Jeffrey Kleinberger
Ph.D. Candidate
Molecular Medicine Program, Genome Biology Track
Graduate Program in Life Sciences, University of Maryland, Baltimore

Contact Information

Medical School Teaching Facility 357
685 West Baltimore St.
Baltimore, MD 21201
jeffrey.kleinberger@som.umaryland.edu

Education

Juniata College
Huntingdon, PA
August 2004 – May 2008

- Cum Laude distinction (GPA: 3.701)
- Biochemistry major
- Coursework included: Organic Chemistry, Bioinorganic Chemistry, Analytical Chemistry, Physical Chemistry, Chemistry Research, Biology, Cell Biology, Genetics, Biostatistics, Molecular Techniques, Physiology, Microbiology, Biochemistry and Molecular Biology (I-III), Calculus (I&II), Physics (I&II), Scientific Glassblowing, College Writing Seminar, Art of Public Speaking, Intro to Psychology, Learning & Conditioning, Mod. Knowledge & The Self, United States to 1877, The American Revolution, Survey of Western Art, Modern Architecture, Intro to Business, and Spanish (I&II).
- Four-year starter for the rugby team (Vice President: 2005-2006)

University of Maryland School of Medicine
Baltimore, MD
August 2011-May 2013, August 2017-present

- Medical School

July 2013-July2017

- Ph.D. in Molecular Medicine, Genome Biology Track
  Discovery and Analysis of Patients with Monogenic Diabetes in Multiple Cohorts to Guide Future Diagnosis
  Defended July 2017

Employment Experience

Pharmacy Technician Assistant
Bradford Pharmacy
Bradford, PA
June 2005 – August 2005

- Learned the functions, mechanisms, and side-effects of commonly prescribed medications.
- Observed pharmacist consultations with customers.

Summer Undergraduate Research Fellow

Juniata College
Huntingdon, PA
June 2006 – August 2006

- Enhanced the protocol for honey bee venom protein separation through size-exclusion chromatography.
- Analyzed protein concentration of honey bee venom using SELDI-TOF mass spectrometry.
- Examined honey bee venom purity by using reverse phase and ion-exchange HPLC.

General Chemistry Laboratory Assistant
Juniata College
Huntingdon, PA
January 2007 – May 2007

- Monitored student laboratory activities to ensure safety and appropriate laboratory behavior.
- Taught proper pipetting, titrating and pH measurement techniques in order to examine water hardness.

Genetics Laboratory Assistant
Juniata College
Huntingdon, PA
January 2008 – May 2008

- Aided in students' experiments involving PCR and genetic crosses of plants.
- Graded laboratory coursework and notebooks.

Research Technician
University of Pittsburgh School of Medicine Division of Endocrinology
Pittsburgh, PA
May 2008 – July 2011

- Studied cell-cycle molecules of beta cells of both human and rodent origin using adenoviral infection, lentiviral infection, RT-PCR, western blot, immunohistological staining, and FACS analysis.
- Developed protocols for fluorescent immunocytology in human beta cells of cell cycle molecules, including cyclin D1, cyclin D2, cyclin D3, p21, p57, p15, p16, p18, p19, p107, and p130.
- Maintained colonies of multiple strains of transgenic mice. Performed experiments including: PCRs for genetic verification, retro-orbital bleeds, islet isolations, islet transplantations, and nephrectomies.
- Aided in the preparation and purification of adenoviruses.

Volunteering/Internships

Microbiology Summer Internship
Bradford Regional Medical Center Microbiology Laboratory
Bradford, PA
June 2007 – August 2007

- Observed and aided in diagnosis of bacterial pathogens.
- Cultured bacterial pathogens using multiple forms of media and environment.
- Stained and microscopically examined bacterial pathogens.

- Performed bacterial identification and susceptibility testing using automated analyzing techniques.

Health Education Seminar Volunteer
Gallagher Mansion
Baltimore, MD
Spring 2012

- Taught seminars focusing on geriatric health issues to residents of a housing facility for low-income senior citizens
- Played games, held conversations, and spent time with elderly residents

Math Tutor
Franklin Square Elementary
Baltimore, MD
Spring 2013

- Tutored students in mathematics in preparation for the Maryland State Assessment exam

Publications

1. Fiaschi-Taesch NM, Salim F, **Kleinberger J**, Troxell R, Cozar-Castellano I, Selk K, Cherok E, Takane KK, Scott DK, and Stewart AF. Induction of Human Beta Cell Proliferation and Engraftment Using a Single G1/S Regulatory Molecule, Cdk6. *Diabetes.* 2010 Aug; 59(8): 1926-36.
2. Karslioglu E, **Kleinberger JW**, Salim FG, Cox AE, Takane KK, Scott DK, Stewart AF. cMyc is a principal upstream driver of beta-cell proliferation in rat insulinoma cell lines and is an effective mediator of human beta-cell replication. *Mol Endocrinol.* 2011 Oct; 25(10):1760-72.
3. Takane KK, **Kleinberger JW**, Salim FG, Fiaschi-Taesch NM, Stewart AF. Regulated and reversible induction of adult human β-cell replication. *Diabetes.* 2012 Feb: 61(2):418-24.
4. Fiaschi-Taesch NM, **Kleinberger JW**, Salim FG, Troxell R, Wills R, Tanwir M, Casinelli G, Cox AE, Takane KK, Scott DK, Stewart AF. Human pancreatic β-cell G1/S molecule cell cycle atlas. *Diabetes*. 2013 Jul;62(7):2450-9.
5. Fiaschi-Taesch NM, **Kleinberger JW**, Salim FG, Troxell R, Wills R, Tanwir M, Casinelli G, Cox AE, Takane KK, Srinivas H, Scott DK, Stewart AF. Cytoplasmic-nuclear trafficking of G1/S cell cycle molecules and adult human β-cell replication: a revised model of human β-cell G1/S control. *Diabetes*. 2013 Jul;62(7):2460-70.
6. **Kleinberger JW**, Pollin TI. Personalized medicine in diabetes mellitus: current opportunities and future prospects. *Ann N Y Acad Sci*. 2015 Jun;1346(1):45-56.
7. Chen H, **Kleinberger JW**, Takane KK, Salim F, Fiaschi-Taesch N, Pappas K, Parsons R, Jiang J, Zhang Y, Liu H, Wang P, Bender AS, Frank SJ, Stewart AF. Augmented Stat5 Signaling Bypasses Multiple Impediments to Lactogen-Mediated Proliferation in Human β-Cells. *Diabetes*. 2015 Nov;64(11):3784-97.
8. **Kleinberger JW**, Pollin TI. Undiagnosed MODY: Time for Action. *Curr Diab Rep*. 2015 Dec;15(12):110.

9. **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 Nov;18(11):1165.
10. **Kleinberger JW**, Maloney KA, Pollin TI. The Genetic Architecture of Diabetes in Pregnancy: Implications for Clinical Practice. *Am J Perinatol*. 2016 Nov;33(13):1319-1326.
11. **Kleinberger JW**, Copeland KC, Gandica RG, Haymond MW, Levitsky LL, Linder B, Shuldiner AR, Tollefsen S, White NH, and Pollin TI. Monogenic Diabetes in Overweight and Obese Youth Diagnosed with Type 2 Diabetes: The TODAY Clinical Trial. *Genetics in Medicine.* In press.

Book Chapters

1. **Kleinberger JW**, Silver KD, and Shuldiner AR. "Genetics of Type 2 Diabetes: From Candidate Genes to Genome-Wide Association Analysis." *Principles of Diabetes Mellitus.* Ed. Poretsky, L. Springer International Publishing 2017. 978-3-319-18742-6.
2. **Kleinberger JW** and Pollin TI. "Sequencing Approaches to Type 2 Diabetes." The Genetics of Type 2 Diabetes and Related Traits: Biology, Physiology and Translation. Ed. Florez, JC. Springer International Publishing 2016. 978-3-319-01573-6.

Abstracts

1. **Kleinberger J**, Ross C, and Mulfinger LM. Quantitative mass spectrometry of honey bee venom proteins. 233rd American Chemical Society National Meeting. Division of Chemical Education Undergraduate Poster Session. March 2007; Chicago, IL.
2. Eckman T, Gamber A, Gonzalez K, **Kleinberger J**, Lau B, Loveitt A, and Buonaccorsi V. Sensitivity to Amplified Fragment Length Polymorphism (AFLP) Genomic Screen to Quality and Quantity of DNA Template and Primer. Juniata College Liberal Arts Symposium. Competitive Poster Session. 2007; Huntingdon, PA.
3. **Kleinberger J**, Ross C, and Mulfinger LM. Quantitative mass spectrometry of honey bee venom proteins. Juniata College Liberal Arts Symposium Competitive Poster Session. 2007; Huntingdon, PA.
4. Cozar-Castellano I, **Kleinberger J**, Selk K, Cherok E, Stewart AF. Combined loss of the cell cycle inhibitors p21cip1 and p27kip1 reveals that both are dispensable for maintaining adult beta cell mass and function. European Association for the Study of Diabetes 45th Conference. September 2009; Vienna, Austria.
5. Karslioglu E, Takane KK, Harb G, **Kleinberger J**, Salim F, Cozar-Castellano I, Garcia-Ocana A, Scott D, and Stewart AF. Comprehensive Upstream Regulatory Exploration Reveals that cMyc is a Key Regulator of Proliferation in Both Commonly Employed Rat Insulinoma Cell Lines. 92nd Endocrine Society Meeting and Expo. June 2010; San Diego, CA.
6. Takane KK, **Kleinberger J**, Salim F, Thomas S, Fiaschi-Taesch NM, Stewart AF. Regulated Induction of Human Beta Cell Replication: Tetracycline-Inducible

Overexpression of Cdk6 and Cyclin D1 in Human Beta Cells. American Diabetes Association 70th Annual Scientific Sessions. Oral presentation. June 2010; Orlando, FL.

7. Cozar-Castellano I, **Kleinberger J**, Selk K, Cherok E, Stewart AF. Combined Genetic Loss of the Cell Cycle Inhibitors, p21cip1 and p27kip1, Reveals That Both Are Dispensable for Adult Beta Cell Proliferation, Mass and Function. American Diabetes Association 70th Annual Scientific Sessions. Oral presentation. June 2010; Orlando, FL.
8. Fiaschi-Taesch NM, Salim F, **Kleinberger J**, Cozar-Castellano I, Selk K, Cherok E, Takane KK, Scott D, and Stewart AF. Induction of Human Beta Cell Proliferation and Engraftment Using A Single G1/S Regulatory Molecule, Cdk6. American Diabetes Association 70th Annual Scientific Sessions. June 2010; Orlando, FL.
9. Fiaschi-Taesch NM, Salim F, **Kleinberger J**, Troxell R, Cox A, Takane KK, Stewart AF. Cytoplasmic-Nuclear Trafficking of G1/S Cell Cycle Molecules: A Critical Regulator of Adult Human β Cell Replication. American Diabetes Association 71st Annual Scientific Sessions. June 2011; San Diego, CA.
10. Karslioglu E, **Kleinberger J**, Salim F, Cox A, Takane KK, Scott DK, Stewart AF. Physiological Expression of cMyc is Responsible for Proliferation in Rodent β Cell Lines. American Diabetes Association 71st Annual Scientific Sessions. June 2011; San Diego, CA.
11. **Kleinberger JW** et al. The Personalized Diabetes Medicine Program. University of Maryland Graduate Research Conference. Baltimore, MD. March 9, 2015.
12. **Kleinberger JW** et al. Implementation of genetic sequencing into clinical practice: the Personalized Diabetes Medicine Program. University of Maryland Graduate Research Conference. Baltimore, MD. March 23, 2016.
13. **Kleinberger JW** et al. Implementation of genetic sequencing into clinical practice: the Personalized Diabetes Medicine Program. Molecular Medicine Retreat. Baltimore, MD. October 6, 2016.
14. Jeng L, **Kleinberger J**. An online tool for applying the ABMGG laboratory training program milestones. American Society of Human Genetics Annual Meeting. Vancouver, CN. October 19, 2016.

Abstract

Title of Dissertation: Discovery and Analysis of Patients with Monogenic Diabetes in Multiple Cohorts to Guide Future Diagnosis

Jeffrey W. Kleinberger, Doctor of Philosophy, 2017

Dissertation directed by:

Toni I. Pollin, M.S., Ph.D., 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 hyperglycemia caused by a variant in a single gene, and it accounts for approximately 1-2% of all diabetes cases. A genetic diagnosis of monogenic diabetes is important because the most common gene etiologies can be effectively managed with treatment regimens other than first line treatments for either type 1 (T1D) or type 2 diabetes (T2D). However, monogenic diabetes can have a similar clinical presentation to either T1D or T2D, leading to clinical misdiagnosis of monogenic diabetes. The goal of this dissertation was to evaluate approaches for identifying patients with monogenic diabetes and a method for functionally testing monogenic diabetes variants to potentially improve diagnosis and treatment of these patients. Monogenic diabetes genetic testing was performed on 488 samples from the Treatment Options for type 2 Diabetes in Adolescents and Youth (TODAY) clinical trial. A total of 4.5% (22/488) of individuals were determined to have pathogenic or likely pathogenic variants. Comparison of clinical characteristics of patients with and without monogenic diabetes discovered statistically, but not clinically, significant lower BMI Z-score, higher fasting glucose, and lower fasting insulin in patients with monogenic diabetes. Treatment

outcomes from the TODAY trial showed that most patients with *HNF4A* monogenic diabetes variants failed treatment therapies rapidly, while none of the patients with *GCK* monogenic diabetes variants failed treatment. In the Personalized Diabetes Medicine Program (PDMP), an implementation study for screening, diagnosis, and return of results for monogenic diabetes, 1,734 participants were screened for monogenic diabetes at four diverse study sites. Of the 138 eligible participants that underwent monogenic diabetes genetic testing, 14 had pathogenic or likely pathogenic monogenic diabetes variants. PDMP patients with monogenic diabetes had a diverse range of ages, races/ethnicities, and previous treatment regimens. Finally, a zebrafish model of *hnf1a*-knockdown and rescue with *HNF1A* monogenic diabetes variants was evaluated to determine that the model could not accurately identify established damaging *HNF1A* genetic variants. The results from these studies have demonstrated the variable presentations of patients with monogenic diabetes as well as the challenges and potential of assessing the function of *HNF1A* variants using an *in vivo* model.

Discovery and Analysis of Patients with Monogenic Diabetes in Multiple Cohorts to Guide Future Diagnosis

by
Jeffrey W. Kleinberger

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
2017

© Copyright 2017 by Jeffrey W. Kleinberger
All rights reserved

Acknowledgements

There are many people that need to be thanked for the completion of this dissertation. First, I would like to thank my mentor, Dr. Toni Pollin, for giving me instruction, guidance, and support during my training. She graciously "adopted" me as a graduate student, despite me being a member of a graduate program different than the one she leads. She has been quick to offer many outstanding opportunities to me, whether it be writing book chapters or traveling for study consortium meetings. I am truly grateful for everything she has allowed me to experience during my graduate training. Beyond giving me so much of her time, effort, and expertise, she has also served as an incredible role model and scientist. She has been an ideal example of how to drive research forward while simultaneously being so considerate of her research teammates, her trainees, and the patients in her studies. I thank her for all she has taught me and given me, and I hope to make her proud with my own research career.

The research of this dissertation would not have been possible without all the people who have played a role in the studies. This dissertation is an encapsulation of team science, and I am extremely grateful for everyone who has given time and effort to move these studies forward. These include all the study leadership, coordinators, and staff of both the TODAY study and the PDMP. I also need to thank the leadership and personnel of the University of Maryland Translational Genomics Lab, where the sequencing was performed. Finally, it is important to thank all of those who went out of their way to help train me in the methods of these studies, from computational analysis to molecular techniques.

I would additionally like to thank the members of my Dissertation Committee for their support and guidance: Dr. Simeon Taylor, Dr. Braxton Mitchell, Dr. Norann Zaghloul, Dr. Magali Fontaine, and Dr. Marc Reitman. It is also important to thank the Molecular Medicine Graduate Program and the Medical Scientist Training Program.

Finally, I need to thank my friends, family, and my wife for all their support. My friends and family have graciously endured my years of training, and I truly appreciate their attempt to understand my research, my training program, and my idiosyncrasies. My wife, Laura, has been the most important person in my life, not just through training, but for over the past 14 years. She has given me peace of mind, motivation, and a daily respite from any troubles during my training. I am eternally grateful for her love, support, and endurance during this process.

Table of Contents

Chapter 1: Introduction 1

A. Diabetes mellitus 1

B. Monogenic diabetes 5

C. Epidemiology of monogenic diabetes 18

D. Monogenic diabetes treatment 21

E. Monogenic diabetes misdiagnosis 23

F. Monogenic diabetes and next-generation sequencing 26

G. Monogenic diabetes functional studies 30

H. Summary and objective 32

Chapter 2: Monogenic Diabetes in Overweight and Obese Youth Diagnosed with Type 2 Diabetes: The TODAY Clinical Trial 36

A. Introduction 36

B. Materials and methods 38

1. Patient Characteristics 38

2. Sequencing methods 41

3. Variant Analysis 43

4. Statistical Analysis 44

C. Results 46

1. Sequencing results 46

2. Monogenic Diabetes Gene Variants 47

3. Characteristics of Patients with Monogenic Diabetes Gene Variants and Study Outcomes 51

4. T2D Risk Polymorphism Analysis 61

D. Discussion 61

Chapter 3: Monogenic Diabetes Screening, Diagnosis, and Pedigree Analysis in the PDMP Study 75

A. Introduction 75

B. Methods 79

1. Recruitment Sites 79

2. Screening Process 80

3. Sequencing Process 84

4. Return of Results 87

5. Family History Machine Learning 88

6. Statistical Analysis 91

C. Results 91

1. Screening Data 91

2. Enrollment 96

3. Genetic Testing 98

4. Monogenic Diabetes Diagnoses 101

5. Pedigree Analysis 105

D. Discussion 107

Chapter 4: Zebrafish as a Model for On-Demand *In Vivo HNF1A* Variant Functional Analysis 116

A. Introduction 116

B. Methods 119

1. Zebrafish, morpholino, and injections 119

2. Rescue mRNA 120

3. Quantification and statistical analysis 123

C. Results 124

1. Dose escalation of mRNA rescue 124

2. Rescue with established *HNF1A* genetic variants 125

D. Discussion 127

Chapter 5: Discussion 133

A. Summary 133

1. Monogenic Diabetes in Overweight and Obese Youth Diagnosed with Type 2 Diabetes: The TODAY Clinical Trial 133

2. Monogenic Diabetes Screening, Diagnosis, and Pedigree Analysis in the PDMP Study 136

3. Zebrafish as a Model for On-Demand *In Vivo HNF1A* Variant Functional Analysis 137

B. Potential Impact 138

C. Limitations 140

D. Opportunities and future directions 145

References 149

List of Tables

Table 1.1: MODY genes 7

Table 1.2 List of genes causing diabetes syndromes 16

Table 1.3 Table of genes with variants causing lipodystrophy[67] 17

Table 1.4 European studies with percentage of *HNF4A*-, *GCK*-, and *HNF1A*-MODY cases out of all studied participants, separated by nation[79] 19

Table 2.1 Patient Clinical Characteristics at screening or baseline 40

Table 2.2: Sex, Race/Ethnicity, and Family History Characteristics 41

Table 2.3: Genes on sequencing panel 42

Table 2.4: ACMG/AMP Standards and Guidelines[119] evidence used to determine variant pathogenicity specifically for monogenic diabetes/MODY 44

Table 2.5: Non-common coding and splice variants in 13 MODY genes 48

Table 2.6: Pathogenic or likely pathogenic MODY gene variants and TODAY study patient data 48

Table 2.7: Previously reported MODY variants not classified as pathogenic or likely pathogenic based on ACMG/AMP criteria 49

Table 2.8: Novel genetic variants in MODY genes 50

Table 2.9: Rare coding and splice variants in non-MODY monogenic diabetes genes ... 51

Table 2.10: Associations between all monogenic diabetes variants, gene subtypes and patient characteristics at earliest study timepoint (screening or baseline), adjusted for age, sex, and race/ethnicity 52

Table 2.11: Linear model effect sizes (adjusted for age, sex, and race/ethnicity) 54

Table 2.12: Linear model effect sizes (adjusted for age, sex, race/ethnicity, and BMI Z-score) 55

Table 2.13: Characteristics of *BLK* p.A71T and *KLF11* p.T220M 56

Table 2.14: Linear model effect sizes 57

Table 3.1: PDMP screening questionnaire* 82

Table 3.2: PDMP inclusion criteria 82

Table 3.3: Distribution of PDMP participants by site and source 94

Table 3.4: Demographics of PDMP participants by site 95

Table 3.5: Screening questionnaire responses of all PDMP participants 95

Table 3.6: Eligibility and enrollment by study site 97

Table 3.7: Demographics of participants that enrolled compared to those who did not .. 97

Table 3.8: Comparison of enrollment categories between African-Americans and Caucasians 98

Table 3.9: Diabetes diagnoses in PDMP screened subset 98

Table 3.10: Total variants reviewed by gene and variant effect 99

Table 3.11: PDMP novel variants 100

Table 3.12: Characteristics of PDMP patients with and without monogenic diabetes ... 102

Table 3.13: Screening questionnaire responses of patients with and without monogenic diabetes 103

Table 3.14: Individual characteristics of PDMP patients with monogenic diabetes 104

Table 3.15: PDMP patients with suspicious monogenic diabetes VUS 105

Table 3.16: Percent of affected family members in patients with and without monogenic diabetes 106

Table 4.1: Established pathogenic HNF1A-MODY variants and cited functional studies
.... 121
Table 4.2: Primers for mutagenesis and confirmation Sanger sequencing 122
Table 4.3: Beta cell number and area of each condition studied 126

List of Figures

Figure 1.1: Summary of genes affected in monogenic diabetes. 6

Figure 1.2: Summary of genetic variants in *HNF1A* gene... 29

Figure 2.1 Failure-free survival curve of MODY gene subtypes and patients without monogenic diabetes.. 59

Figure 2.2: Failure-free survival curve of non-rare cited MODY variants....................... 60

Figure 2.3: Scatterplot of log-adjusted fasting glucose and fasting insulin separated by monogenic diabetes status and gene etiology 64

Figure 2.4: Scatterplot of log-adjusted insulin sensitivity and insulinogenic index separated by monogenic diabetes status and gene etiology 65

Figure 2.5: Violin plots of BMI Z-score, systolic blood pressure, diastolic blood pressure, and DXA fat percentage.. 69

Figure 2.6: Violin plots of log-adjusted values of fasting glucose, fasting insulin, HbA1C, and insulinogenic index. 70

Figure 2.7: Violin plots of unadjusted values of fasting glucose, fasting insulin, HbA1C, and insulinogenic index. 71

Figure 2.8: Violin plots of total cholesterol, LDL cholesterol, HDL cholesterol, and log-adjusted triglycerides. 72

Figure 3.1: PDMP study overview.. 79

Figure 3.2: Example pedigree collected in PDMP... 83

Figure 3.3: Clinical and research process separation of the PDMP................................. 86

Figure 3.4: Algorithm and schematic for pedigree summary for machine learning......... 90

Figure 3.5: Plot of screened patients by site over the study period 93

Figure 3.6: Plot of screened patients by source over the study period 94

Figure 4.1: Sanger sequencing confirmation of mutagenized variants........................... 123

Figure 4.2: Beta cell number of dose response experiment 125

Figure 4.3: Beta cell number and area of each condition studied 127

List of Abbreviations

ACMG: American College of Medical Genetics

ADA: American Diabetes Association

AMP: Association for Molecular Pathologists

ARFF: Attribute-relation file format

BMI: Body mass index

BVAMC: Baltimore Veterans Administration Medical Center

CAP: College of American Pathologists

CGL: Congenital generalized lipodystrophy

CI: Confidence interval

CLIA: Clinical Laboratory Improvement Amendments of 1988

CRISPR: Clustered regularly interspaced short palindromic repeats

DAA: Diabetes autoantibodies

DXA: Dual-energy X-ray absorptiometry scan

ExAC: Exome Aggregation Consortium

ESP: Exome Sequencing Project

EHR: Electronic health record

FPL: Familial partial lipodystrophy

GDM: Gestational diabetes mellitus

GHS: Geisinger Health SystemHbA1c: Hemoglobin A1c

HDL: High-density lipoprotein

HGMD: Human Gene Mutation Database

HIPAA: Health Insurance Portability and Accountability Act of 1996

HR: Hazard ratio

IDF: Indirect diversification factor

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

LADA: Latent autoimmune diabetes of adulthood

LDL: Low-density lipoprotein

MHC: Major histocompatibility complex

MIDD: Maternally-inherited diabetes and deafness

MELAS: Mitochondrial encephalomyopathy lactic acidosis and stroke-like episodes

MODY: Maturity-onset diabetes of the young

NDM: Neonatal diabetes mellitus

NGS: Next-generation sequencing

NHB: Non-Hispanic Black

NHW: Non-Hispanic White

OGTT: Oral glucose tolerance test

PDMP: Personalized Diabetes Medicine Program

PGM: Personalized Genome Machine

PNDM: Permanent neonatal diabetes mellitus

PPV: Positive predictive value

RCAD: Renal cysts and diabetes Syndrome

SD: Standard deviation

SEM: Standard error of the mean

SNV: Single nucleotide variant

T1D: Type 1 diabetes

T2D: Type 2 diabetes

TALEN: Transcription activator-like effector nucleases

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

UK: United Kingdom

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

US: United States

VUS: Variant of uncertain significance

Chapter 1: Introduction

In the current global epidemic of diabetes mellitus, a diagnosis of monogenic diabetes provides one of the few opportunities for genetic information to directly and immediately improve patient care. Although monogenic diabetes accounts for only a small proportion of all cases of diabetes, the large number of patients diagnosed with, and unknowingly affected by, diabetes mellitus present an opportunity to improve healthcare for many people. Because the most common forms of monogenic diabetes can be effectively treated with inexpensive and non-invasive treatment regimens, a correct diagnosis could allow patients to be placed on a more optimal treatment for their disease. At the same time, a misdiagnosis of monogenic diabetes, when the actual disease is of the type 1 or type 2 variety, may delay prescription of an optimal therapy. This dissertation reports the characteristics of patients with monogenic diabetes in multiple cohorts, and describes the testing of an *in vivo* model of functional testing for monogenic diabetes variants to pursue the objective of defining characteristics of patients with monogenic diabetes to discover more patients with monogenic diabetes and potentially improve their treatment.

A. Diabetes mellitus

Before describing specific aspects of monogenic diabetes, I provide a brief overview of the general characteristics, risks, and treatments for diabetes mellitus. Diabetes mellitus is a state of hyperglycemia caused by impaired insulin production and/or response. The clinical triad of diabetes mellitus symptoms includes polyuria, polydipsia, and polyphagia. The two most common forms of diabetes mellitus are type 1 diabetes (T1D) and type 2 diabetes (T2D). T1D is caused by impaired insulin production,

usually due to permanent autoimmune destruction of the pancreatic beta cells.[1] Patients with T1D are generally diagnosed early in life, and T1D accounts for approximately 5% of all diabetes mellitus cases.[2] After diagnosis of T1D, usually with laboratory tests for both blood glucose levels and diabetes autoantibodies (DAAs) such as IA-2, IA-2β, IAA, ZnT8, or GAD65, patients are placed on a regimen of exogenous insulin supplementation either through insulin injections or a continuous insulin pump. In contrast, T2D is caused by a combination of insulin resistance and insulin deficiency, often due to extended exposure to poor diet and/or sedentary lifestyle. Although T2D is a very heterogeneous disease, patients with T2D are generally diagnosed later in life and have accompanying symptoms of metabolic syndrome. T2D accounts for approximately 90% of all diabetes mellitus cases.[3] Those with T2D are diagnosed through measures of glycemic load in the blood, either during a random measure, in the fasted state, during an Oral Glucose Tolerance Test (OGTT), or by measuring the percentage of glycated hemoglobin (HbA1c).[4] After diagnosis of T2D, patients are usually treated with metformin, an insulin sensitizing agent, as first line therapy, and additional medications may be added or substituted until the patient stops producing endogenous insulin and is required to move to exogenous insulin supplementation. The rate of T2D progression is not uniform across affected individuals, and in fact, many patients may never require exogenous insulin or medications beyond metformin.

Another common form of diabetes is gestational diabetes mellitus (GDM). GDM is defined as diabetes first diagnosed in the second or third trimester of pregnancy that is not due to preexisting T1D or T2D.[5] This is a pathologic exacerbation of the physiologic insulin resistance that occurs during pregnancy. GDM has an increased risk of pregnancy

complications, such as cesarean delivery, shoulder dystocia, macrosomia, and neonatal hypoglycemia. Women with GDM are also at higher risk for having T2D after the pregnancy. Universal screening for GDM is common practice due to the high prevalence of GDM risk factors. If patients are unable to control GDM through diet and lifestyle modifications, treatment regimens are usually insulin or select oral antidiabetic medications. In addition to T1D, T2D, and GDM , there are many less frequently occurring forms of diabetes, including monogenic diabetes, latent autoimmune diabetes of adulthood (LADA), diabetes caused by cystic fibrosis, and medication-induced diabetes.[4] These forms of diabetes indicate that diabetes is truly a spectrum of different conditions all with different etiologies and characteristics.

Regardless of the subtype of diabetes, hyperglycemia often leads to dangerous and potentially deadly complications. Diabetes leads to mortality and morbidity through both macrovascular and microvascular disease. Macrovascular effects are manifested as an increased risk for cardiac disease, the most common cause of death in the United States (US), as well as increased risk for stroke. Microvascular effects act on multiple organs, leading to nephropathy, retinopathy, neuropathy, and dermopathy. Diabetes is one of the leading causes of kidney failure, blindness, and non-traumatic amputation in the US.[6] Beyond the likelihood of long-term diabetes complications varying by type, T1D has an acute danger of diabetic ketoacidosis and T2D has an acute danger of hyperglycemic hyperosmolar syndrome.

The genetics of T1D and T2D have been well-studied. The heritability, or proportion of a disease that can be attributed to genetic characteristics, of T2D is estimated to be greater (approximately 30-70%) than that of T1D (approximately 30-

50%).[7,8] Because T1D is a disease caused by an autoimmune mechanism, it has also been associated with genetic variants in genes important for immune function, particularly the major histocompatibility complex (MHC). Specifically, T1D is strongly associated with the HLA-DR3 (DRB1*03-DQB1*0201) and HLA-DR4 (DRB1*04-DQB1*0302) haplotypes, and individuals with those variants are at a higher risk for developing T1D as well as other autoimmune diseases such as Hashimoto's thyroiditis and rheumatoid arthritis.[9] Additionally, there have been over 40 genetic associations between T1D and other gene loci, such as *INS*, *PTPN22*, *ILRA2*, and *CTLA4*.[10] A great deal of effort has also been applied to studying the genetic architecture of T2D, as studies have progressed using different study designs, including: candidate gene studies, small family linkage studies, large agnostic genome-wide association studies, and multi-consortium meta-analyses. As a result, over 100 genetic loci have been associated with either T2D status or hyperglycemic quantitative characteristics.[11] However, T2D is a complex disease with many influencing factors. These genetic associations generally have modest effect sizes (odds ratio range of 1.05-1.6), and aggregation of the genetic associations explain only a small portion of the heritability of T2D.[8] Studies have shown that prediction based on genetic information is rarely more informative than data collected from a standard family medical history.[12] Some of the strongest genetic associations with T2D have been found in or near the genes *TCF7L2*, *KCNQ1*, and *CDKN2A/2B*. While very few of the genetic associations identified to date for T1D and T2D overlap, other subtypes of diabetes have similar mechanistic etiologies, e.g., the *TCF7L2, MTNR1B,* and *CDKAL1* loci shared between T2D and GDM.[13]

B. Monogenic diabetes

While the more common forms of diabetes are complex diseases with numerous genetic associations with ORs ranging from 1.05 to 1.6, monogenic diabetes is defined as hyperglycemia caused by variation in one of many genes (Figure 1.1). This term encompasses multiple subtypes, including: Maturity-Onset Diabetes of the Young (MODY), neonatal diabetes mellitus (NDM), and monogenic syndromes that include diabetes as one of the symptoms. Epidemiological studies have estimated that monogenic diabetes accounts for approximately 1-2% of all cases of diabetes mellitus.[14,15] MODY accounts for the majority of cases of monogenic diabetes, while the other forms are much rarer (less than 3 per million are affected in the general population).[16] Other monogenic conditions can cause or increase the risk of diabetes through effects on metabolism, as is the case for congenital lipodystrophy, familial partial lipodystrophy, and monogenic forms of severe obesity. Finally, genetic variants in the mitochondrial genome can also cause syndromes that include diabetes as part of the phenotype, e.g., Maternally-Inherited Diabetes and Deafness (MIDD). Notably, characteristics of mitochondrial heritability, such as inheritance in all descendants of a female and heteroplasmy, make mitochondrial diseases very different from Mendelian forms of monogenic diabetes.

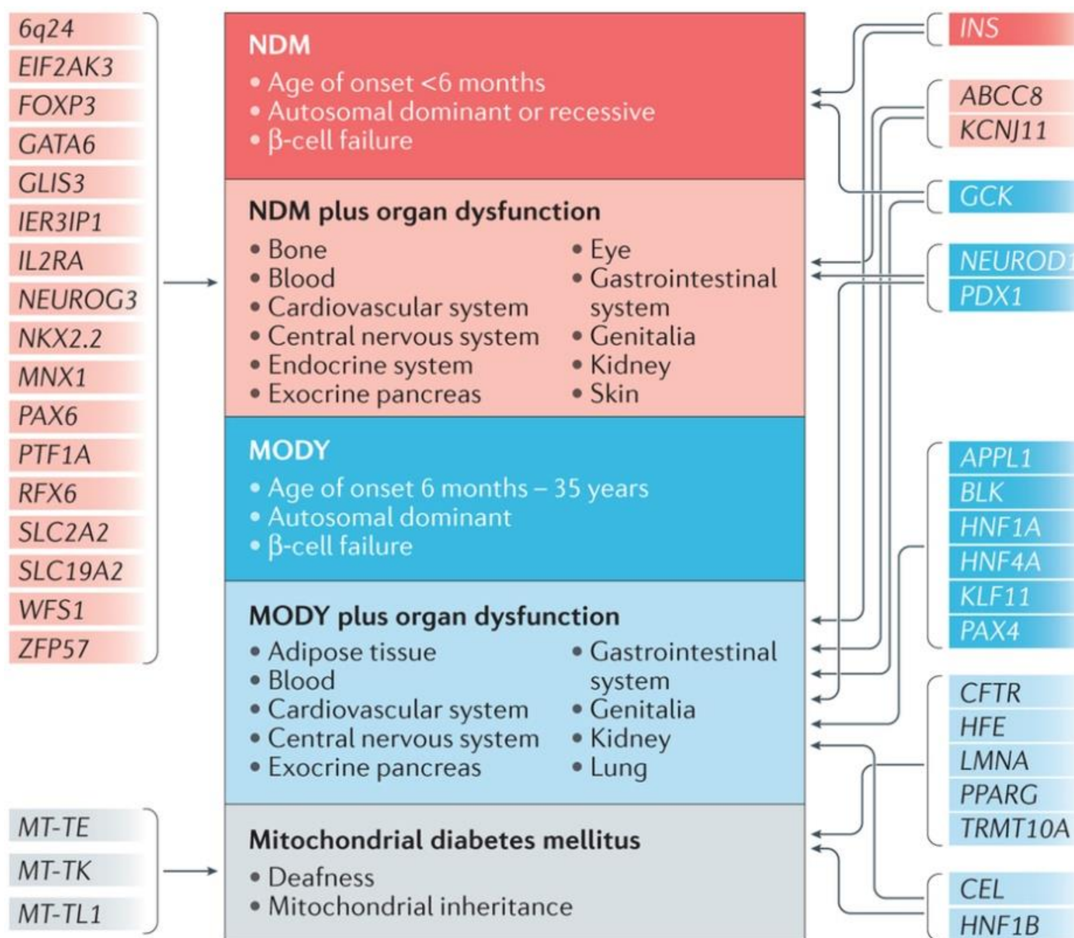


Figure 1.1: Summary of genes affected in monogenic diabetes.
Flow chart depicting the range of effects from some monogenic diabetes genes. Some genes causing diabetes syndromes, such as *INSR*, *ALMS1*, and *CISD2* were not included in this figure from Flannick et al. 2016, *Nat Rev Endo*[17]

MODY is caused by highly penetrant rare variants in 14 known genes (Table 1.1). A clinical diagnosis of MODY can be made if the patient has early-onset (prior to age 25), autosomal dominant diabetes with maintained insulin production, but identification of a causative genetic variant in one of those 14 genes is necessary for a genetic diagnosis of MODY. Originally named and described in the 1970s by Dr. Stephen Fajans and Dr. Robert Tattersall, MODY was used to describe what appeared to be T2D (formerly

referred to as Maturity-Onset Diabetes) found in young patients. At that time, T2D, obesity, and metabolic syndrome were rarely found in young people, as the name indicates. Patients with MODY were described as young (<25 years old), lean individuals with maintained insulin production and an autosomal-dominant inheritance of diabetes in their family. The original family described had a p.Q268X missense variant in *HNF4A*, the genetic etiology that was later classified as MODY type 1 (which will be referred to as *HNF4A*-MODY for clarity in this document).[18] In the 1990s, Vionnet et al. discovered that mutations in *GCK* created a similar phenotype and Yamagata et al. soon after reported that *HNF1A* genetic variants did the same.[19,20] Over the following years, variants in 11 other genes were identified as causes of MODY through studies of single families, studies identifying patients by syndromic features, and studies employing whole genome sequencing. Because of the wide range of clinical features associated with the variants in these newly discovered genes (Table 1.1), some have recommended that the term "MODY" be replaced by more specific designations that incorporate gene function (e.g., "transcription factor monogenic diabetes"[21]), although the more general term, "MODY," continues to persist due to recognition in the general population and to distinguish patients with those characteristics from those with neonatal diabetes or diabetes syndromes. For these reasons, the "MODY" nomenclature will be used throughout this document.

Table 1.1: MODY genes

Gene	**MODY type**	**Clinical features**	**Molecule**	**Class**
HNF4A	MODY 1	Hyperinsulinemic hypoglycemia and macrosomia at birth. Progressive beta cell failure. Sensitivity to sulfonylureas.	Hepatic nuclear factor 4-α	Transcription factor

Table 1.1 Continued

GCK	MODY 2	Mild hyperglycemia. Usually needs no treatment.	Glucokinase	Hexokinase
HNF1A	MODY 3	Progressive beta cell failure. Sensitivity to sulfonylureas.	Hepatic nuclear factor 1-α	Transcription factor
PDX1	MODY 4	Pancreatic agenesis in homozygotes or compound heterozygotes.	Pancreas/duodenum homeobox protein 1	Transcription factor
HNF1B	MODY 5	Renal disease (cysts). Urogenital tract abnormalities in females. Pancreatic atrophy. Exocrine insufficiency.	Hepatic nuclear factor 1-β	Transcription factor
NEUROD1	MODY 6	Very rare, discovered in <5 families. Wide range of age of onset (3-56). Patients may be overweight.	Neurogenic differentiation 1	Transcription factor
KLF11	MODY 7	Very rare, discovered in <5 families. Wide range of age of onset (17-56)	Kruppel-like factor 11	Transcription factor
CEL	MODY 8	Very rare, discovered in <5 families. Pancreatic exocrine dysfunction and pancreatic lipomatosis.	Carboxyl-ester lipase	Lipase
PAX4	MODY 9	Very rare, discovered in <5 families.	Paired box gene 4	Transcription factor
INS	MODY 10	More commonly associated with NDM.	Insulin	Hormone
BLK	MODY 11	Very rare, discovered in <5 families. Reported in a single study. Associated with systemic lupus erythematosus.	B-lymphocyte specific tyrosine kinase	Kinase
KCNJ11	MODY 12	More commonly associated with NDM. Responsive to sulfonylureas.	Inwardly rectifying potassium channel, subfamily J, member 11	Channel protein
ABCC8	MODY 13	More commonly associated with NDM. Responsive to sulfonylureas.	ATP-binding cassette, subfamily C, member 8	Channel protein
APPL1	MODY 14	Very rare, discovered in 2 families. Older age of diagnosis (20-50)	Adaptor protein, phosphotyrosine interaction, pH domain, and leucine zipper-containing protein 1	Kinase

The genes associated with first and third types of MODY, *HNF4A*-MODY (previously MODY 1) and *HNF1A*-MODY (previously MODY 3), were identified in 1996.[18,19] *HNF4A* and *HNF1A* are transcription factors that are strongly expressed in liver, kidney, and pancreatic islet tissues. Patients with mutations in *HNF4A* and *HNF1A* generally have "classic" MODY features (early onset, autosomal dominant inheritance, persistent insulin production) described in the first cases by Tattersall and Fajans. A small number of genetic variants in *HNF1A* have also been identified as a risk factor for renal cell carcinoma and hepatic adenoma.[22,23] Both *HNF1A* and *HNF4A* genes generate multiple tissue-specific isoforms, with *HNF1A* producing 3 isoforms and *HNF4A* producing 9 isoforms.[24,25] *HNF1A* expresses isoform A in the pancreas during fetal development, but transitions to mostly isoform B in the adult pancreas.[26] In the liver, expression of *HNF4A* isoforms 1-6 is reliant on promoter P1, while pancreatic expression of isoforms 7-9 is reliant on the P2 promoter 46kb upstream of the transcription start site.[27] Expression of the P1-dependent isoforms of *HNF4A* occur in the pancreas during fetal development, but only P2-dependent isoforms are expressed in adult pancreas.[28] HNF4A molecules interact with target DNA sites by forming either homodimers, while HNF1A molecules can form homodimers or heterodimerize with HNF1B molecules. The relationship between the HNF1A and HNF4A molecules is complex, with studies suggesting the HNF1A transcription factor regulates the *HNF4A* pancreas-specific P2 promoter, the HNF1A molecule physically interacts with HNF4A protein domains, and the two molecules synergistically regulate target genes.[27,29,30] In pancreatic islets of mice, the knockout of *Hnf1a* results in decreased expression of over 20 genes related to glycolysis, oxidative phosphorylation, and amino acid metabolism, as well as effects on

genes playing a role in beta cell growth. Target genes of the two molecules have a great deal of overlap that relate to the cellular functions of glycolysis, insulin production, and insulin release.[31] Chromatin immunoprecipitation studies have also shown a great deal of overlap between the binding targets of both HNF1A and HNF4A molecules, across pancreatic islets and liver.[32] Single nucleotide variants (SNVs) in both of these genes have been associated with T2D.[11]

Glucokinase, encoded by *GCK*, was identified as the dysfunctional protein in cases of MODY 2 in 1994. Glucokinase is an enzyme that phosphorylates glucose into glucose-6-phospate, which is the rate-limiting step of glycolysis. Because cellular ATP:ADP ratio acts as the trigger for the K^{+}_{ATP} channel hyperpolarization that leads to insulin release in pancreatic beta cells, control of ATP production through the glycolytic pathway is crucial for regulation of insulin release.[33] In that context, glucokinase essentially serves as the blood glucose monitor of the body, and damaging genetic variants in *GCK* cause the "glucose monitor" to have a higher threshold for insulin release. As a result, patients with *GCK*-MODY have mildly elevated blood glucose. These patients usually maintain mildly elevated blood glucose levels, with a hemoglobin A1c (HbA1c) from 5.6-7.6%, throughout their lifetime without increased incidence of macrovascular disease, neuropathy, and nephropathy. However, they do appear to have increased mild background retinopathy that is not sight-threatening, based on a study of 99 patients with *GCK* mutations that had a median age of 48.6 years.[34] Genetic variants in *GCK* can also cause neonatal diabetes when expressed as homozygotes or compound heterozygotes.[35]

Besides being the first types described, MODY caused by mutations in *HNF4A*, *HNF1A*, and *GCK* are also the most common forms of MODY. After their discoveries, mutations in at least 11 other genes now comprising MODY have been identified, although these other classes of MODY each have a much lower prevalence than MODY 1-3. Most of these rare forms of MODY are caused by genetic variants in transcription factors, such as those encoded by the genes *PDX1*, *HNF1B*, *NEUROD1*, *KLF11*, *PAX4* and *BLK*. Mutations in *PDX1* (formerly known as *IPF1*) can cause severe diabetes that may be accompanied by decreased pancreatic volume. In the homozygous or compound heterozygous state, damaging *PDX1* variants can cause complete pancreatic agenesis.[36,37] PDX1 is a transcription factor that is known to be very important for development of pancreatic tissue and crucial for maintenance of the beta cell phenotype.[38] Similarly, *NEUROD1* is a transcription factor important for beta cell development, although *NEUROD1*-MODY is very rare and has not been associated with whole-pancreas development issues like variants in *PDX1*. Homozygous damaging variants in *NEUROD1* can cause permanent NDM with neurological abnormalities.[39,40] *PAX4* is another transcription factor important for development and differentiation of pancreatic tissue. In a Thai population, *PAX4*-MODY was discovered to result from a missense variant (p.R46W) or a canonical splice site variant.[41] Other genetic variants in *PAX4* have been associated with T2D and ketosis-prone diabetes, indicating its importance in glycemic control.[42,43] As its name suggests, the transcription factor encoded by *HNF1B* shares homology with the HNF1A molecule and has some overlapping function and targets in the pancreatic beta cell.[44] However, patients with *HNF1B*-MODY often present with dysmorphic renal features (usually as cysts) in addition to diabetes. *HNF1B*-MODY has

been referred to as Renal Cysts and Diabetes Syndrome (RCAD), and screening for this condition is usually performed by selecting patients with both of those characteristics. Partial or whole gene deletions of *HNF1B* are a common cause *HNF1B*-MODY.[45] *KLF11* is a transcription factor that did not have an obvious function in the beta cell until the discovery of variants that caused *KLF11*-MODY in large families.[46] Since then, studies have determined that the KLF11 molecule regulates transcription of multiple metabolic genes, including insulin, and mutations in the KLF11 transcription factor's target region have been found as a potential cause of NDM in one case.[47,48] Finally, *BLK* encodes a transcription factor strongly expressed in B-lymphocytes, but without a known role in beta cells. Three families were found to have variants either causing a single missense change (p.A71T) within the coding region or different noncoding variants within 100kb of *BLK* that co-segregate with diabetes. Functional studies determined that the variants could cause decreased glucose-stimulated insulin secretion and insulin content in mouse beta cell lines, possibly through an interaction with *PDX1*.[49] However, a later study suggested that the *BLK* p.A71T missense change was unlikely to cause MODY, since 52 normoglycemic patients between the age of 31 and 65 were discovered in a study of a large cohort.[50] The different forms of MODY caused by transcription factors other than *HNF4A* or *HNF1A* are much rarer, although *HNF1B*-MODY appears to be the most common of the rare forms.

