Nagashima M, et al. Mitochondrial gene mutations in the tRNA(Leu(UUR)) region and diabetes: prevalence and clinical phenotypes in Japan. *Clin Chem.* 2001;47(9):1641-1648.

159. Decoux-Poullot AG, Bannwarth S, Procaccio V, et al. Clinical phenotype of mitochondrial diabetes due to rare mitochondrial DNA mutations. *Ann Endocrinol (Paris).* 2020;81(2-3):68-77.

160. Tian LH, Han XY, Huang XT, et al. A Screening Approach for Mitochondrial tRNA(Leu(UUR)) A3243G Mutation in a Hospital-Based Population with Diabetes. *Chin Med J (Engl).* 2018;131(9):1117-1119.

161. Naylor R. Economics of Genetic Testing for Diabetes. *Curr Diab Rep.* 2019;19(5):23.

162. TODD J, KLEINBERGER JW, SRINIVASAN S, et al. Monogenic Diabetes in the Progress for Diabetes Genetics in Youth (ProDiGY) Collaboration. *Diabetes.* 2018;67(Supplement 1):268-OR.

163. Naylor RN, John PM, Winn AN, et al. Cost-effectiveness of MODY genetic testing: translating genomic advances into practical health applications. *Diabetes Care.* 2014;37(1):202-209.

164. Thomas ER, Brackenridge A, Kidd J, et al. Diagnosis of monogenic diabetes: 10-Year experience in a large multi-ethnic diabetes center. *J Diabetes Investig.* 2016;7(3):332-337.

165. Carlsson A, Shepherd M, Ellard S, et al. Absence of Islet Autoantibodies and Modestly Raised Glucose Values at Diabetes Diagnosis Should Lead to Testing for MODY: Lessons From a 5-Year Pediatric Swedish National Cohort Study. *Diabetes Care.* 2020;43(1):82-89.

166. American Diabetes A. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. *Diabetes Care.* 2020;43(Suppl 1):S14-S31.

167. Thanabalasingham G, Pal A, Selwood MP, et al. Systematic assessment of etiology in adults with a clinical diagnosis of young-onset type 2 diabetes is a successful strategy for identifying maturity-onset diabetes of the young. *Diabetes Care.* 2012;35(6):1206-1212.

168. McDonald TJ, Colclough K, Brown R, et al. Islet autoantibodies can discriminate maturity-onset diabetes of the young (MODY) from Type 1 diabetes. *Diabet Med.* 2011;28(9):1028-1033.

169. Shields BM, McDonald TJ, Ellard S, Campbell MJ, Hyde C, Hattersley AT. The development and validation of a clinical prediction model to determine the probability of MODY in patients with young-onset diabetes. *Diabetologia.* 2012;55(5):1265-1272.

170. Yilmaz Agladioglu S, Sagsak E, Aycan Z. Urinary C-Peptide/Creatinine Ratio Can Distinguish Maturity-Onset Diabetes of the Young from Type 1 Diabetes in

Children and Adolescents: A Single-Center Experience. *Horm Res Paediatr.* 2015;84(1):54-61.

171. Pal A, Farmer AJ, Dudley C, et al. Evaluation of serum 1,5 anhydroglucitol levels as a clinical test to differentiate subtypes of diabetes. *Diabetes Care.* 2010;33(2):252-257.

172. Thanabalasingham G, Shah N, Vaxillaire M, et al. A large multi-centre European study validates high-sensitivity C-reactive protein (hsCRP) as a clinical biomarker for the diagnosis of diabetes subtypes. *Diabetologia.* 2011;54(11):2801-2810.

173. McDonald TJ, Shields BM, Lawry J, et al. High-sensitivity CRP discriminates HNF1A-MODY from other subtypes of diabetes. *Diabetes Care.* 2011;34(8):1860-1862.

174. Owen KR, Thanabalasingham G, James TJ, et al. Assessment of high-sensitivity C-reactive protein levels as diagnostic discriminator of maturity-onset diabetes of the young due to HNF1A mutations. *Diabetes Care.* 2010;33(9):1919-1924.

175. Fu J, Wang T, Liu J, et al. Using Clinical Indices to Distinguish MODY2 (GCK Mutation) and MODY3 (HNF1A Mutation) from Type 1 Diabetes in a Young Chinese Population. *Diabetes Ther.* 2019;10(4):1381-1390.

176. Juszczak A, Pavic T, Vuckovic F, et al. Plasma Fucosylated Glycans and C-Reactive Protein as Biomarkers of HNF1A-MODY in Young Adult-Onset Nonautoimmune Diabetes. *Diabetes Care.* 2019;42(1):17-26.

177. Besser RE, Shepherd MH, McDonald TJ, et al. Urinary C-peptide creatinine ratio is a practical outpatient tool for identifying hepatocyte nuclear factor 1-

{alpha}/hepatocyte nuclear factor 4-{alpha} maturity-onset diabetes of the young from long-duration type 1 diabetes. *Diabetes Care.* 2011;34(2):286-291.

178. Fendler W, Borowiec M, Antosik K, et al. HDL cholesterol as a diagnostic tool for clinical differentiation of GCK-MODY from HNF1A-MODY and type 1 diabetes in children and young adults. *Clin Endocrinol (Oxf).* 2011;75(3):321-327.

179. Ellard S, Lango Allen H, De Franco E, et al. Improved genetic testing for monogenic diabetes using targeted next-generation sequencing. *Diabetologia.* 2013;56(9):1958-1963.

180. Ellard S, Colclough K, Patel KA, Hattersley AT. Prediction algorithms: pitfalls in interpreting genetic variants of autosomal dominant monogenic diabetes. *J Clin Invest.* 2020;130(1):14-16.

181. Carmody D, Park SY, Ye H, et al. Continued lessons from the INS gene: an intronic mutation causing diabetes through a novel mechanism. *J Med Genet.* 2015;52(9):612-616.

182. Kwak SH, Jung CH, Ahn CH, et al. Clinical whole exome sequencing in early onset diabetes patients. *Diabetes Res Clin Pract.* 2016;122:71-77.

183. Johansson S, Irgens H, Chudasama KK, et al. Exome sequencing and genetic testing for MODY. *PLoS One.* 2012;7(5):e38050.

184. Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update

(ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med.* 2017;19(2):249-255.

185. Shim YJ, Kim JE, Hwang SK, et al. Identification of Candidate Gene Variants in Korean MODY Families by Whole-Exome Sequencing. *Horm Res Paediatr.* 2015;83(4):242-251.

186. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003;31(13):3812-3814.

187. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248-249.

188. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet.* 2016;99(4):877-885.

189. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-424.

190. Abou Tayoun AN, Pesaran T, DiStefano MT, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat.* 2018;39(11):1517-1524.

191. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* 2019;12(1):3.

192. Tavtigian SV, Harrison SM, Boucher KM, Biesecker LG. Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. *Hum Mutat.* 2020.

193. Rivera-Munoz EA, Milko LV, Harrison SM, et al. ClinGen Variant Curation Expert Panel experiences and standardized processes for disease and gene-level specification of the ACMG/AMP guidelines for sequence variant interpretation. *Hum Mutat.* 2018;39(11):1614-1622.

194. Malikova J, Kaci A, Dusatkova P, et al. Functional Analyses of HNF1A-MODY Variants Refine the Interpretation of Identified Sequence Variants. *J Clin Endocrinol Metab.* 2020;105(4).

195. Raimondo A, Chakera AJ, Thomsen SK, et al. Phenotypic severity of homozygous GCK mutations causing neonatal or childhood-onset diabetes is primarily mediated through effects on protein stability. *Hum Mol Genet.* 2014;23(24):6432-6440.

196. Majithia AR, Tsuda B, Agostini M, et al. Prospective functional classification of all possible missense variants in PPARG. *Nat Genet.* 2016;48(12):1570-1575.

197. Thormaehlen AS, Schuberth C, Won HH, et al. Systematic cell-based phenotyping of missense alleles empowers rare variant association studies: a case for LDLR and myocardial infarction. *PLoS Genet.* 2015;11(2):e1004855.

198. Althari S, Najmi LA, Bennett AJ, et al. Unsupervised Clustering of Missense Variants in HNF1A Using Multidimensional Functional Data Aids Clinical Interpretation. *The American Journal of Human Genetics.* In press.

199. Najmi LA, Aukrust I, Flannick J, et al. Functional Investigations of HNF1A Identify Rare Variants as Risk Factors for Type 2 Diabetes in the General Population. *Diabetes.* 2017;66(2):335-346.

200. Gloyn AL, Odili S, Buettger C, et al. Glucokinase and the regulation of blood sugar. *Glucokinase and Glycemic Disease: From Basics to Novel Therapeutics.* 2004;1:92-109.

201. Loman NJ, Constantinidou C, Chan JZ, et al. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol.* 2012;10(9):599-606.

202. Fisher T. Torrent Suite Software. https://www.thermofisher.cn/cn/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. Accessed April, 2021.

203. Ebbert MT, Wadsworth ME, Staley LA, et al. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC Bioinformatics.* 2016;17 Suppl 7:239.

204. Zhou W, Chen T, Zhao H, et al. Bias from removing read duplication in ultra-deep sequencing experiments. *Bioinformatics.* 2014;30(8):1073-1080.

205. Zook JM, Catoe D, McDaniel J, et al. Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Sci Data.* 2016;3:160025.

206. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155(2):945-959.

207. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655-1664.

208. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 1977;74(12):5463-5467.

209. Metzker ML. Emerging technologies in DNA sequencing. *Genome Res.* 2005;15(12):1767-1776.

210. International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. *Nature.* 2004;431(7011):931-945.

211. Shendure J, Balasubramanian S, Church GM, et al. DNA sequencing at 40: past, present and future. *Nature.* 2017;550(7676):345-353.

212. Baudhuin LM, Lagerstedt SA, Klee EW, Fadra N, Oglesbee D, Ferber MJ. Confirming Variants in Next-Generation Sequencing Panel Testing by Sanger Sequencing. *J Mol Diagn.* 2015;17(4):456-461.

213. De Summa S, Malerba G, Pinto R, Mori A, Mijatovic V, Tommasi S. GATK hard filtering: tunable parameters to improve variant calling for next generation sequencing targeted gene panel data. *BMC Bioinformatics.* 2017;18(Suppl 5):119.