In addition to the forms of MODY caused by genetic variants in transcription factors and *GCK*, there are also other non-transcription factor genes that cause MODY. Genetic variants in *CEL*, *INS*, *KCNJ11*, *ABCC8*, and *APPL1* have also been found to cause MODY. *CEL*-MODY was discovered in two large Norwegian families that both

had different frameshift mutations. Affected members in the families had early onset diabetes in addition to signs of exocrine pancreatic dysfunction.[51] *CEL* encodes carboxyl-ester lipase, a molecule produced in the pancreas and released into the duodenum during digestion where it is eventually activated by bile salts to digest cholesterols and other esters. Genetic variants in *INS*, which codes for preproinsulin, can also cause MODY.[52] As expected, genetic mutations (mostly missense variants) can lead to decreased secretion, production, or effectiveness of insulin, mainly through mechanisms such as decreased proinsulin processing, decreased proinsulin localization or secretion, and ER stress from protein misfolding in pancreatic beta cells.[53] Mutations in *INS* can also cause more severe forms of diabetes, including permanent NDM or hyperproinsulinemia, a condition of elevated insulin precursors in the blood due to defective insulin processing.[54] *KCNJ11* and *ABCC8* encode the molecules that make up the ion channel and receptor components, respectively, of the K^{+}_{ATP} channel, which is necessary for depolarization that triggers calcium influx and insulin release from pancreatic beta cells in a high-ATP state. Activating mutations in both of those genes have been shown to cause NDM, with permanent forms of NDM being caused more commonly by variants in *KCNJ11* and transient NDM caused more commonly by *ABCC8*.[55] While these genes have been known to cause NDM for many years, the advent of next-generation sequencing was utilized to discover that some variants in these genes could cause a MODY phenotype.[56,57] *APPL1* is the most recent addition to the list of genes that cause MODY, and it was similarly discovered using next generation sequencing. The APPL1 molecule is a downstream member of the insulin signaling pathway, and the two families were

discovered to have loss of function *APPL1* genetic variants. Family members with *APPL1*-MODY had later age of onset (into the early 30s).[58]

Another subtype of monogenic diabetes, NDM is defined as hyperglycemia diagnosed within the first 6 months of life, and can occur in either permanent form or transient form, in which the condition resolves on its own accord in less than a year. This condition has a prevalence of approximately 1:400,000 live births (some estimates as high as 1:100,000), and transient NDM accounts for approximately 50% of cases.[59] As previously mentioned, homozygous or compound heterozygous variants in *GCK,* homozygous, compound heterozygous, or heterozygous variants in *INS,* as well as activating mutations in *KCNJ11* and *ABCC8* can all cause NDM. Mutations in *GCK* and *INS* have only been shown to cause permanent NDM, while those in *KCNJ11* and *ABCC8* can cause either transient or permanent forms.[55] Genetic variants in the chromosomal region 6q are the most common cause of transient NDM, accounting for over 50% of variants. While the exact mechanism is unknown, mechanisms have been suggested that include epigenetic modifying effects, effects on the gene *ZFP57*, or altering the ZAC region important for histone acetylase activity.[60-62] The most common causes of permanent NDM are *KCNJ11* mutations, which make up over 25% of permanent NDM.[55] Some cases with *KCNJ11* mutations have syndromic features termed DEND (developmental delay, epilepsy and neonatal diabetes), in which the affected individuals can have severe seizures, neurologic deterioration, and dysmorphic features.

There are also multiple genes that can cause monogenic syndromes that include diabetes as one of the prominent symptoms found in most cases. As mentioned previously, *PDX1*, *HNF1B*, and *KCNJ11* can cause syndromic features that accompany

diabetes. Diabetes syndromes are very rare, have severe phenotypes, and are often inherited in an autosomal recessive fashion (RCAD caused by *HNF1B* variants and DEND caused by *KCNJ11* variants are examples of exceptions in which diabetes syndromes have an autosomal dominant inheritance). For example, mutations in *FOXP3*, a gene that encodes for a transcription factor important for T-cell regulation, can cause IPEX (immunodysregulation, polyendocrinopathy, and enteropathy, X-linked) syndrome. IPEX is a severe autoimmune disease in which the immune system attacks multiple organs, causing severe dermatitis (eczema or pemphous nodules), celiac-like pattern of intestinal villous destruction (causing diarrhea and failure to thrive), thyroid dysregulation (more commonly hypothyroidism), destruction of blood cells (anemia or thrombocytopenia), and/or kidney disease.[63] There are also conditions in which carrier status of genetic variants can also cause symptoms, such as Wolfram syndrome: a disease characterized by diabetes mellitus, hearing loss, vision loss, diabetes insipidus, urinary tract problems, and psychiatric disease.[64] Patients heterozygous for variants in the gene that causes Wolfram syndrome, *WFS1*, have been shown to have a milder phenotype of diabetes mellitus and hearing loss.[65] Table 1.2 provides a list of monogenic syndromes that include diabetes as a prominent part of the phenotype.

Table 1.2 List of genes causing diabetes syndromes

Gene name	Diabetes syndrome
ALMS1	Alström syndrome
CEL	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
GLIS3	Neonatal diabetes mellitus with congenital hypothyroidism
HNF1B	Renal Cysts And Diabetes (RCAD) syndrome
INSR	Type A insulin resistance, Donohue syndrome, Rabson-Mendenhall syndrome
PDX1	Pancreatic agenesis
KCNJ11	PNDM with neurological features
PAX4	Ketosis-prone diabetes
PTF1A	Pancreatic (and cerebellar) agenesis
RFX6	Mitchell-Riley syndrome, Martinez-Frias syndrome
SLC19A2	Thiamine-responsive megaloblastic anemia syndrome (Rogers syndrome)
SLC2A2	Fanconi-Bickel syndrome
WFS1	Wolfram syndrome

Monogenic diseases that affect metabolism may also cause diabetes, as is the case for lipodystrophy and monogenic obesity. Lipodystrophy can present as severe, whole body loss of subcutaneous fat, termed congenital generalized lipodystrophy (CGL), or as partial (usually affecting the lower limbs) loss of subcutaneous fat, referred to as familial partial lipodystrophy (FPL). Lipodystrophy is accompanied by severe insulin resistance, hypertriglyceridemia, fatty liver, and often diabetes. Multiple genes have been identified to cause lipodystrophy (Table 1.3), and FPL is found with a prevalence of 1:1,000,000, while CGL affects about 1:10,000,000 people worldwide. While all forms of CGL are autosomal recessive, most of the FPL forms are inherited in an autosomal dominant fashion. Insulin resistance is more severe in patients with CGL compared to FPL. Monogenic forms of severe obesity have also been described to be caused by genetic variants in *LEP*, *LEPR*, *MC4R*, and *SIM1*, among others.[66] Damaging variants in these

genes cause hyperphagia in patients which leads to obesity, metabolic syndrome, and diabetes mellitus.

Table 1.3 Table of genes with variants causing lipodystrophy[67]

Gene	Disorder	Unique features	Gene Function	Inheritance	Ref.
AGPAT2	CGL Type 1	Appendicular skeletal lesions	Biosynthesis of triglyceride and phospholipids from glycerol-3-phosphate	AR	68
BSCL2	CGL Type 2	Mild mental retardation, cardiomyopathy	Lipid droplet formation and adipocyte differentiation	AR	69
CAV1	CGL Type 3 (Partial CGL)	Short stature, vitamin D deficiency	Caveolae component	AR	70
PTRF	CGL Type 4	Muscular dystrophy, pyloric stenosis	Biogenesis of caveolae	AR	71
LMNA	FPL Type 2	Normal or excess facial/neck fat during puberty	Nuclear lamina component	AD	72
PPARG	FPL Type 3	Normal abdominal fat, hypertension	Hormone receptor in adipose tissue	AD	73,74
PLIN1	FPL Type 4	Reduction in adipocyte size and increased fibrosis	Lipid droplet coating protein	AD	75
CIDEC	FPL Type 5	Pancreatitis, white adipocytes with many small lipid droplets	Promotes lipid droplet formation, may mediate apoptosis	AR	76
LIPE	FPL Type 6	Reduced white adipose tissue with inflammation	Converts cholesteryl ester to cholesterol	AR	77

CGL: Congenital Generalized Lipodystrophy - Berardinelli–Seip syndrome, FPL: Familial Partial Lipodystrophy, AR: autosomal recessive, AD: autosomal dominant
Table from Kleinberger JW et al. Genetics of type 2 diabetes: From candidate genes to genome-wide association analysis. In: Poretsky L, ed. *Principles of diabetes mellitus*. New York, NY: Springer International Pub; 2017.[67]

C. Epidemiology of monogenic diabetes

There have been multiple epidemiological studies examining the prevalence of monogenic diabetes, especially MODY, across Europe, but studies of the prevalence in the rest of the world has lagged. These studies have progressed rapidly in Europe due to multiple factors, such as national patient registries, availability of clinical care, and strong genetic testing resources. One of the leading worldwide centers for monogenic diabetes studies is the Molecular Genetic Laboratory at the Royal Devon and Exeter Hospital in the U.K., as they are the centralized testing center for the United Kingdom (UK) and have been involved in gene discovery, patient screening, and international referral in the field of monogenic diabetes for decades.[14] Research from the Exeter center has shown that monogenic diabetes consistently accounts for approximately 1-2% of all diabetes. Based on referrals for monogenic diabetes testing across geographic regions of the UK, diagnosis rates from 1996 to 2009 projected to an estimate of 68-108 affected individuals per million.[15] Similarly, the Netherlands also has a centralized location for referred monogenic diabetes testing, and a study over 10 years beginning in 2001 concluded that monogenic diabetes has a prevalence of 30 per million in the Netherlands population.[78] Both monogenic diabetes testing centers have shown that *HNF1A*-MODY is the most common form (52% in Exeter and 44% in the Netherlands), followed by *GCK*-MODY (32% and 41%), and then *HNF4A*-MODY (10% and 15%) (Table 1.4). Although centralized testing centers are excellent resources for epidemiological data, they are usually reliant on physicians for suspecting and suggesting testing.

Table 1.4 European studies with percentage of *HNF4A*-, *GCK*-, and *HNF1A*-MODY cases out of all studied participants, separated by nation[79]

Nation	**Ascertainment Method**	**Partici-pants**	**Genetic MODY**	***HNF1A*-MODY**	***GCK*-MODY**	***HNF4A*-MODY**
Norway[80]	MODY criteria‡	>1500†	<31% (458)	<14% (208)	<9% (139)	<3% (40)
Denmark[81]	Physician referral – selected for MODY criteria‡	78	49% (38)	36% (28)	10% (8)	3% (2)
Netherlands[78]	Physician referral	1,319	39% (502)	17% (222)	15% (204)	6% (76)
UK[15]	Physician referral	2,072	27% (564)	14% (293)	9% (180)	3% (56)
Poland[82]	MODY criteria‡	1,351	7% (100)	<0.3% (4*)	6.2% (84)	<0.3% (4*)
Germany/ Austria[83]	MODY criteria‡	272	97% (263)	31% (84)	62% (169)	4% (10)
Czech Rep. [84]	MODY criteria‡	61	48% (29)	11.5% (7)	31% (19)	5% (3)
Italy[85,86]	Hyperglycemia, MODY criteria‡	172, 58	70% (121), 78% (45)	7% (12), 16% (9)	63% (109), 53% (31)	N/A, 5% (3)
Greece[87]	Physician referral – based on hyperglycemia	134	66% (88)	12% (16)	54% (72)	N/A
Spain[88]	MODY criteria‡	95	89% (85)	8% (8)	80% (76)	0% (0)
Pooled	N/A	7,112	32% (2,293)	13% (898)	15% (1091)	3% (194)

Percentages represent the proportion of patients diagnosed with the condition in the study.
‡ MODY criteria are generally defined as diabetes onset before the age of 25 with an autosomal dominant form of inheritance and lack of diabetes autoantibodies. Studies listed may have altered MODY criteria moderately, i.e., altering age of onset to age 30, including individuals with hyperglycemia in the non-diabetic range (e.g., Spain) or defining autosomal dominant inheritance by number of generations affected with diabetes.
† Study did not provide precise number of registry participants
* Study did not differentiate between *HNF1A* and *HNF4A* mutations
Adapted from Kleinberger et al. *Curr Diab Rep* 2015[79]

Large patient registries both for general pediatric diabetes and specifically for monogenic diabetes have also been utilized for epidemiological studies. In Norway, study of patients in the Norwegian Childhood Diabetes Registry found that 1.1% of cases of childhood diabetes were monogenic diabetes, which projects to 31 children per million in the entire childhood population of Norway.[89] A MODY-specific database in Norway

showed that 53% of MODY cases were *HNF1A*-MODY, while *GCK*-MODY accounts for 30%, and *HNF4A*-MODY accounts for 7.5%.[80] In Germany, monogenic diabetes testing of DNA from members of the national DPV-Wiss registry of patients with pediatric diabetes found that 0.83% of that population had monogenic diabetes. However, they found that 62% of the patients had *GCK*-MODY, while 31% had *HNF1A*-MODY, and 4% *HNF4A*-MODY.[83] A similar finding of higher numbers of *GCK*-MODY compared to *HNF1A*-MODY was discovered in a pediatric diabetes registry in Poland.[82] Other smaller studies from either single healthcare centers of small groups of healthcare centers in Italy, Spain, Greece, and the Czech Republic have also shown higher numbers of *GCK*-MODY compared to *HNF1A*-MODY.[84-88] While *HNF1A*-MODY appears to be more common in northern European countries and *GCK*-MODY more common in southern European countries, factors such as ascertainment criteria have a major effect on the ratio of *GCK*-MODY to *HNF1A*-MODY. For example, patients with *HNF1A*-MODY are more likely to be diagnosed by a physician and are therefore more likely to be ascertained in studies based on physician referral. Alternatively, because *GCK*-MODY causes mildly elevated blood glucose, it is more commonly found as cases of incidental hyperglycemia, pre-diabetes, or where screening asymptomatic individuals is more routine.

While numerous studies of monogenic diabetes have been carried out in Europe, there have been much fewer studies performed in other populations across the world. A moderate number of studies in Asian populations has shown that patients with MODY have similar clinical characteristics as their European counterparts, but small study sizes make it very difficult to extrapolate the prevalence rates across the entire populations.

These studies have been performed in Chinese, Japanese, and Korean populations.[90-92] Available data, while sparse, suggest that *GCK*-MODY is nearly non-existent in Indian populations.[93] However, these studies are very small and may be the result of differing medical practice for treatment of incidental pre-diabetes by physicians in India versus countries where early treatment of chronic diabetes is a public health priority. Finally, there have been very few studies of monogenic diabetes in African, Latin American, and Middle Eastern populations.[94-96] Because these populations have been understudied, there is less published data about monogenic diabetes variants that may be specific to these populations. In African-ancestry populations, for example, the greater diversity of genetic variation could make it difficult to determine the pathogenicity of variants both due to the greater number of potential variants and novelty of some variants.[97]

D. Monogenic diabetes treatment

One of the most important aspects of monogenic diabetes is the opportunity for providing individualized treatment based on a genetic diagnosis. The most dramatic example of personalized therapy for patients with monogenic diabetes is the treatment for patients with activating *KCNJ11* or *ABCC8* genetic variants causing NDM. Usually after an early diabetes diagnosis, patients are placed on exogenous insulin therapy immediately. However, insulin therapy is ill-advised for patients with *KCNJ11* or *ABCC8* mutations since one of the older diabetes classes of medications, sulfonylureas, directly target the K^{+}_{ATP} channel. In high doses, sulfonylureas can inhibit the dysfunctional hyperactive channel to induce insulin secretion.[98,99] Patients are therefore able to transition from expensive and inconvenient insulin injections to inexpensive oral medications, which can have a profound effect on their lifestyle and can impact

neurological features in some cases.[100] In addition to specific treatment for patients with NDM caused by K^{+}_{ATP} channel mutations, patients with the three most common forms of MODY can also be treated in an individualized manner. As previously discussed, patients with *GCK*-MODY have mildly elevated blood glucose that does not lead to diabetic complications, possibly due to moderate effects of mild glucose elevation and lower intensity of postprandial insulin spikes. As a result, these patients generally do not require treatment, which could potentially save the costs of unnecessary medications, constant glucose monitoring, and healthcare visits. Finally, patients with *HNF1A*-MODY can be effectively treated with low-dose sulfonylureas, even after years of insulin therapy, as shown in a study in which 24 of 43 patients with *HNF1A*-MODY previously treated with insulin for a median of 4 years were able to effectively transition to sulfonylureas with improved glycemic control.[101] Another study found that 26/51 patients with *HNF1A*-MODY were effectively treated over 7 years with sulfonylureas alone, although the remaining 25 patients needed treatment supplementation with other medications (n =14) or treatment by other means (insulin = 4, insulin and metformin = 1, metformin = 5, and diet = 1). Interestingly, patients on sulfonylureas alone showed improvement of glucose control, as measured by HbA1c, without apparent beta cell exhaustion.[102] Since sulfonylureas are insulin secretagogues, they can take advantage of the fact that patients with *HNF1A*-MODY still make insulin, but it is not properly released in response to blood glucose elevations. This treatment is theoretically more effective than the first-line treatment for T2D, since metformin's mechanism of action is through insulin sensitizing activity. Patients with *HNF1A*-MODY are particularly sensitive to sulfonylureas, which may be because of the effects *HNF1A* genetic variants have on expression of membrane

transporters in liver tissue, where *HNF1A* is strongly expressed and metabolism of sulfonylureas occurs. Studies on Hnf1a(-/-) mice have demonstrated an increased half-life of a sulfonylurea (glibenclamide) compared to wild-type littermates, likely due to decreased hepatic uptake.[103] However, pharmacokinetic and pharmacodynamics studies in human *HNF1A*- and *HNF4A*-MODY patients failed to reproduce the increased half-life reported in the Hnf1a(-/-) mouse model.[104] Sulfonylureas have also been found to be effective in case reports of patients with *HNF4A*-MODY.[105] More robust studies on the effectiveness of sulfonylureas in patients with *HNF4A*-MODY are difficult to perform due to the small number of cases of *HNF4A*-MODY diagnosed. Additionally, long term studies of the effectiveness of sulfonylureas in *HNF1A*-MODY and *HNF4A*-MODY patients are necessary, since sulfonylureas are known to have limited long-term durability as a monotherapy for T2D.[106]

E. Monogenic diabetes misdiagnosis

Although monogenic diabetes presents an excellent opportunity for successful implementation of personalized medicine, misdiagnosis as either T1D or T2D has undercut the number of patients with monogenic diabetes who have been able to switch to alternative treatment regimens. Many patients with monogenic diabetes have a lean body-type and are diagnosed at a young age, characteristics that overlap with a T1D profile. On the other hand, patients with monogenic diabetes often have a strong family history of diabetes and maintain residual insulin production, characteristics that overlap with a T2D profile. It is obvious how a patient could be misdiagnosed with either T1D or T2D, which are each more common than monogenic diabetes. Additional factors that could play a role in misdiagnosis or lack of diagnosis include: cost of genetic testing,

uncertainty over insurance reimbursement for genetic testing, or healthcare provider awareness. The study of monogenic diabetes in the Polish pediatric diabetes registry discussed above demonstrates the influence of physician understanding and decision making on the referral rate for monogenic diabetes testing. That study found an increase in referral rate of monogenic diabetes genetic testing that was temporally related to a nationwide campaign to inform physicians, patients, and educators about *GCK*-MODY.[82,107] The increase in referrals did not lead to increased discovery of monogenic diabetes, suggesting that rate of discovery of *GCK*-MODY may be near maximal in that population. However, since *GCK*-MODY made up 83% of monogenic diabetes cases, the discovery rate of *HNF1A*-MODY (4% of cases) may not be near maximal in that population. Additionally, a study of the regions of the UK showed that although referral rates varied considerably between regions, the referral rates were strongly correlated prevalence of MODY in each region.[15] This data suggests that the limiting factor in diagnosis of monogenic diabetes may be due in part to unequal use of genetic testing among healthcare providers.

One US study has clearly demonstrated the problem with monogenic diabetes misdiagnosis. The SEARCH study for diabetes in youth is a multi-center study designed to characterize the landscape of pediatric diabetes nationwide.[108] This study assessed the prevalence of monogenic diabetes among children with diabetes. From the original study cohort of 5,963 participants with pediatric diabetes, those without signs of T1D (assessed through DAAs or absence of C-peptide) were studied. Of these patients, 586 had genetic testing for the three most common forms of monogenic diabetes (*HNF1A*-MODY, *HNF4A*-MODY, and *GCK*-MODY). The study found 47 participants had variants in

these genes causing monogenic diabetes, making up 8% of the sample studied and projecting to approximately 1.2% of the entire pediatric diabetes population.[109] However, only 3 of these 47 participants had been previously diagnosed with MODY. The majority (24/47 or 51%) of these MODY patients were incorrectly diagnosed with T2D, and 36% (17/47) were incorrectly diagnosed with T1D. Of all the participants with MODY, the majority were on insulin therapy (24/47 or 51%) or metformin (19/47 or 40%), which is unlikely to be the optimal treatment for these patients. Approximately 6% of patients with *HNF1A*- and *HNF4A*-MODY were treated with sulfonylureas, and 55% of patients with *GCK*-MODY were on no pharmacological treatment. This study strongly highlights the problem with misdiagnosis of monogenic diabetes, especially as either T1D or T2D, and how those misdiagnoses can lead to suboptimal treatment.

Other factors restricting the number of monogenic diabetes genetic diagnoses have also played a role in the suboptimal diagnosis rates, especially in the US. There is no centralized testing center in the US, and in the context of skyrocketing healthcare costs, testing for even a small panel (3-5 genes), costs thousands of dollars. Insurers have variable coverage policies for Next-Generation Sequencing (NGS) gene panels, even in well-established examples like *BRCA1/2* and breast cancer, demonstrating the complexity of reimbursement processes for genetic testing in the context of rapidly advancing sequencing methods.[110] Specific to monogenic diabetes, there is also the possibility that a patient could also have either coincident T1D or T2D. For example, 17% of patients with monogenic diabetes were also positive for at least one DAA (ICA, IA2, IAA, or GAD65) in the German/Austrian pediatric database study.[83] Likewise, someone with monogenic diabetes could possibly become insulin resistant or obese. There has even been a case

report of a patient with an *HNF1A* genetic variant, DAAs, and insulin resistance, termed "triple diabetes" by the authors.[111] Finally, family history may not be a sensitive indicator of monogenic diabetes, since the SEARCH study showed that 50% of those with MODY had no parental family history of diabetes.[109]

F. Monogenic diabetes and next-generation sequencing

Progress in the field of genetics has rapidly advanced over the past 20 years, mainly due to technological breakthroughs in DNA genotyping and sequencing. While sequencing was originally performed by painstakingly "walking" along a chromosome with Sanger sequencing, new methods such as paired-end shotgun sequencing, flow-based polymerase detection, and nanopore technology, which are all broadly classified as next-generation sequencing (NGS) have greatly advanced DNA sequencing throughput. These tools are directly applicable to discovery and diagnosis of monogenic diabetes, and NGS has been used to discover variants in *KCNJ11*, *ABCC8*, and *APPL1* which cause MODY.[56-58] A huge challenge brought about by NGS, however, is determining which of the many variants identified by this technology are damaging versus which have little to no effect on the phenotype of the individual. As the entire human knowledgebase of genetic variation grows, many large datasets have been made publicly available for a more complete understanding of the landscape of human variation.[112-114] While this has been extremely valuable, many populations are underrepresented, which can lead to the assumption that some variants are rare, even if those variants are actually common in underrepresented populations. This population imbalance has also impacted databases such as ClinVar, a resource of disease-causing genetic variants across the genome, as studies have shown that ClinVar has a bias to report variants from European-ancestry

populations while often missing or misclassifying variants from minority populations.[97,115] Because these datasets are also used as the training datasets for the growing number of *in silico* prediction tools for the effect of genetic variants, this bias may be propagated to other less-obvious metrics.[116,117] A study searching for monogenic diabetes in an Indian population using an NGS gene panel has demonstrated this bias.[118] The study found numerous variants, many in less-common MODY genes (i.e. *PDX1*, *NEUROD1*, *PAX4*). These variants were assessed according to *in silico* predictors, but no other experiments were performed to assess the effects of the variants. While those variants may in fact contribute to diabetes either alone or in combination with other factors, it is extremely difficult to properly assess the variants without a deeper understanding of the landscape of genetic variation in that population.

Other groups have recognized the potential dangers of misattributing phenotypic effects to genetic variants, and they have produced guidelines to aid the process of determining the pathogenicity of genetic variants. A joint collaboration between the American College of Medical Genetics (ACMG) and the Association for Molecular Pathologists (AMP) published standards and guidelines for the interpretation of sequence variants to address this problem.[119] This method requires data across multiple categories, consisting of: population data, computational and predictive data, functional data, segregation data, *de novo* status, allelic data, miscellaneous database data, and patient information. Based on fulfillment of 28 specific criteria at seven levels of evidence strengths in those categories, the ACMG/AMP committee developed an additive algorithm used to classify genetic variants into one of five categories: pathogenic, likely pathogenic, variant of uncertain significance (VUS), likely benign, or benign. This

process creates a conservative metric for interpretation of sequencing results that can prevent overreliance on single pieces of evidence that may be misleading, such as *in silico* prediction algorithms or co-segregation in small families. However, even this process is susceptible to differences in opinion regarding interpretation of the 28 criteria and how the variant-specific information may fulfill the criteria. Also, private sources of information, such as patient databases, can lead to different interpretations of pathogenicity between groups. Studies have shown that interaction and data sharing between groups with differing interpretations usually lead to consensus.[120]

Interpretation of variant pathogenicity is important for monogenic diabetes, especially the most common forms of *HNF1A*-MODY and *GCK*-MODY. There have been nearly 500 published variants in *HNF1A* and over 600 variants in *GCK*, most of which are rare (Figure 1.2).[121,122] While many of these have been assumed to cause MODY, it is almost certainly true that many others also have little or no phenotypic effects. This was demonstrated in a study by Flannick et al. that examined the prevalence and clinical characteristics of patients in the Framingham Heart Study and Jackson Heart Study with variants in *HNF1A* and *GCK* that were assumed to cause MODY.[123] These variants were assigned pathogenicity status based on presence in a variant database (HGMD), conservation metrics (PhyloP), *in silico* prediction tools (SIFT and Polyphen-2), or putative loss-of-function variants, since this study was released before the availability of ACMG/AMP criteria for variant interpretation. They discovered that numerous variants purported to be pathogenic were discovered in euglycemic individuals into middle and old age, indicating that those variants are either unlikely to be monogenic causes of diabetes, or the variants may have lower penetrance than previously believed. A

later review article by the same group found that many variants previously assumed to be pathogenic for monogenic diabetes would be downgraded to a less-damaging classification when implementing the ACMG/AMP guidelines.[17] Finally, it is important to recognize that the ever-growing knowledgebase of genetic variants can alter variant interpretation for monogenic diabetes variants, as well as variants associated with other diseases. A retrospective analysis of molecular testing from 2009 to 2013 in a US monogenic diabetes genetic testing center, the Seattle Children's Molecular Genetics Laboratory, found that variant interpretation of 30% of 115 reported variants changed from the original report over the 4-year period.[124] This demonstrates the importance of using as many resources as possible for variant interpretation and potentially revisiting previously-reported variants as new information is gained or published.

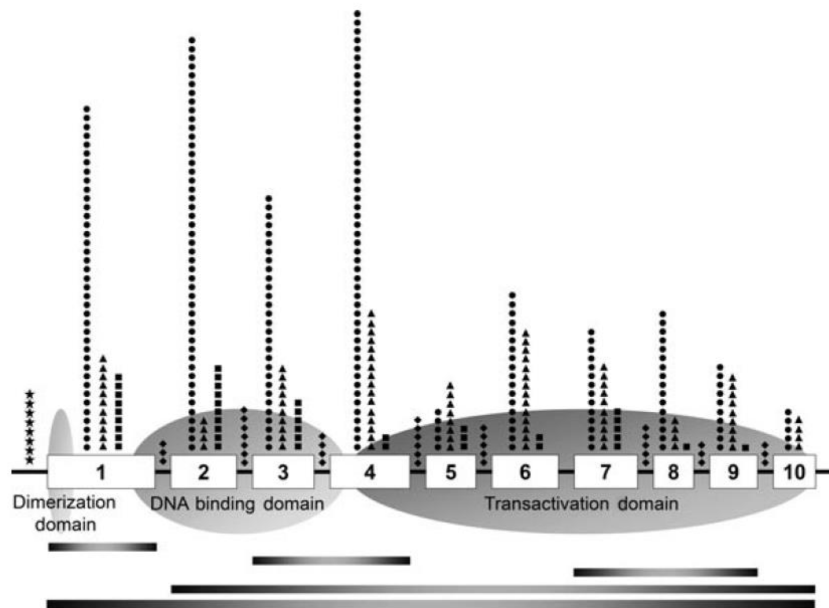


Figure 1.2: Summary of genetic variants in *HNF1A* gene
The exons, functional domains, and known variants in HNF1A are represented. Circles represent missense mutations, triangles represent frameshift mutations, squares represent nonsense mutations, diamonds represent splice-site variants, asterisks represent promoter mutations, and shaded bars represent large deletions. From Colclough et al. 2013, *Hum. Mut.*[121]

G. Monogenic diabetes functional studies

One of the most valuable tools for assessing pathogenicity of genetic variants is experimental functional studies. Unlike other pieces of information that rely on happenstance or "experiments of nature", such as co-segregation data or population data, experimental data can be generated in an on-demand fashion as variants are discovered. The conclusions gained from experimental studies are also very valuable, as demonstrated by the strength of pathogenicity evidence (second only to putative loss-of-function variants) assigned to functional studies by the ACMG/AMP guidelines.[119] However, functional assays have been performed on a relatively small proportion of monogenic diabetes genetic variants. In the case of *GCK*-MODY, this may be due to technical expertise and experience necessary to perform enzymatic analysis. Measuring the stability and enzymatic efficiency of the GCK enzyme is a sensitive process, and the majority of published studies testing the effects of *GCK* genetic variants have relied upon investigators at the Exeter Molecular Testing Laboratory, which has a great deal of experience performing and interpreting the results of GCK enzymatic analyses.[125-127] On the other hand, it is unclear why *HNF1A*-MODY variants, the most or second most common form of monogenic diabetes, have not been functionally studied in an extensive manner. In a review of known *HNF1A* genetic variants, only 52/455 total variants had been analyzed with functional analyses.[121] Only a small number of other *HNF1A* variants have been studied with functional analysis since publication of that review article.[128,129] Functional studies of *HNF1A* variants also have a great deal of variability in experimental design. Many published studied have been performed using luciferase assays measuring expression of a luciferase molecule attached to an HNF1A-target promoter after

transfection of mutated *HNF1A*. These studies have been performed in different cell lines, ranging from beta cell lines of mouse origin to human gastrointestinal, kidney, or hepatic cell lines, as well as using different HNF1A-target promoter molecules, ranging from the insulin promoter to promoters for albumin or sucrase-isomaltase.[121] Additionally, some *HNF1A* variants have shown other dysfunctional molecular characteristics, such as decreased nuclear localization measured microscopically. Interpretation of the experimental evidence is another complication to determining the effect of monogenic diabetes genetic variants. A recent study utilized luciferase and localization studies to determine if rare *HNF1A* variants predispose individuals to T2D.[128] While the T2D-predisposing *HNF1A* variants generally showed less extreme effects than *HNF1A*-MODY positive controls, differentiating between experimental effects of a T2D-predisposing *HNF1A* variant and an *HNF1A*-MODY variant is difficult, especially with differing cell-line and HNF1A-target conditions. These issues could be addressed through the creation of a simple experimental model in which generalized downstream effects of variants could be assessed in an on-demand fashion or through use of a model with high-throughput capacity to assess the entirety of genetic variants for a gene. The latter has been demonstrated in *PPARG*, a genetic cause of FPL type 3, using automated creation, identification, effect measurement, and outcome interpretation methods for every possible amino acid change in *PPARG*.[130] While this prospective approach is extremely valuable for analyzing yet-to-be-discovered variants and gaining insight into the most crucial residues for molecular function, it has the shortfall that the model system may not accurately assess variants with effects not measured by the model. Additionally, not all molecules may be appropriate for prospective functional studies, due to

differences in molecule size and ability to measure the effect of variants in a high-throughput manner.

H. Summary and objective

In summary, monogenic diabetes is an excellent opportunity for implementation of personalized medicine. As a condition strongly controlled by genetic factors, it differs from T1D and T2D, which each have their own multiple genetic factors that play a smaller role in disease etiology. However, besides *GCK*-MODY, these conditions share the trait that chronic, uncontrolled hyperglycemia still poses the same risks of complications causing cardiac, vascular, renal, neural, and retinal complications. While monogenic diabetes mostly consists of *GCK*-MODY, *HNF1A*-MODY, or *HNF4A*-MODY, the term also includes many rare monogenic conditions that can cause hyperglycemia, such as other forms of MODY, NDM, syndromic forms of diabetes, or lipodystrophy. Epidemiologic studies across Europe have concluded that approximately 1-2% of diabetes cases, affecting approximately 30-100 individuals per million of the general population, are actually cases of monogenic diabetes. European epidemiology studies have shown that the most common forms are either *GCK*-MODY or *HNF1A*-MODY, often dependent on patient selection methods. A diagnosis of monogenic diabetes is an opportunity for personalized medicine since, for example, patients with *GCK*-MODY usually have non-progressive mildly elevated blood glucose that does not cause diabetic complications and patients with *HNF1A*-MODY can often be effectively treated with low-doses of inexpensive oral sulfonylureas even after years of insulin therapy.

However, there are many issues regarding monogenic diabetes that need to be addressed before it can be considered a broadly successful example of implementation of personalized medicine. For example, populations of African, Latin American, Asian, and Middle Eastern ancestry have been under-represented in epidemiologic studies and population-specific studies have been minimal. Therefore, little information is known about types of monogenic diabetes variants in those populations, the population-specific prevalence of monogenic diabetes, and whether the patient characteristics may differ in those populations. Characteristics of monogenic diabetes can also resemble those of T1D or T2D, making identification of patients very difficult. The SEARCH study in the US showed that of pediatric patients with confirmed monogenic diabetes, 51% had been misdiagnosed with T2D and 36% had been misdiagnosed as T1D prior to genetic testing.[109] The majority of those pediatric patients were on insulin therapy or metformin, which is unlikely to be the optimal therapy for them. A study in the UK, a country considered to be well-surveyed for monogenic diabetes due to the high volume of testing performed, suggested that monogenic diabetes prevalence by region correlated with referral rates for testing, demonstrating the importance of the physician's role in attaining a genetic diagnosis. In addition to problems with misdiagnosis, technological advances in the form of NGS have introduced concerns over genetic variant interpretation due to the discovery of many genetic variants in monogenic diabetes genes. While interpretation guidelines like those from the ACMG/AMP have provided a conservative metric for assessing pathogenicity of variants, many variants previously labeled as pathogenic have been found not to be pathogenic and reclassification should be considered due to new evidence. One potentially valuable form of evidence, experimental functional studies, has

been underutilized in the example of *HNF1A* genetic variants. Less than approximately 25% of known HNF1A variants have been tested functionally, often with differing experimental conditions that are difficult to interpret. This represents an opportunity for gaining information about *HNF1A* variants through a widely-applicable method for assessing variants in an on-demand fashion.

The overall objective of this dissertation was to evaluate approaches for identifying patients with monogenic diabetes and a method for functionally testing monogenic diabetes variants to potentially improve diagnosis and treatment of these patients.. This was accomplished through three aims that provided information about patients with monogenic diabetes in unique populations with unique phenotypic data, as well as assessing a method to gain insight regarding the pathogenicity of specific variants in *HNF1A*. The first aim tested the hypothesis that MODY (or monogenic diabetes) may be the true cause of diabetes in many overweight and obese adolescent patients who have been (incorrectly) diagnosed with T2D in the Treatment Options for type 2 Diabetes in Adolescents and Youth (TODAY) clinical trial cohort. The second aim tested the hypothesis that a simple screening tool and family history information could identify monogenic diabetes patients through analysis of patient monogenic diabetes status, screening data, and pedigree information in the Personalized Diabetes Medicine Program. Finally, the third aim tested the hypothesis that manipulation of a transgenic zebrafish model could serve as a method for analyzing the pathogenicity of *HNF1A* monogenic diabetes variants through experimental study of known pathogenic *HNF1A* genetic variants. The information gained from these three aims furthered the understanding of monogenic diabetes pathophysiology, the characteristics of those affected by monogenic

diabetes, and methods for analyzing monogenic diabetes variants. This information contributed to the clinical understanding and diagnosis of monogenic diabetes as an example of personalized medicine. More broadly, studies from this dissertation can serve as models for implementation of personalized medicine for more complex conditions, such as common forms of diabetes or other disease.

Chapter 2: Monogenic Diabetes in Overweight and Obese Youth Diagnosed with Type 2 Diabetes: The TODAY Clinical Trial[1]

A. Introduction

Maturity-onset diabetes of the young (MODY), [2] the most common category of monogenic diabetes, results from a single variant in an individual patient in one of 14 known genes. In the 1970s, the term MODY was created to characterize patients described by Dr. Stefan Fajans as having a non-insulin-dependent form of diabetes at a young age.[131] Epidemiological studies across Europe determined that MODY accounts for approximately 1-2% of all diabetes.[14] *GCK*-MODY, *HNF1A*-MODY, and *HNF4A*-MODY account for at least 85% of MODY cases.[78,86,133] MODY generally presents in an autosomal dominant pattern of inheritance early in life as non-insulin requiring hyperglycemia. Correct characterization of monogenic diabetes is important for optimal patient treatment since the most common etiologies can be effectively treated with methods different from first-line treatments for type 1 diabetes (T1D) (insulin) or type 2 diabetes (T2D) (metformin). Patients with *HNF1A*-MODY and *HNF4A*-MODY are effectively treated with oral sulfonylurea therapy.[101,105] Patients with *GCK*-MODY have mildly elevated baseline blood glucose concentrations that commonly do not require

---

[1] Jeffrey W. Kleinberger, Kenneth C. Copeland, Rachelle G. Gandica, Morey W. Haymond, Lynne L. Levitsky, Barbara Linder, Alan R. Shuldiner, Sherida Tollefsen, Neil H. White, and Toni I. Pollin for the TODAY Study Group. Monogenic Diabetes in Overweight and Obese Youth Diagnosed with Type 2 Diabetes: The TODAY Clinical Trial. In preparation for submission.

[2] Since the first characterization in the 1970's, it has been suggested that MODY be changed to "familial young-onset diabetes" because of the discovery of multiple causative genes and better understanding of the gene-specific patient characteristics combined with the increased prevalence of type 2 diabetes (T2D) in childhood.[132] While this nomenclature change is appropriate, the term MODY will be used in this chapter to differentiate from other forms of monogenic diabetes, such a neonatal or syndromic forms of monogenic diabetes, and because the term is well-recognized by some members of the general population.

treatment and do not lead to diabetic complications.[34] Therefore, proper diagnosis of monogenic diabetes can lead to treatment that is more effective, more convenient, and less expensive than insulin therapy for patients and potentially family members with the same variant.