214. Platt CD, Zaman F, Bainter W, et al. Efficacy and economics of targeted panel versus whole-exome sequencing in 878 patients with suspected primary immunodeficiency. *J Allergy Clin Immunol.* 2021;147(2):723-726.

215. Saudi Mendeliome G. Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases. *Genome Biol.* 2015;16:134.

216. Dillon OJ, Lunke S, Stark Z, et al. Exome sequencing has higher diagnostic yield compared to simulated disease-specific panels in children with suspected monogenic disorders. *Eur J Hum Genet.* 2018;26(5):644-651.

217. Lionel AC, Costain G, Monfared N, et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet Med.* 2018;20(4):435-443.

218. Chang YS, Huang HD, Yeh KT, Chang JG. Evaluation of whole exome sequencing by targeted gene sequencing and Sanger sequencing. *Clin Chim Acta.* 2017;471:222-232.

219. LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res.* 2009;37(13):4181-4193.

220. Guo Y, He J, Zhao S, et al. Illumina human exome genotyping array clustering and quality control. *Nat Protoc.* 2014;9(11):2643-2662.

221. Hunter-Zinck H, Shi Y, Li M, et al. Genotyping Array Design and Data Quality Control in the Million Veteran Program. *Am J Hum Genet.* 2020;106(4):535-548.

222. Zhao S, Jing W, Samuels DC, Sheng Q, Shyr Y, Guo Y. Strategies for processing and quality control of Illumina genotyping arrays. *Brief Bioinform.* 2018;19(5):765-775.

223. Yamagata K, Oda N, Kaisaki PJ, et al. Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). *Nature.* 1996;384(6608):455-458.

224. Yamagata K, Furuta H, Oda N, et al. Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). *Nature.* 1996;384(6608):458-460.

225. Vionnet N, Stoffel M, Takeda J, et al. Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. *Nature.* 1992;356(6371):721-722.

226. Horikawa Y, Iwasaki N, Hara M, et al. Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nat Genet.* 1997;17(4):384-385.

227. Malecki MT, Jhala US, Antonellis A, et al. Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. *Nat Genet.* 1999;23(3):323-328.

228. Neve B, Fernandez-Zapico ME, Ashkenazi-Katalan V, et al. Role of transcription factor KLF11 and its diabetes-associated gene variants in pancreatic beta cell function. *Proc Natl Acad Sci U S A.* 2005;102(13):4807-4812.

229. Plengvidhya N, Kooptiwut S, Songtawee N, et al. PAX4 mutations in Thais with maturity onset diabetes of the young. *J Clin Endocrinol Metab.* 2007;92(7):2821-2826.

230. Borowiec M, Liew CW, Thompson R, et al. Mutations at the BLK locus linked to maturity onset diabetes of the young and beta-cell dysfunction. *Proc Natl Acad Sci U S A.* 2009;106(34):14460-14465.

231. Shield JP. Neonatal diabetes: how research unravelling the genetic puzzle has both widened our understanding of pancreatic development whilst improving children's quality of life. *Horm Res.* 2007;67(2):77-83.

232. Delepine M, Nicolino M, Barrett T, Golamaully M, Lathrop GM, Julier C. EIF2AK3, encoding translation initiation factor 2-alpha kinase 3, is mutated in patients with Wolcott-Rallison syndrome. *Nat Genet.* 2000;25(4):406-409.

233. Barrett TG, Bundey SE, Macleod AF. Neurodegeneration and diabetes: UK nationwide study of Wolfram (DIDMOAD) syndrome. *Lancet.* 1995;346(8988):1458-1463.

234. van den Ouweland AM, Dreesen JC, Verdijk M, et al. Mutations in the vasopressin type 2 receptor gene (AVPR2) associated with nephrogenic diabetes insipidus. *Nat Genet.* 1992;2(2):99-102.

235. Chakera AJ, Steele AM, Gloyn AL, et al. Recognition and Management of Individuals With Hyperglycemia Because of a Heterozygous Glucokinase Mutation. *Diabetes Care.* 2015;38(7):1383-1392.

236. Pearson ER, Flechtner I, Njolstad PR, et al. Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *N Engl J Med.* 2006;355(5):467-477.

237. Pihoker C, Gilliam LK, Ellard S, et al. Prevalence, characteristics and clinical diagnosis of maturity onset diabetes of the young due to mutations in HNF1A, HNF4A, and glucokinase: results from the SEARCH for Diabetes in Youth. *J Clin Endocrinol Metab.* 2013;98(10):4055-4062.

238. Donath X, Saint-Martin C, Dubois-Laforgue D, et al. Next-generation sequencing identifies monogenic diabetes in 16% of patients with late adolescence/adult-onset diabetes selected on a clinical basis: a cross-sectional analysis. *BMC Med.* 2019;17(1):132.

239. Karaoglan M, Nacarkahya G. Clinical and laboratory clues of maturity-onset diabetes of the young and determination of association with molecular diagnosis. *J Diabetes.* 2021;13(2):154-163.

240. Liu Y, Xie Z, Sun X, et al. A new screening strategy and whole-exome sequencing for the early diagnosis of maturity-onset diabetes of the young. *Diabetes Metab Res Rev.* 2020:e3381.

241. Kherra S, Blouin JL, Santoni F, Schwitzgebel V. Precision medicine for monogenic diabetes: from a survey to the development of a next-generation diagnostic panel. *Swiss Med Wkly.* 2017;147:w14535.

242. Shepherd M, Shields B, Hammersley S, et al. Systematic Population Screening, Using Biomarkers and Genetic Testing, Identifies 2.5% of the U.K. Pediatric Diabetes Population With Monogenic Diabetes. *Diabetes Care.* 2016;39(11):1879-1888.

243. Kleinberger J, Maloney KA, Pollin TI, Jeng LJ. An openly available online tool for implementing the ACMG/AMP standards and guidelines for the interpretation of sequence variants. *Genet Med.* 2016;18(11):1165.

244. Stanik J, Kusekova M, Huckova M, et al. Impact of Type 2 diabetes on Glucokinase diabetes (GCK-MODY) phenotype in a Roma (Gypsy) family - case report. *Endocr Regul.* 2012;46(2):99-105.

245. Estalella I, Rica I, Perez de Nanclares G, et al. Mutations in GCK and HNF-1alpha explain the majority of cases with clinical diagnosis of MODY in Spain. *Clin Endocrinol (Oxf).* 2007;67(4):538-546.

246. Massa O, Meschi F, Cuesta-Munoz A, et al. High prevalence of glucokinase mutations in Italian children with MODY. Influence on glucose tolerance, first-phase insulin response, insulin sensitivity and BMI. *Diabetologia.* 2001;44(7):898-905.

247. Prisco F, Iafusco D, Franzese A, Sulli N, Barbetti F. MODY 2 presenting as neonatal hyperglycaemia: a need to reshape the definition of "neonatal diabetes"? *Diabetologia.* 2000;43(10):1331-1332.

248. Li X, Ting TH, Sheng H, et al. Genetic and clinical characteristics of Chinese children with Glucokinase-maturity-onset diabetes of the young (GCK-MODY). *BMC Pediatr.* 2018;18(1):101.

249. Zhou Y, Wang S, Wu J, Dong J, Liao L. MODY2 in Asia: analysis of GCK mutations and clinical characteristics. *Endocr Connect.* 2020;9(5):471-478.

250. Wang Z, Diao C, Liu Y, et al. Identification and functional analysis of GCK gene mutations in 12 Chinese families with hyperglycemia. *J Diabetes Investig.* 2019;10(4):963-971.

251. Borowiec M, Antosik K, Fendler W, et al. Novel glucokinase mutations in patients with monogenic diabetes - clinical outline of GCK-MD and potential for founder effect in Slavic population. *Clin Genet.* 2012;81(3):278-283.

252. Beer NL, Osbak KK, van de Bunt M, et al. Insights into the pathogenicity of rare missense GCK variants from the identification and functional characterization of compound heterozygous and double mutations inherited in cis. *Diabetes Care.* 2012;35(7):1482-1484.

253. Liu L, Liu Y, Ge X, et al. Insights into pathogenesis of five novel GCK mutations identified in Chinese MODY patients. *Metabolism.* 2018;89:8-17.

254. Gloyn AL, Odili S, Zelent D, et al. Insights into the structure and regulation of glucokinase from a novel mutation (V62M), which causes maturity-onset diabetes of the young. *J Biol Chem.* 2005;280(14):14105-14113.

255. Arden C, Trainer A, de la Iglesia N, et al. Cell biology assessment of glucokinase mutations V62M and G72R in pancreatic beta-cells: evidence for cellular instability of catalytic activity. *Diabetes.* 2007;56(7):1773-1782.

256. Ellard S, Beards F, Allen LI, et al. A high prevalence of glucokinase mutations in gestational diabetic subjects selected by clinical criteria. *Diabetologia.* 2000;43(2):250-253.

257. Hwang JS, Shin CH, Yang SW, Jung SY, Huh N. Genetic and clinical characteristics of Korean maturity-onset diabetes of the young (MODY) patients. *Diabetes Res Clin Pract.* 2006;74(1):75-81.

258. Sagen JV, Bjorkhaug L, Molnes J, et al. Diagnostic screening of MODY2/GCK mutations in the Norwegian MODY Registry. *Pediatr Diabetes.* 2008;9(5):442-449.

259. Thomson KL, Gloyn AL, Colclough K, et al. Identification of 21 novel glucokinase (GCK) mutations in UK and European Caucasians with maturity-onset diabetes of the young (MODY). *Hum Mutat.* 2003;22(5):417.

260. Vits L, Beckers D, Craen M, et al. Identification of novel and recurrent glucokinase mutations in Belgian and Luxembourg maturity onset diabetes of the young patients. *Clin Genet.* 2006;70(4):355-359.

261. Velho G, Blanche H, Vaxillaire M, et al. Identification of 14 new glucokinase mutations and description of the clinical profile of 42 MODY-2 families. *Diabetologia.* 1997;40(2):217-224.

262. Henderson M, Levy E, Delvin E, Losekoot M, Lambert M. Prevalence and clinical phenotype of the p.Val226Met glucokinase gene mutation in French Canadians in Quebec, Canada. *Mol Genet Metab.* 2007;90(1):87-92.

263. Feigerlova E, Pruhova S, Dittertova L, et al. Aetiological heterogeneity of asymptomatic hyperglycaemia in children and adolescents. *Eur J Pediatr.* 2006;165(7):446-452.