American Diabetes Association (ADA) guidelines suggest a diagnosis of monogenic diabetes be considered when diabetes is diagnosed in the first 6 months of life, when the patient does not have features of T1D (negative for diabetes-associated antibodies) or T2D (nonobese, lacking other metabolic features) especially when there is a strong family history of diabetes, or when there is stable, mildly elevated fasting blood glucose.[4] However, studies indicate these guidelines are either not utilized or fail to detect many cases of monogenic diabetes. The SEARCH study for diabetes in youth discovered that greater than 85% of patients with *GCK*, *HNF1A*, or *HNF4A* variants were misdiagnosed as T1D (36%) or T2D (51%), and only 19% of patients with MODY variants had treatment appropriate to their etiology.[109] Many factors contribute to the underdiagnosis of monogenic diabetes, including: heterogeneity of monogenic diabetes patient characteristics, similarity between monogenic diabetes and the more common forms of diabetes (especially with increasing prevalence of T2D in children and adolescents), cost of genetic testing, lack of insurance reimbursement, and lack of awareness among healthcare providers. With such high rates of overweight and obesity in young people, one might expect common co-occurrence of a T2D phenotype with monogenic diabetes, making currently suggested algorithms for diagnosis of monogenic diabetes even less sensitive.

Although T2D has been historically found mainly in overweight adults over age 40, recent increases in overweight or obese adolescents have led to an increased occurrence of T2D in young populations. Because of the increasing prevalence of T2D in adolescents and the lack of data regarding adolescent-specific T2D treatment methods, the Treatment Options for type 2 Diabetes in Adolescents and Youth (TODAY) study analyzed the effectiveness of metformin alone or in combination with rosiglitazone or lifestyle changes in adolescents with recently-diagnosed T2D.[134] Approximately half (48.3%) of the 699 participants treated with metformin alone maintained glycemic control, while the combination of metformin and rosiglitazone showed a small but significant improvement in the durability of glycemic control (61.4% maintained control, 25.3% decrease in primary outcome occurrence compared to metformin alone, p=0.006) over a relatively short duration of treatment (patients were followed for an average of 3.86 years).[135] In our current study, we analyze the prevalence of monogenic diabetes in the TODAY study participants and their outcomes.

B. Materials and methods

1. Patient Characteristics

The TODAY study participants were adolescents age 10-17 diagnosed with T2D according to ADA criteria within 2 years of study enrollment. Eligibility criteria also included: body-mass index (BMI) above the 85th percentile while accounting for age and sex, negative for diabetes-associated autoantibodies (GAD65 and ICA512), and fasting C-peptide $\geq$ 0.20 nmol/L. Descriptions of diabetes treatment regimens for each patient prior to study enrollment was not available for analysis. Anti-insulin antibody measurements were not performed on the study participants. Following a run-in period in

which glycemic control (HbA1c <8%) on metformin alone was attained, metabolic, glycemic, anthropometric, and lifestyle attributes were collected longitudinally from all TODAY study participants. Patients were followed longitudinally until they lost glycemic control and reached the primary outcome of the study, defined as glycated hemoglobin values of at least 8.0% for 6 consecutive months or the inability to wean the participant from insulin within 3 months after treatment for acute metabolic decompensation. Further descriptions of study protocol, design, methods, and results have been previously reported.[134,135]

This study analyzed DNA from a subset of 488 adolescents (177 males and 311 females) from the total TODAY cohort of 699 participants. Some TODAY participants had no DNA available to analyze because they did not attend a study visit during the DNA collection period or because they refused to consent to provide genetic data. In addition, all participants of undefined race/ethnicity or of a race/ethnicity with a small sample size were excluded from this analysis. Subject data included in this analysis were Hispanic (n=217), non-Hispanic black (NHB, n=166), or non-Hispanic white (NHW, n=105) and showed no obvious differences from the entire TODAY cohort with those race/ethnicities (Tables 2.1, 2.2).

Table 2.1 Patient Clinical Characteristics at screening or baseline

	Total cohort from TODAY study*	**Total providing genetic data**	**Hispanic**	**NHB**	**NHW**
n (females/males)	647 (414/233)	488 (311/177)	217 (133/84)	166 (115/51)	105 (63/42)
Age	14.0 ±2.0	13.9 ±2.0	13.9 ±2.0	13.8 ±2.0	14.0 ±2.0
BMI Z-score	2.30 ±0.42	2.31 ±0.41	2.29 ±0.42	2.40 ±0.38	2.22 ±0.44
DXA fat %†	37.8 ±6.2	38.0 ±6.1	38.8 ±5.8	36.7 ±5.8	38.1 ±6.9
HbA1c (%)	7.48 ±2.0	7.52 ±1.98	7.52 ±1.99	7.47 ±1.96	7.58 ±1.99
Fasting glucose (mmol/L)	6.18 ±1.41	6.13 ±1.29	6.06 ±1.32	6.15 ±1.34	6.26 ±1.16
Fasting insulin (mmol/L)	217.4 ±151.5	215.8 ±153.1	214.8 ±146.7	237.2 ±165.8	183.9 ±140.2
Insulinogenic index (Ins_{30-0}/Glu_{30-0})	1.52 ±2.17	1.55 ±2.32	1.58 ±2.13	1.79 ±2.82	1.12 ±1.66
Systolic blood pressure (mmHg)	114.0 ±11.6	113.7 ±11.6	112.0 ±11.5	115.9 ±11.2	113.9 ±11.6
Diastolic blood pressure (mmHg)	67.9 ±9.0	67.2 ±8.8	66.9 ±8.8	67.4 ±8.4	67.5 ±9.4
Total cholesterol (mmol/L)	4.06 ±0.96	4.01 ±0.86	4.02 ±0.80	4.01 ±0.89	3.97 ±0.89
HDL cholesterol (mmol/L)	1.04 ±0.24	1.03 ±0.24	1.04 ±0.27	1.02 ±0.22	1.02 ±0.23
LDL cholesterol (mmol/L)	2.38 ±0.80	2.35 ±0.71	2.38 ±0.69	2.33 ±0.73	2.34 ±0.68
Triglycerides (mmol/L)	1.42 ±1.05	1.39 ±0.91	1.47 ±0.98	1.31 ±0.83	1.35 ±0.89

Values are presented as mean ± SD to show similarity with the entire TODAY study cohort, although no formal analysis was performed. All measures were taken at screening in the TODAY study, except for fasting glucose, fasting insulin, insulinogenic index, and DXA fat measures, which were taken at the baseline time-point of the trial. At baseline time-point, patients had been titrated to a target dose of 1000mg metformin twice per day and all other diabetes medications had been discontinued.

* Only Hispanic, NHB, and NHW participants from the entire TODAY cohort were included to provide a precise comparison to those providing genetic data.

† DXA data was available for 344 individuals providing genetic data (109 NHB, 163 Hispanic, and 72 NHW).

Table 2.2: Sex, Race/Ethnicity, and Family History Characteristics

	Total providing genetic data	**No monogenic diabetes‡**	**Participants with monogenic diabetes**	**Participants with previously-cited variants unlikely to be damaging**
Total	488	426	22	40
Females	311 (64%)	277 (65%)	12 (55%)	22 (55%)
Males	177 (36%)	149 (35%)	10 (45%)	18 (45%)
Race/Ethnicity*				
His	217 (44%)	200 (47%)	8 (36%)	9 (23%)
NHB	166 (34%)	139 (33%)	3 (14%)	24 (60%)
NHW	105 (22%)	87 (20%)	11 (50%)	7 (18%)
Family history of diabetes				
Maternal diabetes (n=455)	195 (40%)	166 (39%)	6 (27%)	23 (58%)
Paternal diabetes (n=364)	123 (25%)	109 (26%)	8 (36%)	6 (15%)
Biparental diabetes (n=352)	50 (10%)	44 (10%)	1 (5%)	5 (13%)
Insulin resistance				
Acanthosis nigricans† (n=466)	410 (84%)	358 (84%)	15 (68%)	37 (93%)

* Significant Fisher's Exact Test for participants with monogenic diabetes vs participants without monogenic diabetes (p-value=0.007).
† Significant Fisher's Exact Test for participants with monogenic diabetes vs participants without monogenic diabetes (OR=0.28 [95%CI 0.01-0.85] p-value=0.01)
‡ "No Monogenic Diabetes" category refers to patients without monogenic diabetes variants classified as "pathogenic" or likely pathogenic" according to ACMG/AMP criteria for variant interpretation nor any previously-cited MODY variants.

2. Sequencing methods

We performed next-generation sequencing using a customized gene panel on the coding and flanking regions of 40 autosomal genes with variants known or predicted to cause monogenic diabetes, including 13 genes identified to cause MODY at the time of study design (*APPL1* was published as a cause of MODY after design of our study) as well as genes causing neonatal diabetes, diabetes syndromes, lipodystrophy, severe obesity, and hyperinsulinemia (Table 2.3). Genes known to cause hyperinsulinemia were included on the panel based on the theoretical possibility that gain of function mutations could cause hyperglycemia. Sequencing was performed using the Ion Torrent

Personalized Genome Machine; alignment as well as variant calling was performed using the Torrent Mapping and Alignment Program (TMAP) and variantCaller plugin on the Torrent Server. Only samples with ≥20x coverage depth of ≥80% of the target region (139,491bp) were used for analysis. Variants were annotated using a customized pipeline using multiple large population datasets, *in silico* prediction tools, and conservation metrics. All variants in non-coding or non-canonical splice regions, variants with a minor allele frequency (MAF) > 5.0%, or synonymous variants were filtered from further analysis. Additional genotyping was performed on each sample using the Axiom Biobank Genotyping Array, containing over 600,000 genome-wide SNVs. Genotyping data analysis and quality control was performed as specified by Axiom best practice guidelines.

Table 2.3: Genes on sequencing panel

Gene name	**MODY type**	**Neonatal diabetes type**	**Diabetes syndrome**
ABCC8	MODY 12	-	-
AGPAT2	-	-	CGL type 1
ALMS1	-	-	Alström syndrome
BLK	MODY 11	-	-
BSCL2	-	-	CGL type 2
CAV1	-	-	CGL type 3
CEL	MODY 8	-	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	MODY 2	PNDM	-
GLIS3	-	-	Neonatal diabetes mellitus with congenital hypothyroidism
HNF1A	MODY 3	-	-
HNF1B	MODY 5	-	Renal Cysts And Diabetes (RCAD) syndrome
HNF4A	MODY 1	-	-
INS	MODY 10	PNDM	-

Table 2.3 Continued

INSR	-	-	Type A insulin resistance, Donohue syndrome, Rabson-Mendenhall syndrome
PDX1	MODY 4	-	Pancreatic agenesis
KCNJ11	MODY 13	PNDM, TNDM	PNDM with neurological features
KLF11	MODY 7	-	-
LMNA	-	-	FPL type 2
NEUROD1	MODY 6	-	-
PAX4	MODY 9	-	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

PNDM: Permanent Neonatal Diabetes Mellitus, TNDM: Transient Neonatal Diabetes Mellitus, CGL: Congenital Generalized Lipodystrophy, FPL: Familial Partial Lipodystrophy

3. Variant Analysis

Non-common (<5% MAF) coding or splice-site variants were analyzed for pathogenicity according to American College of Medical Genetics/Association for Molecular Pathology (ACMG/AMP) guidelines for variant interpretation.[119] These guidelines were created to standardize the complex process of classifying variants into categories (“pathogenic,” “likely pathogenic,” “uncertain significance,” “likely benign,”

or "benign") based on population data, computational data, functional data, and segregation data. Criteria (Table 2.4) were assessed for each variant and pathogenicity was determined based on the total evidence according to the algorithm of the ACMG/AMP guidelines.[119]

Table 2.4: ACMG/AMP Standards and Guidelines[119] evidence used to determine variant pathogenicity specifically for monogenic diabetes/MODY

Evidence	Pathogenic or Benign	Evidence Strength	Criteria
PVS1	Pathogenic	Very strong	Nonsense, frameshift, or splice site mutations
PS3	Pathogenic	Strong	Literature search for decreased function in experimental studies
PM1	Pathogenic	Moderate	NCBI Protein database search for functional importance of altered amino acid residue
PM2	Pathogenic	Moderate	Variant absent from 1000 Genomes, NHLBI-ESP, and ExAC
PM5	Pathogenic	Moderate	Literature search for novel change in the same residue as a cited MODY variant
PP3	Pathogenic	Supporting	Pathogenic prediction in ≥3 of 10 *in silico* prediction tools*
PP5	Pathogenic	Supporting	ClinVar, LOVD search for previous report as pathogenic for MODY
BS1	Benign	Strong	Variant found in >1% of any subpopulation of 1000 Genomes, NHLBI-ESP, or ExAC
BS2	Benign	Strong	Variant found in greater than 5 controls in T2D-Genes exome database
BS3	Benign	Strong	Literature search for normal function in experimental studies
BP4	Benign	Supporting	Benign prediction in ≥3 of 10 *in silico* prediction tools*
BP6	Benign	Supporting	ClinVar search for previous report as benign

Evidence category names are a combination of the Pathogenic or Benign status, followed by the strength of evidence and a numerical identifier to differentiate between evidence categories at the same strength.
*GERP++, CADD, SIFT, PolyPhen2, MutationTaster, MutationAssessor, LRT, FATHMM, SVM, LR

4. Statistical Analysis

Patients were grouped for statistical analysis. The group in whom monogenic diabetes was presumed to be absent (n=426) did not have any pathogenic, likely pathogenic, or previously-cited monogenic diabetes variants. Other groups included

patients with pathogenic or likely pathogenic variants in any monogenic diabetes gene (n=22), *HNF1A* (n=5), *HNF4A* (n=7), or *GCK* (n=7). Participants with previously-cited monogenic diabetes variants that were not classified as pathogenic or likely pathogenic (n=40) were not included in analysis due to uncertainty over their monogenic diabetes status. Associations between monogenic diabetes subcategories and patient characteristics, including BMI Z-score, HbA1c, fasting glucose, fasting insulin, insulinogenic index, dual energy x-ray absorptiometry (DXA), lipid measures, and blood pressure, were evaluated using linear models accounting for sex, age, race, and BMI Z-score (where indicated). These measures were collected from the earliest available time-point for the trait (screening data for those traits where it was available; otherwise from baseline measures). Baseline measures, consisting of fasting glucose, fasting insulin, insulinogenic index, and DXA measure, were collected at the first study visit after the patient had been titrated to a target dose of 1000mg metformin twice per day and all other diabetes medications had been discontinued. Log transformation was used to normalize datasets with skewed distributions. Because these analyses were hypothesis-driven, a p-value of <0.05 was considered nominally significant and no adjustment for multiple comparison testing was performed.[136] Treatment failure was defined as the primary outcome of the TODAY Study (loss of glycemic control, defined as a glycated hemoglobin values of at least 8.0% for 6 consecutive months or the inability to wean the participant from insulin within 3 months after treatment for acute metabolic decompensation). Treatment failure analyses were performed using a Cox proportional hazards model using the patient outcomes specified in the original TODAY study.[135] The treatment failure analysis accounted for participant sex, age, race/ethnicity, and TODAY

treatment group. The proportional hazards assumption was met by each variable, except for sex, which was treated as a stratifying variable in the model.

C. Results

1. Sequencing results

A search for monogenic diabetes gene variants was performed in 488 adolescents with recently-diagnosed T2D. The average depth of coverage was 147.7±56.9 (mean ± SD), and an average of 92.8% of the target bases reached a coverage depth of ≥20x. Poorly-amplifying regions of a small subset of genes including *GATA6*, *PDX* 1, *PTF1A*, and *ZFP57*, accounted for the majority of the regions with less than 20x coverage depth. Importantly, all bases in *HNF1A*, *HNF4A*, and *GCK* consistently had sufficient coverage across samples. Each participant had a mean of 156.0±19.4 total variants and 5.3±2.8 variants after filtering out common, non-coding, and synonymous variants. A total of 509 missense variants, 4 nonsense variants, 3 splice-site variants, and 12 small insertions/deletions were analyzed according to ACMG/AMP standards and guidelines for variant interpretation. In 13 MODY genes, 132 missense variants were discovered. Only one canonical splice site variant and one nonsense variant were found, both in *HNF4A.* Three novel frameshift mutations were discovered in *HNF4A* and *GCK*, while we detected one rare frameshift mutation in *CEL* that had previously been reported in five African-ancestry individuals in the 1000 Genomes Project (all 1000 Genomes Participants declared to be healthy at the time of sample collection). Novel variants (not represented in 1000 Genomes, NHLBI-ESP, ExAC, dbSNP, T2D-Genes Exome database, nor literature searches) were discovered in nine of the MODY genes. A total of 30 novel variants were found in 31 individuals (Table 2.8).

2. Monogenic Diabetes Gene Variants

The thirteen genes reported to cause MODY at the time of study design were analyzed for non-common coding and splice site variants in this cohort (Table 2.5). Twenty-six previously-cited MODY variants were assessed according to ACMG/AMP criteria; 18 of these were classified pathogenic or likely pathogenic (Table 2.6), while eight were classified as benign, likely benign, or variants of uncertain significance (Table 2.7). Thirty novel variants were discovered, four of which we classified as pathogenic or likely pathogenic and 26 of which we classified as variants of uncertain significance (Table 2.8). Patients with pathogenic or likely pathogenic variants were considered to have monogenic diabetes. In sum, 22 individuals with 22 different pathogenic or likely pathogenic variants were discovered; i.e. 22/488 of those analyzed, or 4.5% of this sample of TODAY participants, had monogenic diabetes. Patients with monogenic diabetes were found across each of the three categories of race/ethnicity and across all treatment arms of the TODAY clinical trial (Table 2.6, 2.2). While variants were found in non-MODY genes in this study, none were determined to be pathogenic or likely pathogenic for monogenic diabetes or monogenic forms of obesity (Table 2.9).

Table 2.5: Non-common coding and splice variants in 13 MODY genes

Gene	MODY Type	Mis-sense	Splice Var.	Non-sense	Frame-shift	Prev. cited as MODY	Novel Var.	Seq. Bases*
HNF4A	1	12	1	1	2	5	8	2.6
GCK	2	12	0	0	1	6	4	2.1
HNF1A	3	17	0	0	0	8	3	2.4
PDX1	4	4	0	0	0	0	2	1.0
HNF1B	5	4	0	0	0	2	0	2.4
NEUROD1	6	6	0	0	0	0	1	1.1
KLF11	7	11	0	0	0	2	0	2.1
CEL	8	14	0	0	1	0	1	2.2
PAX4	9	7	0	0	0	0	0	1.5
INS	10	2	0	0	0	2	0	0.4
BLK	11	15	0	0	0	1	1	2.1
ABCC8	12	20	0	0	0	0	7	6.7
KCNJ11	13	8	0	0	0	0	3	2.2

*Total number of kilobases covered in each gene region

Table 2.6: Pathogenic or likely pathogenic MODY gene variants and TODAY study patient data

Gene	Sex (age)	Race/ ethnicity[a]	Treatment arm[b]	TODAY primary outcome[c]	Amino acid change / Site change[d]	ACMG pathogenicity
HNF4A	M (10)	His.	Met.+Ros.	-	p.R64Q[137]	Likely pathogenic
HNF4A	F (12)	NHW	Met.+Ros.	+	p.R64fs[e]	Likely pathogenic
HNF4A	F (13)	NHB	Met.+Life	+	p.Q86X[e]	Pathogenic
HNF4A	F (13)	His.	Met.+Life	+	p.V105I[138]	Pathogenic
HNF4A	F (14)	His.	Met.+Life	+	Splice-site (c.573+1G>A)[105]	Pathogenic
HNF4A	M (16)	His.	Metformin	+	p.R308H[86]	Likely pathogenic
HNF4A	F (14)	NHW	Met.+Life	+	p.H365fs[e]	Likely pathogenic
GCK	M (10)	His.	Met.+Life	-	p.V62M[139]	Pathogenic
GCK	F (13)	NHW	Met.+Ros.	-	p.R191W[140]	Likely pathogenic
GCK	F (17)	NHW	Met.+Ros.	-	p.T206M[141]	Pathogenic
GCK	M (13)	NHW	Met.+Life	-	p.N254H[122]	Likely pathogenic
GCK	F (12)	NHW	Metformin	-	p.E265K[142]	Pathogenic
GCK	F (13)	NHW	Met.+Life	-	p.R392C[143]	Likely pathogenic
GCK	M (13)	NHW	Met.+Life	-	p.S396fs[e]	Likely pathogenic
HNF1A	M (12)	His.	Metformin	+	p.P112L[144]	Pathogenic

Table 2.6 Continued

HNF1A	F (11)	NHB	Metformin	+	p.R131W[145]	Pathogenic
HNF1A	F (12)	NHW	Metformin	-	p.R271Q[128]	Pathogenic
HNF1A	M (14)	His.	Met.+Life	+	p.P379A[146]	Pathogenic
HNF1A	M (10)	NHW	Met.+Life	-	p.P519L[145]	Pathogenic
KLF11	M (16)	His.	Met.+Ros.	+	p.A347S[46]	Pathogenic
INS	F (12)	NHB	Metformin	-	p.R6H[147]	Pathogenic
INS	M (15)	NHW	Met.+Ros.	-	p.R46Q[148]	Pathogenic

[a]Self-reported race/ethnicity: His.=Hispanic, NHW = Non-Hispanic White, NHB = Non-Hispanic Black
[b]TODAY clinical trial treatment arm: Met.+Ros. = Metformin and rosiglitazone, Met.+Life = Metformin and lifestyle intervention
[c]TODAY clinical trial outcome: (-) = Patient did not reach primary outcome (treatment failure), (+) = Patient reached primary outcome (treatment failure)
[d]Amino acid reported sites are according to the following reference sequences: *HNF4A*-NM_001287183; NP_001274112, *GCK*-NM_000162; NP_000153, *HNF1A*-NM_000545; NP_000536, *KLF11*-NM_003597; NP_003588, and *INS*-NM_000207; NP_000198.1
[e]Novel variant (not reported in dbSNP, 1000Genomes, Exome Sequencing Project, ExAC or in literature searches)

Table 2.7: Previously reported MODY variants not classified as pathogenic or likely pathogenic based on ACMG/AMP criteria

Gene	**Nucleotide change**	**Amino acid change**	**dbSNP**	**ACMG classification**	**Hetero-zygotes**	**Homo-zygotes**
HNF4A	c.430G>A	p.V144I[149]	rs142204928	Benign	1	0
HNF1A	c.29C>T	p.T10M[150]	.	VUS	1	0
HNF1A	c.92G>A	p.G31D[151]	rs137853247	Likely Benign	1	0
HNF1A	c.1165T>G	p. L389V[150]	.	Likely Benign	2	0
HNF1B	c.226G>T	p.G76C[45]	rs144425830	VUS	1	0
HNF1B	c.1108G>A	p.G370S[45]	rs113042313	VUS	4	0
KLF11	c.659C>T	p.T220M[46]	rs34336420	Benign	13	0
BLK	c.211G>A	p.A71T[152]	rs55758736	Benign	18	0

VUS: Variant of Uncertain Significance

Table 2.8: Novel genetic variants in MODY genes

Gene	Nucleotide change	Amino acid change	Effect	ACMG classification	Hetero-zygotes	Homo-zygotes
HNF4A	c.121C>G	p.R41G	Missense	VUS	1	0
HNF4A	c.191delG	p.R64fs	Frameshift	Likely pathogenic	1	0
HNF4A	c.256C>T	p.Q86X	Nonsense	Pathogenic	1	0
HNF4A	c.535T>C	p.Y179H	Missense	VUS	1	0
HNF4A	c.616G>A	p.G206R	Missense	VUS	1	0
HNF4A	c.1093delC	p.H365fs	Frameshift	Likely pathogenic	1	0
HNF4A	c.1167C>A	p.N389K	Missense	VUS	1	0
HNF4A	c.1300C>A	p.P434T	Missense	VUS	1	0
GCK	c.1235T>G	p.V412G	Missense	VUS	1	0
GCK	c.1187_1188 insAG	p.S396fs	Frameshift	Likely pathogenic	1	0
GCK	c.472G>C	p.D158H	Missense	VUS	1	0
GCK	c.152A>G	p.E51G	Missense	VUS	1	0
HNF1A	c.415C>T	p.L139F	Missense	VUS	1	0
HNF1A	c.599G>T	p.R200L	Missense	VUS	1	0
HNF1A	c.611T>C	p.F204S	Missense	VUS	1	0
PDX1	c.161T>C	p.L54P	Missense	VUS	1	0
PDX1	c.212C>A	p.P71H	Missense	VUS	1	0
NEUROD1	c.334A>G	p.N112D	Missense	VUS	1	0
CEL	c.782A>G	p.K261R	Missense	VUS	1	0
BLK	c.46A>G	p.K16E	Missense	VUS	1	0
ABCC8	c.4247C>A	p.T1416N	Missense	VUS	1	0
ABCC8	c.4207A>G	p.I1403V	Missense	VUS	1	0
ABCC8	c.4204A>G	p.I1402V	Missense	VUS	1	0
ABCC8	c.3896G>A	p.R1299K	Missense	VUS	1	0
ABCC8	c.1801G>A	p.V601I	Missense	VUS	1	0
ABCC8	c.893G>C	p.R298P	Missense	VUS	2	0
ABCC8	c.505T>G	p.F169V	Missense	VUS	1	0
KCNJ11	c.20T>G	p.I7S	Missense	VUS	1	0
KCNJ11	c.379G>T	p.V127F	Missense	VUS	1	0
KCNJ11	c.586C>T	p.L196F	Missense	VUS	1	0

VUS: Variant of Uncertain Significance

Table 2.9: Rare coding and splice variants in non-MODY monogenic diabetes genes

Gene	Inheritance	Missense	Splice var.	Nonsense	Indel	Novel SNVs
INSR	AR, AR, AD	23	0	1	0	7
PPP1R3A	AD	29	0	0	1	1
GLIS3	-	28	0	0	0	4
SLC2A2	AR, AD	6	0	0	0	1
ALMS1	AR	91	0	0	1	25
EIF2AK3	AR	11	0	0	0	3
WFS1	AR	55	0	1	2	4
CISD2	AR	1	0	0	0	1
RFX6	AR	19	0	1	0	1
FOXP3	XLR	1	0	0	0	0
PTF1A	AR	4	0	0	0	2
GATA6	AD	6	0	0	0	1
SLC19A2	AR	6	0	0	0	1
AGPAT2	AR	13	1	0	0	1
BSCL2	AR	7	0	0	2	2
CAV1	-	2	0	0	0	0
PTRF	AR	3	0	0	0	1
LMNA	AD	3	0	0	0	0
PPARG	AD	3	0	0	0	1
PLIN1	AD	12	0	0	0	1
LEP	AR	2	0	0	0	0
LEPR	-	18	0	0	0	2
MC4R	AD	9	0	0	0	0
SIM1	AR	10	0	0	0	2
GLUD1	AD	4	1	0	0	2
HADH	AR	9	0	0	0	1

AR: Autosomal recessive, AD: Autosomal dominant, XLR: X-linked recessive

3. Characteristics of Patients with Monogenic Diabetes Gene Variants and Study Outcomes

Characteristics at the earliest available time-point (BMI Z-score, HbA1c, blood pressure, triglycerides, and lipid measures were from the screening visit, while fasting glucose, fasting insulin, insulinogenic index, and DXA measures were from the baseline visit) of subjects with pathogenic or likely pathogenic monogenic diabetes variants (n=22) were compared to those without monogenic diabetes (n=426) (Table 2.10, 2.11). Individuals with monogenic diabetes showed lower BMI Z-scores (2.05 vs. 2.32 p=0.004), higher fasting glucose (6.65 vs. 6.08 mmol/L p=0.02), lower fasting insulin

(152.1 vs. 213.9 pmol/L p=0.03) and higher total cholesterol (4.50 vs. 4.00 mmol/L p=0.003) compared to individuals without monogenic diabetes. These associations were still significant when adjusted for BMI Z-score (Table 2.12). Separated by gene etiology, monogenic diabetes subgroups showed similar trends in patient characteristics, although they were generally underpowered to detect significant differences. Two benign variants that were previously reported to be MODY-causing, *BLK* p.A71T and *KLF11* p.T220M, were found in a large number of study participants (n=18 and n=13, respectively). *BLK* p.A71T was not associated with any differences in patient characteristics, and patients with *KLF11* p.T220M actually showed higher fasting insulin (299.0 vs. 213.9 pmol/L) compared to those without MODY gene variants (Table 2.13, 2.14).

Table 2.10: Associations between all monogenic diabetes variants, gene subtypes and patient characteristics at earliest study timepoint (screening or baseline), adjusted for age, sex, and race/ethnicity

	No mono-genic diabetes[d]	**All mono-genic diabetes**[b]	***HNF4A* -MODY**	***GCK*- MODY**	***HNF1A* - MODY**	***INS*- MODY**	***KLF11*- MODY**
n	426	22	7	7	5	2	1
Age at time of study (Mean ± SD)	13.9 ± 2.0	13.0 ± 1.9	13.1 ± 1.9	13.0 ± 2.1	11.8 ± 1.5	13.5 ± 2.1	16
BMI Z-score	2.32 ±0.42	2.05 ±0.42[a]	2.12 ±0.52	1.91 ±0.33[a]	2.06 ±0.32	2.52 ±0.44	1.49
DXA fat %[c]	38.3 ±6.02	35.2 ±6.9	38.4 ±7.0	38.3 ±5.7	29.9 ±1.5[a]	30.6	23.9
HbA1c (%)	7.51 ±1.97	7.45 ±1.93	7.27 ±2.35	6.99 ±2.23	7.78 ±1.03	9.40 ±0.57	6.4
Fasting glucose (mmol/L)	6.08 ±1.27	6.65 ±1.56[a]	7.15 ±1.09[a]	6.69 ±0.28	5.46 ±1.2	5.22 ±0.24	11.6
Fasting insulin (pmol/L)	213.9 ±145.3	152.1 ±210.0[a]	122.2 ±46.3	87.5 ±32.7[a]	107.6 ±7.5[a]	588.9 ±679.7	164.6
Insulinogenic index (Ins_{30-0}/Glu_{30-0})	1.63 ±2.45	0.81 ±0.84	0.56 ±0.41	1.28 ±1.24	0.49 ±0.17	1.29	0.08

Table 2.10 Continued

Systolic blood pressure (mmHg)	113.8 ±11.4	107.9 ±13.6	113.1 ±16.5	99.4 ±9.4 [a]	107.6 ±7.5	119.8 ±22.3	109
Diastolic blood pressure (mmHg)	67.2 ±8.8	64.3 ±9.7	67.1 ±12.0	60.9 ±11.3	64.7 ±5.4	66.3 ±8.1	63.5
Total cholesterol (mmol/L)	3.98 ±0.86	4.50 ±1.06 [a]	4.77 ±1.15 [a]	4.11 ±0.9	4.38 ±1.08	4.26 ±0.68	6.4
HDL cholesterol (mmol/L)	1.03 ±0.24	1.09 ±0.24	1.19 ±0.21	1.05 ±0.34	1.03 ±0.12	1.18 ±0.20	0.83
LDL cholesterol (mmol/L)	2.33 ±0.72	2.65 ±0.72 [a]	2.82 ±0.96	2.48 ±0.68	2.56 ±0.47	2.51 ±0.84	3.34
Triglycerides (mmol/L)	1.38 ±0.91	1.71 ±1.30	1.66 ±0.77	1.27 ±0.79	1.72 ±1.32	1.23 ±0.07	6.09

Values are presented as mean ±SD. All measures were taken at screening in the TODAY study, except for fasting glucose, fasting insulin, insulinogenic index and DXA fat measures, which were taken at the baseline time-point.

[a] $p<0.05$ for the effect size of the classification (monogenic diabetes status in aggregate or separated by gene) in a linear model accounting for sex, age, and race/ethnicity (effect sizes, BMI Z-score adjusted results, and p-values found in Table 2.11 and Table 2.12)

[b] Monogenic diabetes Pathogenic or Likely Pathogenic Mutation

[c] DXA data was available for 303 individuals assumed to not have monogenic diabetes and 15 patients with monogenic diabetes (5 *HNF4A*-MODY, 5 *GCK*-MODY, 3 *HNF1A*-MODY, 1 *INS*-MODY, and 1 *KLF11*-MODY).

[d] “No Monogenic Diabetes” category refers to patients without monogenic diabetes variants classified as “pathogenic” or likely pathogenic” according to ACMG/AMP criteria for variant interpretation nor any previously-cited MODY variants.

Table 2.11: Linear model effect sizes (adjusted for age, sex, and race/ethnicity)

	Monogenic diabetes*	***HNF4A* -MODY**	***GCK*-MODY**	***HNF1A*- MODY**
n	22	7	7	5
BMI Z-score	-0.26 (-0.45--0.08) p=0.004	-0.17 (-0.55-0.21) p=0.3	-0.36 (-0.75-0.02) p=0.02	-0.3 (-0.81-0.21) p=0.1
DXA fat %[†]	-2.35 (-5.42-0.73) p=0.1	-1.06 (-7.78-5.66) p=0.7	1.21 (-5.56-7.99) p=0.6	-6.77 (-20.11-6.58) p=0.03
HbA1c (%)	-0.03 (-0.76-0.78) p=0.9	-0.26 (-1.6-1.42) p=0.7	-0.58 (-1.86-1.03) p=0.3	0.51 (-1.4-3.1) p=0.5
Fasting glucose (mmol/L)	0.49 (0.09-0.93) p=0.01	0.90 (0.05-1.91) p=0.009	0.61 (-0.20-1.59) p=0.07	-0.23 (-1.10-0.88) p=0.5
Fasting insulin (pmol/L)	-67.6 (-99.2--25.4) p=0.002	-70.4 (-126.3-29.1) p=0.06	-97.7 (-141.9--18.0) p=0.005	-91.0 (-147.4-32.5) p=0.03
Insulinogenic index (Ins_{30-0}/Glu_{30-0})	-0.99 (-1.67-0.12) p=0.06	-1.53 (-2.39- 0.69) p=0.06	0.55 (-1.44- 5.79) p=0.6	-0.62 (-0.99- 0.93) p=0.1
Systolic blood pressure (mmHg)	-4.63 (-9.74-0.47) p=0.06	1.09 (-9.14-11.31) p=0.8	-13.54 (-23.89--3.19) p=0.001	-4.45 (-18.12-9.23) p=0.4
Diastolic blood pressure (mmHg)	-1.83 (-5.83-2.18) p=0.3	0.79 (-7.28-8.86) p=0.8	-5.52 (-13.73-2.7) p=0.1	-0.56 (-11.37-10.26) p=0.9
Total cholesterol (mmol/L)	0.59 (0.19-1.00) p=0.002	0.82 (0.01-1.63) p=0.01	0.19 (-0.63-1.02) p=0.6	0.47 (-0.62-1.56) p=0.2
HDL cholesterol (mmol/L)	0.07 (-0.04-0.18) p=0.2	0.17 (-0.05-0.40) p=0.06	0.04 (-0.19-0.27) p=0.7	0.0 (-0.30-0.31) p=1.0
LDL cholesterol (mmol/L)	0.37 (0.03-0.70) p=0.02	0.50 (-0.17-1.18) p=0.07	0.18 (-0.51-0.86) p=0.5	0.29 (-0.62-1.19) p=0.4
Triglycerides (mmol/L)	0.23 (-0.07-0.61) p=0.1	0.30 (-0.26-1.24) p=0.3	-0.04 (-0.48-0.71) p=0.9	0.21 (-0.45-1.56) p=0.5

Values presented are β (95% CI) p-value from linear models accounting for sex, age, race/ethnicity. Effect size represents the estimate of difference for each group compared to those presumed to not have monogenic diabetes. p-value represents the likelihood of the classification (monogenic diabetes status in aggregate or separated by gene) having an effect size as extreme as observed if the null hypothesis (no effect) is true.

* Monogenic diabetes Pathogenic or Likely Pathogenic Mutation

[†] DXA data was available for 303 individuals assumed to not have monogenic diabetes and 15 patients with monogenic diabetes (5 *HNF4A*-MODY, 5 *GCK*-MODY, 3 *HNF1A*-MODY, 1 *INS*-MODY, and 1 *KLF11*-MODY).

Table 2.12: Linear model effect sizes (adjusted for age, sex, race/ethnicity, and BMI Z-score)

	Monogenic diabetes*	***HNF4A* -MODY**	***GCK*-MODY**	***HNF1A*- MODY**
n	22	7	7	5
HbA1c (%)	-0.01 (-0.72-0.79) p=1.0	-0.25 (-1.57-1.41) p=0.7	-0.55 (-1.82-1.05) p=0.4	0.51 (-1.38-3.09) p=0.5
Fasting glucose (mmol/L)	0.49 (0.05-0.96) p=0.02	0.93 (0.03-2.00) p=0.01	0.60 (-0.26-1.63) p=0.1	-0.28 (-1.20-0.89) p=0.5
Fasting insulin (pmol/L)	-14.9 (-25.4--1.06) p=0.03	-18.8 (-37.4-13.4) p=0.1	-24.7 (-40.8-3.57) p=0.03	-23.0 (-42.7-18.2) p=0.07
Insulinogenic index (Ins_{30-0}/Glu_{30-0})	-0.61 (-0.24-0.14) p=0.3	-0.71 (-0.41-0.18) p=0.1	-0.27 (-0.17-1.38) p=0.2	-0.71 (-0.44-0.63) p=0.3
Systolic blood pressure (mmHg)	-1.81 (-6.57-2.95) p=0.4	2.85 (-6.65-12.35) p=0.5	-9.85 (-19.54--0.16) p=0.01	-1.40 (-14.17-11.38) p=0.8
Diastolic blood pressure (mmHg)	-0.52 (-4.47-3.42) p=0.8	1.61 (-6.29-9.50) p=0.6	-3.78 (-11.86-4.30) p=0.3	0.84 (-9.77-11.46) p=0.8
Total cholesterol (mmol/L)	0.58 (0.17-0.99) p=0.003	0.81 (-0.01-1.62) p=0.02	0.18 (-0.65-1.01) p=0.6	0.45 (-0.65-1.54) p=0.3
HDL cholesterol (mmol/L)	0.07 (-0.05-0.18) p=0.2	0.17 (-0.06-0.39) p=0.07	0.03 (-0.20-0.26) p=0.7	0.0 (-0.31-0.30) p=1.0
LDL cholesterol (mmol/L)	0.37 (0.03-0.71) p=0.02	0.50 (-0.17-1.18) p=0.07	0.18 (-0.51-0.87) p=0.5	0.29 (-0.62-1.20) p=0.4
Triglycerides (mmol/L)	0.24 (-0.10-0.68) p=0.2	0.33 (-0.32-1.42) p=0.3	-0.07 (-0.55-0.77) p=0.8	0.21 (-0.55-1.77) p=0.5

Values presented are β (95% CI) p-value from linear models accounting for sex, age, race/ethnicity, and BMI Z-score. p-value represents the likelihood of the classification (monogenic diabetes status in aggregate or separated by gene) having an effect size as extreme as observed if the null hypothesis (no effect) is true.

* Monogenic diabetes Pathogenic or Likely Pathogenic Mutation

† DXA data was available for 303 individuals assumed to not have monogenic diabetes and 15 patients with monogenic diabetes (5 *HNF4A*-MODY, 5 *GCK*-MODY, 3 *HNF1A*-MODY, 1 *INS*-MODY, and 1 *KLF11*-MODY).

Table 2.13: Characteristics of *BLK* p.A71T and *KLF11* p.T220M

	No monogenic diabetes[†]	***BLK* p.A71T**	***KLF11* p.T220M**
n	426	18	13
BMI Z-score	2.32 ±0.42	2.34 ±0.52	2.29 ±0.35
DXA fat %[c]	38.3 ±6.0	37.2 ±5.9	34.2 ±5.3
HbA1c (%)	7.51 ±1.97	7.46 ±1.8	8.01 ±2.56
Fasting glucose (mmol/L)	6.08 ±1.27	6.77 ±1.41	6.22 ±1.01
Fasting insulin (pmol/L)	213.9 ±145.3	284.4 ±217.1	299.0 ±163.5
Insulinogenic index (Ins_{30-0}/Glu_{30-0})	1.63 ±2.45	1.01 ±0.85	1.49 ±1.08
Systolic blood pressure (mmHg)	113.8 ±11.4	114.2 ±11.1	118.2 ±11.0
Diastolic blood pressure (mmHg)	67.2 ±8.8	66.7 ±8.4	70.2 ±8.4
Total cholesterol (mmol/L)	3.98 ±0.86	3.93 ±0.59	3.92 ±0.54
HDL cholesterol (mmol/L)	1.03 ±0.24	1.01 ±0.22	1.03 ±0.27
LDL cholesterol (mmol/L)	2.33 ±0.72	2.38 ±0.57	2.26 ±0.56
Triglycerides (mmol/L)	1.38 ±0.91	1.17 ±0.54	1.38 ±0.59

Values are presented as mean ± SD. All measures were taken at screening in the TODAY study, except for fasting glucose, fasting insulin, insulinogenic index, oral disposition index, and DXA fat measures, which were taken at the baseline time-point of the trial.

* p<0.05 (effect sizes found in Table 2.14)

† "No Monogenic Diabetes" category refers to patients without monogenic diabetes variants classified as "pathogenic" or likely pathogenic" according to ACMG/AMP criteria for variant interpretation nor any previously-cited MODY variants.

Table 2.14: Linear model effect sizes

	Monogenic diabetes*	***BLK* p.A71T**	***KLF11* p.T220M**
n	22	18	13
BMI Z-score	-0.26 (-0.45--0.08) p=0.004	-0.03 (-0.24-0.17) p=0.7	-0.13 (-0.38-0.12) p=0.3
HbA1c (%)	-0.03 (-0.76-0.78) p=0.9	0.01 (-0.78-0.91) p=1.0	0.53 (-0.5-1.73) p=0.3
DXA fat %	-2.35 (-5.42-0.73) p=0.1	0.16 (-3.05-3.36) p=0.9	-3.52 (-8.06-1.03) p=0.09
Fasting glucose (mmol/L)	0.49 (0.09-0.93) p=0.01	0.39 (-0.04-0.86) p=0.06	0.18 (-0.32-0.76) p=0.4
Fasting insulin (pmol/L)	-67.6 (-99.2--25.4) p=0.002	24.76 (-34.32-107.01) p=0.4	90.98 (-1.53-226.82) p=0.03
Insulinogenic index (Ins_{30-0}/Glu_{30-0})	-0.99 (-1.67-0.12) p=0.06	-0.65 (-1.51-0.82) p=0.3	-0.7 (-1.75-1.31) p=0.4
Systolic blood pressure (mmHg)	-4.63 (-9.74-0.47) p=0.06	-2.11 (-7.77-3.56) p=0.4	2.72 (-4.01-9.46) p=0.4
Diastolic blood pressure (mmHg)	-1.83 (-5.83-2.18) p=0.3	-1.51 (-6-2.99) p=0.5	3.13 (-2.2-8.47) p=0.2
Total cholesterol (mmol/L)	0.59 (0.19-1.00) p=0.002	-0.06 (-0.5-0.37) p=0.8	-0.07 (-0.6-0.47) p=0.8
HDL cholesterol (mmol/L)	0.07 (-0.04-0.18) p=0.2	-0.01 (-0.14-0.11) p=0.8	-0.01 (-0.16-0.14) p=0.9
LDL cholesterol (mmol/L)	0.37 (0.03-0.70) p=0.02	0.04 (-0.33-0.4) p=0.8	-0.08 (-0.52-0.37) p=0.7
Triglycerides (mmol/L)	0.23 (-0.07-0.61) p=0.1	-0.1 (-0.34-0.23) p=0.5	0.12 (-0.25-0.65) p=0.5

Values presented are β (95% CI) p-value from linear models accounting for sex, age, race/ethnicity. p-value represents the likelihood of the classification (monogenic diabetes status in aggregate or separated by gene) having an effect size as extreme as observed if the null hypothesis (no effect) is true.
* Monogenic diabetes Pathogenic or Likely Pathogenic Mutation

Time-to-treatment failure analyses, using the primary outcome of the TODAY Study, compared individuals presumed to not have monogenic diabetes (n=426) with individuals with *HNF1A*-MODY, *HNF4A*-MODY, and *GCK*-MODY (Figure 2.1). No patients with *GCK*-MODY (0/7) failed treatment in the TODAY study. Across all three study arms, 6 of 7 of the subjects with *HNF4A*-MODY failed treatment over the first 2 years of study, with a hazard ratio of 5.03 (2.18-11.58 95%CI) (p=0.0002) compared to

subjects without monogenic diabetes. Three patients with *HNF4A*-MODY lost glycemic control by their first post-baseline study visit. There was no significant difference in treatment response of individuals with *HNF1A*-MODY compared to individuals without monogenic diabetes. Additionally, comparison of treatment outcomes between those with questionable, but previously-cited, MODY variants *BLK* p.A71T or *KLF11* p.T220M showed no significant differences from those without monogenic diabetes (Figure 2.2).