264. Delvecchio M, Ludovico O, Bellacchio E, et al. MODY type 2 P59S GCK mutant: founder effect in South of Italy. *Clin Genet.* 2013;83(1):83-87.

265. Alkorta-Aranburu G, Carmody D, Cheng YW, et al. Phenotypic heterogeneity in monogenic diabetes: the clinical and diagnostic utility of a gene panel-based next-generation sequencing approach. *Mol Genet Metab.* 2014;113(4):315-320.

266. Bennett JT, Vasta V, Zhang M, Narayanan J, Gerrits P, Hahn SH. Molecular genetic testing of patients with monogenic diabetes and hyperinsulinism. *Mol Genet Metab.* 2015;114(3):451-458.

267. Brahm AJ, Wang G, Wang J, et al. Genetic Confirmation Rate in Clinically Suspected Maturity-Onset Diabetes of the Young. *Can J Diabetes.* 2016;40(6):555-560.

268. Costantini S, Malerba G, Contreas G, et al. Genetic and bioinformatics analysis of four novel GCK missense variants detected in Caucasian families with GCK-MODY phenotype. *Clin Genet.* 2015;87(5):440-447.

269. Milenkovic T, Zdravkovic D, Mitrovic K. [Novel glucokinase mutation in a boy with maturity-onset diabetes of the young]. *Srp Arh Celok Lek.* 2008;136(9-10):542-544.

270. Pinterova D, Ek J, Kolostova K, et al. Six novel mutations in the GCK gene in MODY patients. *Clin Genet.* 2007;71(1):95-96.

271. Cao H, Shorey S, Robinson J, et al. GCK and HNF1A mutations in Canadian families with maturity onset diabetes of the young (MODY). *Hum Mutat.* 2002;20(6):478-479.

272. Sagen JV, Odili S, Bjorkhaug L, et al. From clinicogenetic studies of maturity-onset diabetes of the young to unraveling complex mechanisms of glucokinase regulation. *Diabetes.* 2006;55(6):1713-1722.

273. Pruhova S, Dusatkova P, Sumnik Z, et al. Glucokinase diabetes in 103 families from a country-based study in the Czech Republic: geographically restricted distribution of two prevalent GCK mutations. *Pediatr Diabetes.* 2010;11(8):529-535.

274. Yamada S, Nishigori H, Onda H, et al. Identification of mutations in the hepatocyte nuclear factor (HNF)-1 alpha gene in Japanese subjects with IDDM. *Diabetes.* 1997;46(10):1643-1647.

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

276. Cappelli A, Tumini S, Consoli A, et al. Novel mutations in GCK and HNF1A genes in Italian families with MODY phenotype. *Diabetes Res Clin Pract.* 2009;83(3):e72-74.

277. Kaisaki PJ, Menzel S, Lindner T, et al. Mutations in the hepatocyte nuclear factor-1alpha gene in MODY and early-onset NIDDM: evidence for a mutational hotspot in exon 4. *Diabetes.* 1997;46(3):528-535.

278. Frayling TM, Bulamn MP, Ellard S, et al. Mutations in the hepatocyte nuclear factor-1alpha gene are a common cause of maturity-onset diabetes of the young in the U.K. *Diabetes.* 1997;46(4):720-725.

279. Frayling TM, Bulman MP, Appleton M, Hattersley AT, Ellard S. A rapid screening method for hepatocyte nuclear factor 1 alpha frameshift mutations; prevalence in maturity-onset diabetes of the young and late-onset non-insulin dependent diabetes. *Hum Genet.* 1997;101(3):351-354.

280. Delvecchio M, Ludovico O, Menzaghi C, et al. Low prevalence of HNF1A mutations after molecular screening of multiple MODY genes in 58 Italian families recruited in the pediatric or adult diabetes clinic from a single Italian hospital. *Diabetes Care.* 2014;37(12):e258-260.

281. Joshi R, Phatarpekar A. Neonatal diabetes mellitus due to L233F mutation in the KCNJ11 gene. *World J Pediatr.* 2011;7(4):371-372.

282. Mory PB, Crispim F, Freire MB, et al. Phenotypic diversity in patients with lipodystrophy associated with LMNA mutations. *Eur J Endocrinol.* 2012;167(3):423-431.

283. Thong KM, Xu Y, Cook J, et al. Cosegregation of focal segmental glomerulosclerosis in a family with familial partial lipodystrophy due to a mutation in LMNA. *Nephron Clin Pract.* 2013;124(1-2):31-37.

284. Le Dour C, Wu W, Bereziat V, Capeau J, Vigouroux C, Worman HJ. Extracellular matrix remodeling and transforming growth factor-beta signaling abnormalities induced by lamin A/C variants that cause lipodystrophy. *J Lipid Res.* 2017;58(1):151-163.

285. Vollbach H, Brandt S, Lahr G, et al. Prevalence and phenotypic characterization of MC4R variants in a large pediatric cohort. *Int J Obes (Lond).* 2017;41(1):13-22.

286. Ustaoglu M, Onder F, Karapapak M, Taslidere H, Guven D. Ophthalmic, systemic, and genetic characteristics of patients with Wolfram syndrome. *Eur J Ophthalmol.* 2020;30(5):1099-1105.