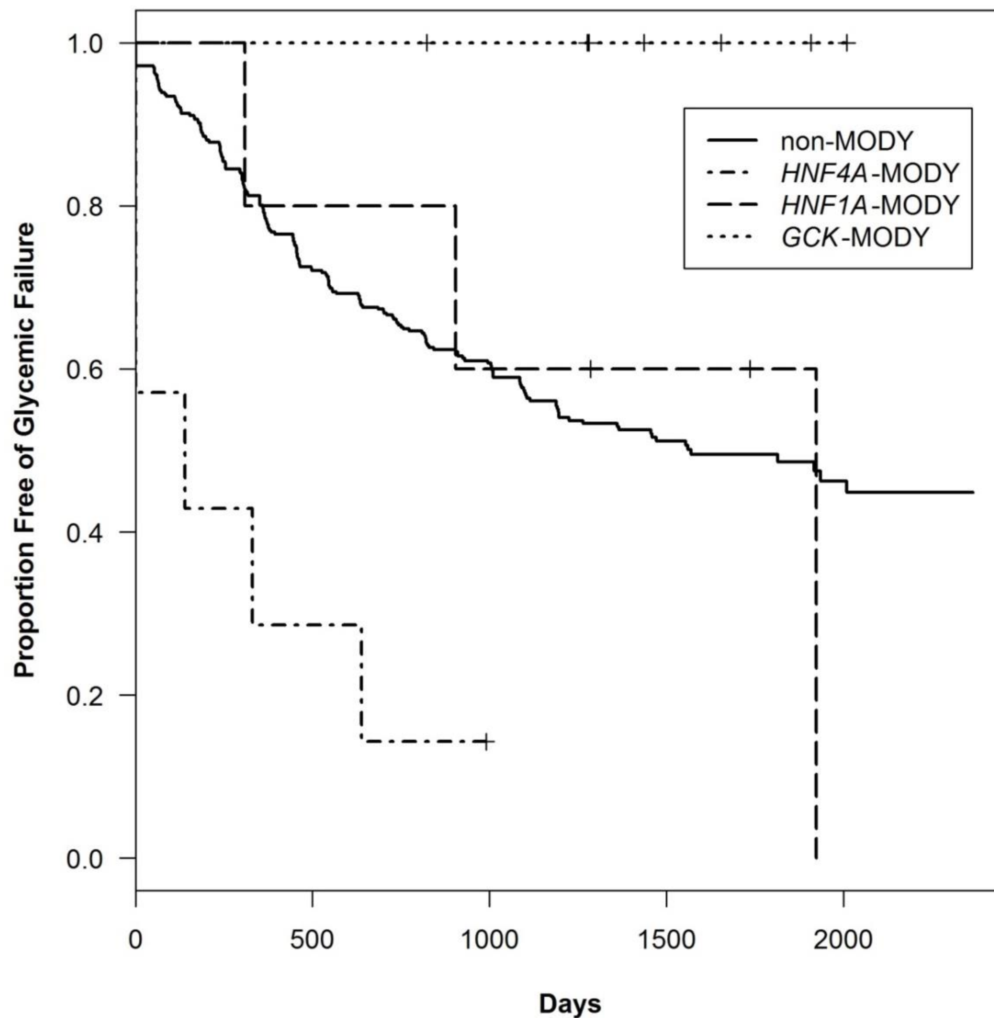


Figure 2.1 Failure-free survival curve of MODY gene subtypes and patients without monogenic diabetes.

Plot of event-free survival of specific MODY gene subtypes. X-axis is defined as post-baseline-visit days until start of failure interval. Events are defined as elevated glycated hemoglobin (>8.0%) over a period of 6 months or the inability to wean the participant from insulin within 3 months after treatment initiation in the TODAY study. Hazard ratio for each subtype: *GCK*-MODY HR= undefined (no events), *HNF1A*-MODY HR=1.26 (0.40-4.02 95%CI) p=0.7, and *HNF4A*-MODY HR=5.03 (2.18-11.58 95%CI) p=0.0002.

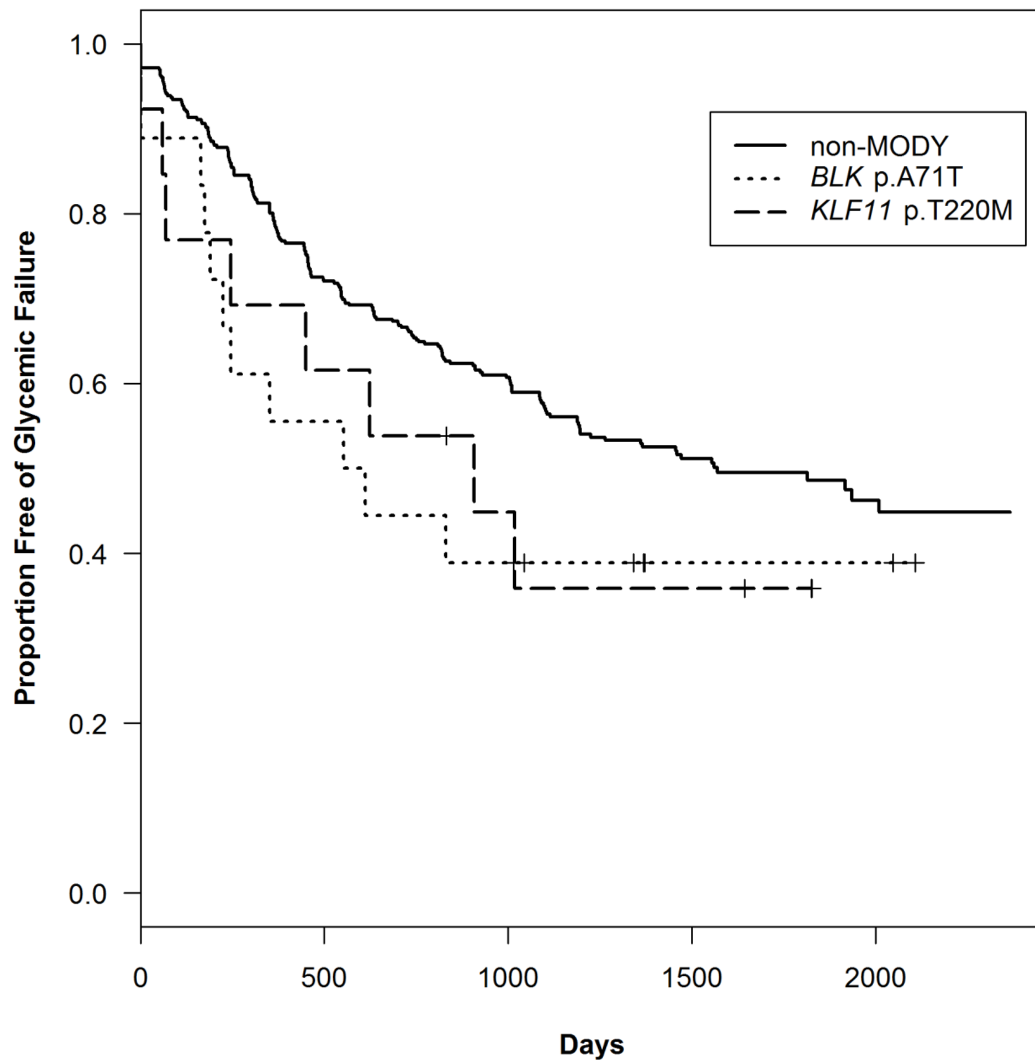


Figure 2.2: Failure-free survival curve of non-rare cited MODY variants
Plot of event-free survival of non-rare cited MODY variants. Events are defined as elevated glycated hemoglobin (>8.0%) over a period of 6 months or the inability to wean the participant from insulin within 3 months after treatment initiation in the TODAY study. Hazards ratios for each MODY variant: *BLK* p.A71T HR= 1.51 (0.82-2.78 95%CI) p=0.2 and *KLF11* p.T220M HR=1.36 (.066-2.78 95%CI) p=0.4.

4. T2D Risk Polymorphism Analysis

Using data from genotyping analysis of the TODAY subjects, both weighted and unweighted analysis of T2D risk allele burden was performed. This analysis attempted to determine if those patients with monogenic diabetes had a lower T2D risk allele burden than those with T2D in the study. Thirty-nine SNVs associated with increased T2D risk on the genotyping array were analyzed in the TODAY subjects. Total risk allele burden was not different between those with and without monogenic diabetes (45.0±4.6 SD vs. 45.7±4.2 SD variants per individual, p=0.5). When weighting risk variants by effect sizes, there was still no difference between groups (-0.04±0.77 SD vs 0.002±1.000 SD of z-score for total variant burden effect size, p=0.8).

D. Discussion

Based on our findings, an appreciable number of youth diagnosed with T2D may, in fact, have undiagnosed monogenic diabetes. Individuals with monogenic diabetes participating in TODAY displayed subtle, but significant, differences in select characteristics from those without pathogenic or likely pathogenic monogenic diabetes variants. However, in this adolescent population sample selected for being overweight or obese and having non-autoimmune, C-peptide positive diabetes, it was not possible to reliably distinguish between T2D and monogenic diabetes based on clinical features at baseline in this study. Patients were on metformin therapy at baseline in this study, which is likely to alter clinical features compared to patients that have not initiated therapy. Importantly, patients with monogenic diabetes were found in each race/ethnicity examined, although in a higher proportion of NHW participants (Table 2.2) likely because previous monogenic diabetes studies have mainly focused on European-ancestry

populations and the variant classification process relies heavily on that previously published information. We also confirmed the importance of a genetic diagnosis of monogenic diabetes to inform treatment protocols since metformin, while the first line treatment of T2D, was ineffective in those who turned out to have *HNF4*-MODY. Further emphasizing the known relationship between genetic diagnosis and treatment protocols, those with *GCK*-MODY did not reach the primary outcome of the study, since patients with *GCK*-MODY do not generally have HbA1c that exceeds the study endpoint (>8.0%) and typically do not require any drug therapy. These findings have strong implications for clinical practice.

At least 4.5% of TODAY participants (22/488) had pathogenic or likely pathogenic monogenic diabetes variants (Table 2.6). It is estimated that there are approximately 3,700 new diagnoses of T2D in youth each year in the US, and our findings indicate that as many as 160 of those cases or more could be misdiagnosed cases of monogenic diabetes.[108] The SEARCH study for diabetes in youth previously found 8% of participants with diabetes-associated autoantibody-negative, C-peptide positive diabetes had monogenic diabetes variants in *GCK*, *HNF1A*, and *HNF4A*.[109] Similar to the SEARCH study, we found no differences in family history of diabetes, as measured by reported maternal and paternal diabetes status, between TODAY patients with and without monogenic diabetes, and we found comparable racial/ethnic distributions of patients with monogenic diabetes between TODAY and SEARCH (Table 2.2). Compared to the SEARCH study, we probably found a lower percentage of patients with monogenic diabetes because the TODAY study eligibility criteria required adolescents with diabetes to be overweight or obese. Additionally, it is possible for individuals to have monogenic

diabetes and coincident insulin resistance or T2D, which can further complicate both diagnosis and treatment. Comparison of patients with monogenic diabetes to those without for both insulin and glucose graphically showed no obvious segregation of the groups (Figure 2.3). Performing the same graphical comparison of insulinogenic index and insulin sensitivity as an indication of insulin response does not have any obvious segregation between groups either (Figure 2.4). These analyses could either indicate that patients with monogenic diabetes cannot be identified by these characteristics or that the patients in this cohort have monogenic diabetes in addition to insulin resistance or T2D.

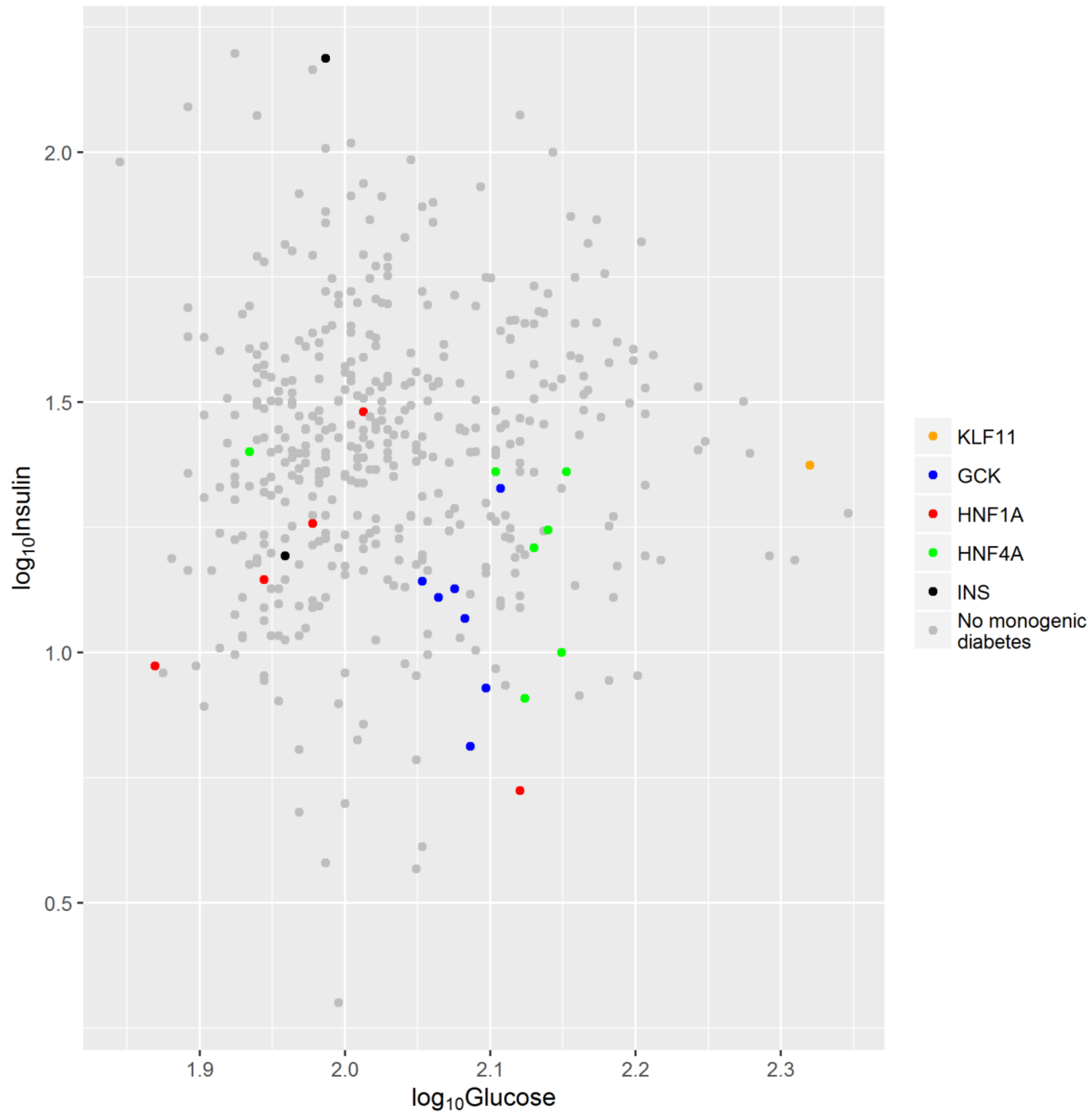


Figure 2.3: Scatterplot of log-adjusted fasting glucose and fasting insulin separated by monogenic diabetes status and gene etiology

Figure displays the log-adjusted fasting glucose values on the X-axis and the log-adjusted fasting insulin values on the Y-axis for each TODAY participant sequenced in the study. Patients without monogenic diabetes are marked with gray points, while those with monogenic diabetes are colored based on gene etiology found to the right of the scatterplot.

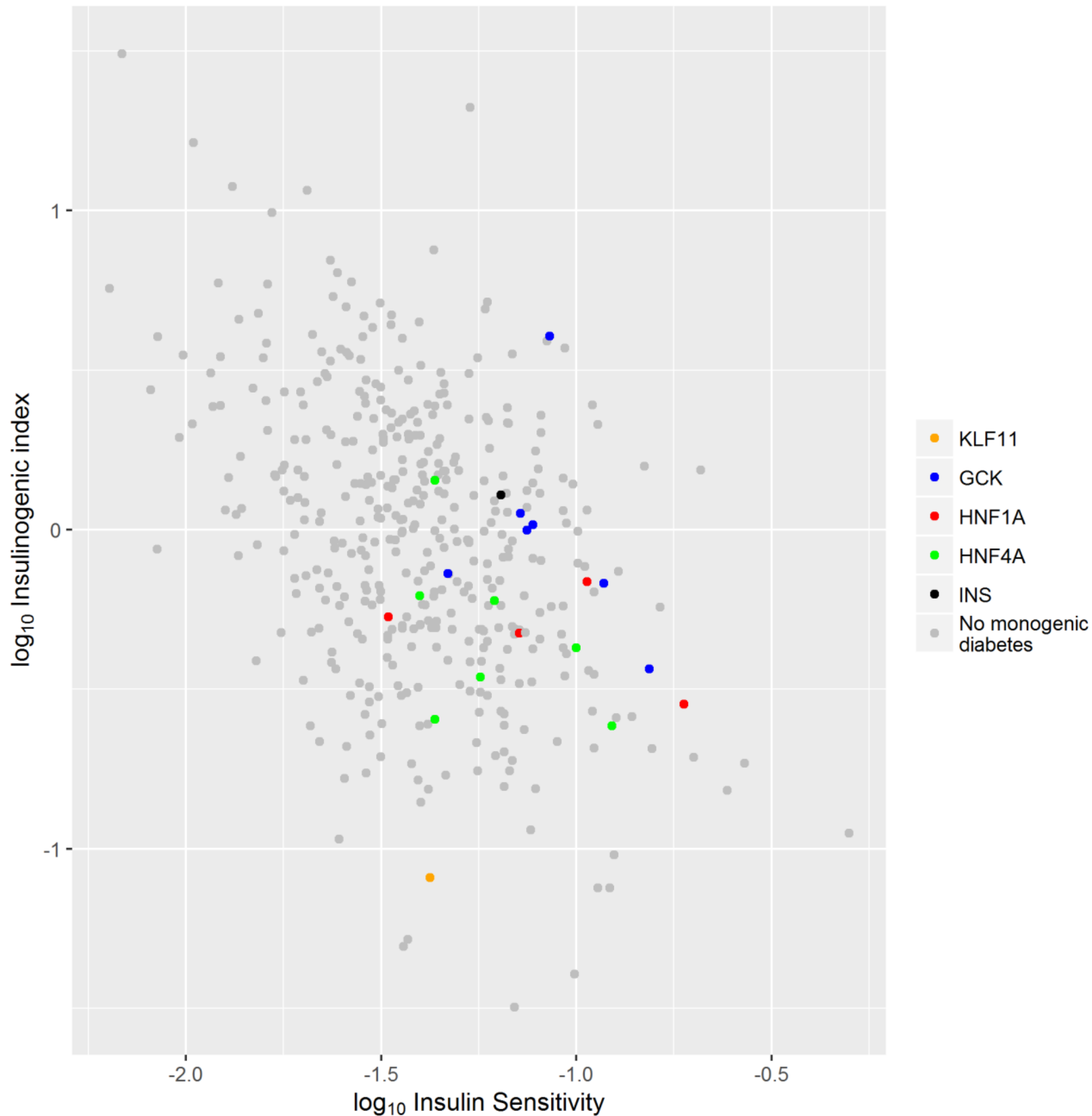


Figure 2.4: Scatterplot of log-adjusted insulin sensitivity and insulinogenic index separated by monogenic diabetes status and gene etiology

Figure displays the log-adjusted insulinogenic index (change in insulin over the first 30 minutes of an oral glucose tolerance test divided by the change in glucose over the same timeframe) values on the X-axis and the log-adjusted insulin sensitivity (defined as the inverse of the fasting insulin concentration) values on the Y-axis for each TODAY participant sequenced in the study. Patients without monogenic diabetes are marked with gray points, while those with monogenic diabetes are colored based on gene etiology found to the right of the scatterplot.

By following the ACMG/AMP guidelines for variant annotation, our study has potentially limited our estimation of the prevalence of monogenic diabetes in the cohort. As shown in previous studies, this standardized variant classification process often reclassifies variants previously assumed to be pathogenic as benign, likely benign, or variants of uncertain significance (VUS) due to increases in genetic and phenotypic database information.[124] Patients in our cohort that had these "reclassified" variants were excluded from our statistical analysis, but separate analysis of clinical characteristics of those with the *BLK* p.A71T and *KLF11* p.T220M variants did not show differences from those without monogenic diabetes (Table 2.13, 2.14). Since their original discovery, these two variants have been found to be relatively common in specific populations (e.g., minor allele frequencies of 2.6% and 4.4% in the African subpopulation of 1000G respectively). Neither variant showed the same trends as the other gene-specific MODY groups, and *KLF11* p.T220M was associated with higher fasting insulin (p=0.03). A previous study has questioned the *BLK* p.A71T variant as a cause for MODY.[50] While these variants may have associations with diabetes or diabetes symptoms, they are unlikely to be causative for MODY. However, the ACMG/AMP guidelines often classify novel variants as VUS due to the lack of evidence to suggest the variant is either pathogenic or benign. Further study of the novel variants found in this study is a potentially fruitful topic for future research and could increase the estimate of monogenic diabetes prevalence in this study. Additionally, multiple variants were discovered in genes that cause either NDM or diabetes syndromes, but did not fulfill criteria for pathogenicity according to ACMG/AMP guidelines. These variants are likely to still influence the patient's phenotype. For example, a nonsense variant (p.R1054X) was discovered in the insulin

receptor gene (*INSR*) which is likely to affect the patient's insulin response, but did not fulfill enough criteria to classify as pathogenic or likely pathogenic in part due to the established pattern of inheritance. Although these types of variants are not clear-cut monogenic diabetes variants suitable for return to patients, they represent an opportunity for discovering new forms of monogenic diabetes or learning more about the effects that variants in monogenic diabetes genes can have on T2D.

We found a larger proportion of individuals with *HNF4A*-MODY than expected in our cohort (Table 2.6). While *HNF1A*-MODY accounts for approximately 30-50% of MODY diagnoses and *HNF4A*-MODY accounts for less than 10%, we observed more patients with *HNF4A*-MODY than with *HNF1A*-MODY.[15] Interestingly, *HNF4A*-MODY has been associated with increased birth weight and macrosomia in the neonatal stage, regardless of maternal genotype but exacerbated by the mother having the same mutation and the associated hyperglycemic intrauterine environment.[153] Macrosomia has been correlated with overweight or obese status through adolescence and adulthood.[154] Thus there could be an association between *HNF4A*-MODY and higher BMI that could cause *HNF4A*-MODY to be misdiagnosed as T2D when using BMI as a criterion for monogenic diabetes. We hypothesize that the TODAY study inclusion criteria (BMI ≥ 85th percentile for age and sex) may have created a selection bias toward *HNF4A*-MODY compared to the other gene-specific subgroups; however, this remains to be demonstrated. Further studies incorporating birthweight and prevalence of T2D misdiagnosis of patients with *HNF4A*-MODY are necessary to test this hypothesis.

We did not discover any clinical criteria to differentiate overweight or obese adolescents with monogenic diabetes from those with T2D in this cohort selected for

overweight or obese status and non-autoimmune, C-peptide positive diabetes. Individuals with monogenic diabetes had lower BMI Z-score (p=0.004), but all 22 adolescents with monogenic diabetes were still overweight (>85th percentile by age) per the TODAY study design. For each of the clinical characteristics measured in the TODAY study, values of patients with monogenic diabetes could be found throughout the range of values for patients with T2D (Table 2.10, Figures 2.5-2.8). The functional effects of genetic defects leading to an insulin secretion deficit were demonstrated as adolescents with monogenic diabetes had lower fasting plasma insulin (p=0.03) compared to those without monogenic diabetes. They also had higher fasting blood glucose (p=0.02) concentration. Similar to published studies on populations with broader patient demographics, our patients with monogenic diabetes have greater insulin sensitivity (defined as the inverse of the fasting insulin) compared to those in our cohort with apparent T2D, but characteristics such as lipid profiles and beta-cell function (insulinogenic index) were not different between the groups in this specific cohort selected by overweight/obese status and non-autoimmune diabetes.[155-157] In summary, this young cohort of overweight/obese autoantibody-negative, C-peptide positive diabetes, there were differences in the means of some metabolic characteristics, but it was not possible to identify any characteristics to reliably distinguish between those with and without monogenic etiologies. Other markers shown to help identify patients with monogenic diabetes, such as high-sensitivity C-reactive protein and plasma glycan profiles, were not collected in the TODAY study.[158,159]

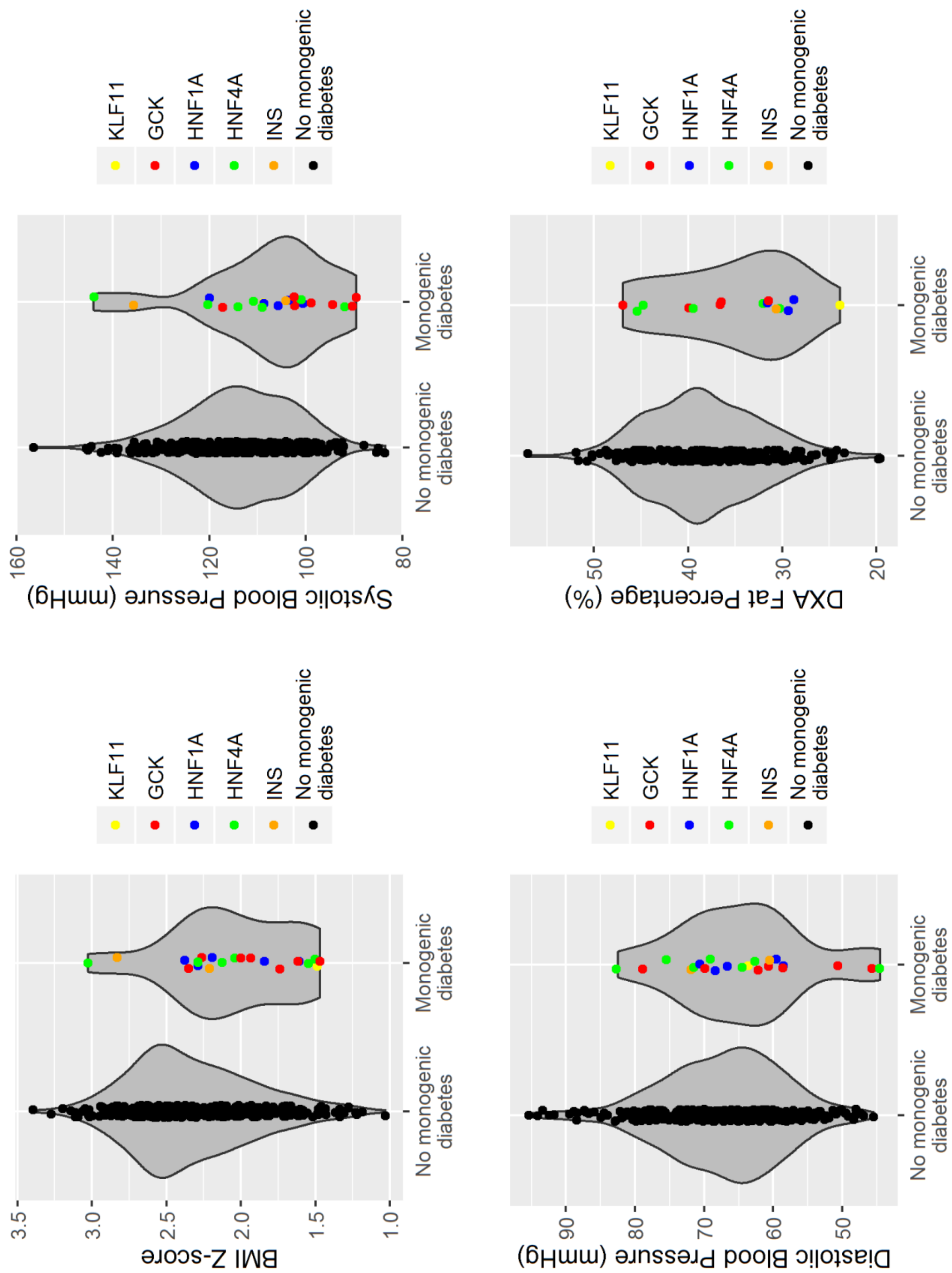


Figure 2.5: Violin plots of BMI Z-score, systolic blood pressure, diastolic blood pressure, and DXA fat percentage.

Plots of the distribution and data points for each of the listed characteristics separated for those TODAY study participants with or without monogenic diabetes. Genetic etiology of each participant with monogenic diabetes is defined by the color scheme listed to the right of each plot.

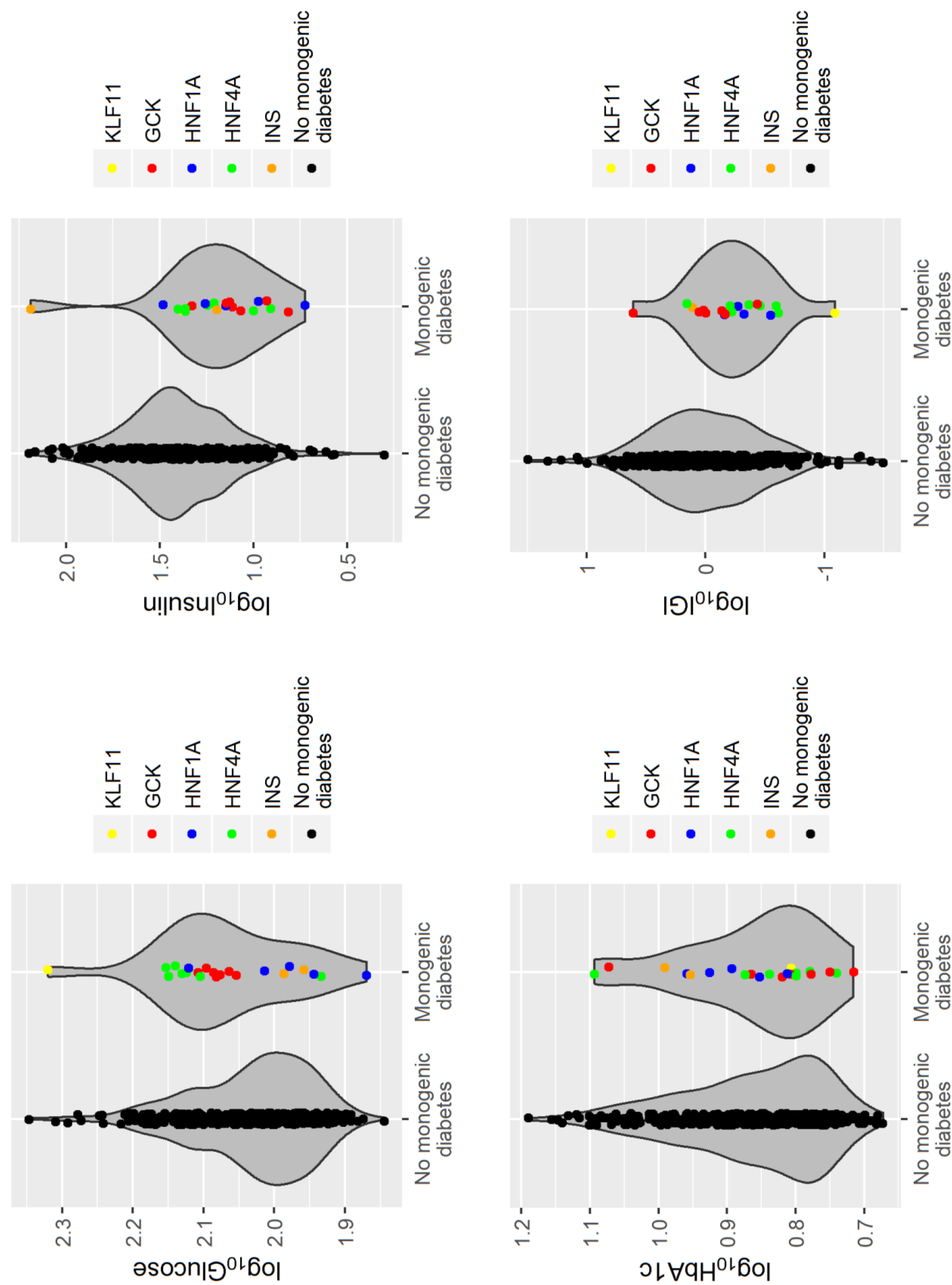


Figure 2.6: Violin plots of log-adjusted values of fasting glucose, fasting insulin, HbA1C, and insulinogenic index.

Plots of the distribution and data points for each of the listed log-adjusted glycemic characteristics separated for those TODAY study participants with or without monogenic diabetes. Genetic etiology of each participant with monogenic diabetes is defined by the color scheme listed to the right of each plot.

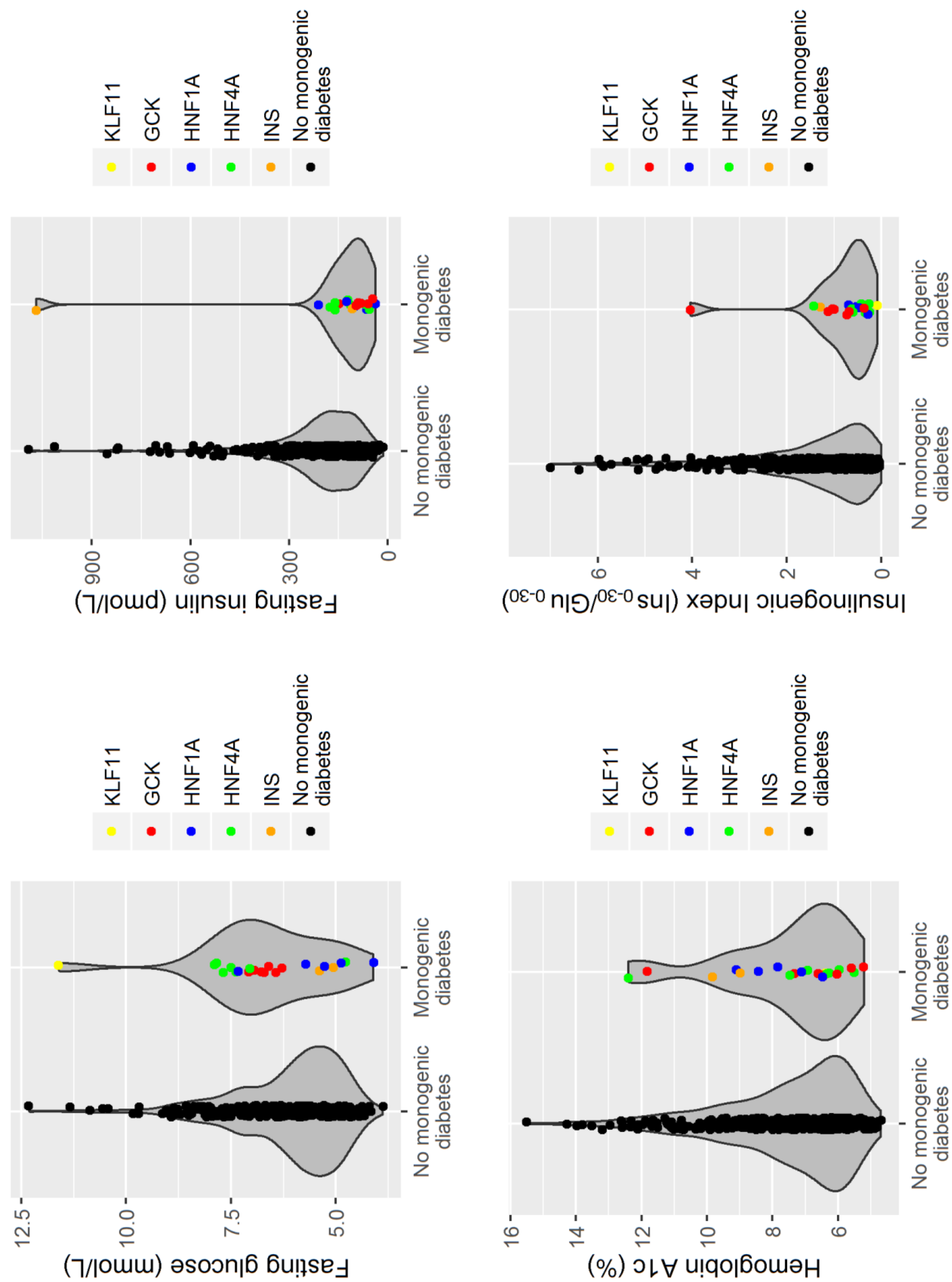


Figure 2.7: Violin plots of unadjusted values of fasting glucose, fasting insulin, HbA1C, and insulinogenic index.

Plots of the distribution and data points for each of the listed unadjusted glycemic characteristics separated for those TODAY study participants with or without monogenic diabetes. Genetic etiology of each participant with monogenic diabetes is defined by the color scheme listed to the right of each plot.

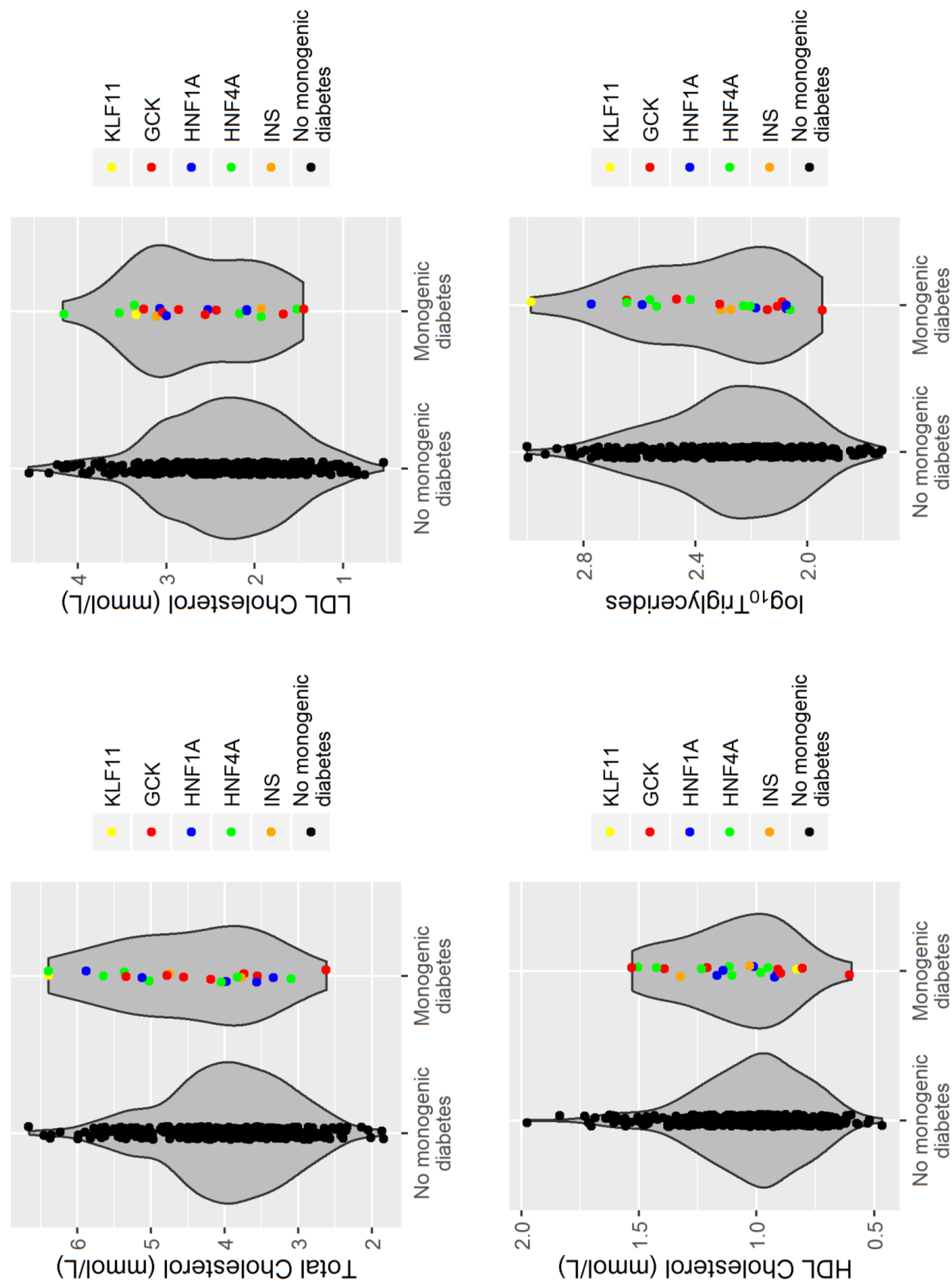


Figure 2.8: Violin plots of total cholesterol, LDL cholesterol, HDL cholesterol, and log-adjusted triglycerides.

Plots of the distribution and data points for each of the listed unadjusted cholesterol or log-adjusted triglyceride characteristics separated for those TODAY study participants with or without monogenic diabetes. Genetic etiology of each participant with monogenic diabetes is defined by the color scheme listed to the right of each plot.

Although the TODAY study provided unique information about the effectiveness of metformin alone or in combination with rosiglitazone or lifestyle changes, the small numbers of patients with each gene subtype of monogenic diabetes reduced our ability to draw conclusions regarding response to each of the treatment arms. However, none of the patients with *GCK*-MODY failed the treatment regimens in the TODAY study. This is consistent with the *GCK*-MODY phenotype of mildly elevated fasting blood glucose (fasting glucose of 5.49-8.66 mmol/L and HbA1c of 5.6-7.6%) that usually needs no treatment to avoid chronic complications of diabetes. In contrast, 6 of 7 patients with *HNF4A*-MODY failed treatment across study arms (HR=5.03 p=0.0002), indicating poor response regardless of therapies offered in the TODAY study. Similarly, though not statistically significant, 3/5 patients with *HNF1A*-MODY failed the TODAY study treatments. These results would be expected since the established treatment for *HNF1A*- and *HNF4A*-MODY are sulfonylurea drugs, rather than metformin and/or rosiglitazone, which have different mechanisms of action.[101,105] However, it is possible that in addition to sulfonylureas, metformin and/or thiazolidinediones may be appropriate for some monogenic diabetes patients with concomitant obesity and insulin resistance. While an analysis of interaction between monogenic diabetes subcategory and clinical trial treatment group would be of interest, this study is underpowered to perform such an analysis.

Genetic testing is not commonly implemented in diabetes clinical care due to current costs of testing, uncertainty over insurance reimbursement, and difficulty of interpretation of sequencing results. However, clinical care is moving into an era of genomic medicine, and monogenic diabetes provides a unique opportunity for immediate

implementation of personalized genomic medicine. Under specific conditions of cost and discovery rate, genetic testing for MODY has been modeled to be as cost-effective as current medical practices and potentially cost-saving.[133] As the cost and throughput of genetic testing continues to improve, the knowledgebase of rare genetic variation will continue to grow to inform clinical practice. Although current impediments to genetic testing such as cost, availability, and reimbursement may limit genetic testing for monogenic diabetes in large populations, this study has shown monogenic diabetes should be strongly considered as a possible diagnosis in young people with antibody-negative and C-peptide positive diabetes.

The findings from this study have strong implications for informing the practice of managing diabetes in youth. We discovered individuals with monogenic diabetes across all races/ethnicities in a cohort of overweight and obese adolescents diagnosed with T2D. While it is likely that more patients with monogenic diabetes could have been found in a non-overweight cohort, our discovery demonstrates that overweight or obese status does not exclude the possibility of having monogenic diabetes. Therefore, the currently recommended use of body type to select patients for genetic testing may lead to missing potential diagnoses of monogenic diabetes. We suggest that with secular trends of increasing obesity in children and adolescents, monogenic diabetes be considered as a potential etiology in diabetes-associated autoantibody-negative and C-peptide-positive adolescents regardless of BMI. Despite the small sample size of our cohort, treatment response based on monogenic diabetes diagnosis was consistent with predicted results, indicating the importance of monogenic diabetes genetic testing and proper genetic interpretation for providing optimal treatment to youth with diabetes.

Chapter 3: Monogenic Diabetes Screening, Diagnosis, and Pedigree Analysis in the PDMP Study[1]

A. Introduction

Monogenic diabetes is hyperglycemia caused by one or more genetic variants in a single gene and accounts for approximately 1-2% of all diabetes.[14] The most common monogenic form of diabetes is historically known as Maturity-Onset Diabetes of the Young (MODY); other categories include neonatal diabetes mellitus (NDM) and syndromic forms of diabetes. MODY has classically been diagnosed in non-insulin-requiring lean individuals before the age of 25, is inherited in an autosomal dominant fashion and has been attributed to mutations in 14 different genes.[131,160] NDM is hyperglycemia diagnosed within the first months of life and may be transient or permanent. Syndromic forms of monogenic diabetes include diseases such as Wolfram syndrome (*WFS1* gene), Alström syndrome (*ALMS1* gene), and severe insulin resistance (*INSR* gene), among several others, and comprise both diabetes and extrapancreatic features.

Based on prevalence estimates in European populations, monogenic diabetes is estimated to account for approximately 300,000 of the 29.1 million individuals with diabetes in the US. However, due to barriers such as lack of provider awareness, overlapping clinical characteristics with common forms of diabetes, as well as complexity, cost, and difficulty in gaining access to genetic testing, monogenic diabetes

---

[1] Jeffrey W. Kleinberger, Haichen Zhang, Tim O'Connor, Kristin Maloney, Linda Jeng, Kathleen Palmer, Devon Nwaba, Mickaela Nicholson, Yue Guan, Alan R. Shuldiner, and Toni I. Pollin. Monogenic Diabetes Screening, Diagnosis, and Pedigree Analysis in the PDMP Study. In preparation for submission.

is often misdiagnosed or underdiagnosed. A recent study found that over 90% of youths with the three most common forms of MODY were misdiagnosed as having more common forms of diabetes: type 1 (T1D) or type 2 diabetes (T2D).[109]

Failure to diagnose monogenic forms of diabetes can result in suboptimal therapeutic treatment. Patients with *GCK*-MODY have mildly elevated blood glucose without progression to microvascular or macrovascular diabetic complications, and therefore usually need no therapeutic intervention. Thus, those incorrectly diagnosed with T1D or T2D are often unnecessarily treated with insulin and/or oral medications and are exposed to their potential side effects.[34] Those with *HNF1A*-MODY can be effectively treated with low-dose sulfonylureas, even after years of insulin treatment, though some eventually require insulin.[101] Likewise, patients with *HNF4A*-MODY display a similar response to low-dose sulfonylureas, although it has not been studied to the same extent as *HNF1A*-MODY.[105,161] Additionally, NDM caused by genetic variants in *ABCC8* and *KCNJ11* has been effectively treated with high-dose sulfonylureas.[98,162] These oral medications are potentially more effective than metformin, the standard first line treatment for T2D, and they are potentially safer, more convenient, and often more effective than insulin injections, required for T1D, demonstrating the benefits of accurate genetic diagnosis of monogenic diabetes.

Multiple groups have suggested monogenic diabetes screening protocols are valuable for identifying individuals with monogenic diabetes. Some of these studies have suggested narrowing the clinical profile of those tested for monogenic diabetes based on common measures collected during diabetes treatment (e.g. insulin sensitivity), while other studies broaden the range of those who should undergo genetic testing.[163] For

example, a study comparing strict genetic testing criteria for MODY phenotype (age of diabetes diagnosis ≤ 25years, family history of diabetes, evidence of noninsulin dependence) with extended criteria (young onset, C-peptide-positive diabetes with a predominantly beta cell defect) found that half of those individuals with *HNF1A*- or *HNF4A*-MODY would not have been tested under strict criteria. Understandably, expanding the criteria increased the genetic testing rate (six-fold).[164] Similarly, the second chapter of this dissertation provides evidence for widening diagnostic testing criteria of those considered for monogenic diabetes testing to include overweight and obese adolescents diagnosed with non-autoimmune diabetes. One interesting opportunity for screening for *GCK*-MODY would be through glucose tolerance tests performed in pregnant women being screened for gestational diabetes as part of routine care during pregnancy. Chakera et al., in fact, demonstrated that 1 of 3 patients with incidental hyperglycemia discovered during gestational diabetes screening while having a pre-pregnancy Body Mass Index (BMI) less than 25 had *GCK* genetic variants.[165] Screening using biomarkers such as plasma glycan profiles, high-sensitivity C-reactive peptide, and CD36, among others, have been suggested.[159,166,167] However, there is no consensus yet regarding standardized application of biomarker testing and how to apply them clinically.

Multiple screening algorithms have been suggested for clinicians to identify patients likely to have monogenic diabetes.[168-170] These algorithms rely on published epidemiological data (most often from European ancestry cohorts) to suggest gene-specific tests while reducing the amount of unnecessary testing for monogenic diabetes. Since diagnosing MODY relies on multiple clinical criteria, one particularly interesting tool, a MODY Calculator, was developed by the University of Exeter to aid physicians

who may lack experience identifying patients with monogenic diabetes. The calculator, which is available online or as a smart phone application, requires input about the patient's glycemic characteristics, age of diagnosis, BMI, and family history and returns the likelihood that the patient has a MODY variant.[171,172] While this tool may be useful for physicians, it is based on the European Caucasian cohort of the University of Exeter, and may not be accurate for populations of other races/ethnicities common in the US.

The Personalized Diabetes Medicine Program (PDMP) at the University of Maryland is an implementation study to create, disseminate and evaluate a sustainable program for screening, diagnosing, and promoting individualized treatment for patients suspected to have highly-penetrant monogenic forms of diabetes (Figure 3.1). The PDMP was designed to measure the impact of this genetic diagnosis on the patient, physician, and healthcare resource utilization. This specific project within the PDMP tested the hypothesis that a simple screening protocol and family history data could effectively enrich and potentially identify patients with monogenic diabetes from a diabetes clinic population.

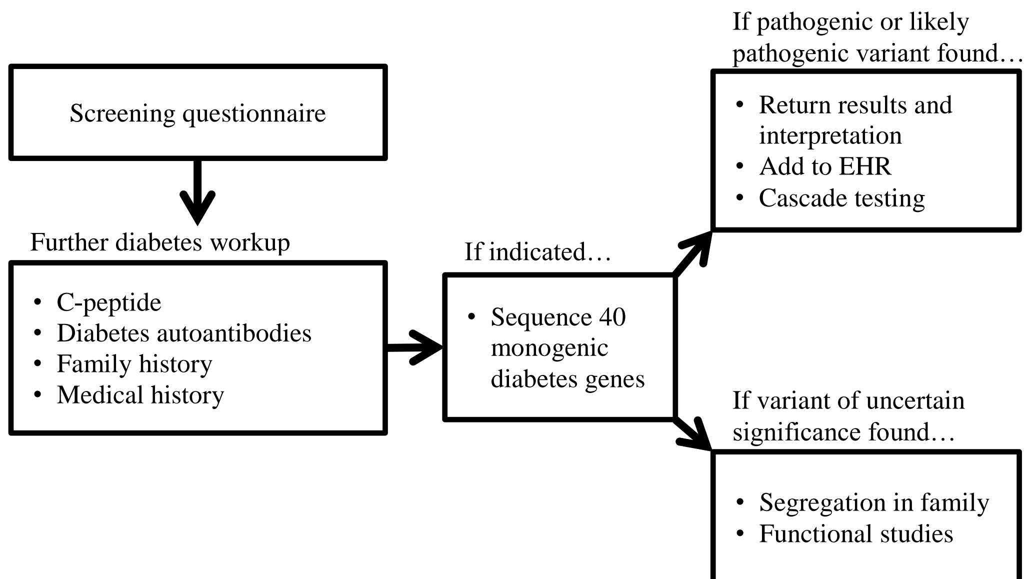


Figure 3.1: PDMP study overview
Flow chart of screening, visit 1, genetic testing, and results process for each patient. EHR: Electronic Health Record.

B. Methods

1. Recruitment Sites

The PDMP is centered at the University of Maryland Center for Diabetes and Endocrinology (UM-CDE), an academic center providing multidisciplinary treatment for children and adults with diabetes and other endocrine conditions in Baltimore, MD. Satellite study locations were selected to include different types of practices (an integrated health system, a large suburban private practice and a Veteran's administration hospital), caring for patients of a variety of races, ages, and locations. The Baltimore Veterans Administration Medical Center (BVAMC) is a study site adjacent to the University of Maryland Medical Center that provides inpatient, outpatient, and primary care services. The Geisinger Health System (GHS), spanning both rural and metropolitan communities of central and northeastern Pennsylvania, is an integrated healthcare system

with a comprehensive electronic health record (EHR) to recruit and track patients. Finally, Bay West Endocrinology Associates is a large, suburban private endocrinology practice in the Mid-Atlantic region working with MODEL Clinical Research, LLC to recruit participants for the PDMP. By enrolling patients from each of these sites, the PDMP draws from a patient population of nearly 40,000 patients currently being treated for diabetes, and potentially 2,500 new patients each year. The overall patient population is racially diverse (56% Caucasian, 39% African-American, and 5% other).

2. Screening Process

Patients are screened using a simple screening questionnaire provided in the waiting area of each site. The questionnaire includes seven questions to gain general information about the patient's age of diabetes diagnosis, family history of diabetes, and conditions suggestive of syndromic forms of diabetes (Table 3.1). Questionnaires are collected by the administrative staff at each site and are evaluated by on-site PDMP study coordinators in consultation with site principal investigators (PIs) and in some cases, the study PI. When available, the patient's electronic health record (EHR) is reviewed under a HIPAA waiver for prior diabetes laboratory test results including: GAD65 (glutamic acid decarboxylase autoantibody), IAA (insulin autoantibodies), IA-2 (insulinoma-2-associated autoantibodies), ICA (islet cell cytoplasmic autoantibodies), ZnT8A (zinc transporter 8 autoantibodies), and fasting C-peptide measures. The questionnaires and diabetes auto-antibodies/C-peptide results are analyzed in the context of the inclusion criteria (Table 3.2) to identify patients suspected to have monogenic diabetes, who are then invited to the first visit of the study. If the diabetes laboratory testing results are not available, the patient is enrolled, and their blood sample is sent off for commercial

laboratory testing for GAD65, IA-2, and C-peptide (if the patient is using insulin therapy). Additionally, there is a referral option available to physicians caring for patients they suspect to have monogenic diabetes. Patient characteristics of those referred by physicians are reviewed by study coordinators to determine if they fulfill one of the categories of inclusion criteria. Because clinical judgement based on extensive physician experience working with many patients with diabetes is potentially a valuable tool for diagnosis of monogenic diabetes, our study includes a provision for enrolling patients based on physician intuition. Patients that do not fit any other inclusion criteria may instead be enrolled based on "high suspicion of monogenic diabetes."

Table 3.1: PDMP screening questionnaire*

1. Were you diagnosed with diabetes or high blood sugar before 1 year of age?
2. Were you diagnosed with diabetes or high blood sugar at age 30 or younger?
   How old were you when you were diagnosed?
3. Were you extremely overweight when you were diagnosed?
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?
5. Do you have type 1 diabetes (if unsure, were you on insulin at diagnosis and have been ever since)?
6. Do you have a parent or a child with type 1 diabetes?
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.

*The newest version of the screening questionnaire also allows the patient to select their form of diabetes (T1D, T2D, GDM, etc.).

Table 3.2: PDMP inclusion criteria

1. Diagnosed with diabetes $\leq$1 year of age
2. Diagnosed with type 1 diabetes and has a parent or child with type 1 diabetes
3. Diagnosed with non-type 1 diabetes $\leq$ 30 years old
4. Diagnosed with type 2 diabetes $\leq$ 45 years old, not obese at diagnosis, and 2+ relatives with diabetes diagnosed $\leq$ 50 years old
5. Diabetes plus extra-pancreatic features of a genetic diabetes syndrome
6. Fasting glucose $\geq$ 100 mg/dl during a pregnancy without pre-existing diabetes and a pre-pregnancy BMI < 25
7. Clinical suspicion of highly penetrant genetic form of diabetes

Patients invited to the first study visit meet with a study coordinator and a genetic counselor. The study coordinator informs the patient of the study processes, reviews the potential outcomes of the study, obtains informed consent for enrollment, and collects a detailed medical history. A genetic counselor collects a detailed family history of the patient, comprising a three-generation pedigree focused on characteristics related to diabetes, diabetes complications, and extrapancreatic features found in monogenic diabetes syndromes (Figure 3.2). During the visit, blood samples are collected for multiple purposes: to measure diabetes autoantibodies and C-peptide (if not available in the patient chart or from the referring provider); to rule out the possibility of T1D; to provide a sample for targeted multi-gene next generation sequencing (as well as for

banking blood serum and plasma); and to obtain a sample to extract and hold DNA for confirmation of clinically actionable (pathogenic or likely pathogenic) variants in the CLIA/CAP accredited component of the Translational Genomics Laboratory (TGL).

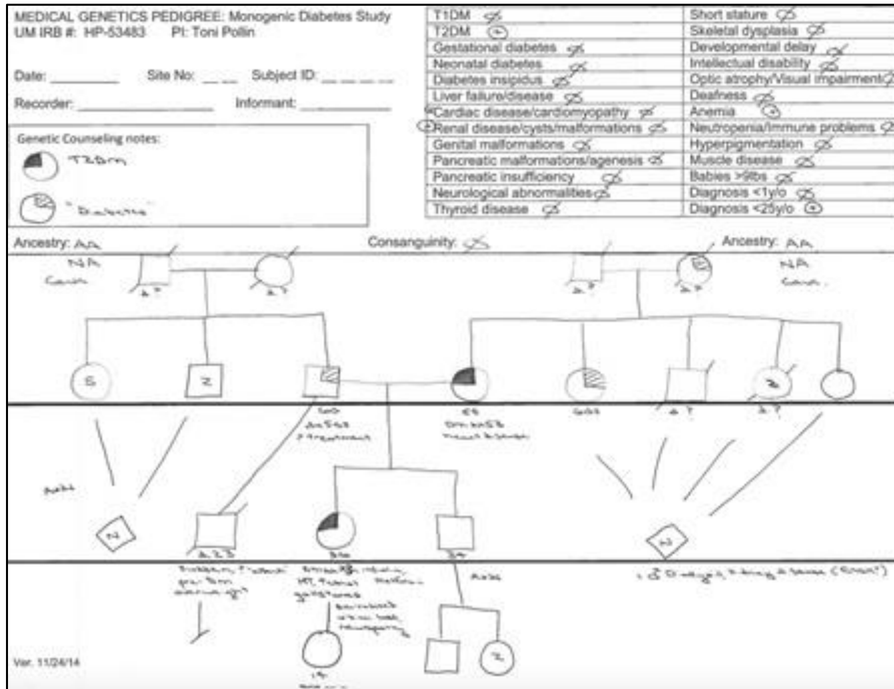


Figure 3.2: Example pedigree collected in PDMP
Figure represents an example pedigree collected by a genetic counselor during the enrollment study visit. The upper left corner contains the pedigree legend. The upper right corner contains a checklist for 26 characteristics that could inform about monogenic diabetes status. The body of the document contains the 3+ generation pedigree representative for the level of detail available for each family member.

For data collection and analysis, study coordinators deposit patient data into RedCap, a secure database web application compliant with HIPAA standards. Each patient entry extensively documents that patient's study, including responses to screening questionnaires, dates of study milestones, and pertinent EHR information for study eligibility. The entire RedCap PDMP dataset was downloaded on 4/25/2017 for analysis in this study. Additionally, pedigree information from the family medical history

collected by the study genetic counselor is input into Progeny software for computational analysis. Progeny is a risk modeling and pedigree software system that provides pedigree-drawing capability with a relational database for collecting phenotype data. A customized Progeny data format was created based on the monogenic diabetes-specific information collected by the genetic counselor. The Progeny output includes a separate entry for each individual in the pedigree, including their monogenic diabetes-specific data as well as basic information (age, age of death, etc.) and pedigree relationship information (mother, father, maternal/paternal relation, etc.). Progeny data was collected as patients underwent genetic testing, with 152 family entries included at the time of analysis.

3. Sequencing Process

Genetic testing for monogenic diabetes is initially performed using a next-generation sequencing platform. Separate blood samples from patients are delivered to the University of Maryland Biobank and the TGL. The Biobank uses an automated system for DNA extraction, quantification, and storage of research samples. Samples from the Biobank are used for initial genetic testing. Testing is performed using a custom gene panel of 795 amplicons covering the coding and flanking intronic regions of 40 monogenic diabetes genes. The genes are those known to cause forms of monogenic diabetes, including: MODY, NDM, syndromic forms of monogenic diabetes (e.g., Wolfram Syndrome, Roger's Syndrome, etc.), lipodystrophy, or hyperinsulinemic hypoglycemia (where mutations have the potential to alternatively cause hyperglycemia), as well as monogenic forms of obesity (Table 2.3). Each sample is amplified and ligated with short oligonucleotide barcodes to allow multiplexed sequencing. Pooled samples

undergo emulsion PCR followed by sequencing on an Ion Torrent Personalized Genome Machine. Raw data is processed, aligned, variants are called using TMAP software and Ion Torrent Server software. All samples with >10x coverage on more than 90% of the total designed regions are analyzed. This entire process is performed in the TGL under a research protocol.

Patient variants are initially analyzed through a customized pipeline incorporating ANNOVAR annotations (allele frequencies-1000 Genomes, ExAC, and Exome Sequencing Project, in silico predictors – SIFT, PolyPhen2, Provean, MutationTaster, MutationAssessor, LRT, MetaLR, MetaSVM, FATHMM, CADD, GERP++), as well as comparisons to both openly-available databases (type2diabetesgenetics.org, HGMD-public version) and private datasets sequenced using the same panel (i.e. the TODAY cohort).[173-175] Rare/uncommon (<5% MAF) protein-altering variants are then manually evaluated according the American College of Medical Genetics/Association for Molecular Pathology Standards and Guidelines for the Interpretation of Sequence Variants (ACMG/AMP Guidelines) using a tool we have created to calculate pathogenicity based on manual input of individual criteria.[119,176] This process is performed through in-person meetings including those performing the sequencing/analysis, the TGL Director, a genetic counselor, and the PDMP PI. Variants classified as pathogenic or likely pathogenic according to ACMG/AMP guidelines are then clinically confirmed in the TGL. The process for maintaining separation between clinical and research activities during sample analysis is depicted in Figure 3.3.

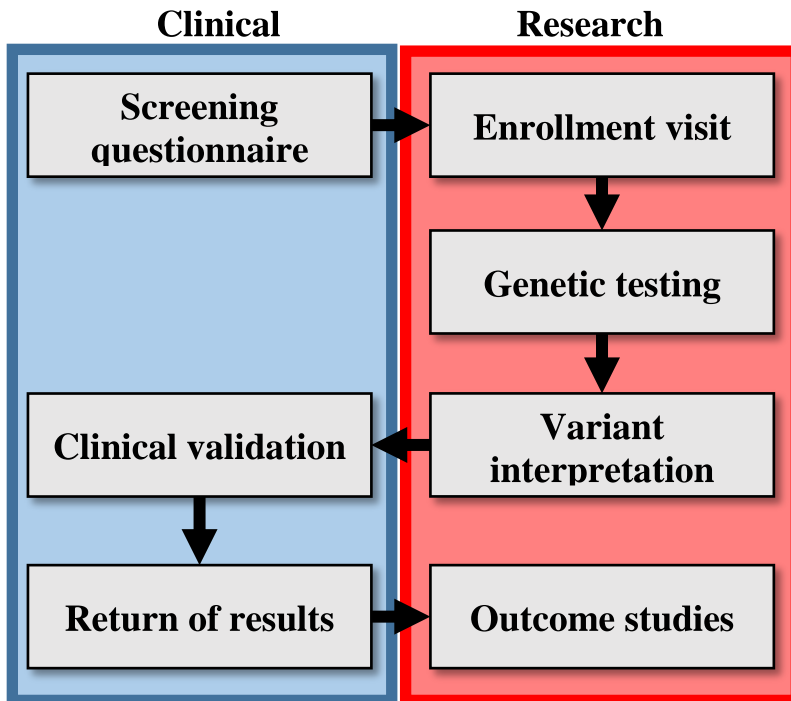


Figure 3.3: Clinical and research process separation of the PDMP
Flow chart representing the process for crossing the barrier between research and clinical activities in the PDMP. This separation is necessary for protecting study participant identifying information. Note that the screening questionnaire is a part of the research project that physically takes place in the clinic.

At the time of the patient's first study visit, blood samples are sent directly to the TGL for potential future clinical validation. DNA is extracted from the samples and stored in the TGL for 12 months under an "Extract and Hold" protocol. When a pathogenic or likely pathogenic variant is discovered through the research sequencing process, a request for a "Confirmation of Research Finding" request is made. The DNA extracted and stored in the TGL is then analyzed for the variant of interest through Sanger sequencing. A sample of the same individual's DNA extracted under a research protocol from the Biobank is also provided as a positive control for sequencing. Throughout this process, no PDMP study identifiers are included to maintain patient confidentiality in the study.

4. Return of Results

After clinical confirmation of a pathogenic or likely pathogenic variant, the TGL Director and TGL Genetic Counselor write a comprehensive laboratory report regarding the variant, with input from the PDMP Principal Investigator. This report describes how the patient's individual and family medical histories indicated testing for monogenic diabetes. It then reports the variant using standard nomenclature and describes the information used to determine the pathogenicity of the variant, such as previous citations, prevalence in (or absence from) major population databases, or pathogenicity predictions from *in silico* tools. Finally, the report explains the potential clinical and therapeutic implications of the variant and recommends genetic counseling and possible referral to an endocrinologist.

Patients receiving notification of a clinically significant genetic test result have an in-person (ideally) or telephone visit with a genetic counselor and a medical geneticist and/or endocrinologist. The purpose of this meeting is to disclose, explain, and discuss the test result and its potential implications for the patient's diagnosis, prognosis, and treatment as well as family members' risks and testing options. The patient is told not to make any changes in treatment but to return to his/her referring or other diabetes care provider. The patient is provided the TGL Laboratory Report, a copy of the clinical consult note, and a letter written to the patient by the genetic counselor in lay terms reviewing the information provided. The medical geneticist then adds the clinical consult note, genetic counseling letter, and test result to the patient's EHR. For patients referred from outside the main study sites, copies of these documents are mailed to the referring or treating provider.

Study participants without any clinically significant variants (i.e. all variants are classified as benign, likely benign, or variants of uncertain significance) receive a letter from the PDMP indicating that no clinically relevant variants were discovered. This letter is provided to assure patients that the testing has been completed. However, the letter is carefully written to indicate that the sequencing assay was research grade, the result should be considered uninformative rather than negative, and future clinical genetic tests for monogenic diabetes should not be precluded based on the PDMP results. The list of genes that were analyzed is not included in the patient's report to prevent interpretation of the test as a negative clinical result. Additionally, those with variants of interest may be invited to take part in family studies to determine how the variant segregates in their family. These patients without clinically significant variants are told this information through a telephone conversation with the PDMP Genetic Counselor in addition to receiving a summary letter.

5. Family History Machine Learning

Pedigrees were assessed for the overall affected status and percentage of affected family members, as well as using a machine learning algorithm to determine the ability to separate those with variants causing monogenic diabetes from those with no pathogenic or likely pathogenic variants. Using the Progeny pedigree software output, an original software program was written using Python computer language to collect data from the affected proband and to summarize their family data. Each pedigree's "Affectedness" metrics (A_i and A_g) were collected for each collected status trait according to previously published pedigree-description methods proposed by Gay et al.[177] Because those methods summarize pedigrees from a "top-down" approach utilizing a single ancestor as a starting

point, each primary ancestor (those with no further described ancestors) of the proband was classified as a “node”, and A_g and A_i was collected based on the descendants of that node (Figure 3.4). A_g is defined as the proportion of individuals affected in the node, and A_i is defined as the sum of the δ multiplied by the indirect diversification factor (IDF) for each branch of the node. The δ for each node member is defined as the affected status (0 or 1) multiplied by 2 raised to the power of generation number (zero-based numbering), and the IDF is the generation-based proportion of the node’s leaves. Data from each of the multiple nodes in each pedigree was collected and summarized by maximum and mean statistics, as suggested by Gay et al. Additionally, a customized metric, D_i, was created based on the previously cited algorithm (Figure 3.4). The D_i metric uses a “proband-out” recursive pedigree algorithm, and it is also defined as the sum of the δ for everyone in the branch multiplied by the IDF(Figure 3.4). For the D_i however, δ is defined as the affected status (0 or 1) multiplied by the inverse of 2 raised to the power of the degree of relatedness ($1 \times \frac{1}{2^0}$ for an affected proband, $1 \times \frac{1}{2^1}$ for an affected first-degree relative, etc.), and IDF is the degree-based proportion of the proband’s leaves. A D_g metric was also created to measure the proportion of proband-related individuals affected. This information, along with proband-specific data was converted into Attribute-Relation File Format (ARFF) format, and it was analyzed using the Naïve Bayes classifier machine learning algorithm in the Weka platform.[178] The machine-learning model was developed using 10-fold cross-validation on 152 pedigrees (13 cases and 139 controls) that had undergone genetic testing. The number of families is greater (152 versus 138) for this analysis because more patients had undergone sequencing than were input into the RedCap screening database at the same time.

A.

A_I **(node,era,idf,delta)**[177]	D_I **(proband,degree,idf,delta)**
1: **if** node is affected **then**	1: **if** proband is affected **then**
2: $newDelta = delta + 2^{era}$	2: $newDelta = delta + \frac{1}{2^{(degree)}}$
3: **else**	3: **else**
4: $newDelta = delta$	4: $newDelta = delta$
5: **end if**	5: **end if**
6: **if** node has offspring **then**	6: **if** proband has unassessed 1° relatives **then**
7: $indicator = 0$	7: $indicator = 0$
8: $newSegment = \frac{idf}{offspring\ size}$	8: $newSegment = \frac{idf}{number\ of\ 1°relatives}$
9: $newEra = era + 1$	9: $newDegree = degree + 1$
10: **forall** p in node's offspring **do**	10: **forall** p in proband's 1° relatives **do**
11: $indicator += A_I(p, newEra, newSegment, newDelta)$	11: $indicator += D_I(p, newDegree, newSegment, newDelta)$
12: **end for**	12: **end for**
13: **return** *indicator*	13: **return** *indicator*
14: **else**	14: **else**
15: **return** *newDelta* x *idf*	15: **return** *newDelta* x *idf*
16: **end if**	16: **end if**

B.

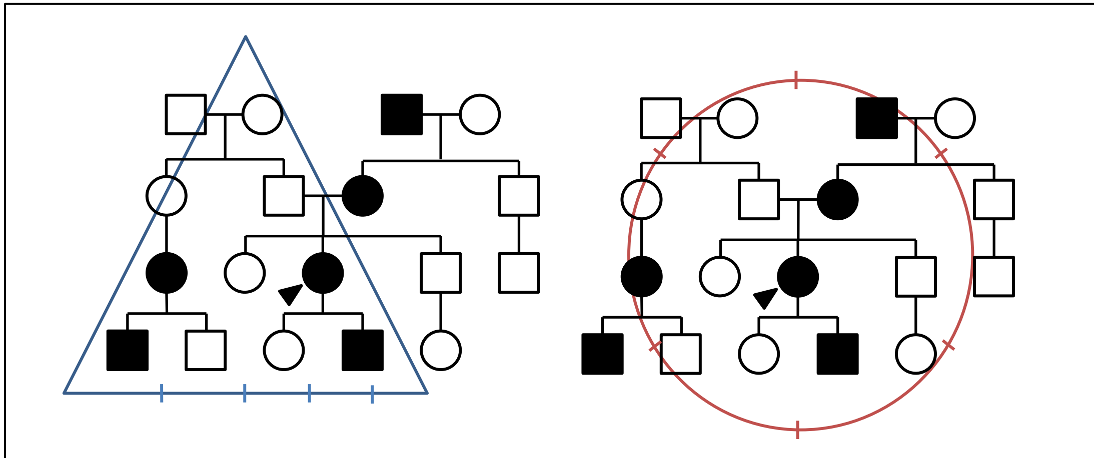


Figure 3.4: Algorithm and schematic for pedigree summary for machine learning

Figure displays A.) Recursive algorithms for measuring "top-down" A_i metric for each family from Gay et al. (left) compared to the adapted algorithm (right) for measuring "proband-out" D_i metric in this study and B.) a schematic of sampling effect for "top-down" versus "proband-out" methods for the same example pedigree. The left image represents one of the 4 A_i "nodes" (from the paternal grandmother), the group encompassed (the triangle), and the scale of effect size of each leaves' indirect diversification factor (IDF, ticked line at bottom of triangle). The right image represents the single "proband-out" D_i metric for the family, with the entirety of pedigree encompassed (the circle), and the first-degree of the indirect diversification factor (IDF, ticked edge of the circle).

6. Statistical Analysis

Summaries of PDMP patient populations based on study site, patient source, and screening responses were not quantitatively assessed for statistical differences since they were used only to describe the general characteristics of the PDMP study population. Chi-square test for independence was performed to compare distributions of participant enrollment or reason for lack of enrollment between Caucasian and African-American participants since there were sufficient numbers in each category to fulfill the assumption of the statistical test. Other races/ethnicities were not included in the analysis because the number of eligible participants was too small for analysis. Comparisons between distributions of individuals with and without monogenic diabetes were performed for responses to screening questionnaires and pedigree characteristics using Fisher's exact test to account for the small number of those with monogenic diabetes. For each statistical test, $p<0.05$ was considered significant.

C. Results

1. Screening Data

The PDMP has screened a total of 1,734 individuals across the four study sites over 2.5 years the study has been underway (Figure 3.5). Most of the patients have been from UM-CDE (34.8%) and GHS (41.5%) (Table 3.3). Most of the patients have been screened using paper questionnaires (Clinic screen, 51.3%) or through GHS's electronic medical record system that selects patients based on diabetes status (Targeted screen, 41.1%). There have been a relatively low proportion of individuals that have either been referred by physicians from outside institutions (Physician referral, 4.6%) or by independently inquiring about the study (Self-referral, 3.1%) (Figure 3.6). UM-CDE and

BVAMC have screened the largest proportion of African-American participants (60.2% and 63.6%), while GHS has screened the largest proportion of Caucasian participants (94.3%) (Table 3.4). BVAMC's screened population is the oldest (60.9 ± 10.4 years), while the UM-CDE population is the youngest (47.7 ± 16.5 years). According to screening questionnaire responses, most participants confirmed that they had more than two blood relatives with diabetes (60.8%) (Table 3.5). Approximately a third of those screened had been diagnosed with T1D (32.8%), been diagnosed before the age of 30 (32.2%), or been obese at the age of diagnosis (31.5%). Very few (1.6%) participants had been diagnosed before 1 year of age.

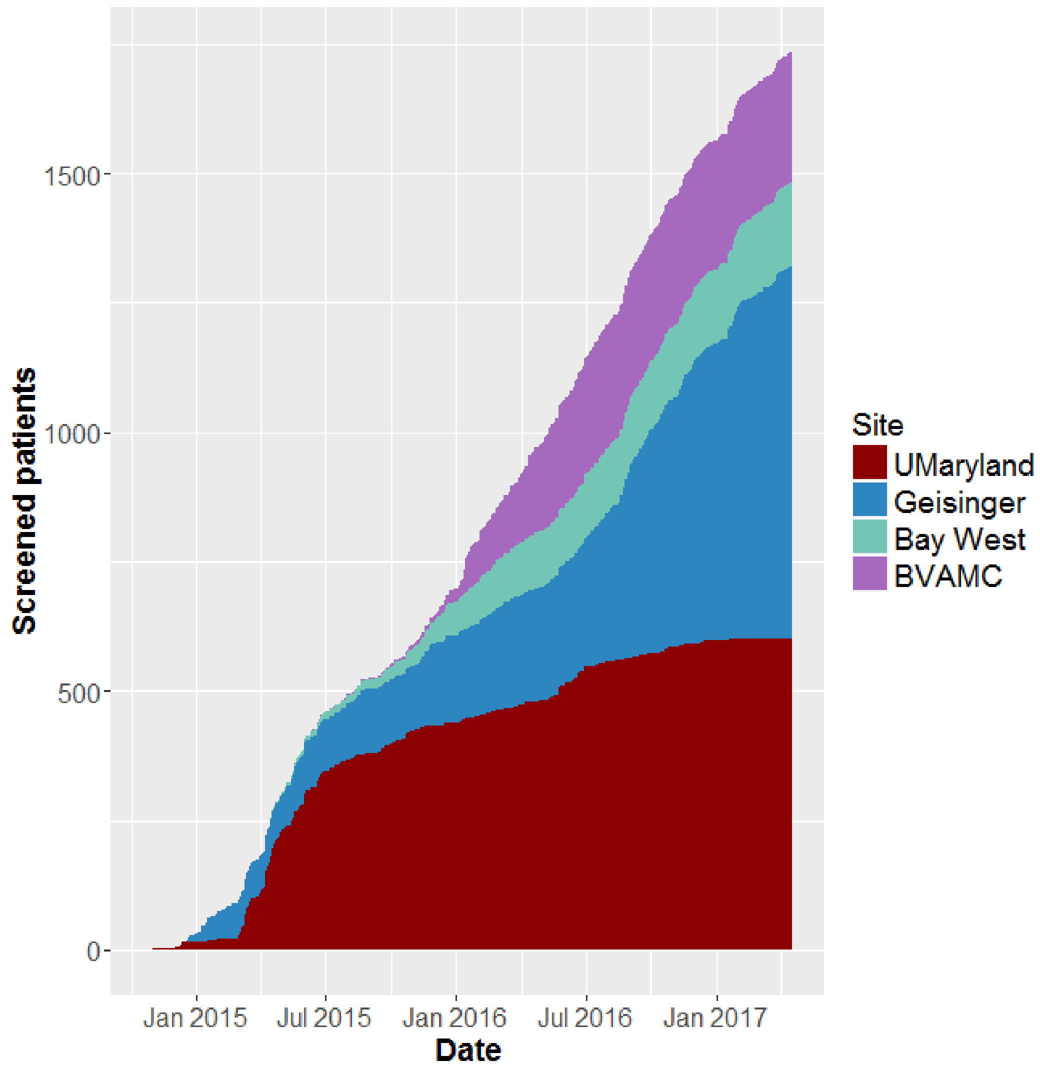


Figure 3.5: Plot of screened patients by site over the study period
Area plot of the number of screened patients separated by study site. UMaryland; University of Maryland Center for Diabetes and Endocrinology, BVAMC; Baltimore Veterans Affairs Medical Center.

Table 3.3: Distribution of PDMP participants by site and source

	UM-CDE	Geisinger	Bay West	BVAMC	Total
Physician referral	58	8	10	4	80
Self-referral	48	0	5	0	53
Clinic screen	497	0	146	246	889
Targeted screen	0	712	0	0	712
Total	603	720	161	250	1734

UM-CDE; University of Maryland Center for Diabetes and Endocrinology, BVAMC; Baltimore Veterans Affairs Medical Center.

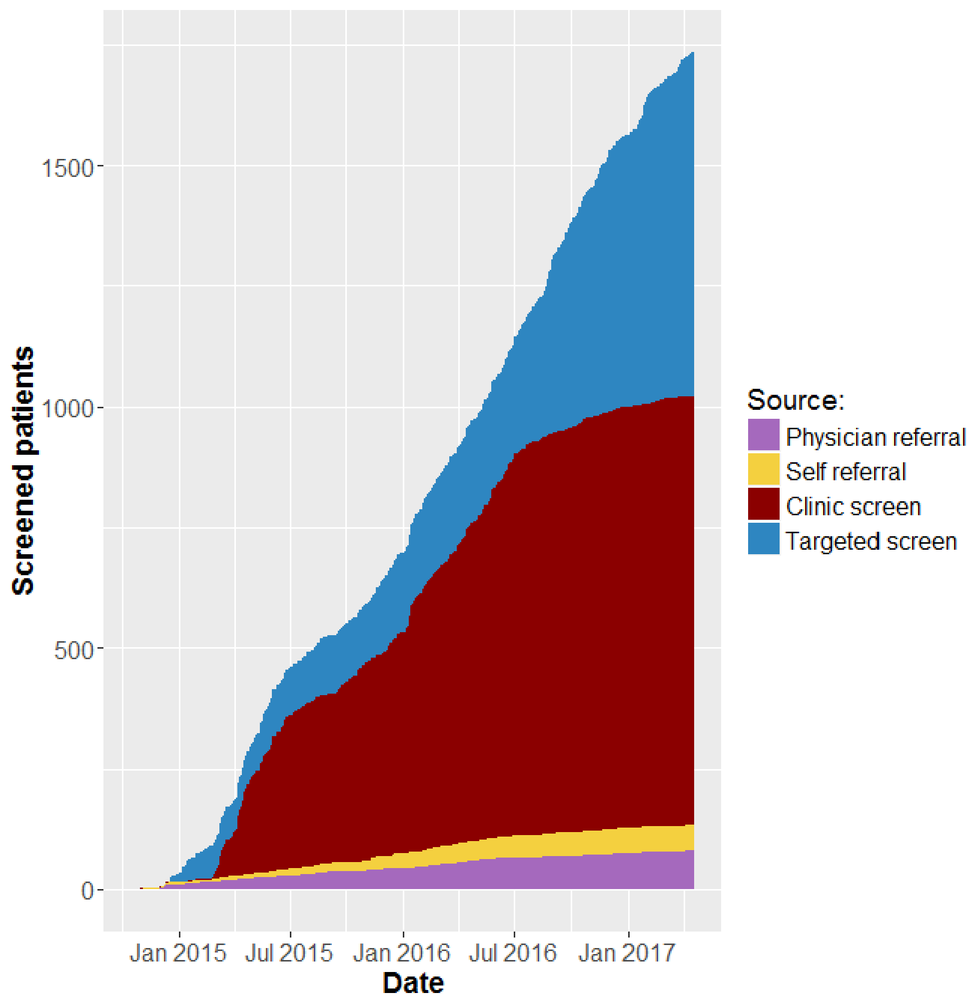


Figure 3.6: Plot of screened patients by source over the study period
Area plot of the number of screened patients separated by patient source.

Table 3.4: Demographics of PDMP participants by site

	UM-CDE (n=603)	Geisinger (n=720)	Bay West (n=161)	BVAMC (n=250)	Total (n=1734)
Age ± SD	47.7 ± 16.5	51.9 ± 19.0	57.1 ± 15.4	60.9 ± 10.4	52.2 ± 17.4
Female (%)	59.4%	56.8%	64.2%	8.5%	51.4%
Hispanic	2.4%	2.0%	1.4%	4.5%	2.3%
African-American	60.2%	2.8%	18.5%	63.6%	33.0%
Caucasian	35.7%	94.3%	79.0%	34.0%	63.8%
Asian	2.2%	1.1%	2.5%	0.4%	1.5%
Pacific islander	0.3%	0.3%	0%	0.4%	0.3%
Native American	1.0%	1.2%	2.5%	0.8%	1.2%

UM-CDE; University of Maryland Center for Diabetes and Endocrinology, BVAMC; Baltimore Veterans Affairs Medical Center.

Table 3.5: Screening questionnaire responses of all PDMP participants

PDMP Screening question	Percent Answered Yes
1. Were you diagnosed with diabetes or high blood sugar before 1 year of age?	1.6% (28/1732)
2. Were you diagnosed with diabetes or high blood sugar at age 30 or younger? How old were you when you were diagnosed?	32.8% (567/1729)
3. Were you extremely overweight when you were diagnosed?	31.5% (539/1713)
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?	16.4% (284/1728)
5. Do you have type 1 diabetes (if unsure, were you on insulin at diagnosis and have been ever since)?	32.2% (555/1724)
6. Do you have a parent or a child with type 1 diabetes?	13.6% (233/1717)
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	60.8% (1049/1726)

2. Enrollment

After using the screening information to determine eligibility, only 17.7% of all participants fulfilled one of the inclusion criteria for enrollment (Table 3.6). Of those that were eligible, 52% were enrolled in the study. Those who did not enroll were mostly unable to be contacted (17.1%) or declined participation (19.1%). While most of the participants that enrolled were Caucasian (64.7%) and/or female (61.5%), most people who declined participation or were unable to be contacted were African-American (58.6% and 53.8%) (Table 3.7). Compared to Caucasians, African-Americans were more likely to be unable to be contacted (odds ratio=1.6), declined participation (OR=2.3), or canceled their visit (OR=4.9) (overall distribution Chi-Square statistic = 29.5, $p<0.001$, Table 3.8). An updated version of the screening questionnaire in which patients provided their diagnosed type of diabetes showed that most of the patients were diagnosed with T2D (61.7%), followed by T1D (26.5%) (Table 3.9). Of the 1,734 participants screened in the PDMP, 156 individuals were enrolled in the study at the time of this analysis.