287. Lieber DS, Vafai SB, Horton LC, et al. Atypical case of Wolfram syndrome revealed through targeted exome sequencing in a patient with suspected mitochondrial disease. *BMC Med Genet.* 2012;13:3.

288. Chaussenot A, Rouzier C, Quere M, et al. Mutation update and uncommon phenotypes in a French cohort of 96 patients with WFS1-related disorders. *Clin Genet.* 2015;87(5):430-439.

289. Riddle MC, Philipson LH, Rich SS, et al. Monogenic Diabetes: From Genetic Insights to Population-Based Precision in Care. Reflections From a Diabetes Care Editors' Expert Forum. *Diabetes Care.* 2020;43(12):3117-3128.

290. Liu X, Han S, Wang Z, Gelernter J, Yang BZ. Variant callers for next-generation sequencing data: a comparison study. *PLoS One.* 2013;8(9):e75619.

291. Wai HA, Lord J, Lyon M, et al. Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genet Med.* 2020.

292. Xiao B, Qiu W, Ji X, et al. Marked yield of re-evaluating phenotype and exome/target sequencing data in 33 individuals with intellectual disabilities. *Am J Med Genet A.* 2018;176(1):107-115.

293. Won D, Kim SH, Kim B, Lee ST, Kang HC, Choi JR. Reanalysis of Genomic Sequencing Results in a Clinical Laboratory: Advantages and Limitations. *Front Neurol.* 2020;11:612.

294. James KN, Clark MM, Camp B, et al. Partially automated whole-genome sequencing reanalysis of previously undiagnosed pediatric patients can efficiently yield new diagnoses. *NPJ Genom Med.* 2020;5:33.

295. Wenger AM, Guturu H, Bernstein JA, Bejerano G. Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. *Genet Med.* 2017;19(2):209-214.

296. Al-Nabhani M, Al-Rashdi S, Al-Murshedi F, et al. Reanalysis of exome sequencing data of intellectual disability samples: Yields and benefits. *Clin Genet.* 2018;94(6):495-501.

297. Costain G, Jobling R, Walker S, et al. Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing. *Eur J Hum Genet.* 2018;26(5):740-744.

298. Ewans LJ, Schofield D, Shrestha R, et al. Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders. *Genet Med.* 2018;20(12):1564-1574.

299. Nambot S, Thevenon J, Kuentz P, et al. Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: substantial interest of prospective annual reanalysis. *Genet Med.* 2018;20(6):645-654.

300. Wright CF, McRae JF, Clayton S, et al. Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genet Med.* 2018;20(10):1216-1223.

301. Liu P, Meng L, Normand EA, et al. Reanalysis of Clinical Exome Sequencing Data. *N Engl J Med.* 2019;380(25):2478-2480.

302. Baker SW, Murrell JR, Nesbitt AI, et al. Automated Clinical Exome Reanalysis Reveals Novel Diagnoses. *J Mol Diagn.* 2019;21(1):38-48.

303. Sun Y, Xiang J, Liu Y, et al. Increased diagnostic yield by reanalysis of data from a hearing loss gene panel. *BMC Med Genomics.* 2019;12(1):76.

304. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research.* 2010;38(16):e164-e164.

305. Type 2 Diabetes Knowledge Portal. http://www.type2diabetesgenetics.org/informational/data. Accessed.

306. Zhang H, Maloney KA, Barbetti F, et al. 1453-P: Adaption of the ACMG/AMP Variant Interpretation Guidelines for GCK, HNF1A, HNF4A-MODY: Recommendations from the ClinGen Monogenic Diabetes Expert Panel. In: Am Diabetes Assoc; 2020.

307. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell.* 2019;176(3):535-548 e524.

308. Team RC. R: A language and environment for statistical computing. In: Vienna, Austria; 2013.

309. JOHANSSON BB, PAVITHRAM A, ZHANG H, et al. 1638-P: Functional Characterization of HNF1B Variants Can Enhance Diabetes Precision Medicine. *Diabetes.* 2020;69(Supplement 1):1638-P.

310. Leman R, Gaildrat P, Le Gac G, et al. Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico/in vitro studies: an international collaborative effort. *Nucleic Acids Research.* 2018;46(15):7913-7923.

311. Hwang S, Kim E, Lee I, Marcotte EM. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci Rep.* 2015;5:17875.

312. Tiulpakov A, Zubkova N, Makretskaya N, et al. Minigene splicing assessment of 20 novel synonymous and intronic glucokinase gene variants identified in patients with maturity-onset diabetes of the young. *Human Mutation.* 2020;41(1):129-132.

313. Glaser B, Blech I, Krakinovsky Y, et al. ABCC8 mutation allele frequency in the Ashkenazi Jewish population and risk of focal hyperinsulinemic hypoglycemia. *Genet Med.* 2011;13(10):891-894.

314. Thomas PM, Cote GJ, Wohllk N, et al. Mutations in the sulfonylurea receptor gene in familial persistent hyperinsulinemic hypoglycemia of infancy. *Science.* 1995;268(5209):426-429.

315. Tarasov AI, Nicolson TJ, Riveline JP, et al. A rare mutation in ABCC8/SUR1 leading to altered ATP-sensitive K+ channel activity and beta-cell glucose sensing is associated with type 2 diabetes in adults. *Diabetes.* 2008;57(6):1595-1604.

316. Riveline JP, Rousseau E, Reznik Y, et al. Clinical and metabolic features of adult-onset diabetes caused by ABCC8 mutations. *Diabetes Care.* 2012;35(2):248-251.

317. Sokolova EA, Bondar IA, Shabelnikova OY, Pyankova OV, Filipenko ML. Replication of KCNJ11 (p.E23K) and ABCC8 (p.S1369A) Association in Russian Diabetes Mellitus 2 Type Cohort and Meta-Analysis. *PLoS One.* 2015;10(5):e0124662.

318. Schwanstecher C, Meyer U, Schwanstecher M. K(IR)6.2 polymorphism predisposes to type 2 diabetes by inducing overactivity of pancreatic beta-cell ATP-sensitive K(+) channels. *Diabetes.* 2002;51(3):875-879.

319. Hamming KSC, Soliman D, Matemisz LC, et al. Coexpression of the type 2 diabetes susceptibility gene variants KCNJ11 E23K and ABCC8 S1369A alter the ATP and sulfonylurea sensitivities of the ATP-sensitive K(+) channel. *Diabetes.* 2009;58(10):2419-2424.

320. LaDuca H, Farwell KD, Vuong H, et al. Exome sequencing covers >98% of mutations identified on targeted next generation sequencing panels. *PLoS One.* 2017;12(2):e0170843.

321. Ankala A, da Silva C, Gualandi F, et al. A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield. *Ann Neurol.* 2015;77(2):206-214.

322. Consugar MB, Navarro-Gomez D, Place EM, et al. Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet Med.* 2015;17(4):253-261.

323. Szopa M, Ludwig-Galezowska A, Radkowski P, et al. Genetic testing for monogenic diabetes using targeted next-generation sequencing in patients with maturity-onset diabetes of the young. *Pol Arch Med Wewn.* 2015;125(11):845-851.

324. Irgens HU, Molnes J, Johansson BB, et al. Prevalence of monogenic diabetes in the population-based Norwegian Childhood Diabetes Registry. *Diabetologia.* 2013;56(7):1512-1519.

325. Park SS, Jang SS, Ahn CH, et al. Identifying Pathogenic Variants of Monogenic Diabetes Using Targeted Panel Sequencing in an East Asian Population. *J Clin Endocrinol Metab.* 2019.

326. Lin DC, Huang CY, Ting WH, et al. Mutations in glucokinase and other genes detected in neonatal and type 1B diabetes patient using whole exome sequencing may lead to disease-causing changes in protein activity. *Biochim Biophys Acta Mol Basis Dis.* 2019;1865(2):428-433.

327. Glotov OS, Serebryakova EA, Turkunova ME, et al. Wholeexome sequencing in Russian children with nontype 1 diabetes mellitus reveals a wide spectrum of genetic variants in MODYrelated and unrelated genes. *Mol Med Rep.* 2019;20(6):4905-4914.

328. Bonnefond A, Durand E, Sand O, et al. Molecular diagnosis of neonatal diabetes mellitus using next-generation sequencing of the whole exome. *PLoS One.* 2010;5(10):e13630.

329. Dusatkova P, Fang M, Pruhova S, et al. Lessons from whole-exome sequencing in MODYX families. *Diabetes Res Clin Pract.* 2014;104(3):e72-74.

330. Today Study Group, Zeitler P, Epstein L, et al. Treatment options for type 2 diabetes in adolescents and youth: a study of the comparative efficacy of metformin alone or in combination with rosiglitazone or lifestyle intervention in adolescents with type 2 diabetes. *Pediatr Diabetes.* 2007;8(2):74-87.

331. Search Study Group. SEARCH for Diabetes in Youth: a multicenter study of the prevalence, incidence and classification of diabetes mellitus in youth. *Control Clin Trials.* 2004;25(5):458-471.

332. Hamman RF, Bell RA, Dabelea D, et al. The SEARCH for Diabetes in Youth study: rationale, findings, and future directions. *Diabetes Care.* 2014;37(12):3336-3344.

333. Srinivasan S, Chen L, Todd J, et al. The First Genome-Wide Association Study for Type 2 Diabetes in Youth: The Progress in Diabetes Genetics in Youth (ProDiGY) Consortium. *Diabetes.* 2021;70(4):996-1005.

334. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.

335. Regalado A. More than 26 million people have taken an at-home ancestry test. *MIT Technology Reivew.* 2019.

336. Tandy-Connor S, Guiltinan J, Krempely K, et al. False-positive results released by direct-to-consumer genetic tests highlight the importance of clinical confirmation testing for appropriate patient care. *Genet Med.* 2018;20(12):1515-1521.

337. Ekins R, Chu FW. Microarrays: their origins and applications. *Trends Biotechnol.* 1999;17(6):217-218.

338. Wright CF, West B, Tuke M, et al. Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. *Am J Hum Genet.* 2019;104(2):275-286.

339. MN W, L J, JW H, et al. Use of SNP chips to detect rare pathogenic variants: retrospective, population based diagnostic evaluation. *BMJ.* 2021;372:n214.

340. Stewart KFJ, Wesselius A, Schreurs MAC, Schols A, Zeegers MP. Behavioural changes, sharing behaviour and psychological responses after receiving direct-to-consumer genetic test results: a systematic review and meta-analysis. *J Community Genet.* 2018;9(1):1-18.