Table 3.6: Eligibility and enrollment by study site

	UM-CDE (n=603)	Geisinger (n=720)	Bay West (n=161)	BVAMC (n=250)	Total (n=1734)
Eligible	29.5%	11.5%	9.9%	12.4%	17.7%
n=	198	83	16	39	336
Unable to contact	16.6%	7.2%	0%	30.8%	17.1%
Declined participation	15.7%	15.7%	0%	33.3%	19.1%
Cancelled visit	12.1%	1.2%	0%	5.1%	9.0%
Enrolled	53.5%	28.9%	93.8%	28.2%	46.4%
n=	106	24	15	11	156
Tested (n=)	97	18	15	8	138

"Declined participation" also includes patients who were offered enrollment, but did not follow up (passive decline). "Cancelled visits" also includes scheduled visits that the patient did not attend without notice. UM-CDE; University of Maryland Center for Diabetes and Endocrinology, BVAMC; Baltimore Veterans Affairs Medical Center.

Table 3.7: Demographics of participants that enrolled compared to those who did not

	Enrolled (n=156)	Unable to contact (n=52)	Declined participation (n=53)	Cancelled visit (n=26)
Age (mean ± SD)	43.4 ± 16.2	49.8 ± 17.0	48.0 ± 17.4	45.0 ± 14.3
Female (%)	61.5%	40.4%	49.1%	59.3%
Hispanic	2.6%	0%	2.6%	5.9%
Afr. American	31.4%	53.8%	58.6%	70.4%
Caucasian	64.7%	40.4%	36.2%	18.5%
Asian	3.8%	0%	1.7%	0%
Pacific Islander	0%	0%	0%	3.7%
Native American	0.6%	0%	0%	0%

"Declined participation" also includes patients who were offered enrollment, but did not follow up (passive decline). "Cancelled visits" also includes scheduled visits that the patient did not attend without notice.

Table 3.8: Comparison of enrollment categories between African-Americans and Caucasians

	African-American (n=129)	Caucasian (n=147)
Unable to contact	27	21
Declined participation	34	20
Cancelled visit	19	5
Enrolled	49	101

Chi-square statistic is 29.5 ($p<0.001$) for difference between groups. Odds ratios of African-Americans compared to Caucasians: "Unable to contact" =1.6, "Declined participation" =2.3, "Cancelled visit" =4.9, "Enrolled" =0.3. "Declined participation" also includes patients who were offered enrollment, but did not follow up (passive decline). "Cancelled visits" also includes scheduled visits that the patient did not attend without notice.

Table 3.9: Diabetes diagnoses in PDMP screened subset

	UM-CDE (n=117)	Geisinger (n=450)	Bay West (n=30)	BVAMC (n=58)	Total (n=656)
Age at diabetes diagnosis	35.5 ± 16.7	35.0 ± 19.0	40.8 ± 16.7	44.7 ± 11.4	37.1 ± 17.4
Type 1	19.7%	29.6%	16.7%	22.4%	26.5%
Type 2	62.4%	59.8%	73.3%	69.0%	61.7%
Gestational	2.6%	3.8%	6.7%	0%	3.4%
Pre-diabetes	14.5%	5.6%	3.3%	8.6%	7.3%
Other	0.9%	1.3%	0%	0%	1.1%

3. Genetic Testing

After reviewing patient information for enrolled participants in the Redcap database, 138 of 156 enrolled in the study underwent genetic testing. Those who did not undergo genetic testing had present DAAs, undetectable C-peptide, or otherwise did not fulfill the inclusion criteria. The average depth of coverage for each nucleotide was 367±131x (mean ± SD) and an average of 96.6 ± 2.0% of the target nucleotides had a coverage depth of ≥20x. Each participant had an average of 170±23 total variants and 9.2±5.1 of variants after filtering out common, non-coding, and synonymous variants. A total of 222 missense variants, 2 nonsense variants, 4 splice-site variants, and 7 small

insertions/deletions were analyzed according to the ACMG/AMP guidelines for variant interpretation the variant classification meetings (Table 3.10). A total of 38 novel variants were discovered (not found in 1000 Genomes, NHLBI-ESP, ExAC, dbSNP, or T2D-Genes Exome database) in 35 PDMP participants (Table 3.11).

Table 3.10: Total variants reviewed by gene and variant effect

Gene	Total variants	Missense	Nonsense	Splice	Insertion/ Deletion
LEPR	7	7	0	0	0
SLC19A2	3	3	0	0	0
KLF11	7	6	0	0	1
ALMS1	46	46	0	0	0
EIF2AK3	4	4	0	0	0
NEUROD1	1	1	0	0	0
PPARG	1	1	0	0	0
SLC2A2	2	2	0	0	0
WFS1	28	28	0	0	0
HADH	5	5	0	0	0
SIM1	1	1	0	0	0
RFX6	9	9	0	0	0
GCK	14	7	1	2	2
PPP1R3A	14	14	0	0	0
CAV1	1	1	0	0	0
PAX4	4	4	0	0	0
LEP	1	1	0	0	0
BLK	6	6	0	0	0
GLIS3	13	13	0	0	0
CEL	4	4	0	0	0
AGPAT2	7	6	0	1	0
PTF1A	1	1	0	0	0
GLUD1	1	1	0	0	0
INS	1	1	0	0	0
KCNJ11	3	3	0	0	0
ABCC8	4	3	1	0	0
BSCL2	6	4	0	1	1
HNF1A	8	7	0	0	1
PDX1	4	3	0	0	1
PLIN1	8	8	0	0	0
HNF1B	1	1	0	0	0
PTRF	2	1	0	0	1
GATA6	3	3	0	0	0
MC4R	5	5	0	0	0
INSR	7	7	0	0	0
HNF4A	2	2	0	0	0
FOXP3	1	1	0	0	0

Table 3.11: PDMP novel variants

Gene	mRNA	Exon	Nucleotide change	Base change	ACMG pathogenicity
LEPR	NM_001198687	exon8	c.1055G>T	p.Cys352Phe	VUS
KLF11	NM_003597	exon3	c.608_609AG	.	VUS
ALMS1	NM_015120	exon1	c.112G>A	p.Asp38Asn	Likely Benign
ALMS1	NM_015120	exon8	c.2209G>A	p.Glu737Lys	Likely Benign
ALMS1	NM_015120	exon8	c.6034G>T	p.Ala2012Ser	Likely Benign
ALMS1	NM_015120	exon10	c.9109C>G	p.Pro3037Ala	Likely Benign
ALMS1	NM_015120	exon20	c.12292A>G	p.Arg4098Gly	Likely Benign
EIF2AK3	NM_004836	exon3	c.440T>G	p.Val147Gly	VUS
WFS1	NM_006005	exon2	c.172G>A	p.Ala58Thr	VUS
WFS1	NM_006005	exon5	c.578A>G	p.Lys193Arg	VUS
WFS1	NM_006005	exon6	c.667C.A	p.Leu223Met	VUS
WFS1	NM_006005	exon8	c.1868T>A	p.Met623Lys	VUS
WFS1	NM_006005	exon8	c.1896G>T	p.Met632Ile	VUS
SIM1	NM_005068	exon11	c.1894C>T	p.His632Tyr	VUS
RFX6	NM_173560	exon12	c.1319A>G	p.Tyr440Cys	VUS
GCK	NM_000162	exon10	c.1361C>A	p.Aal454Glu	Likely Path.
GCK	NM_000162	exon10	c.1344delC	p.Gly448fs	Pathogenic
GCK	NM_000162	.	c.1253+2T>A	.	Pathogenic
GCK	NM_000162	exon9	c.1113C>A	p.Cys371Ter	Pathogenic
GCK	NM_000162	exon8	c.1016A>G	p.Glu339Gly	Pathogenic
GCK	NM_000162	exon8	c.918_919AT	.	VUS
GCK	NM_000162	exon7	c.748C>T	p.Arg250Cys	Likely Path.
GCK	NM_000162	.	c.680-1G>A	.	Pathogenic
GCK	NM_000162	exon2	c.122T>C	p.Met41Thr	Likely Path.
GCK	NM_000162	exon2	c.113A>C	p.Gln38Pro	Likely Path.
GCK	NM_000162	exon2	c.74T>A	p.Leu25Gln	VUS
PPP1R3A	NM_002711	exon1	c.416C>G	p.Ser139Cys	Likely Benign
PAX4	NM_006193	exon1	c.104T>C	p.Ile35Thr	VUS
BLK	NM_001715	exon2	c.2T>G	p.Met1Arg	VUS
GLIS3	NM_001042413	exon11	c.2765A>T	p.Gln922Leu	VUS
INS	NM_000207	exon3	c.278A>G	p.Glu93Gly	Likely Path.
KCNJ11	NM_000525	exon1	c.794G>T	p.Ser265Ile	VUS
BSCL2	NM_001122955	.	c.631-1G>C	.	VUS
BSCL2	NM_001122955	exon2	c.199A>C	p.Asn67His	VUS
HNF1A	NM_000545	exon3	c.616T>A	p.Trp206Arg	VUS
HNF1A	NM_000545	exon3	c.694dupC	p.Thr231fs	Pathogenic
INSR	NM_000208	exon8	c.1741C>T	p.Arg581Trp	VUS
FOXP3	NM_014009	exon4	c.409G>A	p.Ala137Thr	VUS

4. Monogenic Diabetes Diagnoses

A total of 14 participants were discovered to have variants that were either pathogenic or likely pathogenic for monogenic diabetes. The patients spanned a wide spectrum of demographic and clinical characteristics. Only 8/14 had been diagnosed before the age of 30 (Table 3.12). While most patients with monogenic diabetes were Caucasian, there were also two African-American patients, one Asian patient, and one patient of Hispanic ethnicity. Importantly, only one participant with monogenic diabetes was discovered through the clinical screening process, but instead, 5/14 enrolled in the PDMP study through physician referral and 8/14 through self-referral. While patients with monogenic diabetes did not have major differences from those without monogenic diabetes in terms of screening questionnaire answers, it is important to note that none (0/14) self-reported obesity at diagnosis (Table 3.13). However, subsequent self-report of BMI at diagnosis revealed one of the patients to have apparently been overweight or obese. Two of the patients were siblings who enrolled in the study separately. Most of the genetic variants were discovered in *GCK* (n=11), although two patients had variants in *HNF1A* and one had a variant in *INS* (Table 3.14). There were 7 additional variants of uncertain significance in the three most common genes causing monogenic diabetes (*GCK*, *HNF1A*, and *HNF4A*) (Table 3.15).

Table 3.12: Characteristics of PDMP patients with and without monogenic diabetes

	Monogenic diabetes (n=14)	**No monogenic diabetes (n=124)**
Age of diagnosis	26.7 ± 17.4	27.0 ± 12.9
Enrolled at UM-CDE (%)	92.9% (13/14)	67.7% (84/124)
Percent female (%)	85.7% (12/14)	61.3% (76/124)
Screened (%)*	7.1% (1/14)	43.5% (54/124)
Physician referral (%)	35.7% (5/14)	42.7% (53/124)
Self-referral (%)*	57.1% (8/14)	13.7% (17/124)
Hispanic (%)	7.1% (1/14)	1.6% (2/122)
African-American (%)	21.4% (3/14)	32.3% (40/124)
Caucasian (%)	71.4% (10/14)	63.7% (79/124)
Asian (%)	7.1% (1/14)	4.0% (5/124)

*$p<0.01$ for Fisher's Exact Test for proportion of patients with monogenic diabetes versus patients without monogenic diabetes

Table 3.13: Screening questionnaire responses of patients with and without monogenic diabetes

PDMP Screening question	Monogenic diabetes	No monog. diabetes
1. Were you diagnosed with diabetes or high blood sugar before 1 year of age?	0% (0/14)	0.8% (1/124)
2. Were you diagnosed with diabetes or high blood sugar at age 30 or younger? How old were you when you were diagnosed?	57.1% (8/14)	63.4% (78/123)
3. Were you extremely overweight when you were diagnosed?	0% (0/14)	19.4% (24/124)
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?	7.1% (1/14)	26.6% (33/124)
5. Do you have type 1 diabetes (if unsure, were you on insulin at diagnosis and have been ever since)?	7.1% (1/14)	26.6% (33/124)
6. Do you have a parent or a child with type 1 diabetes?	21.4% (3/14)	23.1% (28/121)
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	85.7% (12/14)	82.9% (102/123)

None of the distributions of responses were significantly different between the groups using Fisher's Exact Test.

Table 3.14: Individual characteristics of PDMP patients with monogenic diabetes

Gene	Age (sex)	Race / ethnicity	Characteristics*	Amino acid change / Site change	ACMG pathogenicity
GCK	27 (F)	Caucasian	GDM and T2D diagnosed at 26. Thin. Treated with insulin.	p.Ala454Glu	Likely Pathogenic
GCK	28 (F)	Caucasian	Prediabetes since age 19.	p.Gly448fs	Likely Pathogenic
GCK	51 (F)	African-American	Prediabetes in 30's. Diagnosed with DM at 45. Treated with combination therapies.	Splice site (c.1253+2T>A)	Pathogenic
GCK	34 (F)	Caucasian	Prediabetes since age 15. T2D diagnosed at 21. Treated with insulin during pregnancy.	p.Glu339Gly	Pathogenic
GCK	66 (F)	Caucasian	High blood sugar since age 20. T2D diagnosed at 65.	p.Arg250Cys	Likely Pathogenic
GCK	22 (F)	Caucasian	T2D diagnosed at age 22. Thin. Treated with sulfonylureas.	Splice site (c.680-1G>A)	Pathogenic
GCK	26 (F)	Caucasian	Overweight. Not on medications.	Splice site (c.680-1G>A)	Pathogenic
GCK	60 (F)	Caucasian	Diagnosed with diabetes type "1.5" at 54.	p.Arg43His	Likely Pathogenic
GCK	41 (M)	Caucasian	T2D diagnosed at 26.	p.Met41Thr	Likely Pathogenic
GCK	32 (M)	Caucasian	Prediabetes since age 5. Clinical diagnosis of MODY at 32. Thin.	p.Gln38Pro	Likely Pathogenic
GCK	27 (F)	Asian	T2D diagnosed at 27. Clinical suspicion of MODY. Thin.	p.Cys371*	Pathogenic
INS	5 (F)	Caucasian / Hispanic	T1D diagnosed at age 4.	p.Glu93Gly	Likely Pathogenic
HNF1A	45 (F)	Caucasian	T2D diagnosed at 16. Clinical suspicion of MODY.	p.Thr231fs	Pathogenic
HNF1A	29 (F)	African-American	T2D diagnosed at 19yr	p.Arg272His	Pathogenic

*Characteristics were collected from descriptions collected by genetic counselors during pedigree collection.

Table 3.15: PDMP patients with suspicious monogenic diabetes VUS

Gene	**Age (sex)**	**Race**	**Characteristics**	**Amino acid change**
GCK	30 (F)	Caucasian	Clinically diagnosed with MODY.	p.Leu307Phe
GCK	31 (F)	African-American	Elevated HbA1c. Normal fasting glucose. Normal BMI.	p.Leu25Gln
GCK	23 (F)	Caucasian	High blood sugar found at age 13. Thin. Clinical suspicion of MODY at 23.	p.Thr49Asn
HNF1A	61 (M)	Caucasian	T2D diagnosed at age 42.	p.His514Arg
HNF1A	42 (F)	Caucasian	T2D diagnosed at 15. Initially on insulin, but no longer.	p.Trp206Arg
HNF1A	52 (F)	African-American	T2D diagnosed at 19. Overweight at diagnosis. Treated with oral medications.	p.Tyr322Cys
HNF4A	30 (M)	Caucasian	T1D diagnosed at 29. Thin. Treated with insulin.	p.Asp304Asn

VUS – variants of uncertain significance

5. Pedigree Analysis

Analysis of the pedigrees showed very few differences between those with monogenic diabetes variants and those without. Families of patients with monogenic diabetes were more likely to have a clinical diagnosis of MODY in their family or their first-degree relatives (Table 3.16). Interestingly, patients with monogenic diabetes were less likely to have first degree relatives with cardiac disease, although the proportion of families with cardiac disease in the entire pedigree was not different between those with and without monogenic diabetes. Even using machine learning software, patient classifiers could only correctly predict the monogenic diabetes status in 71.1% of the

cases, with a positive predictive value (PPV) of 0.46 and a negative predictive value of 0.73. The machine learning software was more sensitive when the customized D_i/D_g metric ("proband-out" method) was used to summarize the pedigree affectedness status rather than the maximum A_i/A_g metric (PPV = 0.31) or the mean A_i/A_g metric (PPV=0.15). Methods to further target monogenic diabetes patients, such as excluding those with suspicious VUS (Table 3.15) or limiting analysis to those with *GCK* variants only, did not improve yield of the machine learning method.

Table 3.16: Percent of affected family members in patients with and without monogenic diabetes

Condition	**Monogenic diabetes (n=13)**	**No monogenic diabetes (n=139)**
MODY – entire pedigree*	2.0±3.5% (4)	0.3±1.0% (8)
First-degree relatives*	4.5±11.1% (2)	0.1±1.7% (1)
Second-degree relatives	0 (0)	0.0±0.4% (1)
T2D – entire pedigree	10.3±11.1% (11)	14.2±11.4% (121)
First-degree relatives	12.3±18.1% (5)	22.6±24.0% (85)
Second-degree relatives	6.5±10.0% (5)	11.3±12.5% (86)
T1D – entire pedigree	3.5±6.5% (5)	3.9±6.1% (63)
First-degree relatives	3.5±8.5% (2)	6.4±13.0% (34)
Second-degree relatives	3.6±7.8% (3)	1.6±5.5% (21)
GDM – entire pedigree	1.7±3.9% (3)	1.4±2.6% (40)
First-degree relatives	4.7±9.2% (3)	4.1±10.0% (24)
Second-degree relatives	0.9±8.5% (1)	0.6±2.4% (9)
Pre-diabetes – entire pedigree	5.1±7.6% (7)	1.6±2.9% (44)
First-degree relatives	13.1±24.0% (4)	4.6±10.8% (28)
Second-degree relatives	2.2±5.7% (2)	0.7±2.6% (13)
Cardiac disease – entire pedigree	5.3±5.1% (9)	8.1±8.2% (106)
First-degree relatives*	2.6±9.2% (1)	13.2±18.2% (66)
Renal disease – entire pedigree	1.7±2.8% (4)	2.0±3.3% (55)
First-degree relatives	0 (0)	4.8±10.2% (32)
Visual impairment - entire pedigree	2.5±5.6% (4)	2.3±3.8% (56)
First-degree relatives	2.6±9.2% (1)	2.5±6.9% (19)

Data reported is mean ± s.d. of the percent of affected individuals either in the entire pedigree, the first-degree relatives, or the second-degree relatives across each of the pedigrees of PDMP participants with or without monogenic diabetes. The value in parentheses represents the number of pedigrees positive for the condition either in the entire pedigree or in the specified degree of relation.
*p<0.05 for Fisher's Exact Test for patients with monogenic diabetes versus patients without monogenic diabetes

D. Discussion

The PDMP was designed to create, disseminate, and evaluate a sustainable program for screening, diagnosing, and providing individualized treatment for patients suspected to have highly-penetrant monogenic forms of diabetes. As part of the initial phase of PDMP, this study focused on the analysis and potential for screening for patients with monogenic diabetes based on patient-reported information. To date, 1,734 patients across four diverse study sites have been screened, 138 patients have undergone genetic testing, and 14 have been diagnosed with monogenic diabetes. This process has revealed insights about the population of patients with monogenic diabetes and suggested potential methods for improving monogenic screening practices in the US. It has also replicated previous observations, such as lower enrollment of minority populations and difficulty distinguishing Mendelian and complex forms of disease.[179] Studies to assess treatment regimens, treatment efficacy, and patient opinions of genetic testing before and after monogenic diabetes diagnosis as well as economic impact assessment are currently underway as part of the PDMP, but are beyond the scope of the study for this dissertation.

The PDMP was designed to cover a wide demographic of patients in rural, urban, academic, private-practice, and veteran-focused medical centers to assess implementation results in different settings. Demographics of patients screened matched those expected for each institution (Table 3.4). UM-CDE (an urban medical center) screened a larger proportion of African-Americans, GHS (a rural medical center) screened a larger proportion of Caucasians, and BVAMC screened a predominantly older male population. The majority of screened individuals came from the study hub, UM-CDE, and GHS, which is able to utilize their integrated EHR and large, stable patient population to

identify and screen a large number of individuals with diabetes.[180] As the study hub, UM-CDE also received the majority of referrals from external physicians and patient self-referrals, some of whom traveled across the country to take part in the study.

The screening information from the PDMP provides insights about the community examined and frames the question about the characteristics of individuals found to have monogenic diabetes. Many individuals screened in this study noted that they were diagnosed with T1D (32.2%, Table 3.5). Additionally, the PDMP average age of diabetes diagnosis (37.1 years old, Table 3.9) was lower than the national mean (54.2 years old).[181] This information could indicate that the population of this study tended to be younger patients with T1D in need of medical supervision for the complex management of issues like insulin regimen. Patients with standard care for T2D may instead be treated by primary care physicians, and could be less likely to be included in this population. Since patients with *GCK* usually do not exhibit diabetic symptoms and their blood glucose levels remain in the pre-diabetic range, those patients may be more likely to be unaware of their mildly elevated blood glucose or be effectively treated with straightforward glucose management regimens by their primary care provider. Of the PDMP patients screened, 31.5% were extremely overweight at diagnosis and 60.8% had multiple family members with diabetes diagnosed before age 50 (Table 3.5), suggesting evidence of the current epidemics of obesity and diabetes in the US. Importantly, the screening process for the PDMP narrowed the population to be considered for enrollment approximately 10-fold (156 enrolled / 1734 screened, Table 3.6), which is necessary for maximizing the cost-effectiveness of monogenic diabetes genetic testing.[133]

Multiple patients screened in the PDMP who were eligible to enroll did not participate in the study. Of the 336 eligible, less than half (156) enrolled in the study (Table 3.6). Patients not enrolling were categorized as "unable to be contacted," "declined participation," or "cancelled visit." Interestingly, the majority of participants who enrolled (61.5%) or cancelled their enrollment visit (59.5%) were female, while fewer females (40.4%) made up the group who were unable to be contacted (Table 3.6). The finding that women were more likely to enroll than men is supported by a 2012 report on enrollment by gender in all NIH-supported Clinical Research that showed most patients enrolled were women (57%).[182] Compared to African-Americans, Caucasians had a higher proportion of enrolled individuals (101/147 or 68.7% of Caucasians versus 49/129 or 38.0% of African-Americans) and lower proportions of participants that declined participation (20/147 or 13.6% of Caucasians versus 34/129 or 26.3% of African-Americans) (Table 3.7). This finding is in line with previous studies that have demonstrated that minorities are less likely to take part in genetic testing studies.[179] An important caveat to note is that 23 eligible Caucasian participants were not enrolled or grouped into one of the categories because their reason for non-enrollment was either not collected or the patients were still in the process of being contacted for enrollment. This represents a large number of individuals that could alter the distribution of Caucasians in the different categories.

Of the 156 patients enrolled in the PDMP, 138 were eligible to undergo genetic testing, and 14 individuals had pathogenic or likely pathogenic monogenic diabetes variants (Table 3.6, Table 3.14). Pathogenic or likely pathogenic variants were mostly discovered in *GCK* (11/14), with only two in *HNF1A* and one in *INS*. Characteristics

common to multiple participants with *GCK* variants were thin body habitus, diagnosis with pre-diabetes, and diagnosis with T2D (Table 3.14). Patients with *GCK* variants ranged in age from their 20s to their 60s, although multiple individuals had been diagnosed with prediabetes at a young age (Table 3.14). The two patients with *HNF1A* variants were both diagnosed with T2D in their teens, and the patient with an *INS* variant was diagnosed with T1D at age 4 (Table 3.14). The group of patients with monogenic diabetes was heterogeneous in terms of age, diagnosis, body habitus, and treatment regimens (Table 3.14). Studies are currently underway to measure the effect a diagnosis of monogenic diabetes has on treatment, behavior, and patient viewpoints of genetic testing. Since none of the variants discovered outside of *HNF1A*, *GCK*, or *INS* could be classified as pathogenic or likely pathogenic, it could be cost effective to limit the gene panel to genes with well-studied monogenic diabetes variants. While this could save costs, it would limit the ability to discover variants in less-common (and less studied) monogenic diabetes genes. There were seven patients that with variants of uncertain significance in the three most common monogenic diabetes genes (*GCK*: 3, *HNF1A*: 3, *HNF4A*: 1) (Table 3.14). Although these variants are interesting for further analysis with methods like co-segregation studies in family members, at the time of review, there was not enough information according to the conservative guidelines of the ACMG/AMP criteria to classify them as pathogenic or benign. Some of the patients with VUS demonstrate classic characteristics of monogenic diabetes, such as early onset, lean body-type, and effective response to oral medications (Table 3.14).

In comparing the screening questionnaire responses of those with and without monogenic diabetes, no responses were significantly different between the groups (Table

3.13). However, none of those with monogenic diabetes self-reported being extremely overweight at diagnosis, while 19.2% of the other patients that underwent genetic testing were overweight at diagnosis (Table 3.13). Although we have demonstrated through the TODAY study that obesity and monogenic diabetes can occur simultaneously in adolescents, BMI measures at diabetes diagnosis could potentially be a useful metric for improving screening specificity in select circumstances for adults. Patients with monogenic diabetes were identified across multiple races/ethnicities, indicating the importance of considering monogenic diabetes as a diagnosis even in an individual of a race/ethnicity at high risk for T2D.

The patient source of those with and without monogenic diabetes did demonstrate significant differences (Table 3.12). While self-referrals only accounted for 13.7% (17/124) of those without monogenic diabetes, most of those with monogenic diabetes were self-referred (57.1%, 8/14, $p < 0.01$). Likewise, patients from the general screened population only accounted for 7.1% (1/14) of those with monogenic diabetes, while it made up 43.5% (54/124) of the group without monogenic diabetes ($p < 0.01$). Those that were physician- and self-referred would have also been picked up by screening criteria. However, since so few patients with monogenic diabetes were discovered from the general patient populations at the four study sites, it calls into question if patients with monogenic diabetes would be found in those patient populations. Since patients with monogenic diabetes may be misdiagnosed as pre-diabetes patients (if caused by *GCK* variants) or may be effectively treated with oral medications (if caused by *HNF1A*/*HNF4A* variants), those patients could possibly be found in patient populations of primary care physicians rather than being referred to diabetes clinics. For this reason, it

may be effective to screen patients with diabetes that are treated by primary care physicians, but this remains to be seen.

Pedigrees of patients who underwent genetic testing were analyzed for trends in family medical history data both through basic quantitative measures and through use of a machine learning algorithm. More patients with monogenic diabetes had family tree members or first degree relatives with a clinical diagnosis of MODY than those without monogenic diabetes ($p < 0.05$, Table 3.16). Since the majority (8/14) of patients with monogenic diabetes were self-referred, this result is expected (those with suspected MODY in their family would seek out genetic testing). Comparing proportions of affected relatives by degree of relation from the proband also showed that fewer patients with monogenic diabetes had first-degree relatives with cardiac disease compared to those without monogenic diabetes ($p < 0.05$, Table 3.16). Since most of those discovered to have monogenic diabetes had *GCK* variants, it is possible that immediate family members with the same variant may not progress to cardiovascular complications of diabetes. However, the proportion of individuals with cardiac disease in their entire family was not different between those with and without monogenic diabetes (Table 3.16), indicating that those with monogenic diabetes could still have more distant family members (potentially those on the unaffected side of their family) still affected by cardiac disease. It is important to note that this outcome needs to be validated in larger studies of patients with *GCK*-MODY, since factors, such as average family age, could have a major effect. Additionally, other diabetic complications, such as visual impairments or renal disease, did not show the same pattern as cardiac disease.

Analysis of the pedigree dataset was also performed using a machine-learning algorithm and customized statistics to summarize the affected members in a pedigree. The pedigrees were summarized across all collected family medical history status categories (Figure 3.2) using computational algorithms for both previously-published "top-down" metrics (A_i/A_g) and customized "proband-out" metrics (D_i/D_g) (Figure 3.3). Using a Naïve Bayes Classifier, this resulted in poor discrimination between those with and without monogenic diabetes (71.1% overall accuracy). However, the novel "proband-out" summarizing methodology demonstrated more sensitivity than the published "top-down" approach (0.46 PPV vs 0.31 PPV). This indicates that the "proband-out" metrics could be useful for situations where family medical history data is provided by a single individual, resulting in a multi-nodal (multiple-ancestor) pedigree. While other methods to improve overall prediction were implemented to improve accuracy (such as removing/reclassifying patients with suspicious VUS or limiting analysis to patients with *GCK* pathogenic or likely pathogenic variants), none improved the accuracy or the sensitivity of the model. It is important to note that the differences between patient sources for those with monogenic diabetes compared to those without may have introduced bias into the pedigree analysis, and the summarizing statistics as well as the machine learning algorithm would need further study using other datasets. Utilizing a larger dataset or a larger proportion of patients with monogenic diabetes could potentially improve the machine-learning model, and could be a future possibility for the PDMP. Patient-enabled approaches to pedigree collection could provide an opportunity for assessment of a larger number of pedigrees and could present a consistent method to limit potential bias introduced by variability of collection methods for family medical history

data. A web-based family health history collection tool and clinical support program, MeTree, is an example of such an approach that could be useful.[183]

Screening for monogenic diabetes could potentially be used to limit the number of genetic tests necessary to detect cases of monogenic diabetes in populations with diabetes. This study has demonstrated that individuals with monogenic diabetes have variable ranges of ages, ages of diabetes diagnosis, body types, races/ethnicities, and treatment regimens. As a result, it is challenging to narrow the clinical criteria necessary to increase specificity of genetic testing for monogenic diabetes. Some options, like using family medical history to aid identification of patients with monogenic diabetes, are potential opportunities in need of further study. Finally, the large proportion of individuals that were physician- or self-referred for the study indicates that those are potentially valuable resources that could aid diagnosis of monogenic diabetes. While this study can inform the direction of future efforts, it is important to recognize that this study is generally underpowered (only 138 genetic tests performed) for complete analysis of study sites with such variable characteristics. Additionally, the stringent criteria used for determining variant pathogenicity has resulted in multiple VUS. While this is appropriate criteria for returning information clinically, this could underestimate the prevalence of monogenic diabetes in this cohort and result in multiple individuals that may have monogenic diabetes variants in need of further study. Cascade screening, or genetic testing of family members of a proband with a monogenic diabetes variant, is currently being undertaken to potentially diagnose more patients as part of the PDMP. Finally, PDMP outcome studies are currently underway to assess the qualitative effects on patients and perform economic modeling to estimate the impact of attaining a genetic

diagnosis. The PDMP has already provided important information for improving the diagnosis of patients with monogenic diabetes, and will continue to inform about the effects of attaining a genetic diagnosis of monogenic diabetes. The findings from this study could inform future genetic screening studies focused on other monogenic conditions as well, improving personalized medicine beyond the field of monogenic diabetes.

Chapter 4: Zebrafish as a Model for On-Demand *In Vivo HNF1A* Variant Functional Analysis[1]

A. Introduction

One of the challenges of clinical testing for genetic disease is the question as to whether a discovered variant is causative for the disease. This is especially difficult for Mendelian forms of disease that have similar phenotypic characteristics to complex forms of disease. The ACMG/AMP Guidelines for Genetic Variant Interpretation have provided a method that relies on multiple evidence categories, including population data, computational data, functional data, and segregation data, to comprehensively evaluate the pathogenicity of each genetic variant.[119] Some pieces of evidence rely on happenstance, such as population data and segregation data, and pieces of evidence that can be assessed across all variants (e.g. computational data) are weakly weighted because of inconsistent correlations with disease status. However, functional studies on genetic variants are one of most strongly-weighted pieces of evidence that could be generated as variants are discovered. As such, rapid, on-demand experimental testing for function of genetic variants is a valuable commodity for assessment of pathogenicity.

Monogenic diabetes is one of many genetic diseases that could benefit from an on-demand functional testing model. Monogenic diabetes is defined as hyperglycemia caused by one or more genetic variants in a single gene. The most common form of monogenic diabetes is Maturity-Onset Diabetes of the Young (MODY), which usually presents clinically as early onset (before age 25), autosomal-dominant diabetes in patients

[1] Jeffrey W. Kleinberger, Elizabeth O'Hare, Carmen Leitch, Toni I. Pollin, and Norann Zaghloul. Zebrafish as a Model for On-Demand In Vivo HNF1A Variant Functional Analysis. In preparation for submission.

with maintained insulin production.[160] However, clinical suspicion of monogenic diabetes is often complicated by the prevalence of common forms of diabetes, such as type 1 or type 2 diabetes (T1D, T2D), as well as clinical overlap of patient characteristics. For example, the early-onset of MODY is similar to T1D, while the strong family history and maintained insulin production is similar to T2D. There are also numerous examples, including examples in this dissertation, of patients with monogenic diabetes that have characteristics outside those stereotypical for the disease.[109,184] Accordingly, when assessing monogenic diabetes genetic variants, the similarity between MODY and other forms of diabetes often make it impossible to use phenotype data in support of pathogenic status (ACMG/AMP Supporting Pathogenic Evidence 4: patient's phenotype or family history is highly specific for a disease with a single genetic etiology).

One of the most common causes of monogenic diabetes is genetic variants in *HNF1A*. In the UK, mutations in *HNF1A* account for 52% of all cases of MODY.[15] *HNF1A* encodes HNF1A, a transcription factor highly expressed in liver, kidney, and pancreatic islets that forms both homodimers and heterodimers to bind with its target regions. Three major functional domains make up the HNF1A molecule: dimerization domain (amino acids 1-32), DNA-binding domain (amino acids 91-281), and transactivation domain (amino acids 282-631). Over 500 variants have been discovered in *HNF1A* across all regions of the gene.[121] A study of the type and position of *HNF1A* variants in patients with genetically confirmed *HNF1A*-MODY demonstrated that sequence-terminating variants have an earlier age of diagnosis compared to missense variants, and individuals with missense variants in the dimerization and DNA-binding domains have an earlier age of diagnosis than those in the transactivation domain.[150]

Additionally, there have been multiple protein-altering polymorphisms that have a modest impact on diabetes risk.[185] Like genetic variants in many genes, the variants in *HNF1A* can have a spectrum of different effects that must be evaluated carefully to determine the phenotypic effects of the variants.

Even though the ACMG/AMP guidelines classify experimental functional data as strong evidence to suggest pathogenicity, there has been relatively little published data experimentally assessing the function of *HNF1A* genetic variants. Of the > 500 *HNF1A* genetic variants discovered to date, fewer than 100 have been functionally evaluated.[121,186] All of the experimental studies have been *in vitro*, which can provide rapid results, but varying conditions (cell lines, target DNA sequence, functional measures) make interpretation of experimental results difficult to assess mutant *HNF1A* function. Additionally, many of the functional studies have been performed in cell lines that are not of beta cell origin and use a variety of *HNF1A* targets to assess function, adding additional layers of complexity to interpretation of results. Assessment of the consequences of *HNF1A* genetic variants could potentially be improved by using beta-cell-specific measures in an *in vivo* model that would provide more physiologically-relevant environment.

One animal model used for decades in genetic studies is the zebrafish, *Danio rerio*. Zebrafish are an excellent animal model because they develop rapidly, have high fecundity, and can be easily visualized due to translucency during development. For genetic studies, zebrafish can be easily genetically manipulated using random mutagenesis or targeted gene silencing via injections into the single-cell stage. Zebrafish have been used to study metabolic disease through many different mechanisms, and they

have proven to be a good model for basic metabolic principles shared by higher organisms, including humans.[187] For diabetes studies, the zebrafish pancreatic beta cells have been visualized using insulin-promoter driven fluorescence, and the number, size, and fluorescence of the beta cells have been shown to respond to caloric excess and high-fat diet.[188] Zebrafish have also been used to model monogenic diabetes: a specific mutation in the *INS* gene was shown to cause defective processing and endoplasmic-reticulum accumulation in zebrafish beta cells, and pdx1-mutant zebrafish (a model for *PDX1*-MODY or IPEX) have decreased beta cells and insulin levels.[189,190]

A recent study demonstrated that morpholino-knockdown of many T2D-associated genes causes reduced pancreatic beta cell number and mass in zebrafish.[191] Knockdown of the genes that have been shown to cause MODY all showed decreased beta cell number and mass, with the exceptions of *INS* and *ABCC8*. The goal of this study was to build off the finding that *HNF1A* knockdown causes decreased beta cell number and mass to create an *in vivo* model to test the function of *HNF1A* genetic variants in an on-demand fashion. This study assessed if beta cell number and mass could be rescued with *HNF1A* mRNA, and if that system could be used to model the effects of site-specific mutations in *HNF1A*.

B. Methods

1. Zebrafish, morpholino, and injections

Transgenic zebrafish used for this study carried the mCherry fluorescent molecule under control of the preproinsulin promoter (*ins*:mCherry). Adult zebrafish were maintained at 28-30°C water temperature, and embryos were maintained at 28.5°C in the

dark. The splice-blocking morpholino for hnf1a (5'-CCTCTCTAACACACATTAATACACC) has been previously described to be target-gene specific at 4.0 ng/uL.[191] For this study, the hnf1a splice-blocking morpholino was used at a concentration of 3.5 ng/uL to accommodate the co-injection of rescue mRNA (ranging from 25pg to 150pg). Phenol red served as a visual marker of injection into the embryo. Injections of 1.2uL of morpholino or morpholino in combination with rescue mRNA were performed using embryos at the 1- to 2-cell stage. A total of approximately 150-300 embryos were injected for each study condition to yield at least 50 viable fish for assessment. At 5 days post-fertilization (dpf), embryos were fixed in 4% paraformaldehyde for quantification.

2. Rescue mRNA

Six *HNF1A* variants previously demonstrated to be pathogenic for MODY were selected for study. These variants all have published *in vitro* data indicating decreased molecular function (Table 4.1) in addition to reports of the presence in patients with MODY. The variants encompass the three functional domains of *HNF1A* (DNA-binding, dimerization, and transactivation) and also include a well-established frameshift mutation to observe the effects of different variant types. A common protein-altering variant, p.I27L (29.9% minor allele frequency in 1000 Genomes), was selected as a control. Previously, the complete *HNF1A* coding sequence (GenBank: BC104910.1) was cloned into the pCS2+ plasmid for amplification. Site-directed mutagenesis was performed using a Q5 Site-Directed Mutagenesis Kit (New England Biolabs) with specific mutagenesis primers (Table 4.2) to create the 7 variants (6 pathogenic variants and 1 control polymorphism) of this study. Mutagenesis products were amplified using site-specific

primers (Table 4.2) and confirmed via Sanger sequencing (Table 4.2, Figure 4.1). Plasmids were linearized with Not1 restriction enzymes, and mRNA was created using the mMessage mMachine SP6 Transcription Kit (Thermo-Fisher). mRNA quantification was performed using a NanoDrop spectrophotometer before storage at -80°C.

Table 4.1: Established pathogenic HNF1A-MODY variants and cited functional studies

Variant	**Domain**	**Studies (*in vitro*)**	**Cited *HNF1A* Target: Cell line**
p.G20R	Dimerization	Co-IP, Luciferase	Insulin promoter: HeLa[192]
p.P112L	DNA-binding	EMSA, Luciferase	Albumin and Insulin promoter: HeLa[193,194]
p.R131W	DNA-binding	Localization, EMSA, Luciferase	Albumin and sucrose-isomaltase promoter: HeLa and Caco-2[193,195]
p.R272H	DNA-binding	EMSA, Luciferase	Sucrase-isomaltase and beta-fibrinogen promoter: Caco-2 and C33[195,196]
p.P379fs	Transactivation	Localization, EMSA, Luciferase	Albumin and sucrose-isomaltase promoter: HeLa and Caco-2[193,195]
p.P447L	Transactivation	Immunoblot, EMSA, Luciferase	beta-fibrinogen promoter: C33[196]

Table 4.2: Primers for mutagenesis and confirmation Sanger sequencing

Name	Sequence
p.G20R Mutagenesis Forward	5’-GCTCGAGTCAcGGCTGAGCAAAGAGGC-3’
p.G20R Mutagenesis Reverse	5’-AGGGCCGCCAGGAGCTCC-3’
p.I27L Mutagenesis Forward	5’-AGAGGCACTGcTCCAGGCACT-3’
p.I27L Mutagenesis Reverse	5’-TTGCTCAGCCCTGACTCG-3’
p.P112L Mutagenesis Forward	5’-CAGGAGGACCtGTGGCGTGTG-3’
p.P112L Mutagenesis Reverse	5’-CAGAAGGGTCTCCACCACG-3’
p.R131W Mutagenesis Forward	5’-CATCCCACAGtGGGAGGTGGT-3’
p.R131W Mutagenesis Reverse	5’-TTGTGCTGCTGCAGGTAG-3’
p.R272H Mutagenesis Forward	5’-GCCAACCGGCaCAAAGAAGAAG-3’
p.R272H Mutagenesis Reverse	5’-AAACCAGTTGTAGACACGC-3’
p.P379fs Mutagenesis Forward	5’-GTCAGCACCCTGACAGCACTGCAC-3’
p.P379fs Mutagenesis Reverse	5’-GGGGGAGGGGGCCCCCAG-3’
p.P447L Mutagenesis Forward	5’-CAGAGTGTGCtGGTCATCAACAGCATG-3’
p.P447L Mutagenesis Reverse	5’-TGCCTGCGTGGAGGCCAG-3’
p.G20R Sequencing Forward	5’-ATGGTTTCTAAACTGAGC-3’
p.G20R Sequencing Reverse	5’-GTGAAGTCTTCCCCATCGTC-3’
p.I27L Sequencing Forward	5’-ATGGTTTCTAAACTGAGC-3’
p.I27L Sequencing Reverse	5’-GTGAAGTCTTCCCCATCGTC-3’
p.P112L Sequencing Forward	5’-CCATCCTCAAAGAGCTGGAG-3’
p.P112L Sequencing Reverse	5’-GTTGAGGTGTTGGGACAGGT-3’
p.R131W Sequencing Forward	5’-CCATCCTCAAAGAGCTGGAG-3’
p.R131W Sequencing Reverse	5’-GTTGAGGTGTTGGGACAGGT-3’
p.R272H Sequencing Forward	5’-AGCGAGAGACGCTAGTGGAG-3’
p.R272H Sequencing Reverse	5’-CCCGCTGTACGTGTCCAT-3’
p.P379fs Sequencing Forward	5’-TACACCCCTCCACCAAGTGT-3’
p.P379fs Sequencing Reverse	5’-AGTGAGGCCATGATGAGGTT-3’
p.P447L Sequencing Forward	5’-ACCTCATCATGGCCTCACTT-3’
p.P447L Sequencing Reverse	5’-CACATGGCTCTGCACAGGT-3’

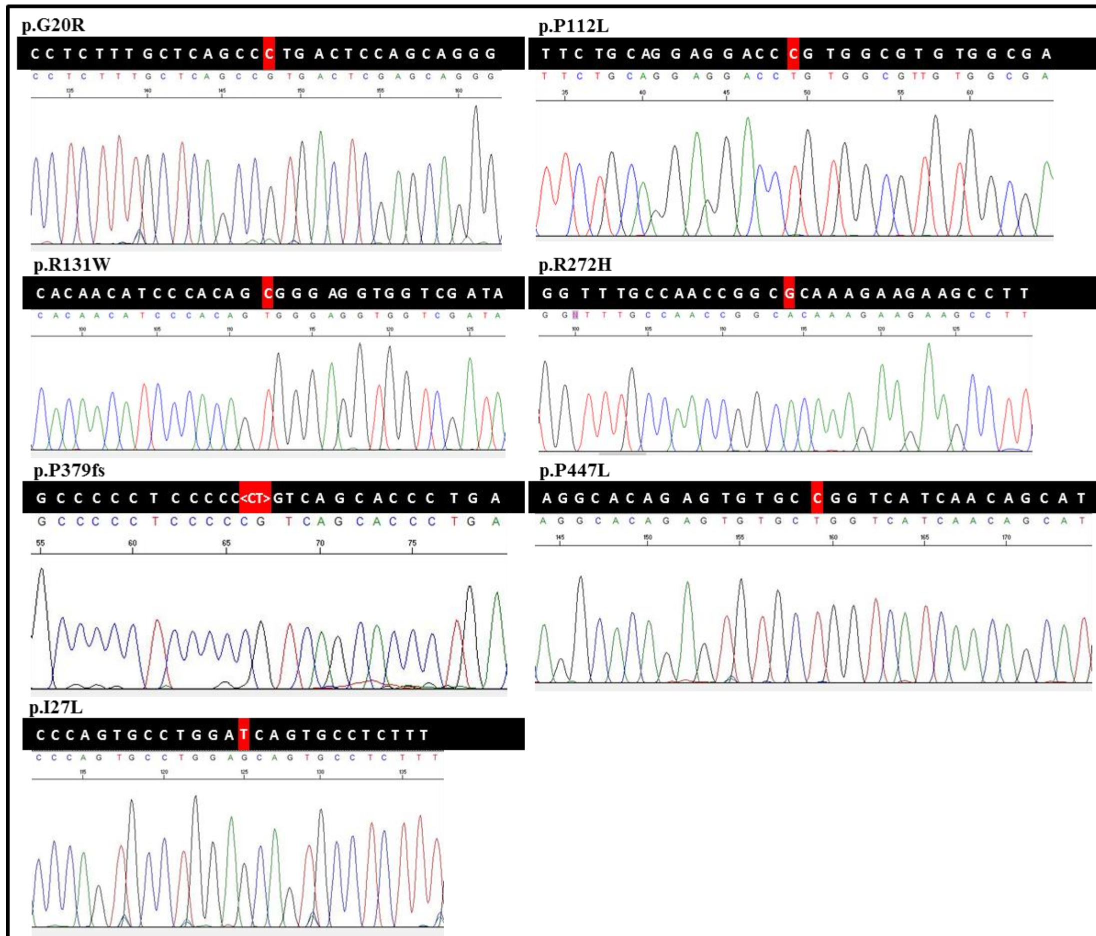


Figure 4.1: Sanger sequencing confirmation of mutagenized variants
Figure shows chromatograms of 6 pathogenic *HNF1A*-MODY variants and 1 *HNF1A* polymorphism mutagenized for creation of mRNA. DNA sequence found in black bars represents reference sequence, and the DNA sequence below the black bars are the automated sequence predictions from the chromatogram (Applied Biosystems Sequence Scanner Software 2). Highlighted red nucleotides are the altered nucleotides in the chromatogram.

3. Quantification and statistical analysis

Each 5 dpf larva was mounted on a glass slide using ProLong Gold Antifade Reagent (Life Technologies) and covered with a glass coverslip. Imaging of beta cell mCherry fluorescence was collected using a Zeiss SteREO Lumar.V12 microscope at 120x magnification. Each larva was then compressed to disperse beta cells for single-cell resolution and manual quantification. Quantification of beta cell area and fluorescence

intensity was performed on each collected image using ImageJ software (National Institutes of Health). Measures for morpholino and wild-type-rescued morpholino conditions were compared to uninjected wild-type larvae, while beta cell number and area measures for each variant were compared to the wild-type-rescued morpholino condition. To assess statistical significance, two-tailed t-tests were performed.

C. Results

1. Dose escalation of mRNA rescue

To assess the effect of *HNF1A* genetic variants on the ability to rescue morpholino-induced knockdown effects in zebrafish, demonstration of rescue with wild-type *HNF1A* is necessary. This was performed by measuring beta cell number at 5 dpf of wild-type, morpholino, and titrated rescue conditions (morpholino plus 25pg, 50pg, 100pg, or 150pg of wild-type *HNF1A* mRNA). Injection of the two highest concentrations of *HNF1A* mRNA did not produce viable embryos that survived for 5 days. Beta cell count was statistically different between wild-type zebrafish (27.6±6.9 beta cells) and morpholino knockdown (20.4±6.1 beta cells, $p=2x10^{-6}$) as well as morpholino-knockdown zebrafish rescued with 25pg wild-type *HNF1A* mRNA (21.2±7.6 beta cells, $p=2x10^{-4}$). However, there was no significant difference between beta cell count of wild-type and those morpholino-knockdown zebrafish rescued with 50pg wild-type *HNF1A* (25.1±7.3 beta cells, p=0.08) (Figure 4.2). Based on this data, 50pg *HNF1A* mRNA was used for injections of *HNF1A* mRNA containing genetic variants.

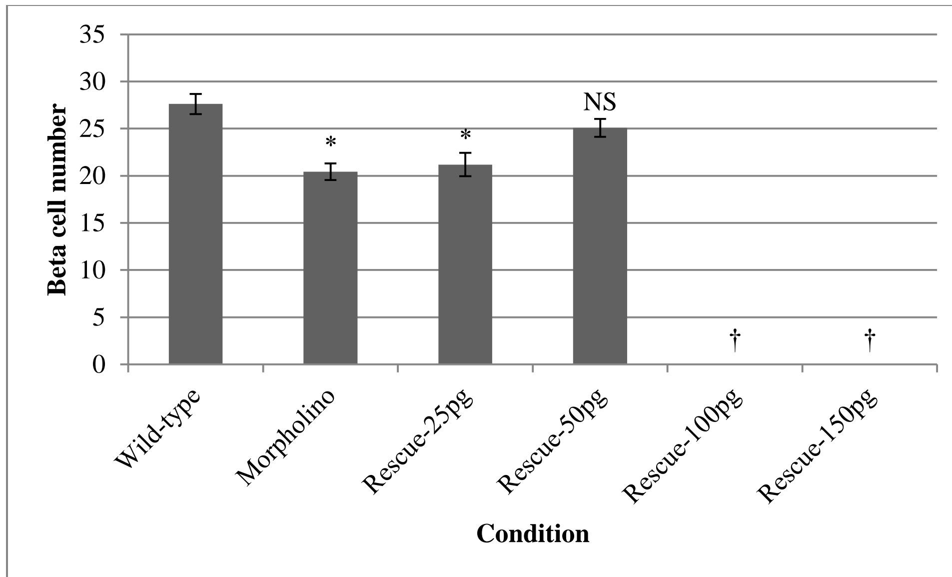


Figure 4.2: Beta cell number of dose response experiment
Bar graph showing the beta cell number for wild type, morpholino, and each mRNA amount used to rescue morpholino knockdown. Error bars represent standard error. Number of larvae per condition range from 35 to 60. * $p<0.05$ for comparison to wild-type zebrafish. NS = not significantly different from wild-type zebrafish. † No viable larvae at 5dpf.

2. Rescue with established *HNF1A* genetic variants

After demonstrating that morpholino-induced knockdown of *hnf1a* could be rescued with wild-type *HNF1A* mRNA, we examined the effect of rescuing with *HNF1A* containing established *HNF1A*-MODY variants and one protein-altering *HNF1A* polymorphism (p.I27L). While knockdown of *hnf1a* caused a decrease in beta cell number and area compared to wild type islets (number ± SD: 29.7±5.9 vs. 21.7±6.4 p=5e-23, area ± SD: 28.0±8.9 vs 22.4±5.4um^2, p=2e-6), no significant difference was observed between wild-type and the wild-type rescue group (number: 28.2±6.8 p=0.05, area: 26.6±7.2um^2 p=0.3). Comparisons between beta cell number and area in wild-type rescue and variant rescue conditions only showed decreased beta cell number and area using p.G20R (number: 26.0±5.4 p=0.01, area: 23.3±5.1um^2 p=0.001) and p.R131W (number: 26.3±5.4 p=0.03, area: 24.3±6.1um^2 p=0.03) variants. All of the other *HNF1A*-MODY variants (p.P112L, p.R272H, p.P379fs, p.P447L) and the common *HNF1A*

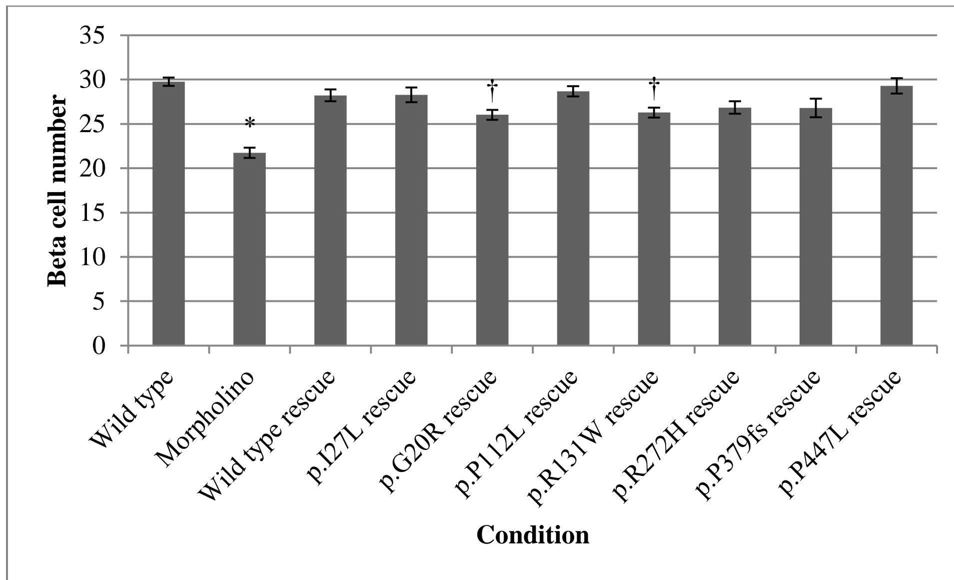
polymorphism (p.I27L), showed no significant differences in beta cell number or area compared to the wild-type rescue condition (Table 4.3, Figure 4.3).

Table 4.3: Beta cell number and area of each condition studied

Condition	Beta cell number	p-value	Beta cell area (um^2)	p-value	Larvae n=
Wild type	29.7±5.9	-	28.0±8.9	-	84
Morpholino*	21.7±6.4	**5e-23**	22.4±5.4	**2e-06**	84
Wild type rescue*	28.2±6.8	0.05	26.6±7.2	0.3	73
p.I27L rescue	28.3±6.9	1.0	26.1±7.7	0.7	71
p.G20R rescue	26.0±5.4	**0.01**	23.3±5.1	**0.001**	96
p.P112L rescue	28.7±5.2	0.6	25.6±5.3	0.3	87
p.R131W rescue	26.3±5.4	**0.03**	24.3±6.1	**0.03**	93
p.R272H rescue	26.8±6.5	0.2	25.6±7.8	0.4	95
p.P379fs rescue	26.8±8.0	0.3	25.4±6.8	0.3	57
p.P447L rescue	29.3±6.8	0.3	25.1±6.7	0.2	63

* p-value represents comparison to wild type zebrafish. All other p-values represent comparisons to the wild type rescue conditions.

A.



B.

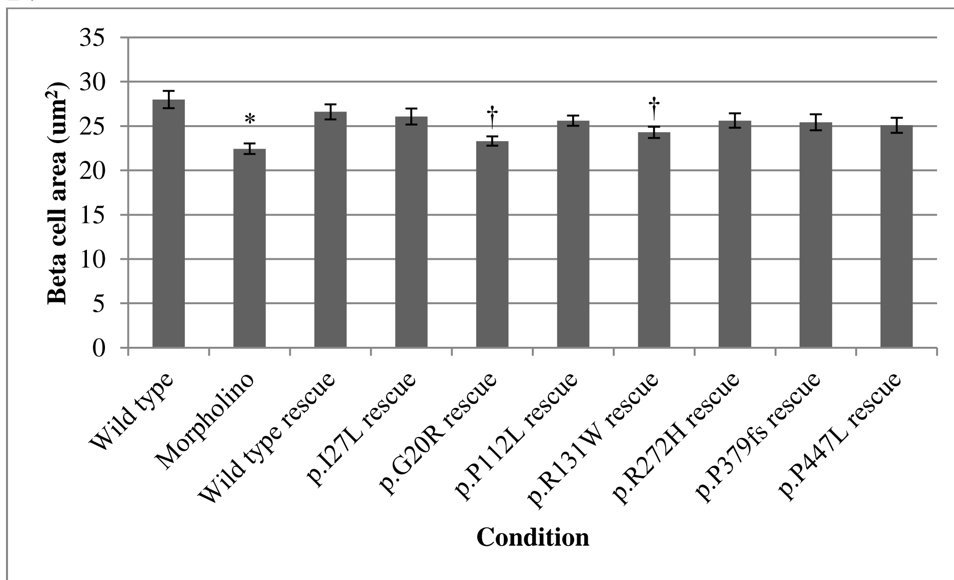


Figure 4.3: Beta cell number and area of each condition studied
Bar graphs showing the A.) beta cell number and B.) beta cell area for each condition. Error bars represent standard error. * $p<0.05$ for comparison to wild-type zebrafish. † $p<0.05$ for comparison to wild type rescue condition.

D. Discussion

We analyzed the ability to use measures of beta cell number and area in zebrafish to model the functional effects of genetic variants through *hnf1a* morpholino-knockdown and mRNA rescue with *HNF1A*. Although the decrease in beta cell number and area

induced by morpholino-knockdown of *hnf1a* could be effectively rescued with *HNF1A* mRNA, the model system was unable to demonstrate decreases in number or area of beta cells with mRNA rescue for all but two of six established *HNF1A*-MODY variants. Unfortunately, the effective dose window for the rescue mRNA is small (25pg did not rescue, but 100pg caused morphological defects), which does not lend to creation of dose-response curves to enhance subtle differences between response conditions. This indicates that, at this time, this model requires further exploration of optimal conditions to serve as a functional assay of *HNF1A* genetic variants. However, demonstration of significantly diminished ability to rescue using two select variants (p.G20R and p.R131W) could indicate potential for further studies, either with altered conditions or for analysis of select variants.

The explosion in genetic variant data due to next-generation sequencing capabilities has left a substantial need for effective methods to assess the functional effects of the large number of new genetic variants. Some groups have recognized this demand, and created multiplexed reporter assays to assay the functional effects of many variants in a high-throughput manner.[197,198] These assays have mostly measured the effects of noncoding variants in regulatory regions, but one recent study prospectively created and functionally analyzed all possible amino acid changes in *PPARG*.[130] *PPARG* codes for the transcription factor peroxisome proliferator-activated receptor gamma, and genetic variants in *PPARG* cause familial partial lipodystrophy type 3 (FPL type 3) as well as being implicated as one of two genetic variants in a case of digenic diabetes.[199,200] That prospective study used a combination of target expression and ligand-response functional assays followed by supervised machine learning based on variants known to

cause FPL type 3.[130] A published database is available to view the functional output scores and the predictions as to whether they are damaging or not. While this approach is unique and novel, *PPARG* is particularly well-suited for this type of study due to its moderate size and measurable response to multiple ligands. Additionally, while the functional assays are used to create the prediction, the computational output analysis of this project makes the predictions a form of functional study-*in silico* hybrid that may not carry the confidence of traditional functional studies. Therefore, rapid on-demand functional variant studies may be more practical and informative compared to prospective genetic variant analysis.

It is important to also recognize the potential pitfalls and difficulties of on-demand variant testing. Genetic disorders are often characterized broadly as Mendelian (strongly controlled by a single locus) or multifactorial (controlled by many weakly associated loci), but genetic disorders lie on a spectrum that could include diseases moderately controlled by one or more loci. One recent study examined the effect of 27 rare *HNF1A* variants as measured by luciferase assays (HeLa cells, rat albumin target) and localization experiments.[186] They showed that the variants have a smooth distribution of function (with p.P112L and p.P447L *HNF1A*-MODY variants serving as negative controls and p.I27L serving as a positive control), and an arbitrary cutoff (<60% of wild-type function) could select a group of 11 variants that were associated with diabetes in the general population (OR 5.04, p=0.0007). This finding demonstrates the difficulty in interpretation of functional studies, since it would be difficult to assume the same functional distribution using different cell lines, target regions, and cutoffs for MODY compared to diabetes-associated or non-damaging variants in *HNF1A*.

In this study, we tested the capability of six established *HNF1A*-MODY variants to rescue the beta cell number and area defects observed after morpholino-induced knockdown of *HNF1A*. Only the p.G20R and p.R131W variants showed significantly diminished ability to rescue the knockdown phenotype. Each of the variants tested had previously been shown to decrease function of HNF1α in multiple *in vitro* studies (Table 4.1). A previous study has demonstrated that in patients with *HNF1A*-MODY, the region and type of mutation can significantly alter the age of diagnosis, with missense mutations in the DNA-binding and dimerization domains having earlier onset than those with missense variants in the transactivation domain.[150] The study also found that individuals with truncating mutations had earlier onset than those with missense mutations. Although the only truncating mutation we tested, p.P379fs, did not show any difference from the wild-type rescue, the p.G20R missense variant is found in the dimerization domain (amino acids 1-32) and the p.R131W missense variant is found in the DNA-binding domain (amino acids 91-281). However, the p.P112L and p.R272H are also found in the DNA-binding domain, but they showed no differences from the wild type rescue. While it would be interesting to follow up this discovery to map the most important residues of *HNF1A* in this model system, doing so is beyond the scope and the goal of this study.

It is important to note that while zebrafish provide a convenient model with many advantages for genetic studies, they may not be an appropriate model for studying *HNF1A* variants in particular. Zebrafish were selected to provide an *in vivo* model that could provide more consistency than the current paradigm of using luciferase experiments utilizing many different *HNF1A* targets across cell lines of different tissue and species origins to assess the function of *HNF1A* variants. However, zebrafish are

particularly effective for studying organism development. While individuals with *HNF1A*-MODY are often discovered early in life, little is known about their beta cell number or mass at birth or during early development. It is known that individuals with *HNF1A*-MODY do have reduced pancreatic volume in adulthood.[201] It is possible that variants causing monogenic diabetes may not dramatically affect beta cell number or mass in the earliest stages of development, and therefore are not detected with these analyses. In support of this theory, mouse models of Hnf1a knockout have shown that heterozygous Hnf1a (+/-) mice do not develop diabetes and have normal pancreatic islet morphology and beta cell mass. However, Hnf1a (-/-) mice develop diabetes and have reduced beta cell mass.[202] Estimating the effects of human *HNF1A* variants using zebrafish may also be an ineffective model due to physiological differences between the species. Despite effectively rescuing the beta cell phenotype with human *HNF1A*, the zebrafish have different cofactors, target binding regions, and relevant protein-protein interactions that may not be conserved between species. These differences between species could potentially be crucial for expression of the phenotype in humans, but may not be as important for the phenotype of beta cell number and area in zebrafish.

There are multiple other factors that could be altered from the current model that may improve the ability to differentiate between wild-type rescue and rescue with one of the *HNF1A*-MODY variants. Firstly, beta cell number and mass may not be the most sensitive quantitative measure for variant function. Indeed, while a sample size of 50 larvae is predicted to have 90% power to detect a 5% difference in beta cell number, detecting a difference 1% difference in beta cell number would require 200 larvae for 90% power. More technical measures, such as quantifying insulin or glucose levels, may

be more apt to detect defective *HNF1A* variant effects. Secondly, morpholino knockdown could potentially be substituted with newer knockout methods, such as transcription activator-like effector nucleases (TALEN) or clustered regularly interspaced short palindromic repeats (CRISPR). While there is debate over the potential off-target effects of morpholinos versus these methods, permanent consistent knockdown of *HNF1A* could improve the precision of this assay.[203,204] Finally, conditions such as high-glucose or high-caloric conditions have been demonstrated to increase beta cell mass and numbers in zebrafish, and those conditions could potentially increase the window of measures between wild-type and mutant rescue conditions.[188]

In summary, we were unable to use morpholino-knockdown and subsequent rescue with mutated *HNF1A* as an *in vivo* model to effectively distinguish damaging variants in *HNF1A*. Although unlikely, it is possible that this model could be accurately reflecting the minimal effect of variants that are not pathogenic for monogenic diabetes. While some of the established *HNF1A* variants tested did show significantly less rescue than wild-type *HNF1A*, those variants affected the established important functional domains of *HNF1A*. Using this model to further study other variants in those domains, as well as altering the model to improve the ability to definitively separate pathogenic and non-pathogenic variants, are potential directions of further research.

Chapter 5: Discussion

A. Summary

The goal of this dissertation was to evaluate approaches for identifying patients with monogenic diabetes and a method for functionally testing monogenic diabetes variants with the ultimate goal of potentially improving diagnosis of monogenic diabetes. We have addressed this goal through a variety of methods: genetic testing and analysis of a selected clinical trial cohort of overweight and obese children and adolescents with recently diagnosed T2D in the TODAY clinical trial, screening and genetic testing of diabetes clinic patients at four diverse study sites in the PDMP genetic testing implementation study, and development of a model using zebrafish knockdown and rescue to assess the effects of *HNF1A* variants. The results from these studies have provided information about potential misdiagnosis of overweight and obese children and adolescents with monogenic diabetes, the wide-ranging characteristics of patients with monogenic diabetes from the PDMP screening study, and the difficulty of using zebrafish to demonstrate the damaging effects of *HNF1A*-MODY variants. The results of the three studies of this dissertation are summarized below.

1. Monogenic Diabetes in Overweight and Obese Youth Diagnosed with Type 2 Diabetes: The TODAY Clinical Trial

As described in Chapter 2, we discovered that 4.5% (22/488) of the TODAY clinical trial participants had variants pathogenic or likely pathogenic for monogenic diabetes. This demonstrated that overweight or obese children and adolescents can have monogenic diabetes, and a diagnosis of monogenic diabetes should at least be considered in children and adolescents with diabetes autoantibody-negative and C-peptide positive

diabetes, regardless of overweight or obese status. Variants were discovered mostly in the three most common monogenic diabetes genes (5 in *HNF1A*, 7 in *GCK*, and 7 in *HNF4A*), but there were also discovered in less common MODY genes as well (2 in INS, 1 in KLF11). Finding patients with monogenic diabetes variants outside the three most common genes demonstrates the value of using gene panel sequencing for monogenic diabetes testing, although no pathogenic or likely pathogenic variants (according to ACMG/AMP guidelines for variant interpretations) were found in the 27 non-MODY monogenic diabetes genes on the panel. However, genetic variants in the 27 non-MODY that did not fulfill ACMG/AMP criteria for being pathogenic or likely pathogenic could potentially be damaging or at least warrant further study. Additionally, the high proportion of *HNF4A*-MODY patients discovered (32% of TODAY monogenic diabetes cases compared to published rates of approximately 10%), is hypothesized to be due to selection for overweight or obese adolescents that may accompany established paradoxical neonatal hyperinsulinemia, hypoglycemia and macrosomia of *HNF4A*-MODY patients.[153] Importantly, monogenic diabetes patients were discovered across the three races/ethnicities studied (Hispanic, NHB, NHW), although in higher proportion of the NHW population. Past studies of monogenic diabetes have mostly been focused on European populations, which may have enhanced the ability to classify monogenic diabetes variants in European-ancestry populations as pathogenic or likely pathogenic.[79] Since Hispanic and NHB populations are at higher risk for polygenic T2D, this finding could also indicate that a lower proportion of individuals in these minorities have monogenic diabetes misdiagnosed as T2D, even if the absolute number of monogenic diabetes cases is similar between groups. However, it is most important to note that

monogenic diabetes can be found in any race/ethnicity, and consideration of monogenic diabetes as a diagnosis should not be limited by race/ethnicity in any manner.

Comparison of clinical characteristics and treatment outcomes between patients with and without monogenic diabetes in the TODAY clinical trial was also performed. In this population selected for overweight or obese status, patients with monogenic diabetes had statistically, but not clinically, significant lower BMI Z-score, lower fasting insulin, and higher fasting blood glucose. While those characteristics are expected, patients with monogenic diabetes unexpectedly had higher total cholesterol and LDL cholesterol, a finding that needs further study in other similar cohorts (e.g. patients with monogenic diabetes that are overweight or obese). Separated by gene type of monogenic diabetes, each subgroup showed similar trends in clinical characteristics, but were generally underpowered to make definitive conclusions. Additionally, analysis of two variants previously reported to cause MODY but were reclassified as benign based on ACMG/AMP criteria, *BLK* p.A71T and *KLF11* p.T220M, showed none of the expected characteristics of monogenic diabetes. This finding indicates the importance of thorough evaluation of discovered sequencing variants using multiple categories of evidence, as indicated by the ACMG/AMP guidelines for variant interpretation.[119] When comparing treatment outcomes of patients with *HNF1A*-, *GCK*-, or *HNF4A*-MODY to those without monogenic diabetes, most patients with *HNF4A*-MODY failed metformin or metformin combination therapy across all study treatment arms rapidly (6/7 failed treatment, hazard ratio=5.03, p=0.0002), while no patients with *GCK*-MODY failed treatment. This finding serves as a demonstration of the effect of not attaining a genetic diagnosis of *HNF4A*-MODY, since those patients should theoretically respond better to insulin secretagogue

therapy rather than an insulin sensitizing agent like metformin. Similarly, though not statistically significant, 3/5 patients with *HNF1A*-MODY also failed treatment. Since patients with *GCK*-MODY have mildly-elevated fasting blood glucose (HbA1c between 5.6% and 7.6%) that should not reach the TODAY study treatment failure definition (HbA1c >8.0% for >6months), none were expected to fail treatment. Together, the findings of this study demonstrate that a significant proportion of overweight and obese children and adolescents with monogenic diabetes may be misdiagnosed as T2D, some of which may lead to inappropriate treatment.

2. Monogenic Diabetes Screening, Diagnosis, and Pedigree Analysis in the PDMP Study

As described in Chapter 3, we identified 14 participants with pathogenic or likely pathogenic variants for monogenic diabetes in a total of 1,734 screened individuals in the PDMP . Demographics of participants from the four study sites demonstrated the expected characteristics of each population: UM-CDE (urban) consisted of 60.2% African-Americans, GHS (rural) consisted of 94.3% Caucasians, BVAMC consisted of 91.5% males with the oldest mean age (60.9 ±10.4 years old). Of the 336 patients eligible for enrollment, less than half (46.4%) enrolled, while 19.1% declined participation and 17.1% were not able to be contacted. A larger proportion of African-Americans were unable to be contacted (OR=1.6), declined participation (OR=2.3), and cancelled visits (OR=4.9) compared to Caucasians, while Caucasians were more likely than African-Americans to enroll in the study (OR=3.6). This is another demonstration of the notion that minority populations are less likely to participate in genetic testing.[179] The 14 patients with pathogenic or likely pathogenic variants for monogenic diabetes displayed a range of characteristics: diabetes onset ranging from age 4 to age 54, body habitus

ranging from “thin” to “overweight,” and treatment regimens including lifestyle interventions, metformin, sulfonylureas, or insulin. More than half of those with monogenic diabetes (8/14) self-referred into the PDMP, while only one participant was discovered through the screening protocol (1/14). Participants who were physician-referred or self-referred for the PDMP still fit screening criteria. However, the low number of participants with monogenic diabetes discovered from screening the diabetes clinic populations may indicate that screening those populations may not yield large numbers of patients with monogenic diabetes. Pedigrees of patients with and without monogenic diabetes were compared using quantitation of percent of affected relatives and through customized pedigree-summary metrics followed by use of a machine-learning algorithm. Due to small number of pedigrees available for analysis (n=152) neither method could effectively differentiate patients with monogenic diabetes, although the customized pedigree-summary metrics performed better than previously suggested metrics.[177] Further studies on larger and more diverse pedigree datasets would be necessary to confirm the effectiveness of our customized pedigree-summary metric. This study demonstrates that patients with monogenic diabetes can have a broad range of individual and family medical history characteristics that are not easily differentiated from patients without monogenic diabetes, although patient self-referral for monogenic diabetes genetic testing may present potential for further study.

3. Zebrafish as a Model for On-Demand *In Vivo HNF1A* Variant Functional Analysis

As described in Chapter 3, this study assessed the potential use of a zebrafish model of *hnf1a* knockdown and rescue with mRNA containing *HNF1A* genetic variants as a method for analyzing the effect of *HNF1A* genetic variants in an *in vivo* system.

Through quantification of beta cell number and area, previously-reported effects of morpholino-induced zebrafish *hnf1a* knockdown could effectively be rescued with simultaneous injection of 50pg of wild-type *HNF1A* mRNA.[191] However, analysis of beta cell number and area in zebrafish rescued with six variants established to be pathogenic for *HNF1A*-MODY with previously-published *in vitro* studies did not demonstrate a consistent pattern of inhibited rescue compared to wild-type rescue conditions. Fewer beta cells and smaller beta cell area compared to wild-type *HNF1A* rescue were observed in zebrafish rescued with *HNF1A* variants p.G20R and p.R131W, indicating the opportunity for future study of the mechanism for inhibited rescue using these variants compared to the other tested variants. This study demonstrated that, at this time, zebrafish *hnf1a* knockdown and rescue with mutant *HNF1A* variants is not an effective model for identifying pathogenic *HNF1A*-MODY variants.

B. Potential Impact

Monogenic diabetes has been estimated to account for approximately 1-2% of all diabetes cases.[15] In the US, where there are approximately 29.1 million individuals with diabetes, this could account for up to 300,000 individuals. It is estimated that there are approximately 3,700 new diagnoses of T2D in children and adolescents each year.[108] The finding from Chapter 2 of this dissertation that 4.5% of TODAY participants have monogenic diabetes indicates that as many as 160 of those new cases may be due to monogenic diabetes. In the current epidemic of childhood obesity that affects as many as one in five school-aged children, assumption of a diagnosis of T2D in obese children without considering monogenic diabetes is a valid concern.[205] Chapter 2 of this

dissertation demonstrates the effect on treatment outcomes when this misdiagnosis is made. Additionally, the estimate that 1-2% of diabetes cases are actually monogenic diabetes comes from epidemiological studies in Europe, which has different rates of diabetes as well as obesity when compared to the US.[206] Chapter 3 is one of the few screening studies of monogenic diabetes in US populations,[109,124] and it suggests that further studies are necessary to delineate how the diabetes populations in diabetes clinics, under primary care, and that are undiagnosed in the US may differ from those in Europe. Finally, interpretation of genetic variants from sequencing platforms is a difficult process that requires a great deal of a broad spectrum of evidence in a publicly available or on-demand manner. Chapter 4 emphasizes the need for quality experimental models to aid in genetic variant interpretation, as well as the difficulty of establishing such models.

In this dissertation, we evaluated screening methods and practices in multiple cohorts to guide future efforts of diagnosis. While a nationwide screening program for monogenic diabetes could potentially discover many cases, the results from these studies have demonstrated the difficulty of assigning screening criteria to such a heterogeneous patient population. This dissertation has shown that patients with monogenic diabetes can unexpectedly have increased BMI, little family history of monogenic diabetes, older age of diagnosis, and a wide range of diabetes symptoms. Therefore, using these pieces of information to set criteria would either require testing a large number of patients or result in missed cases of monogenic diabetes. Improvements in NGS may provide one prospect for the future of genetic testing for monogenic diabetes since the cost of whole genome or exome sequencing is constantly becoming more affordable and efficient. This provides an opportunity to automatically assess prospectively-collected whole genome or exome data

for monogenic diabetes variants when signs of diabetes are observed. Successful implementation of this process would require financial, structural, and organizational support for attaining, storing, and accessing large amounts of prospective sequencing data, which is an enormous task. However, the findings from this dissertation could inform the process of monogenic diabetes variant interpretation, if the sequencing data is available. We discovered a large number of variants of uncertain significance in monogenic diabetes genes, demonstrating the need for future methods of gaining information about the effects of those variants. While this dissertation cannot provide directives for effective screening practices at this time, it informs about the difficulty of interpretation of variants outside the most common monogenic diabetes genes due to the lack of published information about those variants.

C. Limitations

While genetic testing has rapidly advanced over the past two decades, pitfalls and weaknesses are constantly being discovered and improved. Our sequencing methodology using a customized monogenic diabetes gene panel and the Ion Torrent Personalized Genome Machine (PGM) sequencing platform has multiple limitations. This panel limits discovery to established genes designed into the panel. *APPL1* is an example of a monogenic diabetes gene that was discovered after the design of our panel, and therefore is not included.[58] Additionally, this panel does not allow for discovery of novel genes that may cause monogenic diabetes. Similarly, the panel only includes the protein-coding and flanking regions of monogenic diabetes genes. Although the panel could be expanded, in the current state it precludes the possibility of identification of variants in promoter,

intronic, or intergenic regions, as well as large (partial or whole-gene) insertions or deletions. This is particularly problematic for *HNF1B*-MODY which has been shown to be caused by partial or whole-gene deletions or rearrangements.[45] Additionally, each sequencing platform has its own weaknesses, and the PGM is particularly weak at accurately detecting homo-oligomer stretches and generally has a higher error rate than the more-commonly used Illumina sequencing platforms.[207] However, the PGM does more easily accommodate lower sample volume and is generally less expensive than other sequencing platforms. While each sequencing method has individual strengths and weaknesses, it is important to recognize the potential pitfalls of the methods used for this dissertation.

For any genetic study, sample size is important to sufficiently power statistical analyses. The analysis of monogenic diabetes in the TODAY population was sufficiently powered to compare characteristics of those with and without monogenic diabetes as well as the treatment outcomes of the gene-specific subtypes of monogenic diabetes. However, larger sample sizes could lead to discovery of more patients with monogenic diabetes that could potentially allow for further dissection of gene-specific monogenic diabetes clinical characteristics as well as analyses of interactions between study arm and monogenic diabetes subtypes. Likewise, the PDMP study is currently underpowered to make definitive assertions of meaningful differences in patient characteristics between those with and without monogenic diabetes, especially with extremely variable assessments such as data collected from family medical histories. Since the overall goal of the PDMP is to measure implementation of a method for monogenic diabetes screening, diagnosis, and return of results, the design of the study is not ideally powered

for comparisons of patient characteristics, and the study is generally underpowered for the analyses presented in this dissertation.

An additional factor that could potentially affect findings of genetic sequencing projects could be interpretation of discovered genetic variants. In this dissertation, we evaluated variants based on the guidelines for variant interpretation put forth by the ACMG/AMP.[119] These guidelines require evidence across multiple categories (such as population data, functional study data, *in silico* prediction tools, family segregation data, etc.), which results conservative estimates of variant pathogenicity. As a result, many genetic variants are classified as "Variants of Uncertain Significance" (VUS). However, at least some of these genetic variants may be causative for monogenic diabetes, but insufficient information is available to classify the variants as pathogenic or likely pathogenic. This potential misclassification of genetic variants could weaken or possibly abolish associations with patient traits. Although this is a concern for both the TODAY and PDMP studies, a greater concern is potential misclassification of variants as causative for monogenic diabetes, and the eventual propagation of that misclassification. Therefore, we believe that conservative estimates of variant pathogenicity are more appropriate, even if it results in loss of potentially significant associations.

Another limitation that can hamper genetic studies is lacking quality or depth of phenotype data. The studies presented in this dissertation are no different, and monogenic diabetes can have such a range of different characteristics that it may be difficult to capture all aspects of the monogenic diabetes phenotype. For example, patients with *HNF1A*-MODY have been shown to have altered lipoprotein profiles, plasma glycan profiles, and C-reactive protein (CRP) levels, in addition to general characteristics such

as early diabetes onset and strong family history of diabetes.[157-159] Therefore, it can be difficult to capture the complete profile of patients with monogenic diabetes, especially considering the subtle phenotype differences of monogenic diabetes of varying genetic etiologies. The TODAY study phenotype data was gathered as part of a clinical trial that collected general clinical measures as well as characteristics focused on T2D traits. However, measures previously associated with monogenic diabetes, such as CRP, were not collected as part of the clinical trial. Family history of diabetes was assessed in the TODAY study as maternal and paternal diabetes statuses, but it does not explore number of consecutive diabetes-affected generations in each family, as is common practice for studies designed to specifically target monogenic diabetes. As a screening and implementation study for monogenic diabetes genetic testing and return of results, the PDMP study has a stronger focus on educating participants and monitoring the potential effects of attaining a monogenic diabetes diagnosis than collecting rich patient phenotype data to aid diagnosis. As a result, much of the patient phenotype data is self-reported information, including answers to the screening questionnaire and pedigree descriptions. Self-reported information is potentially unreliable, since individuals may confuse information or be reticent about revealing potentially embarrassing information (e.g. obesity status at diabetes diagnosis). In the PDMP, clinical measures, such as diabetes autoantibodies and C-peptide levels, have been collected either as part of this study or pulled from the electronic medical record, greatly complicating the prospect of analyzing clinical characteristics of the patients. Both the TODAY study and the PDMP have limitations to phenotypic data available, and it is important to recognize that as a result,

they can only describe a portion of the characteristics of patients with monogenic diabetes.

As described in Chapter 4, there are many limitations to the evaluation of zebrafish as a model to assess *HNF1A* genetic variant pathogenicity. First, the measures (beta cell number and area) used to assess pathogenicity may not directly quantify defects resulting from *HNF1A* genetic variants. Other measures, such as blood glucose or blood insulin measures may more directly measure the effects of *HNF1A* variants. Also, zebrafish physiology may not be as sensitive to the same *HNF1A* genetic variants as humans. For example, an *HNF1A* variant that may alter a residue in the DNA-binding domain could lead to altered binding efficiency to target regions crucial for glucose regulation, but zebrafish may have complementary or backup mechanisms to ameliorate the effects of partially inhibited *HNF1A* function. Likewise, human transcription factor may not have the same affinity for proteins or protein orthologs in a different species like zebrafish. Finally, stressors, such as nutritional excess, may be necessary to induce beta cell expansion to highlight subtle differences in beta cell number or area between the wild-type rescue condition and the *HNF1A* mutant rescue conditions. The measures of beta cell number and area used for this study may not be sensitive enough to detect rescue with *HNF1A* that has partial function. Further study would be necessary to tease out these specific limitations, but may not improve the model sufficiently for confident prediction of the effects of *HNF1A* variants. Additionally, other quality control measures, such as quantifying protein expression of recombinant genes or determining dose-response curves of specific variants, would be necessary for reliable implementation of the model.

In practice, it may be both more efficient and effective to continue to assess the function of *HNF1A* through traditional methods such as manipulation of cultured human cells.

D. Opportunities and future directions

The TODAY monogenic diabetes study was originally conceived to be a post-trial analysis using available DNA from a subset of study participants. However, after the discovery of 22 patients with pathogenic or likely pathogenic monogenic diabetes variants, it was determined to be necessary to act on these findings to potentially improve care for patients in the study. A process was developed to offer the patients the opportunity to have their genetic variants clinically confirmed in the CLIA/CAP accredited University of Maryland Translational Genomics Lab. Patients from the TODAY study with pathogenic or likely pathogenic variants are currently being invited to enroll so their variants can be confirmed. After confirmation, the patients and their physicians can receive the results. This transition from a post-clinical trial research study to potential improvement in patient care is a unique and exciting opportunity for the implementation of personalized medicine.

The PDMP continues to enroll and perform genetic testing on study participants. Processes are currently underway to expand genetic testing to increase the number of patients being tested as part of the study. This will be accomplished by testing subjects who wish to enroll despite not fitting the original enrollment criteria for monogenic diabetes testing, as well as pursuing telemedicine recruitment protocols, and inviting potential referrals from private medical practices through educational seminars about the PDMP and monogenic diabetes. Patient outcome studies are also underway to assess the

qualitative impact that a monogenic diabetes diagnosis has on the patients. It is likely that thorough description of the effects of a genetic diagnosis of monogenic diabetes may be one of the most important pieces of information to come from the PDMP.

There is currently a rich opportunity for study of monogenic diabetes in the US. Although the discovery of different forms of monogenic diabetes have occurred all over the world, the University of Exeter Molecular Genetics Lab has made consistent progress in monogenic diabetes research for years by leveraging the universal healthcare system of the UK and their status as the UK nationwide referral center for monogenic diabetes testing.[14] While they have made a great deal of progress in the diagnosis of monogenic diabetes, their research has largely been focused on UK and European cohorts. Since the US population has different characteristics in terms of diabetes prevalence, obesity prevalence, and genetic ancestry, the diagnostic criteria based on studies from Exeter may not be as effective in US patient populations.[206] Monogenic diabetes studies in the US have been constrained by a fractured, often-changing healthcare system and lack of a designated referral center for monogenic diabetes testing. However, the National Institute of Diabetes and Digestive and Kidney Diseases have recently put forth a request for applications for a U54 Specialized Center grant to fund a Center for Identification and Study of Individuals with Atypical Diabetes Mellitus.[208] The group awarded this grant would identify individuals with rare and uncharacterized diabetes (including monogenic diabetes) as well as creating a database and repository for specimens to be analyzed. This center could effectively serve as a referral center for monogenic diabetes testing in the US, and they could make great strides in defining diagnostic principles for monogenic diabetes in the US as well as determining how that may differ from European

populations. Although the center likely will not be fully active for a couple years, it is an exciting time and opportunity for monogenic diabetes research in the US.

The information gained from this dissertation can potentially go on to inform about implementation of personalized medicine beyond the field of monogenic diabetes. Monogenic diabetes represents an excellent opportunity for personalized medicine because a genetic diagnosis can inform treatment regimens that can lead to better patient care. Using monogenic diabetes as a model for providing personalized medicine, it can serve as a roadmap for other, more complex diseases. For example, other forms of diabetes have multiple genetic elements that can influence factors in patient care, such as susceptibility, disease progression, or medication response. For those variants, it is necessary to examine methods for variant identification via genetic testing, interpreting the variant impact, and effectively returning that information to the patient and healthcare providers to properly utilize the information. The studies in this dissertation can provide a model of how genetic variants that may affect patient care, including those related to more common forms of diabetes as well as any other diseases under genetic influence, can be identified, interpreted, and returned to patients as a form of personalized medicine.