341. Roberts JS, Gornick MC, Carere DA, Uhlmann WR, Ruffin MT, Green RC. Direct-to-Consumer Genetic Testing: User Motivations, Decision Making, and Perceived Utility of Results. *Public Health Genomics.* 2017;20(1):36-45.

342. Roberts JS, Ostergren J. Direct-to-Consumer Genetic Testing and Personal Genomics Services: A Review of Recent Empirical Studies. *Curr Genet Med Rep.* 2013;1(3):182-200.

343. Weitzel KW, Alexander M, Bernhardt BA, et al. The IGNITE network: a model for genomic medicine implementation and research. *BMC Med Genomics.* 2016;9:1.

344. Greshake B, Bayer PE, Rausch H, Reda J. openSNP--a crowdsourced web resource for personal genomics. *PLoS One.* 2014;9(3):e89204.

345. Beachy SH, Alper J, Addie S, Hackmann M. Exploring the Current Landscape of Consumer Genomics: Proceedings of a Workshop. 2020.

346. Sanderson SC, Brothers KB, Mercaldo ND, et al. Public Attitudes toward Consent and Data Sharing in Biobank Research: A Large Multi-site Experimental Survey in the US. *Am J Hum Genet.* 2017;100(3):414-427.

347. Genomes Project C, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010;467(7319):1061-1073.

348. Conrad DF, Jakobsson M, Coop G, et al. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet.* 2006;38(11):1251-1260.

349. Harris DN, Song W, Shetty AC, et al. Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. *Proc Natl Acad Sci U S A.* 2018;115(28):E6526-E6535.

350. Moscarello T, Murray B, Reuter CM, Demo E. Direct-to-consumer raw genetic data and third-party interpretation services: more burden than bargain? *Genet Med.* 2019;21(3):539-541.

351. Wynn J, Chung WK. 23andMe Paves the Way for Direct-to-Consumer Genetic Health Risk Tests of Limited Clinical Utility. *Ann Intern Med.* 2017;167(2):125-126.

352. Turrini M, Prainsack B. Beyond clinical utility: The multiple values of DTC genetics. *Appl Transl Genom.* 2016;8:4-8.

353. Gandica RG, Chung WK, Deng L, Goland R, Gallagher MP. Identifying monogenic diabetes in a pediatric cohort with presumed type 1 diabetes. *Pediatr Diabetes.* 2015;16(3):227-233.

354. Xu A, Lin Y, Sheng H, et al. Molecular diagnosis of maturity-onset diabetes of the young in a cohort of Chinese children. *Pediatr Diabetes.* 2020;21(3):431-440.

355. Delvecchio M, Mozzillo E, Salzano G, et al. Monogenic Diabetes Accounts for 6.3% of Cases Referred to 15 Italian Pediatric Diabetes Centers During 2007 to 2012. *J Clin Endocrinol Metab.* 2017;102(6):1826-1834.

356. Hameed S, Ellard S, Woodhead HJ, et al. Persistently autoantibody negative (PAN) type 1 diabetes mellitus in children. *Pediatr Diabetes.* 2011;12(3 Pt 1):142-149.

357. Yaghootkar H, Abbasi F, Ghaemi N, et al. Type 1 diabetes genetic risk score discriminates between monogenic and Type 1 diabetes in children diagnosed at the age of <5 years in the Iranian population. *Diabet Med.* 2019;36(12):1694-1702.

358. Santana LS, Caetano LA, Costa-Riquetto AD, et al. Clinical application of ACMG-AMP guidelines in HNF1A and GCK variants in a cohort of MODY families. *Clin Genet.* 2017;92(4):388-396.

359. Chambers C, Fouts A, Dong F, et al. Characteristics of maturity onset diabetes of the young in a large diabetes center. *Pediatr Diabetes.* 2016;17(5):360-367.

360. Johnson SR, Ellis JJ, Leo PJ, et al. Comprehensive genetic screening: The prevalence of maturity-onset diabetes of the young gene variants in a population-based childhood diabetes cohort. *Pediatr Diabetes.* 2019;20(1):57-64.

361. Davis TM, Makepeace AE, Ellard S, et al. The prevalence of monogenic diabetes in Australia: the Fremantle Diabetes Study Phase II. *Med J Aust.* 2017;207(8):344-347.

362. Bansal V, Gassenhuber J, Phillips T, et al. Spectrum of mutations in monogenic diabetes genes identified from high-throughput DNA sequencing of 6888 individuals. *BMC Med.* 2017;15(1):213.

363. Fendler W, Borowiec M, Baranowska-Jazwiecka A, et al. Prevalence of monogenic diabetes amongst Polish children after a nationwide genetic screening campaign. *Diabetologia.* 2012;55(10):2631-2635.

364. Malachowska B, Borowiec M, Antosik K, et al. Monogenic diabetes prevalence among Polish children-Summary of 11 years-long nationwide genetic screening program. *Pediatr Diabetes.* 2018;19(1):53-58.

365. Yorifuji T, Fujimaru R, Hosokawa Y, et al. Comprehensive molecular analysis of Japanese patients with pediatric-onset MODY-type diabetes mellitus. *Pediatr Diabetes.* 2012;13(1):26-32.