Finally, worldwide technological advancements have provided the opportunity for people to connect with each other more easily than ever in history. For patients who have or suspect they have a genetic condition like monogenic diabetes, the internet can provide the opportunity to learn about the disorder and potentially connect with others for support. One website, patientslikeme.com is an example of a virtual community where a patient can connect with others having the same medical condition to share information and support.[209] It should be recognized that PatientsLikeMe is a for-profit company that

sells information from the patients' experiences collected through the community web platform to companies developing or selling products to patients. Additionally, it is easier than ever to travel large distances for medical treatment. Some PDMP participants traveled hundreds of miles to enroll because they discovered that the PDMP provided a means to attain genetic testing for their suspected monogenic diabetes unavailable through their local healthcare providers. Additionally, multiple individuals with monogenic diabetes have demonstrated that they would not only go to extreme lengths to attain a genetic diagnosis for themselves, but that they would serve as enthusiastic advocates for others to also gain access to genetic testing. The willingness to travel and advocate for monogenic diabetes genetic testing demonstrates the enormous effect of attaining a genetic diagnosis of monogenic diabetes. Even though improving diagnosis of monogenic diabetes may not save millions of lives, the passion of patients establishes the value and impact that genetic testing can have on quality of life, potential diagnosis of family members, and disease progression outlook at an individual level and expresses the importance of continuing to improve methods and recommendations for diagnosis of monogenic diabetes.

References

1. PubMed Health. **Type 1 diabetes: Overview**. https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0072523/. Accessed 4/11, 2017.

2. Centers for Disease Control and Prevention. *National diabetes statistics report: Estimates of diabetes and its burden in the united states, 2014*. *US Department of Health and Human Services*. 2014.

3. PubMed Health. **Type 2 diabetes: Overview**. https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0072693/. Accessed 4/11, 2017.

4. American Diabetes Association. 2. classification and diagnosis of diabetes. *Diabetes Care*. 2017;40(Suppl 1):S11-S24. doi: 40/Supplement_1/S11 [pii].

5. Chiefari E, Arcidiacono B, Foti D, Brunetti A. Gestational diabetes mellitus: An updated overview. *J Endocrinol Invest*. 2017. doi: 10.1007/s40618-016-0607-5 [doi].

6. American Diabetes Association (ADA). Statistics about diabetes. http://www.diabetes.org/diabetes-basics/statistics/. Accessed 4/11, 2017.

7. Redondo MJ, Yu L, Hawa M, et al. Heterogeneity of type I diabetes: Analysis of monozygotic twins in great britain and the united states. *Diabetologia*. 2001;44(3):354-362. doi: 10.1007/s001250051626 [doi].

8. Almgren P, Lehtovirta M, Isomaa B, et al. Heritability and familiality of type 2 diabetes and related quantitative traits in the botnia study. *Diabetologia*. 2011;54(11):2811-2819. doi: 10.1007/s00125-011-2267-5 [doi].

9. Erlich H, Valdes AM, Noble J, et al. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: Analysis of the type 1 diabetes genetics consortium families. *Diabetes*. 2008;57(4):1084-1092. doi: 10.2337/db07-1331 [doi].

10. Torn C, Hadley D, Lee HS, et al. Role of type 1 diabetes-associated SNPs on risk of autoantibody positivity in the TEDDY study. *Diabetes*. 2015;64(5):1818-1829. doi: 10.2337/db14-1497 [doi].

11. Prasad RB, Groop L. Genetics of type 2 diabetes-pitfalls and possibilities. *Genes (Basel)*. 2015;6(1):87-123. doi: 10.3390/genes6010087 [doi].

12. InterAct Consortium, Scott RA, Langenberg C, et al. The link between family history and risk of type 2 diabetes is not explained by anthropometric, lifestyle or genetic risk factors: The EPIC-InterAct study. *Diabetologia*. 2013;56(1):60-69. doi: 10.1007/s00125-012-2715-x [doi].

13. Kleinberger JW, Maloney KA, Pollin TI. The genetic architecture of diabetes in pregnancy: Implications for clinical practice. *Am J Perinatol*. 2016;33(13):1319-1326. doi: 10.1055/s-0036-1592078 [doi].

14. 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. doi: dc160645 [pii].

15. Shields BM, Hicks S, Shepherd MH, Colclough K, Hattersley AT, Ellard S. Maturity-onset diabetes of the young (MODY): How many cases are we missing? *Diabetologia*. 2010;53(12):2504-2508. doi: 10.1007/s00125-010-1799-4 [doi].

16. Genetics Home Reference. **permanent neonatal diabetes mellitus**. https://ghr.nlm.nih.gov/condition/permanent-neonatal-diabetes-mellitus. Accessed 4/11, 2017.

17. Flannick J, Johansson S, Njolstad PR. Common and rare forms of diabetes mellitus: Towards a continuum of diabetes subtypes. *Nat Rev Endocrinol*. 2016;12(7):394-406. doi: 10.1038/nrendo.2016.50 [doi].

18. 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. doi: 10.1038/384458a0 [doi].

19. 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. doi: 10.1038/384455a0 [doi].

20. 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. doi: 10.1038/356721a0 [doi].

21. Klupa T, Skupien J, Malecki MT. Monogenic models: What have the single gene disorders taught us? *Curr Diab Rep*. 2012;12(6):659-666. doi: 10.1007/s11892-012-0325-0 [doi].

22. Rebouissou S, Vasiliu V, Thomas C, et al. Germline hepatocyte nuclear factor 1alpha and 1beta mutations in renal cell carcinomas. *Hum Mol Genet*. 2005;14(5):603-614. doi: ddi057 [pii].

23. Bluteau O, Jeannot E, Bioulac-Sage P, et al. Bi-allelic inactivation of TCF1 in hepatic adenomas. *Nat Genet*. 2002;32(2):312-315. doi: 10.1038/ng1001 [doi].

24. Bach I, Yaniv M. More potent transcriptional activators or a transdominant inhibitor of the HNF1 homeoprotein family are generated by alternative RNA processing. *EMBO J*. 1993;12(11):4229-4242.

25. Sladek F.M. SSD. **Hepatocyte nuclear factor 4α**. In: T. Burris ERBM, ed. *Nuclear receptors and genetic diseases.* London, England: Academic Press; 2001:309.

26. Harries LW, Ellard S, Stride A, Morgan NG, Hattersley AT. Isomers of the TCF1 gene encoding hepatocyte nuclear factor-1 alpha show differential expression in the pancreas and define the relationship between mutation position and clinical phenotype in monogenic diabetes. *Hum Mol Genet*. 2006;15(14):2216-2224. doi: ddl147 [pii].

27. Thomas H, Jaschkowitz K, Bulman M, et al. A distant upstream promoter of the HNF-4alpha gene connects the transcription factors involved in maturity-onset diabetes of the young. *Hum Mol Genet*. 2001;10(19):2089-2097.

28. Dean S, Tang JI, Seckl JR, Nyirenda MJ. Developmental and tissue-specific regulation of hepatocyte nuclear factor 4-alpha (HNF4-alpha) isoforms in rodents. *Gene Expr*. 2010;14(6):337-344.

29. Boj SF, Petrov D, Ferrer J. Epistasis of transcriptomes reveals synergism between transcriptional activators Hnf1alpha and Hnf4alpha. *PLoS Genet*. 2010;6(5):e1000970. doi: 10.1371/journal.pgen.1000970 [doi].

30. Eeckhoute J, Formstecher P, Laine B. Hepatocyte nuclear factor 4alpha enhances the hepatocyte nuclear factor 1alpha-mediated activation of transcription. *Nucleic Acids Res*. 2004;32(8):2586-2593. doi: 10.1093/nar/gkh581 [doi].

31. Servitja JM, Pignatelli M, Maestro MA, et al. Hnf1alpha (MODY3) controls tissue-specific transcriptional programs and exerts opposed effects on cell growth in pancreatic islets and liver. *Mol Cell Biol*. 2009;29(11):2945-2959. doi: 10.1128/MCB.01389-08 [doi].

32. Odom DT, Zizlsperger N, Gordon DB, et al. Control of pancreas and liver gene expression by HNF transcription factors. *Science*. 2004;303(5662):1378-1381. doi: 10.1126/science.1089769 [doi].

33. Iynedjian PB. Molecular physiology of mammalian glucokinase. *Cell Mol Life Sci*. 2009;66(1):27-42. doi: 10.1007/s00018-008-8322-9 [doi].

34. 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. doi: 10.1001/jama.2013.283980 [doi].

35. Njolstad PR, Sovik O, Cuesta-Munoz A, et al. Neonatal diabetes mellitus due to complete glucokinase deficiency. *N Engl J Med*. 2001;344(21):1588-1592. doi: 10.1056/NEJM200105243442104 [doi].

36. 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. doi: 10.1038/ng0197-106 [doi].

37. 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. doi: 10.1210/jc.2003-030046 [doi].

38. Ohlsson H, Karlsson K, Edlund T. IPF1, a homeodomain-containing transactivator of the insulin gene. *EMBO J*. 1993;12(11):4251-4259.

39. 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. doi: 10.1038/15500 [doi].

40. Rubio-Cabezas O, Minton JA, Kantor I, Williams D, Ellard S, Hattersley AT. Homozygous mutations in NEUROD1 are responsible for a novel syndrome of permanent neonatal diabetes and neurological abnormalities. *Diabetes*. 2010;59(9):2326-2331. doi: 10.2337/db10-0011 [doi].

41. 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. doi: jc.2006-1927 [pii].

42. Shimajiri Y, Sanke T, Furuta H, et al. A missense mutation of Pax4 gene (R121W) is associated with type 2 diabetes in japanese. *Diabetes*. 2001;50(12):2864-2869.

43. Mauvais-Jarvis F, Smith SB, Le May C, et al. PAX4 gene variations predispose to ketosis-prone diabetes. *Hum Mol Genet*. 2004;13(24):3151-3159. doi: ddh341 [pii].

44. Maestro MA, Cardalda C, Boj SF, Luco RF, Servitja JM, Ferrer J. Distinct roles of HNF1beta, HNF1alpha, and HNF4alpha in regulating pancreas development, beta-cell function and growth. *Endocr Dev*. 2007;12:33-45. doi: 109603 [pii].

45. Bellanne-Chantelot C, Clauin S, Chauveau D, et al. Large genomic rearrangements in the hepatocyte nuclear factor-1beta (TCF2) gene are the most frequent cause of maturity-onset diabetes of the young type 5. *Diabetes*. 2005;54(11):3126-3132. doi: 54/11/3126 [pii].

46. 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. doi: 0409177102 [pii].

47. 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. doi: 10.1074/jbc.M110.215822 [doi].

48. Lomberk G, Grzenda A, Mathison A, et al. Kruppel-like factor 11 regulates the expression of metabolic genes via an evolutionarily conserved protein interaction domain functionally disrupted in maturity onset diabetes of the young. *J Biol Chem*. 2013;288(24):17745-17758. doi: 10.1074/jbc.M112.434670 [doi].

49. 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. doi: 10.1073/pnas.0906474106 [doi].

50. Bonnefond A, Yengo L, Philippe J, et al. Reassessment of the putative role of BLK-p.A71T loss-of-function mutation in MODY and type 2 diabetes. *Diabetologia*. 2013;56(3):492-496. doi: 10.1007/s00125-012-2794-8 [doi].

51. 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. doi: ng1708 [pii].

52. 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. doi: db07-1405 [pii].

53. 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. doi: 10.1016/j.mam.2014.12.001 [doi].

54. Stoy J, Steiner DF, Park SY, Ye H, Philipson LH, Bell GI. Clinical and molecular genetics of neonatal diabetes due to mutations in the insulin gene. *Rev Endocr Metab Disord*. 2010;11(3):205-215. doi: 10.1007/s11154-010-9151-3 [doi].

55. McCarthy MI, Hattersley AT. Learning from molecular genetics: Novel insights arising from the definition of genes for monogenic and type 2 diabetes. *Diabetes*. 2008;57(11):2889-2898. doi: 10.2337/db08-0343 [doi].

56. Bowman P, Flanagan SE, Edghill EL, et al. Heterozygous ABCC8 mutations are a cause of MODY. *Diabetologia*. 2012;55(1):123-127. doi: 10.1007/s00125-011-2319-x [doi].

57. 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. doi: 10.1371/journal.pone.0037423 [doi].

58. 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. doi: 10.1016/j.ajhg.2015.05.011 [doi].

59. Polak M, Cave H. Neonatal diabetes mellitus: A disease linked to multiple mechanisms. *Orphanet J Rare Dis*. 2007;2:12. doi: 1750-1172-2-12 [pii].

60. Arima T, Drewell RA, Oshimura M, Wake N, Surani MA. A novel imprinted gene, HYMAI, is located within an imprinted domain on human chromosome 6 containing ZAC. *Genomics*. 2000;67(3):248-255. doi: 10.1006/geno.2000.6266 [doi].

61. Gardner RJ, Mackay DJ, Mungall AJ, et al. An imprinted locus associated with transient neonatal diabetes mellitus. *Hum Mol Genet*. 2000;9(4):589-596. doi: ddd067 [pii].

62. Mackay DJ, Callaway JL, Marks SM, et al. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nat Genet*. 2008;40(8):949-951. doi: 10.1038/ng.187 [doi].

63. Gambineri E, Perroni L, Passerini L, et al. Clinical and molecular profile of a new series of patients with immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome: Inconsistent correlation between forkhead box protein 3 expression and disease severity. *J Allergy Clin Immunol*. 2008;122(6):1105-1112.e1. doi: 10.1016/j.jaci.2008.09.027 [doi].

64. Rigoli L, Lombardo F, Di Bella C. Wolfram syndrome and WFS1 gene. *Clin Genet*. 2011;79(2):103-117. doi: 10.1111/j.1399-0004.2010.01522.x [doi].

65. Sandhu MS, Weedon MN, Fawcett KA, et al. Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet*. 2007;39(8):951-953. doi: ng2067 [pii].

66. Kaur Y, de Souza RJ, Gibson WT, Meyre D. A systematic review of genetic syndromes with obesity. *Obes Rev*. 2017. doi: 10.1111/obr.12531 [doi].

67. Kleinberger JW, Brown K, Silver KD, Shuldiner AR. **Genetics of type 2 diabetes: From candidate genes to genome-wide association analysis**. In: Poretsky L, ed. *Principles of diabetes mellitus.* New York, NY: Springer International Pub; 2017:1-24.

68. Agarwal AK, Arioglu E, De Almeida S, et al. AGPAT2 is mutated in congenital generalized lipodystrophy linked to chromosome 9q34. *Nat Genet*. 2002;31(1):21-23. doi: 10.1038/ng880 [doi].

69. Magre J, Delepine M, Khallouf E, et al. Identification of the gene altered in berardinelli-seip congenital lipodystrophy on chromosome 11q13. *Nat Genet*. 2001;28(4):365-370. doi: 10.1038/ng585 [doi].

70. Kim CA, Delepine M, Boutet E, et al. Association of a homozygous nonsense caveolin-1 mutation with berardinelli-seip congenital lipodystrophy. *J Clin Endocrinol Metab*. 2008;93(4):1129-1134. doi: 10.1210/jc.2007-1328 [doi].

71. Hayashi YK, Matsuda C, Ogawa M, et al. Human PTRF mutations cause secondary deficiency of caveolins resulting in muscular dystrophy with generalized lipodystrophy. *J Clin Invest*. 2009;119(9):2623-2633. doi: 10.1172/JCI38660 [doi].

72. Shackleton S, Lloyd DJ, Jackson SN, et al. LMNA, encoding lamin A/C, is mutated in partial lipodystrophy. *Nat Genet*. 2000;24(2):153-156. doi: 10.1038/72807 [doi].

73. Barroso I, Gurnell M, Crowley VE, et al. Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature*. 1999;402(6764):880-883. doi: 10.1038/47254 [doi].

74. Hegele RA, Cao H, Frankowski C, Mathews ST, Leff T. PPARG F388L, a transactivation-deficient mutant, in familial partial lipodystrophy. *Diabetes*. 2002;51(12):3586-3590.

75. Gandotra S, Le Dour C, Bottomley W, et al. Perilipin deficiency and autosomal dominant partial lipodystrophy. *N Engl J Med*. 2011;364(8):740-748. doi: 10.1056/NEJMoa1007487 [doi].

76. Rubio-Cabezas O, Puri V, Murano I, et al. Partial lipodystrophy and insulin resistant diabetes in a patient with a homozygous nonsense mutation in CIDEC. *EMBO Mol Med*. 2009;1(5):280-287. doi: 10.1002/emmm.200900037 [doi].

77. Carboni N, Brancati F, Cocco E, et al. Partial lipodystrophy associated with muscular dystrophy of unknown genetic origin. *Muscle Nerve*. 2014;49(6):928-930. doi: 10.1002/mus.24157 [doi].

78. Weinreich SS, Bosma A, Henneman L, et al. A decade of molecular genetic testing for MODY: A retrospective study of utilization in the netherlands. *Eur J Hum Genet*. 2015;23(1):29-33. doi: 10.1038/ejhg.2014.59 [doi].

79. Kleinberger JW, Pollin TI. Undiagnosed MODY: Time for action. *Curr Diab Rep*. 2015;15(12):110-015-0681-7. doi: 10.1007/s11892-015-0681-7 [doi].

80. Søvika O, Irgens HU, Molnes J, et al. **Monogenic diabetes mellitus in norway**. *Norwegian Journal of Epidemiology*. 2013;23(1):55-60.

81. Johansen A, Ek J, Mortensen HB, Pedersen O, Hansen T. Half of clinically defined maturity-onset diabetes of the young patients in denmark do not have mutations in HNF4A, GCK, and TCF1. *J Clin Endocrinol Metab*. 2005;90(8):4607-4614. doi: jc.2005-0196 [pii].

82. 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. doi: 10.1007/s00125-012-2621-2 [doi].

83. Schober E, Rami B, Grabert M, et al. Phenotypical aspects of maturity-onset diabetes of the young (MODY diabetes) in comparison with type 2 diabetes mellitus (T2DM) in children and adolescents: Experience from a large multicentre database. *Diabet Med*. 2009;26(5):466-473. doi: 10.1111/j.1464-5491.2009.02720.x [doi].

84. Pruhova S, Ek J, Lebl J, et al. Genetic epidemiology of MODY in the czech republic: New mutations in the MODY genes HNF-4alpha, GCK and HNF-1alpha. *Diabetologia*. 2003;46(2):291-295. doi: 10.1007/s00125-002-1010-7 [doi].

85. Lorini R, Klersy C, d'Annunzio G, et al. Maturity-onset diabetes of the young in children with incidental hyperglycemia: A multicenter italian study of 172 families. *Diabetes Care*. 2009;32(10):1864-1866. doi: 10.2337/dc08-2018 [doi].

86. 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):e258-60. doi: 10.2337/dc14-1788 [doi].

87. Tatsi C, Kanaka-Gantenbein C, Vazeou-Gerassimidi A, et al. The spectrum of HNF1A gene mutations in greek patients with MODY3: Relative frequency and identification of seven novel germline mutations. *Pediatr Diabetes*. 2013;14(7):526-534. doi: 10.1111/pedi.12032 [doi].

88. 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. doi: CEN2921 [pii].

89. 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. doi: 10.1007/s00125-013-2916-y [doi].

90. 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. doi: 10.1111/j.1399-5448.2011.00827.x [doi].

91. Xu JY, Dan QH, Chan V, et al. Genetic and clinical characteristics of maturity-onset diabetes of the young in chinese patients. *Eur J Hum Genet*. 2005;13(4):422-427. doi: 5201347 [pii].

92. Lim DM, Huh N, Park KY. Hepatocyte nuclear factor 1-alpha mutation in normal glucose-tolerant subjects and early-onset type 2 diabetic patients. *Korean J Intern Med*. 2008;23(4):165-169. doi: 10.3904/kjim.2008.23.4.165 [doi].

93. Kanthimathi S, Jahnavi S, Balamurugan K, et al. Glucokinase gene mutations (MODY 2) in asian indians. *Diabetes Technol Ther*. 2014;16(3):180-185. doi: 10.1089/dia.2013.0244 [doi].

94. Elbein SC, Teng K, Eddings K, Hargrove D, Scroggin E. Molecular scanning analysis of hepatocyte nuclear factor 1a (TCF1) gene in typical familial type 2 diabetes in african americans. *Metab Clin Exp*. 2000;49(2):280 <last_page> 284. doi: 10.1016/S0026-0495(00)91663-9.

95. Furuzawa GK, Giuffrida FM, Oliveira CS, Chacra AR, Dib SA, Reis AF. Low prevalence of MODY2 and MODY3 mutations in brazilian individuals with clinical MODY phenotype. *Diabetes Res Clin Pract*. 2008;81(3):e12-4. doi: 10.1016/j.diabres.2008.06.011 [doi].

96. Stern E, Strihan C, Potievsky O, et al. Four novel mutations, including the first gross deletion in TCF1, identified in HNF4a, GCK and TCF1 in patients with MODY in israel. *Journal of Pediatric Endocrinology and Metabolism*. 2007;20(8):909 <last_page> 922. doi: 10.1515/JPEM.2007.20.8.909.

97. Kessler MD, Yerges-Armstrong L, Taub MA, et al. Challenges and disparities in the application of personalized genomic medicine to populations with african ancestry. *Nat Commun*. 2016;7:12521. doi: 10.1038/ncomms12521 [doi].

98. 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. doi: 355/5/467 [pii].

99. Mak CM, Lee CY, Lam CW, Siu WK, Hung VC, Chan AY. Personalized medicine switching from insulin to sulfonylurea in permanent neonatal diabetes mellitus dictated by a novel activating ABCC8 mutation. *Diagn Mol Pathol*. 2012;21(1):56-59. doi: 10.1097/PDM.0b013e318220bb0e [doi].

100. Shepherd M, Hattersley AT. 'I don't feel like a diabetic any more': The impact of stopping insulin in patients with maturity onset diabetes of the young following genetic testing. *Clin Med*. 2004;4(2):144-147.

101. 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. doi: 10.1111/j.1464-5491.2009.02690.x [doi].

102. Bacon S, Kyithar MP, Rizvi SR, et al. Successful maintenance on sulphonylurea therapy and low diabetes complication rates in a HNF1A-MODY cohort. *Diabet Med*. 2016;33(7):976-984. doi: 10.1111/dme.12992 [doi].

103. Boileau P, Wolfrum C, Shih DQ, Yang TA, Wolkoff AW, Stoffel M. Decreased glibenclamide uptake in hepatocytes of hepatocyte nuclear factor-1alpha-deficient mice: A mechanism for hypersensitivity to sulfonylurea therapy in patients with maturity-onset diabetes of the young, type 3 (MODY3). *Diabetes*. 2002;51 Suppl 3:S343-8.

104. Urbanova J, Andel M, Potockova J, et al. Half-life of sulfonylureas in HNF1A and HNF4A human MODY patients is not prolonged as suggested by the mouse Hnf1a(-/-) model. *Curr Pharm Des*. 2015;21(39):5736-5748. doi: CPD-EPUB-70946 [pii].

105. Pearson ER, Pruhova S, Tack CJ, et al. Molecular genetics and phenotypic characteristics of MODY caused by hepatocyte nuclear factor 4alpha mutations in a large european collection. *Diabetologia*. 2005;48(5):878-885. doi: 10.1007/s00125-005-1738-y [doi].

106. Ekstrom N, Svensson AM, Miftaraj M, et al. Durability of oral hypoglycemic agents in drug naive patients with type 2 diabetes: Report from the swedish national diabetes register (NDR). *BMJ Open Diabetes Res Care*. 2015;3(1):e000059. doi: 10.1136/bmjdrc-2014-000059 [doi].

107. Borowiec M, Fendler W, Antosik K, et al. Doubling the referral rate of monogenic diabetes through a nationwide information campaign--update on glucokinase gene mutations in a polish cohort. *Clin Genet*. 2012;82(6):587-590. doi: 10.1111/j.1399-0004.2011.01803.x [doi].

108. Writing Group for the SEARCH for Diabetes in Youth Study Group, Dabelea D, Bell RA, et al. Incidence of diabetes in youth in the united states. *JAMA*. 2007;297(24):2716-2724. doi: 297/24/2716 [pii].

109. 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. doi: 10.1210/jc.2013-1279 [doi].

110. Clain E, Trosman JR, Douglas MP, Weldon CB, Phillips KA. Availability and payer coverage of BRCA1/2 tests and gene panels. *Nat Biotechnol*. 2015;33(9):900-902. doi: 10.1038/nbt.3322 [doi].

111. Bowden SA, Hoffman RP. Triple diabetes: Coexistence of type 1 diabetes mellitus and a novel mutation in the gene responsible for MODY3 in an overweight adolescent. *Pediatr Diabetes*. 2008;9(2):162-164. doi: 10.1111/j.1399-5448.2007.00335.x [doi].

112. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-291. doi: 10.1038/nature19057 [doi].

113. Tennessen JA, Bigham AW, O'Connor TD, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012;337(6090):64-69. doi: 10.1126/science.1219240 [doi].

114. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi: 10.1038/nature15393 [doi].

115. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(Database issue):D980-5. doi: 10.1093/nar/gkt1113 [doi].

116. Shihab HA, Rogers MF, Gough J, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics*. 2015;31(10):1536-1543. doi: 10.1093/bioinformatics/btv009 [doi].

117. Bendl J, Musil M, Stourac J, Zendulka J, Damborsky J, Brezovsky J. PredictSNP2: A unified platform for accurately evaluating SNP effects by exploiting the different characteristics of variants in distinct genomic regions. *PLoS Comput Biol*. 2016;12(5):e1004962. doi: 10.1371/journal.pcbi.1004962 [doi].

118. Chapla A, Mruthyunjaya MD, Asha HS, et al. Maturity onset diabetes of the young in india - a distinctive mutation pattern identified through targeted next-generation sequencing. *Clin Endocrinol (Oxf)*. 2015;82(4):533-542. doi: 10.1111/cen.12541 [doi].

119. 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):405-423. doi: 10.1038/gim.2015.30 [doi].

120. Harrison SM, Dolinsky JS, Knight Johnson AE, et al. Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. *Genet Med*. 2017. doi: 10.1038/gim.2017.14 [doi].

121. 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. doi: 10.1002/humu.22279 [doi].

122. 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. doi: 10.1002/humu.21110 [doi].

123. Flannick J, Beer NL, Bick AG, et al. Assessing the phenotypic effects in the general population of rare variants in genes for a dominant mendelian form of diabetes. *Nat Genet*. 2013;45(11):1380-1385. doi: 10.1038/ng.2794 [doi].

124. Chambers C, Fouts A, Dong F, et al. Characteristics of maturity onset diabetes of the young in a large diabetes center. *Pediatr Diabetes*. 2015. doi: 10.1111/pedi.12289 [doi].

125. 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. doi: 10.2337/dc11-2420 [doi].

126. Gloyn AL, Noordam K, Willemsen MA, et al. Insights into the biochemical and genetic basis of glucokinase activation from naturally occurring hypoglycemia mutations. *Diabetes*. 2003;52(9):2433-2440.

127. Ding SY, Tribble ND, Kraft CA, Markwardt M, Gloyn AL, Rizzo MA. Naturally occurring glucokinase mutations are associated with defects in posttranslational S-nitrosylation. *Mol Endocrinol*. 2010;24(1):171-177. doi: 10.1210/me.2009-0138 [doi].

128. Balamurugan K, Bjorkhaug L, Mahajan S, et al. Structure - function studies of HNF1A (MODY3) gene mutations in south indian patients with monogenic diabetes. *Clin Genet*. 2016. doi: 10.1111/cge.12757 [doi].

129. Magana-Cerino JM, Luna-Arias JP, Labra-Barrios ML, Avendano-Borromeo B, Boldo-Leon XM, Martinez-Lopez MC. Identification and functional analysis of c.422_423InsT, a novel mutation of the HNF1A gene in a patient with diabetes. *Mol Genet Genomic Med*. 2016;5(1):50-65. doi: 10.1002/mgg3.261 [doi].

130. 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. doi: 10.1038/ng.3700 [doi].

131. Tattersall RB, Fajans SS. A difference between the inheritance of classical juvenile-onset and maturity-onset type diabetes of young people. *Diabetes*. 1975;24(1):44-53.

132. 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. doi: 10.1038/ncpendmet0778 [doi].

133. 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. doi: 10.2337/dc13-0410 [doi].

134. 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. doi: PDI237 [pii].

135. TODAY Study Group, Zeitler P, Hirst K, et al. A clinical trial to maintain glycemic control in youth with type 2 diabetes. *N Engl J Med*. 2012;366(24):2247-2256. doi: 10.1056/NEJMoa1109333 [doi].

136. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology*. 1990;1(1):43-46.

137. Forlani G, Zucchini S, Di Rocco A, et al. Double heterozygous mutations involving both HNF1A/MODY3 and HNF4A/MODY1 genes: A case report. *Diabetes Care*. 2010;33(11):2336-2338. doi: 10.2337/dc10-0561 [doi].

138. Monney CT, Kaltenrieder V, Cousin P, Bonny C, Schorderet DF. Large family with maturity-onset diabetes of the young and a novel V121I mutation in HNF4A. *Hum Mutat*. 2002;20(3):230-231. doi: 10.1002/humu.9050 [doi].

139. Gloyn AL, Noordam K, Willemsen MA, et al. Insights into the biochemical and genetic basis of glucokinase activation from naturally occurring hypoglycemia mutations. *Diabetes*. 2003;52(9):2433-2440.

140. 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. doi: 10.1007/s001250050038 [doi].

141. Bertini C, Maioli M, Fresu P, Tonolo G, Pirastu M, Maioli M. A new missense mutation in the glucokinase gene in an italian mody family. *Diabetologia*. 1996;39(11):1413-1414.

142. Galan M, Vincent O, Roncero I, et al. Effects of novel maturity-onset diabetes of the young (MODY)-associated mutations on glucokinase activity and protein stability. *Biochem J*. 2006;393(Pt 1):389-396. doi: BJ20051137 [pii].

143. 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. doi: 10.1038/953 [doi].

144. Bjorkhaug L, Ye H, Horikawa Y, Sovik O, Molven A, Njolstad PR. MODY associated with two novel hepatocyte nuclear factor-1alpha loss-of-function mutations (P112L and Q466X). *Biochem Biophys Res Commun*. 2000;279(3):792-798. doi: 10.1006/bbrc.2000.4024 [doi].

145. 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.

146. Ekholm E, Nilsson R, Groop L, Pramfalk C. Alterations in bile acid synthesis in carriers of hepatocyte nuclear factor 1alpha mutations. *J Intern Med*. 2013;274(3):263-272. doi: 10.1111/joim.12082 [doi].

147. Boesgaard TW, Pruhova S, Andersson EA, et al. Further evidence that mutations in INS can be a rare cause of maturity-onset diabetes of the young (MODY). *BMC Med Genet*. 2010;11:42-2350-11-42. doi: 10.1186/1471-2350-11-42 [doi].

148. 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. doi: 10.2337/db07-1467 [doi].

149. Gragnoli C, Menzinger Von Preussenthal G, Habener JF. Triple genetic variation in the HNF-4alpha gene is associated with early-onset type 2 diabetes mellitus in a philippino family. *Metabolism*. 2004;53(8):959-963. doi: S002604950400160X [pii].

150. Bellanne-Chantelot C, Carette C, Riveline J-, et al. The type and the position of HNF1A mutation modulate age at diagnosis of diabetes in patients with maturity-onset diabetes of the young (MODY)-3. *Diabetes*. 2007;57(2):503 <last_page> 508. doi: 10.2337/db07-0859.

151. Chevre JC, Hani EH, Boutin P, et al. Mutation screening in 18 caucasian families suggest the existence of other MODY genes. *Diabetologia*. 1998;41(9):1017-1023. doi: 10.1007/s001250051025 [doi].

152. 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. doi: 10.1073/pnas.0906474106 [doi].

153. 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. doi: 06-PLME-RA-0685R2 [pii].

154. Wang Y, Gao E, Wu J, et al. Fetal macrosomia and adolescence obesity: Results from a longitudinal cohort study. *Int J Obes (Lond)*. 2009;33(8):923-928. doi: 10.1038/ijo.2009.131 [doi].

155. Stride A, Ellard S, Clark P, et al. Beta-cell dysfunction, insulin sensitivity, and glycosuria precede diabetes in hepatocyte nuclear factor-1alpha mutation carriers. *Diabetes Care*. 2005;28(7):1751-1756. doi: 28/7/1751 [pii].

156. Fendler W, Rizzo M, Borowiec M, et al. Less but better: Cardioprotective lipid profile of patients with GCK-MODY despite lower HDL cholesterol level. *Acta Diabetol*. 2014;51(4):625-632. doi: 10.1007/s00592-014-0567-1 [doi].

157. McDonald TJ, McEneny J, Pearson ER, et al. Lipoprotein composition in HNF1A-MODY: Differentiating between HNF1A-MODY and type 2 diabetes. *Clin Chim Acta*. 2012;413(9-10):927-932. doi: 10.1016/j.cca.2012.02.005 [doi].

158. 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. doi: 10.2337/dc11-0323 [doi].

159. Thanabalasingham G, Huffman JE, Kattla JJ, et al. Mutations in HNF1A result in marked alterations of plasma glycan profile. *Diabetes*. 2013;62(4):1329-1337. doi: 10.2337/db12-0880 [doi].

160. Ellard S, Bellanne-Chantelot C, Hattersley AT, European Molecular Genetics Quality Network (EMQN) MODY group. Best practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young. *Diabetologia*. 2008;51(4):546-553. doi: 10.1007/s00125-008-0942-y [doi].

161. Fajans SS, Brown MB. Administration of sulfonylureas can increase glucose-induced insulin secretion for decades in patients with maturity-onset diabetes of the young. *Diabetes Care*. 1993;16(9):1254-1261.

162. 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. doi: dc07-1785 [pii].

163. Stride A, Ellard S, Clark P, et al. Beta-cell dysfunction, insulin sensitivity, and glycosuria precede diabetes in hepatocyte nuclear factor-1alpha mutation carriers. *Diabetes Care*. 2005;28(7):1751-1756. doi: 28/7/1751 [pii].

164. 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. doi: 10.2337/dc11-1243 [doi].

165. 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. doi: 10.2337/dc13-2248 [doi].

166. 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. doi: 10.1007/s00125-011-2261-y [doi].

167. Bacon S, Kyithar MP, Schmid J, Costa Pozza A, Handberg A, Byrne MM. Circulating CD36 is reduced in HNF1A-MODY carriers. *PLoS One*. 2013;8(9):e74577. doi: 10.1371/journal.pone.0074577 [doi].

168. Pinelli M, Acquaviva F, Barbetti F, et al. Identification of candidate children for maturity-onset diabetes of the young type 2 (MODY2) gene testing: A seven-item clinical flowchart (7-iF). *PLoS One*. 2013;8(11):e79933. doi: 10.1371/journal.pone.0079933 [doi].

169. Carroll RW, Murphy R. Monogenic diabetes: A diagnostic algorithm for clinicians. *Genes (Basel)*. 2013;4(4):522-535. doi: 10.3390/genes4040522 [doi].

170. Fajans SS, Bell GI. MODY: History, genetics, pathophysiology, and clinical decision making. *Diabetes Care*. 2011;34(8):1878-1884. doi: 10.2337/dc11-0035 [doi].

171. 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. doi: 10.1007/s00125-011-2418-8 [doi].

172. 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. doi: 10.1111/jdi.12432 [doi].

173. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The human gene mutation database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet*. 2014;133(1):1-9. doi: 10.1007/s00439-013-1358-4 [doi].

174. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. doi: 10.1093/nar/gkq603 [doi].

175. T2D-GENES consortium, GoT2D consortium, DIAGRAM consortium. type2diabetesgenetics.org. Accessed 2015 Aug 11, 2015.

176. 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. doi: 10.1038/gim.2016.13 [doi].

177. Gay P, Lopez B, Pla A, Saperas J, Pous C. Enabling the use of hereditary information from pedigree tools in medical knowledge-based systems. *J Biomed Inform*. 2013;46(4):710-720. doi: 10.1016/j.jbi.2013.06.003 [doi].

178. Frank E, Hall M, Trigg L, Holmes G, Witten IH. Data mining in bioinformatics using weka. *Bioinformatics*. 2004;20(15):2479-2481. doi: 10.1093/bioinformatics/bth261 [doi].

179. Suther S, Kiros GE. Barriers to the use of genetic testing: A study of racial and ethnic disparities. *Genet Med*. 2009;11(9):655-662. doi: 10.1097/GIM.0b013e3181ab22aa [doi].

180. Paulus RA, Davis K, Steele GD. Continuous innovation in health care: Implications of the geisinger experience. *Health Aff (Millwood)*. 2008;27(5):1235-1245. doi: 10.1377/hlthaff.27.5.1235 [doi].

181. CDC's Division of Diabetes Translation in the National Center for Chronic Disease Prevention and Health Promotion. Mean and median age at diagnosis of diabetes among adult incident cases aged 18–79 years, united states, 1997–2011. https://www.cdc.gov/diabetes/statistics/age/fig2.htm. Accessed 5/16, 2017.

182. NIH Office of Research on Women's Health. Monitoring adhere to the NIH policy on the inclusion of women and minorities as subject in clinical research. . 2013.

183. Wu RR, Himmel TL, Buchanan AH, et al. Quality of family history collection with use of a patient facing family history assessment tool. *BMC Fam Pract*. 2014;15:31-2296-15-31. doi: 10.1186/1471-2296-15-31 [doi].

184. 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. doi: 10.2337/dc11-1243 [doi].

185. Winckler W, Burtt NP, Holmkvist J, et al. Association of common variation in the HNF1alpha gene region with risk of type 2 diabetes. *Diabetes*. 2005;54(8):2336-2342. doi: 54/8/2336 [pii].

186. 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. doi: 10.2337/db16-0460 [doi].

187. Schlegel A, Gut P. Metabolic insights from zebrafish genetics, physiology, and chemical biology. *Cell Mol Life Sci*. 2015. doi: 10.1007/s00018-014-1816-8 [doi].

188. Maddison LA, Chen W. Nutrient excess stimulates beta-cell neogenesis in zebrafish. *Diabetes*. 2012;61(10):2517-2524. doi: db11-1841 [pii].

189. Eames SC, Kinkel MD, Rajan S, Prince VE, Philipson LH. Transgenic zebrafish model of the C43G human insulin gene mutation. *J Diabetes Investig*. 2013;4(2):157-167. doi: 10.1111/jdi.12015 [doi].

190. Kimmel RA, Dobler S, Schmitner N, Walsen T, Freudenblum J, Meyer D. Diabetic pdx1-mutant zebrafish show conserved responses to nutrient overload and anti-glycemic treatment. *Sci Rep*. 2015;5:14241. doi: 10.1038/srep14241 [doi].

191. O'Hare EA, Yerges-Armstrong LM, Perry JA, Shuldiner AR, Zaghloul NA. Assignment of functional relevance to genes at type 2 diabetes-associated loci through investigation of beta-cell mass deficits. *Mol Endocrinol*. 2016;30(4):429-445. doi: 10.1210/me.2015-1243 [doi].

192. Rose RB, Bayle JH, Endrizzi JA, Cronk JD, Crabtree GR, Alber T. Structural basis of dimerization, coactivator recognition and MODY3 mutations in HNF-1alpha. *Nat Struct Biol*. 2000;7(9):744-748. doi: 10.1038/78966 [doi].

193. Bjorkhaug L, Sagen JV, Thorsby P, Sovik O, Molven A, Njolstad PR. Hepatocyte nuclear factor-1 alpha gene mutations and diabetes in norway. *J Clin Endocrinol Metab*. 2003;88(2):920-931. doi: 10.1210/jc.2002-020945 [doi].

194. Xu JY, Chan V, Zhang WY, Wat NM, Lam KS. Mutations in the hepatocyte nuclear factor-1alpha gene in chinese MODY families: Prevalence and functional analysis. *Diabetologia*. 2002;45(5):744-746. doi: 10.1007/s00125-002-0814-9 [doi].

195. Gu N, Suzuki N, Takeda J, et al. Effect of mutations in HNF-1alpha and HNF-1beta on the transcriptional regulation of human sucrase-isomaltase in caco-2 cells. *Biochem Biophys Res Commun*. 2004;325(1):308-313. doi: S0006-291X(04)02327-7 [pii].

196. Vaxillaire M, Abderrahmani A, Boutin P, et al. Anatomy of a homeoprotein revealed by the analysis of human MODY3 mutations. *J Biol Chem*. 1999;274(50):35639-35646.

197. Tewhey R, Kotliar D, Park DS, et al. Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell*. 2016;165(6):1519-1529. doi: 10.1016/j.cell.2016.04.027 [doi].

198. Ulirsch JC, Nandakumar SK, Wang L, et al. Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell*. 2016;165(6):1530-1545. doi: 10.1016/j.cell.2016.04.048 [doi].

199. Savage DB, Agostini M, Barroso I, et al. Digenic inheritance of severe insulin resistance in a human pedigree. *Nat Genet*. 2002;31(4):379-384. doi: 10.1038/ng926 [doi].

200. Garg A, Agarwal AK. Lipodystrophies: Disorders of adipose tissue biology. *Biochim Biophys Acta*. 2009;1791(6):507-513. doi: 10.1016/j.bbalip.2008.12.014 [doi].

201. Vesterhus M, Haldorsen IS, Raeder H, Molven A, Njolstad PR. Reduced pancreatic volume in hepatocyte nuclear factor 1A-maturity-onset diabetes of the young. *J Clin Endocrinol Metab*. 2008;93(9):3505-3509. doi: 10.1210/jc.2008-0340 [doi].

202. Pontoglio M, Sreenan S, Roe M, et al. Defective insulin secretion in hepatocyte nuclear factor 1alpha-deficient mice. *J Clin Invest*. 1998;101(10):2215-2222. doi: 10.1172/JCI2548 [doi].

203. Bill BR, Petzold AM, Clark KJ, Schimmenti LA, Ekker SC. A primer for morpholino use in zebrafish. *Zebrafish*. 2009;6(1):69-77. doi: 10.1089/zeb.2008.0555 [doi].

204. Cho SW, Kim S, Kim Y, et al. Analysis of off-target effects of CRISPR/cas-derived RNA-guided endonucleases and nickases. *Genome Res*. 2014;24(1):132-141. doi: 10.1101/gr.162339.113 [doi].

205. Ogden CL, Carroll MD, Lawman HG, et al. Trends in obesity prevalence among children and adolescents in the united states, 1988-1994 through 2013-2014. *JAMA*. 2016;315(21):2292-2299. doi: 10.1001/jama.2016.6361 [doi].

206. World Health Organization. Global report on diabetes. . 2016.

207. Liu L, Li Y, Li S, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol*. 2012;2012:251364. doi: 10.1155/2012/251364 [doi].

208. National Institute of Health. Center for identification and study of individuals with atypical diabetes mellitus (U54). https://grants.nih.gov/grants/guide/rfa-files/RFA-DK-17-006.html. Updated 2017. Accessed 6/13, 2017.

209. Brownstein CA, Brownstein JS, Williams DS,3rd, Wicks P, Heywood JA. The power of social networking in medicine. *Nat Biotechnol*. 2009;27(10):888-890. doi: 10.1038/nbt1009-888 [doi].